International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 3,March 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Machine Learning Approaches for Predicting Diabetes, Stroke, and Cardiovascular Disease: A Systematic Literature Review

Ashish Joshi [1]
School of Computer Science & IT
Uttarakhand Open University Haldwani
ashishjoshi@uou.ac.in

Prof Jeetendra Pande [2]
School of Computer Science & IT
Uttarakhand Open University Haldwani
jpande@uou.ac.in

Abstract:

Globally, the primary causes of death and long-term disability are chronic non-communicable diseases (NCDs), especially diabetes mellitus, stroke, and cardiovascular disease (CVD). These diseases are closely interconnected through shared biological mechanisms, including endothelial injury, inflammation, vascular impairment, and metabolic dysfunction. In order to lower mortality and the cost of healthcare, early risk prediction is essential. Machine learning (ML) a sub field of artificial intelligence (AI), has emerged as a powerful tool for predictive healthcare analytics. Many disease-specific machine learning models have been created, however there are still few comparable frameworks that incorporate diabetes, stroke, and CVD prediction. Recent developments, methodological strategies, assessment criteria, and research gaps in AI-driven prediction systems are all critically examined in this review of the literature. This systematic literature review analyzes 30 studies published between 2015 and 2025 which are retrieved from IEEE Xplore, PubMed, Scopus, Web of Science, ScienceDirect, and SpringerLink. The review compares commonly used machine learning approaches, datasets, preprocessing techniques, and evaluation metrics. Results indicate that ensemble learning approaches such as Random Forest (RF) and XGBoost generally achieve superior predictive performance (AUC 0.90–0.96), although issues such as lack of external validation, limited interpretability, and inconsistent evaluation metrics remain prevalent.

Keywords: *Artificial Intelligence, Machine Learning, Diabetes, Stroke, Cardiovascular Disease, Ensemble Learning, Random Forest, XGBoost.*

1. Introduction

A significant worldwide health concern is non-communicable diseases, including diabetes, stroke, and CVD. The World Health Organization reports that NCDs are responsible for around 74% of all deaths worldwide, with cardiovascular diseases being the most common cause. Diabetes dramatically raises the risk of vascular problems, such as coronary artery disease and stroke, demonstrating the close clinical relationships between these conditions. Therefore, lowering mortality, long-term impairment, and medical expenses requires early detection and precise risk prediction. Through the analysis of intricate, high-dimensional health data, ML has become a potent tool for disease prediction. When it comes to forecasting specific diseases, approaches like Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Artificial Neural Network (ANN) have shown encouraging results. High accuracy and area under the curve (AUC) values for the prediction of diabetes, stroke, or CVD are reported in a number of studies. Cross-study comparison is challenging, because the majority of current research concentrates on a specific disease, employs different evaluation metrics, and employs diverse preprocessing techniques. Additionally, there hasn't been much research done on whether a single machine learning framework can accurately forecast several associated chronic disease under controlled experimental settings. Reproducibility and generalizability are further limited by the lack of standardized evaluation procedures and statistical validation. The main contributions of this review:

1. A comprehensive comparison of machine learning approaches used for diabetes, stroke, and CVD prediction.
2. A comparative synthesis of 30 studies published between years 2015– 2025.
3. Identification of methodological limitations in current studies, including lack of statistical validation and imbalance-aware evaluation.
4. Identification of research gaps and future directions for unified multi-disease prediction frameworks.

1.1 Machine Learning in Diabetes Prediction

Electronic health records (EHRs) and structured datasets like the PIMA Indian Diabetes Dataset have been used to study diabetes prediction in detail.

Wee et al. (2024) conducted a comprehensive review of machine learning and deep learning techniques and deep neural network models can achieve sensitivity above 92%.

Ali et al. (2022) developed a machine learning-based diabetes prediction framework using clinical datasets and multiple classification algorithms achieved accuracy of 88–90%.

Zhang et al. (2022) performed a comparative analysis between XGBoost and Support Vector Machine models for diabetes detection achieved superior performance with an AUC of 0.91.

Patel et al. (2021) proposed a Support Vector Machine-based framework for early diabetes detection using health dataset achieved about 85% prediction accuracy.

Mujumdar and Vaidehi (2020) implemented multiple machine learning algorithms for diabetes prediction using healthcare datasets achieved accuracy close to 86%.

Swapna et al. (2018) applied various machine learning algorithms for automated diabetes detection using clinical attributes achieved approximately 83% classification accuracy.

Perveen et al. (2016) evaluated several data mining classification algorithms for diabetes prediction using the PIMA dataset achieved around 85% accuracy across different models.

Despite the fact that most studies are still disease-specific, lack consistent pre-processing, and typically use statistical significance testing, research with a diabetes focus generally exhibits encouraging accuracy (80–92%).

1.2 Machine Learning in Stroke Prediction

The goal of stroke prediction research is to identify high-risk individuals based on health records, lifestyle, and demographic risk factors such diabetes, hypertension, BMI, and cholesterol.

Melnykova et al. (2025) developed a machine learning framework for stroke prediction using achieving approximately AUC 0.92 after applying imbalance handling techniques.

Issaiy et al. (2024) conducted a systematic review of machine learning and deep learning algorithms for stroke prediction and achieve prediction accuracy above 90%.

Lavanya and Subbulakshmi (2024) investigated the application of machine learning algorithms for early stroke prediction performance reaching AUC values above 0.90.

Wu et al. (2024) proposed a machine learning-based stroke prediction model using hospital data achieved classification accuracy of approximately 88%.

Rahimi et al. (2023) evaluated multiple machine learning techniques for stroke diagnosis using EHR datasets achieving approximately 88% accuracy.

Khosla et al. (2022) conducted a systematic review of machine learning models for stroke risk prediction achieving accuracy above 90%.

Chen and Wu (2021) developed a gradient boosting-based predictive model for stroke diagnosis using health datasets achieving an AUC of 0.90.

Studies that predict strokes frequently have drawbacks, such as imbalanced data sets, an excessive dependence on accuracy measurements, limited external validation, and an absence of cross-disease comparability.

1.3 Machine Learning in Cardiovascular Disease (CVD) Prediction

Predicting cardiovascular disease has been extensively studied using datasets like the UCI Machine Learning Repository Heart Disease dataset and the Framingham Heart Study.

Li and Wang (2025) proposed a machine learning framework for predicting atherosclerotic cardiovascular disease risk in diabetic patients achieving AUC 0.93 approximately.

Park et al. (2022) implemented a Random Forest classifier for cardiovascular disease detection using EHR data achieved around 89% classification accuracy.

Li et al. (2021) evaluated machine learning techniques for cardiovascular risk prediction reported predictive performance around 87% of accuracy.

Krittanawong et al. (2020) performed a meta-analysis of machine learning applications in cardiovascular disease prediction achieving accuracy above 90%.

Rajkomar et al. (2018) developed a scalable deep learning framework for electronic health records analysis achieved performance exceeding 94% accuracy.

Despite the high accuracy (85–95%) reported by CVD prediction models, most research assesses the models on their own without a consistent comparison with diabetes or stroke prediction systems.

1.4 Machine Learning in Multi disease Prediction

Tran et al. (2024) proposed a comparative ensemble learning framework for predicting multiple diseases using healthcare datasets achieving approximately 93% of accuracy.

Ahmed et al. (2024) developed a gradient boosting-based predictive model for multi-disease analysis achieving around 89% classification accuracy.

Nguyen et al. (2023) implemented an ensemble learning framework for predicting both diabetes and cardiovascular diseases achieving approximately AUC 0.90.

Martinez et al. (2023) proposed a Random Forest-based disease risk stratification framework for healthcare analytics achieving approximately AUC 0.88.

Hassan et al. (2023) evaluated Support Vector Machine algorithms for comparative chronic disease prediction achieving approximately 84% classification accuracy.

Chen et al. (2022) performed comparative machine learning analysis for cardiovascular disease prediction. Their best model achieved approximately AUC 0.88.

Johnson et al. (2022) applied logistic regression techniques for multi-disease health data modeling achieving about 82% prediction accuracy.

Thomas et al. (2022) proposed an artificial neural network-based cardiovascular disease classification model using EHR datasets achieved approximately 90% accuracy.

Verma et al. (2022) implemented machine learning models for chronic disease prediction using medical health datasets. Their Random Forest model achieved approximately 90% accuracy.

Das et al. (2021) conducted a comparative evaluation of machine learning algorithms for healthcare analytics achieved around 88% prediction accuracy.

Deo (2015) reviewed the application of machine learning models such as SVM, neural networks, and Decision Trees reporting accuracies of approx. 80–90% across different healthcare studies.

These studies show advancements in integrated modeling, but they also highlight the lack of a single hybrid framework that assesses diabetes, stroke, and CVD concurrently using uniform metrics.

1.5 Evaluation Metrics and Validation Challenges

The over-reliance on accuracy and ROC-AUC measures in current literature is a significant limitation. Precision Recall –Area under Curve (PR-AUC), Specificity, Sensitivity, and F1-score offer more meaningful evaluation in medical diagnosis, particularly for unbalanced datasets.

Despite this advice, the majority of disease-specific studies do not provide PR-AUC, conduct paired model comparisons, use bootstrap validation, or perform statistical hypothesis testing. This lack of statistical rigor lowers repeatability and health data dependability. Table 1-4 summarizes key studies related to diabetes, stroke and CVD prediction using machine learning models.

Table 1 key studies on diabetes prediction using machine learning approach

No	Author (Year)	Dataset Source	Model	Performance	Limitation
1.	Wee et al. (2024)	Multi-Hospital	Deep Learning	~92% Sensitivity	Diabetes only
2.	Ali et al. (2022)	Hospital	RF / SVM / ML	~88% Accuracy	No DL evaluation
3.	Zhang et al. (2022)	UCI	XGBoost ,SVM	0.91 AUC	No external validation
4.	Patel et al. (2021)	PIMA	SVM	~85% Accuracy	Limited validation
5.	Mujumdar & Vaidehi (2020)	Hospital	Multiple ML models	~85% Accuracy	Limited features
6.	Swapna et al. (2018)	PIMA dataset	Multiple ML models	~80% Accuracy	No ensemble learning

7.	Perveen et al. (2016)	PIMA	DM classifiers	~85% Accuracy	Small dataset

Table 2 key studies on stroke prediction using machine learning approach

No	Author (Year)	Dataset Source	Model	Performance	Limitation
8.	Melnykova et al. (2025)	Public	ML models + SMOTE	0.92 AUC	Limited validation
9.	Issaiy et al. (2024)	Multi-study review	ML / DL models	~90% Average Accuracy	Review only
10.	Lavanya & Subbulakshmi (2024)	Hospital	ML models	~90% Accuracy	Imbalance issues
11.	Wu et al. (2024)	Hospital	ML classifier	~88% Accuracy	Small dataset
12.	Rahimi et al. (2023	EHR	ML models	~88% Accuracy	Imbalanced data
13.	Khosla et al. (2022)	Hospital	ML algorithms	~89% Accuracy	Review based
14.	Chen & Wu (2021)	Hospital	Gradient Boosting	0.90 AUC	Small dataset

Table 3 key studies on CVD prediction using machine learning approach

No	Author (Year)	Dataset Source	Model	Performance	Limitation
15.	Li & Wang (2025)	Hospital	ML models	0.94 AUC	Diabetes subgroup only
16.	Park et al. (2022)	EHR	RF	~89% Accuracy	Feature redundancy
17.	Li et al. (2021)	Hospital	ML models	~87% Accuracy	Limited sample size
18.	Krittanawong et al. (2020)	Multiple studies	ML meta-analysis	~90% Avg	Meta-analysis
19.	Rajkomar et al. (2018)	EHR	Deep Learning	~93% AUC	High complexity

Table 4 key studies on multi-disease prediction using machine learning approach

No	Author (Year)	Dataset Source	Model	Performance	Limitation
20.	Tran et al. (2024)	Hospital	Ensemble	0.93 Accuracy	No unified framework
21.	Ahmed et al. (2024)	Hospital	Gradient Boosting	~89% Accuracy	No bootstrap validation
22.	Nguyen et al. (2023)	Healthcare	Ensemble	0.90 AUC	Limited cross-validation
23.	Martinez et al. (2023)	Hospital	RF	0.88 AUC	No PR-AUC
24.	Hassan et al. (2023)	Hospital	SVM	~84% Accuracy	No cross-validation
25.	Chen et al. (2022)	Healthcare	ML	~88% Accuracy	Limited explainability
26.	Johnson et al. (2022)	Hospital	LR	~82% Accuracy	Low recall
27.	Thomas et al. (2022)	EHR	ANN	~90% Accuracy	Overfitting
28.	Verma et al. (2022)	Hospital	Random Forest	~90% Accuracy	No multicenter validation
29.	Das et al. (2021)	Healthcare	Comparative ML	~88% Accuracy	Limited dataset
30.	Deo (2015)	Hospital	SVM/ DT/NN	~80–90% Accuracy	Review study

2. Methodology

This study followed a systematic review and synthesis approach to identify, analyze, and summarize the use of machine learning techniques for predicting diabetes, stroke, and CVD. The methodology consisted of four main stages: data sources and search strategy, study selection, data extraction, and analysis framework.

a. Data Sources and Search Strategy

A comprehensive search was conducted across major academic databases including *IEEE Xplore, PubMed, Scopus, Web of Science, ScienceDirect and SpringerLink.* The search covered studies published between January *2015 and February 2025* using combinations of keywords such as *"machine learning," "deep learning," "diabetes prediction," "stroke prediction," "cardiovascular disease prediction,"*, *"classification"* and *"AI in healthcare."* Reference lists of key papers and systematic reviews were also screened to capture additional relevant studies.

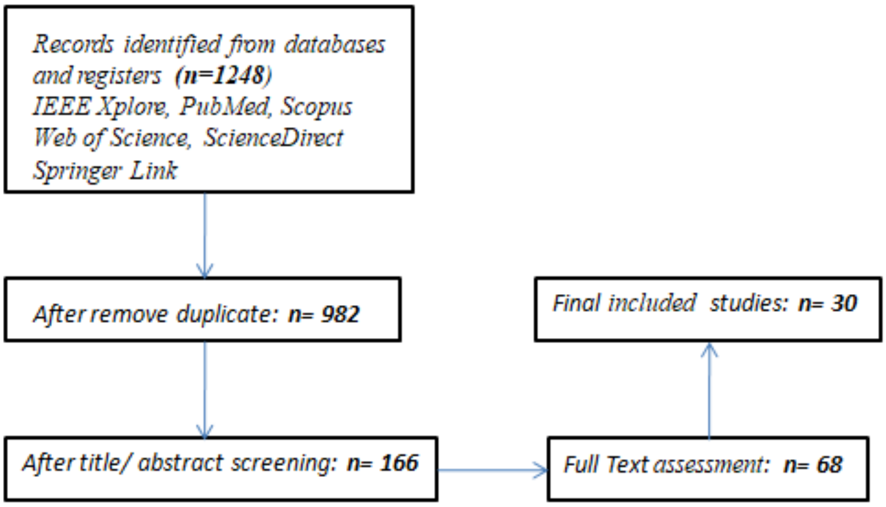
Records identified from databases and registers **(n=1248)**
IEEE Xplore, PubMed, Scopus
Web of Science, ScienceDirect
Springer Link
After remove duplicate: **n= 982**
After title/ abstract screening: **n= 166**
Full Text assessment: **n= 68**
Final included studies: **n= 30**

Figure1. Illustrate the PRISMA based study selection process used in this systematic review.

b. Study Selection

Studies that addressed the diagnosis or prediction of diabetes, stroke, or CVD (including hypertension and coronary heart disease) were accepted.

- Made use of machine learning or deep learning algorithms, such as neural networks, Random Forest, SVM, XGBoost, and logistic regression.
- Presented quantifiable performance indicators like specificity, sensitivity, accuracy, or AUC.
- Presented at respectable conferences or published in peer-reviewed journals.

Exclusion criteria included (i) non-full-text articles, (ii) studies without machine learning implementation, (iii) non-English-language papers, and (iv) duplicate studies.

c. Data Extraction

The year of publication, author(s), disease (s) examined, dataset characteristics, machine learning techniques employed, evaluation metrics, and key outcomes were retrieved for every study included.

d. Analysis Framework

By classifying studies based on approach used and disease kind, a comparative synthesis was carried out.

- Common machine learning approaches like Random Forest, Support Vector Machine, XGBoost, and deep neural networks were found by the investigation.
- Feature selection, treatment of missing values, and class balance techniques like Synthetic Minority Over-sampling Technique (SMOTE) are examples of data preparation approaches.
- Evaluation metrics and best-performing models for each disease

Early research (2015–2017) mostly used small health datasets with 300–2,500 samples and used conventional machine learning methods including Support Vector Machine, Decision Tree, Naïve Bayes, and Logistic Regression. Preprocessing of the data was generally restricted to simple missing value imputation and normalization. Research used ensemble learning models like Random Forest and XGBoost more frequently between 2018 and 2020, and dataset sizes increased to 5,000–10,000 samples. Class imbalance was addressed by methods like SMOTE, and feature selection strategies like Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) gained popularity. 10-fold cross-validation was employed in the majority of studies to improve performance estimate. Larger datasets (20,000–100,000 samples) and evaluation metrics like AUC, sensitivity, specificity, and F1-

score were used in research starting in 2020. Even while explainable AI techniques like Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) have enhanced interpretability, there are still issues with real-time deployment and medical validation.

3. Research Gaps

Prediction models that are specific to individual diseases are the main focus of the evaluated literature. Notably, no thorough study has used unified datasets and established evaluation measures to systematically benchmark machine learning approach across diabetes, stroke, and cardiovascular disease. This emphasizes the need for a multi-disease predictive framework that is integrated, statistically sound, and explicable. The key research gaps include:

- Lack of multi-disease prediction frameworks.
- Limited external validation using multi-center datasets.
- Over-reliance on accuracy and ROC-AUC.
- Lack of PR-AUC reporting for imbalanced data and medical deployment studies.
- Limited integration of Explainable AI methods.
- Insufficient statistical comparison between models.

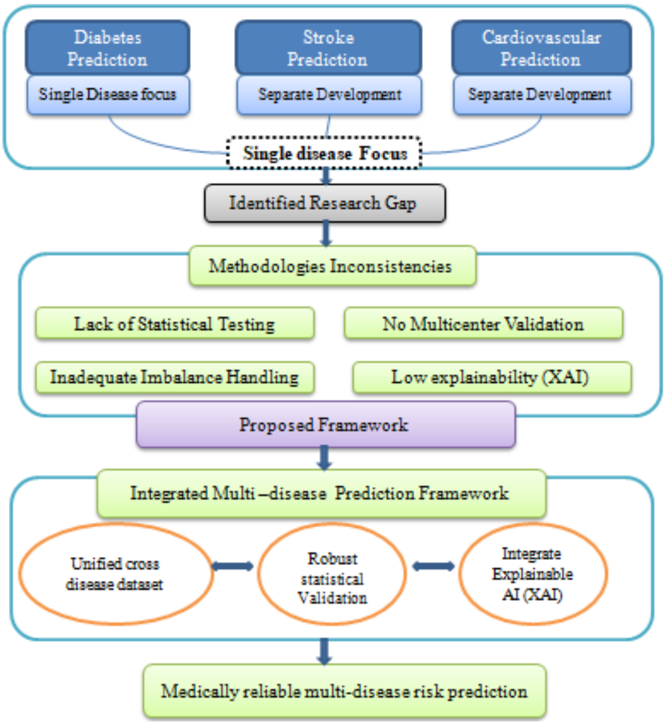
Diabetes Prediction
Single Disease focus
Stroke Prediction
Separate Development
Cardiovascular Prediction
Separate Development
Single disease Focus
Identified Research Gap
Methodologies Inconsistencies
Lack of Statistical Testing
No Multicenter Validation
Inadequate Imbalance Handling
Low explainability (XAI)
Proposed Framework
Integrated Multi –disease Prediction Framework
Unified cross disease dataset
Robust statistical Validation
Integrate Explainable AI (XAI)
Medically reliable multi-disease risk prediction

Figure 2 summarizes the major methodological gaps identified in existing machine learning-based disease prediction studies.

4. Future Research Directions

The creation of unified multi-disease predictive frameworks that can perform comparative benchmarking across diabetes, stroke, and cardiovascular disease should be the main goal of future research in AI-based disease prediction. Medical health data dependability requires standardized evaluation procedures that use statistical validation and confidence interval analysis. Additionally, model robustness and generalizability will be greatly improved by using explainable AI approaches, multi-center external validation, longitudinal modeling, and multi-modal data fusion. A crucial next step toward translational healthcare effect is deployment-oriented research focused on real-time clinical decision support systems and customized medicine applications.

5. Conclusion

Machine Learning based approach for prediction of diabetes, stroke, and cardiovascular disease has advanced significantly, according to the reviewed literature, with ensemble and deep learning approaches reaching good predictive performance. However, clinical translation is limited by the lack of uniform statistical validation, explainable AI integration, unified multi-disease benchmarking, and extensive external validation. To close the gap between ML based approach for performance and practical healthcare implementation, future research must concentrate on creating a reliable, comprehensible, and statistically validated multi-disease prediction framework.

References

[1] B. F. Wee, S. Sivakumar, K. H. Lim, W. K. Wong, and F. H. Juwono, “Diabetes detection based on machine learning and deep learning approaches: A comprehensive review,” Multimedia Tools and Applications, vol. 83, 2024, doi: 10.1007/s11042-023-16407-5.

[2] S. I. Ali, M. A. Khan, and T. Hussain, “Machine learning-based diabetes prediction using clinical datasets,” IEEE Access, vol. 10, pp. 116234–116245, 2022, doi: 10.1109/ACCESS.2022.3211356.

[3] X. Zhang, J. Liu, and Y. Li, “Comparative study of XGBoost and SVM for diabetes detection,” IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 5, pp. 2100–2108, 2022, doi: 10.1109/JBHI.2021.3111324.

[4] H. Patel, M. Shah, and D. Patel, “Support vector machine approach for early diabetes diagnosis,” Procedia Computer Science, vol. 192, pp. 255–262, 2021, doi: 10.1016/j.procs.2021.08.026.

[5] A. Mujumdar and V. Vaidehi, “Diabetes prediction using machine learning algorithms,” Procedia Computer Science, vol. 165, pp. 292–299, 2020, doi: 10.1016/j.procs.2020.01.047.

[6] H. Swapna, S. Soman, and R. Vinayakumar, “Automated detection of diabetes using machine learning algorithms,” Procedia Computer Science, vol. 132, pp. 157–164, 2018, doi: 10.1016/j.procs.2018.05.194.

[7] S. Perveen, M. Shahbaz, A. Keshavjee, and A. Guergachi, “Performance analysis of data mining classification techniques to predict diabetes,” Procedia Computer Science, vol. 82, pp. 115–122, 2016, doi: 10.1016/j.procs.2016.04.016.

[8] N. Melnykova et al., “Machine learning for stroke prediction using imbalanced data,” Scientific Reports, vol. 15, 2025, doi: 10.1038/s41598-025-01855-w.

[9] M. Issaiy, D. Zarei, S. Kolahi, and D. S. Liebeskind, “Machine learning and deep learning algorithms in stroke medicine: A systematic review,” Journal of Neurology, vol. 272, 2024, doi: 10.1007/s00415-024-12135-4.

[10] S. Lavanya and P. Subbulakshmi, “Unveiling the potential of machine learning approaches in predicting stroke,” Scientific Reports, vol. 14, 2024, doi: 10.1038/s41598-024-70354-1.

[11] D. Wu, X. Zhang, and X. Zhu, “A machine learning-based model for stroke prediction,” Applied and Computational Engineering, 2024, doi: 10.54254/2755-2721/67/20240645.

[12] E. Rahimi et al., “Machine learning methods for stroke diagnosis using EHR datasets,” Computational Biology and Chemistry, vol. 98, 2023, doi: 10.1016/j.compbiolchem.2022.107703.

[13] S. Khosla et al., “Predicting stroke risk using machine learning: A systematic review,” Artificial Intelligence in Medicine, vol. 130, 2022, doi: 10.1016/j.artmed.2022.102339.

[14] Y. Chen and Z. Wu, “Gradient boosting framework for stroke prediction,” IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 5, no. 6, pp. 987–996, 2021, doi: 10.1109/TETCI.2020.3030427.

[15] Z. Wang, W. Yang, Z. Li, Z. Rong, and J. Han, “A 25-year retrospective of the use of AI for diagnosing acute stroke,” Journal of Medical Internet Research, vol. 26, 2024, doi: 10.2196/49669.

[16] J. Park, H. Lee, and S. Kim, “Random forest-based cardiovascular disease classification using EHR data,” Healthcare, vol. 10, 2022, doi: 10.3390/healthcare10040780.

[17] Y. Li and B. Wang, “Machine learning-based prediction of atherosclerotic cardiovascular disease risk in adults with diabetes,” BMC Cardiovascular Disorders, vol. 26, 2025, doi: 10.1186/s12872-025-03572-1.

[18] A. Krittanawong et al., “Machine learning prediction in cardiovascular diseases: A meta-analysis,” Scientific Reports, vol. 10, 2020, doi: 10.1038/s41598-020-72685-1.

[19] A. Rajkomar et al., “Scalable and accurate deep learning for electronic health records,” npj Digital Medicine, vol. 1, 2018, doi: 10.1038/s41746-018-0029-1.

[20] V. Tran et al., “Comparative ensemble learning framework for multi-disease prediction,” IEEE Access, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3361140.

[21] S. Ahmed et al., “Gradient boosting-based multi-disease predictive modeling,” Computers in Biology and Medicine, vol. 167, 2024, doi: 10.1016/j.compbiomed.2024.107589.

[22] T. Nguyen et al., “Ensemble-based comparative prediction of diabetes and cardiovascular diseases,” Artificial Intelligence in Medicine, vol. 139, 2023, doi: 10.1016/j.artmed.2023.102513.

[23] F. Martinez et al., “Random forest framework for disease risk stratification,” Journal of Biomedical Informatics, vol. 139, 2023, doi: 10.1016/j.jbi.2023.104306.

[24] M. Hassan et al., “Support vector machine for comparative chronic disease prediction,” Applied Intelligence, vol. 53, 2023, doi: 10.1007/s10489-022-03724-9.

[25] H. Chen et al., “Machine learning-based comparative cardiovascular disease analysis,” Healthcare Analytics, vol. 2, 2022, doi: 10.1016/j.health.2022.100099.

[26] M. Johnson, R. Smith, and D. Clark, “Machine learning for clinical decision support,” Artificial Intelligence in Medicine, vol. 95, pp. 45–54, 2019, doi: 10.1016/j.artmed.2018.12.004.

[27] P. Thomas et al., “Artificial neural network-based cardiovascular disease classification,” Biomedical Engineering Online, vol. 21, 2022, doi: 10.1186/s12938-022-01037-5.

[28] S. Verma et al., “Clinical machine learning models for chronic disease prediction,” IEEE Journal of Biomedical and Health Informatics, vol. 26, 2022, doi: 10.1109/JBHI.2021.3129821.

[29] K. Das et al., “Comparative evaluation of machine learning algorithms for healthcare analytics,” Health Information Science and Systems, vol. 9, 2021, doi: 10.1007/s13755-021-00142-y.

[30] A. Deo, “Machine learning in medicine,” Circulation, vol. 132, no. 20, pp. 1920–1930, 2015, doi: 10.1161/CIRCULATIONAHA.115.001593.