Journal of Artificial Intelligence and Metaheuristic (JAIM)

♦ ISSN: 2833-5597 Vol: 08, No. 02, 2024 ♦ PP. 14–19

Article DOI: https://doi.org/10.54216/JAIM.080203



FULL LENGTH ARTICLE

Heart Attack Diagnosis System Based on Artificial Intelligence and Optimization Algorithms

Bahaa El-Din Waleed[1,*] **El-Sayed M. El-Kenawy**[2,3,4,5] **Sherif Ibrahim**[6]
Asmaa H. Rabie[7] **Hossam El-Din Moustafa**[8]

[1] Department of Applied Health Sciences, Higher Technological Institute of Applied Health Sciences, Mansoura, Egypt
[2] Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura 35111, Egypt
[3] School of ICT, Faculty of Engineering, Design and Information & Communications Technology (EDICT), Bahrain Polytechnic, PO Box 33349, Isa Town, Bahrain
[4] Applied Science Research Center, Applied Science Private University, Amman, Jordan
[5] Jadara University Research Center, Jadara University, Jordan
[6] Faculty of Medicine, Mansoura University, Egypt
[7] Department of Computers and Control, Faculty of Engineering, Mansoura University, Mansoura, Egypt
[8] Department of Communications and Electronics, Faculty of Engineering, Mansoura University, Egypt

Emails: bahaaeldinwaleed@gmail.com; skenawy@ieee.org; Dr_sherifarafa1981@hotmail.com; asmaahamdy@mans.edu.eg; hossam_moustafa@mans.edu.eg

Received: March 27, 2024 Revised: June 03, 2024 Accepted: November 05, 2024 ⋆ Corresponding author

ABSTRACT

Heart attacks, or myocardial infarctions, are a primary cause of mortality worldwide, underscoring the importance of early and accurate diagnosis to improve patient outcomes. This paper reviews various Artificial Intelligence (AI) and Machine Learning (ML) techniques for heart attack diagnosis, focusing on both traditional algorithms and more complex models. The traditional algorithms are Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (LR), and Decision Trees (DT). More complex models are Convolutional Neural Networks (CNN), Extreme Gradient Boosting (XGBoost), Auto-encoders, Artificial Neural Networks (ANN), and TSK Fuzzy Inference System (TANFIS). Additionally, the integration of optimization techniques, including the Grey Wolf Optimizer (GWO), Particle Swarm Optimization (PSO), and Jellyfish Optimization Algorithm (JOA), is explored to enhance model accuracy by selecting the most important features. Our findings indicate that ensemble and hybrid models, which combine ML with metaheuristic optimization, show significant potential in improving diagnostic performance and reducing overfitting. However, challenges remain, particularly regarding computational complexity and interpretability. This study provides insights into the strengths and limitations of different AI-based diagnostic models, contributing to the advancement of automated heart disease prediction systems.

Keywords: Diagnosis system ▪ Feature selection ▪ Heart attack ▪ Optimization algorithms ▪ Artificial Intelligence

1. INTRODUCTION

Heart attacks, also known as myocardial infarctions, are a leading cause of mortality worldwide. Early and accurate detection of heart attacks is crucial for timely medical intervention and improving patient outcomes. Machine learning techniques have shown great potential in assisting healthcare professionals in the early diagnosis and classification of heart attacks, enabling personalized treatment plans for patients. Given the increasing availability of electronic health records

© 2024 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

and advancements in data science, machine learning has the potential to revolutionize the field of cardiovascular medicine by providing accurate and automated classification models for heart attacks.

In conclusion, machine learning algorithms offer promising tools for the accurate classification of heart attacks, aiding in early diagnosis and personalized treatment. While challenges persist, ongoing research and advancements in this field hold great potential for improving patient outcomes and reducing the burden of heart disease on the healthcare system [1]. Traditional diagnostic methods for heart disease have typically relied on statistical models and rule-based systems, but these approaches face limitations in handling complex, high-dimensional datasets. Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) offer a transformative potential to enhance diagnostic accuracy and efficiency by automating feature selection and learning directly from patient data. New AI methods, such as deep learning, can leverage vast amounts of medical data to provide real-time insights and predictions, supporting early intervention and improving patient outcomes [2, 3]. Given the global rise in heart disease, there is an urgent need for AI-powered diagnostic tools that can handle complex clinical scenarios and contribute to better resource allocation and treatment planning [4].

AI techniques combined with optimization approaches are increasingly being used to refine diagnostic models for heart disease. AI models, including Convolutional Neural Networks (CNNs), Artificial Neural Networks (ANNs), and Support Vector Machines (SVMs), are often paired with optimization algorithms like Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) to enhance model performance by prioritizing the most relevant features and reducing computational complexity [5, 6, 7]. These optimization methods enable more accurate predictions, helping healthcare providers better identify patients at risk. The integration of AI and optimization also addresses challenges such as overfitting and interpretability, making these tools highly valuable for precise and reliable medical diagnosis [8, 9].

This paper is structured as follows. Section 2 reviews recent studies on hybrid and ensemble models that integrate various AI and optimization methods. Section 3 provides an overview of heart attack diseases. Section 4 describes AI models commonly used for heart disease diagnosis. Section 5 concludes the paper and discusses current limitations and future research directions.

2. RELATED WORK

Kolukisa and Bakir-Gungor [10] evaluated seven computational and one domain knowledge-based feature-selection methods with six classifiers and four voting ensembles for coronary artery disease diagnosis. They also proposed exhaustive and probabilistic ensemble feature-selection methods. On the Z-Alizadeh Sani, Statlog, and Cleveland datasets, the reported accuracies were 91.78%, 85.55%, and 85.47%, respectively. The study compares many feature-selection and classification configurations across three datasets.

In [11], Machine Learning (ML) algorithms such as SVM, in combination with the Jellyfish Optimization Algorithm (JOA) for feature selection, were proposed to predict heart disease. ML techniques offer low-cost and accurate diagnosis of heart disease, and feature selection using JOA helps improve model performance by reducing overfitting. Despite these advantages, disadvantages include the need for expertise in implementing ML algorithms and optimization techniques, and the challenge of interpreting the ML model. Additionally, the generalizability of the model to different populations or datasets might be limited. Overall, the study demonstrates the effectiveness of employing ML techniques such as SVM with innovative optimization algorithms for heart disease prediction, showcasing high accuracy and performance.

Almutairi [12] proposed an automated heart-disease diagnosis framework combining machine-learning classifiers, Aquila-optimizer feature selection, and grid-search hyperparameter tuning. The evaluation used three Kaggle datasets. The Histogram-based Gradient Boosting model achieved a Weighted Sum Measure (WSM) of 98.65% on the heart-failure prediction dataset, while the Decision Tree achieved 87.64% on the Heart Disease Health Indicators dataset. The reported results therefore depend on both the dataset and the evaluation measure.

Gupta and Singh [13] proposed EDL-NSGA-II, an ensemble deep-learning framework for heart-disease prediction that uses NSGA-II for feature selection. Evaluated with 10-fold cross-validation on the Comprehensive Heart Disease Dataset, the framework achieved 97.32% accuracy, with reported sensitivity, specificity, precision, and F-measure of 92.84%, 92.60%, 91.25%, and 92.17%, respectively. The study also compares the framework with several machine-learning classifiers.

According to [14], a Tuned Adaptive Neuro-Fuzzy Inference System (TANFIS) has been used to predict heart disease efficiently after optimizing TANFIS using Moth Flame Optimization (MFO) and the Grasshopper Optimization Algorithm (GOA). The proposed method represents a new hybrid model that offers high accuracy in predicting heart disease. It depends on population diversity to avoid premature convergence and local optima issues. Additionally, it used the Internet of Things (IoT) for data acquisition and fog computing for storage to contribute to real-time monitoring and alerts in abnormal situations. There are many issues in implementing this proposed method: hybrid AI models may introduce complexity in understanding and tuning the system, and optimization algorithms such as MFO and GOA may be computationally intensive. Hybrid models might also lack interpretability compared with simpler models such as decision trees or logistic regression.

In [15], a hybrid diagnosis model that includes feature selection and Naive Bayes (NB) as a diagnosis model was used to enhance accuracy in diagnosing heart attack disease. The implementation of this model is simple and easy to interpret. Additionally, backward elimination can enhance the accuracy of the classification model. However, NB assumes that features are independent, which may not hold true in real-world datasets. NB may not capture complex relationships in the data as effectively as more advanced algorithms and may not be suitable for large datasets. NB may also perform poorly on very large datasets or datasets with highly correlated features.

According to [16], wavelet denoising and Artificial Neural

Networks (ANN) have been used for heart condition detection. This work aims to improve detection accuracy in the presence of noise in ECG signals. The use of wavelet denoising can enhance the accuracy of heart abnormality detection, especially in the presence of noise. Additionally, wavelet denoising helps reduce noise interference in ECG signals, leading to more precise classification results. The integration of ANN for signal classification enables the system to learn and distinguish between different heart conditions based on ECG signal features. However, selecting the optimal parameters for the wavelet denoising algorithm may require extensive experimentation. ANN performance is highly dependent on the quality and quantity of training data available, and insufficient data may lead to suboptimal performance.

3. OVERVIEW OF HEART ATTACK DISEASES

Heart attack diseases are a significant global health concern, affecting millions of individuals worldwide. These conditions involve the disruption of blood flow to the heart, leading to potentially devastating consequences if not promptly addressed. A heart attack, also known as a myocardial infarction, occurs when the blood flow to a part of the heart is blocked [17]. Without blood flow, part of the heart muscle cannot get oxygen and begins to die. Common heart attack diseases include coronary artery disease, myocardial infarction, and aortic dissection [18]. Coronary artery disease is the most common type of heart attack and is caused by plaque buildup in the coronary arteries that supply the heart muscle with oxygen and nutrients. Myocardial infarction occurs when blood flow to part of the heart muscle is blocked, causing damage or death of heart tissue. Aortic dissection is a serious condition in which the inner layer of the aorta, the large blood vessel branching off the heart, tears. This can lead to life-threatening bleeding [19].

Common warning signs and symptoms of a potential heart attack include clutching the chest in apparent discomfort, one of the primary symptoms. Other symptoms include chest discomfort, indigestion, anxiety, irregular heartbeat, dizziness or fainting, upper body pain, shortness of breath, cold sweat, nausea or vomiting, and fatigue. These symptoms vary in type and intensity and serve as important indicators. Recognizing these signs early is crucial for timely intervention and could be life-saving, as some individuals may experience only a few symptoms without the classic chest pain, as shown in Figure 1.



Figure 1. The common symptoms of heart attacks.

4. OVERVIEW OF DIAGNOSIS MODELS

This section describes different AI techniques used in heart attack diagnosis systems: Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB), Decision Tree (DT), Extreme Gradient Boosting (XGBoost), Convolutional Neural Network (CNN), autoencoder, TSK Fuzzy Inference System (TANFIS), and Artificial Neural Network (ANN).

4.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm that excels at binary classification tasks by identifying the optimal hyperplane that separates different classes. In heart attack diagnosis, SVM uses patient data features, such as cholesterol levels, blood pressure, and age, to distinguish between patients at high risk and low risk. SVM works by selecting data points (support vectors) that define the boundaries of each class, maximizing the margin between them to improve accuracy. In practice, SVM can handle high-dimensional data and works well when there is a clear margin of separation, making it effective for analysing complex medical datasets. However, it can be computationally intensive and may struggle with overlapping data [20].

4.2 K-Nearest Neighbours (KNN)

K-Nearest Neighbours (KNN) is an instance-based learning algorithm used for both classification and regression. For heart attack diagnosis, KNN operates by finding the k most similar cases within a dataset and predicting the patient's diagnosis based on the majority class among these neighbours. KNN is easy to understand and implement, especially in a clinical setting, as it classifies new cases by directly comparing them to past instances. Its main drawback is computational complexity: as the dataset size grows, it becomes more challenging to calculate the nearest neighbours efficiently. Still, KNN remains widely used due to its simplicity and effectiveness in cases with sufficient labelled examples [21].

4.3 Logistic Regression (LR)

Logistic Regression is one of the most popular statistical models for binary classification, where it estimates the probability of a binary outcome using a logistic (sigmoid) function. In heart attack diagnosis, Logistic Regression helps predict the likelihood of a patient having a heart attack based on various clinical features. It is widely used because it is computationally efficient, interpretable, and well-suited for initial screenings where risk assessment is essential. Logistic Regression models are relatively simple but can effectively handle linearly separable data. However, they may struggle with complex, non-linear relationships without additional modifications [22].

4.4 Random Forest (RF)

Random Forest is an ensemble learning algorithm that constructs multiple decision trees and merges their predictions to improve classification accuracy and reduce overfitting.

Table 1. Many Recent Medical Diagnostic Models

Paper	AI Techniques	Optimization/Feature Selection	Accuracy	Advantages	Disadvantages
Ensemble Feature Selection for Coronary Artery Disease [10]	Six classifiers; four voting ensembles	Seven computational and one domain-based FS methods	Z-Alizadeh Sani: 91.78%; Statlog: 85.55%; Cleveland: 85.47%	Three datasets; ensemble feature selection	Many method configurations
Prediction of Heart Disease with Jellyfish Optimization [11]	SVM	Jellyfish optimization	98.47%	Low-cost; high sensitivity and specificity	Requires expertise; limited interpretability
Optimized Feature Selection and Hyperparameter Tuning [12]	HGB, DT, other ML classifiers	Aquila FS; grid-search tuning	HFP: 98.65%; HDHI: 87.64% (WSM)	Joint feature selection and tuning	Dataset-specific metrics
EDL-NSGA-II Heart Disease Prediction [13]	Ensemble deep learning; KNN, SVM, BBN, RF, NB	NSGA-II feature selection	97.32% accuracy	10-fold validation; multiple metrics	Evaluated on one dataset
TANFIS Classifier [14]	TANFIS	Moth Flame, Grasshopper Optimization	99.76%	High efficiency; IoT integration	High computational overhead; interpretability challenges
Naive Bayes with Feature Selection [15]	Naive Bayes	Backward Elimination	94.89%	Simplicity; ease of interpretability	Assumption of feature independence; less expressive
Wavelet Denoising with ANN [16]	ANN	Wavelet Denoising	97.05	Improved noise reduction in ECG signals	Sensitive to SNR changes; tuning complexity

In heart attack diagnosis, Random Forest leverages various patient data features to predict the likelihood of heart disease. It builds each tree on a random subset of data and combines their outcomes through majority voting, making it robust against overfitting and enhancing generalizability. This method works well with complex data, where multiple factors contribute to the diagnosis, as it can capture feature interactions. However, the model complexity makes it challenging to interpret, which can be a drawback in healthcare settings [23].

4.5 Naive Bayes (NB)

Naive Bayes is a probabilistic classifier based on Bayes' theorem. It calculates the probability of each class given the input features and classifies based on the highest posterior probability. Despite the "naive" assumption of feature independence, Naive Bayes is effective in many applications, including heart attack prediction. It processes input features, such as age, cholesterol, and blood pressure, independently to determine a patient's risk of heart attack. Naive Bayes is computationally efficient and suitable for real-time diagnosis but may struggle with more complex, dependent feature relationships [24].

4.6 Decision Tree (DT)

Decision Trees are interpretable, tree-structured models used for classification and regression tasks. Each internal node represents a decision based on a feature (e.g., blood pressure), each branch represents the outcome, and each leaf node represents a class label (e.g., heart disease present or absent). Decision Trees are easily understood by clinicians and can highlight key contributing factors to heart disease. However, they are prone to overfitting on complex data. Combining Decision Trees into ensembles, such as Random Forests, can mitigate this limitation [25].

4.7 Extreme Gradient Boosting (XGBoost)

XGBoost is an advanced gradient boosting algorithm that iteratively builds decision trees to minimize errors from previous models. It is especially valued in heart attack diagnosis for its high accuracy, efficiency, and ability to handle missing values, which are common in medical data. XGBoost optimizes feature importance, capturing complex feature interactions and improving the model's predictive performance. While effective, it requires significant computational resources and careful tuning of hyperparameters to avoid overfitting [26].

4.8 Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) are widely used for image-based medical diagnosis, especially in analysing ECG and echocardiogram images. CNNs use convolutional layers to detect features such as edges and shapes in images, helping identify abnormalities indicative of heart disease. CNNs excel at learning hierarchical features, capturing fine details that traditional algorithms might miss, making them highly effective for image diagnostics. However, CNNs require large amounts of labelled data and significant computational resources, which can be a challenge in some healthcare settings [27].

4.9 Autoencoder

Autoencoders are neural networks designed for unsupervised learning, typically used for dimensionality reduction and feature learning. In heart attack diagnosis, autoencoders can help extract the most important features from large, noisy patient datasets, removing irrelevant information. By encod-ing and then reconstructing data, autoencoders can capture essential patterns and identify anomalies. These models are particularly useful in identifying rare cases or abnormalities that traditional methods might overlook [28].

4.10 TSK Fuzzy Inference System (TANFIS)

The Takagi–Sugeno–Kang (TSK) Fuzzy Inference System, or TANFIS, applies fuzzy logic to manage uncertainty in data, making it well-suited for medical applications where data may be imprecise. In diagnosing heart attacks, TANFIS models patient data as a set of fuzzy rules, accommodating uncertainties such as approximate cholesterol levels or bor-derline blood pressure readings. It combines fuzzy logic with rule-based decision-making, enhancing interpretability and accuracy in uncertain clinical environments [29].

4.11 Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs) are inspired by the structure of the human brain, consisting of interconnected layers of neurons that learn complex patterns in data. In heart attack diagnosis, ANNs analyse diverse patient data inputs, such as blood pressure, age, and ECG readings, to predict heart disease risk. ANNs can model complex relationships between features, offering high accuracy but often lacking interpretability. They are versatile and can be fine-tuned for various diagnostic tasks, but training them requires extensive data and computational power [30].

5. CONCLUSION

The integration of AI and ML techniques in heart attack diagnosis represents a significant advancement in the field of cardiovascular medicine. As this paper has demonstrated, various algorithms, including Support Vector Machines (SVM), Random Forests (RF), and Convolutional Neural Networks (CNN), have shown promising results in improving diagnostic accuracy and efficiency. The combination of these algorithms with optimization techniques, such as Particle Swarm Optimization (PSO) and Jellyfish Optimization Algorithm (JOA), enhances feature selection and reduces overfitting, thereby yielding more robust predictive models. Despite the progress made, challenges remain, particularly in terms of computational complexity and the interpretability of models. As AI systems become more complex, the need for transparency in decision-making processes becomes paramount, especially in clinical settings where understanding model behaviour is crucial for healthcare professionals. Future research should focus on developing hybrid models that leverage the strengths of various algorithms while addressing their limitations. Additionally, the exploration of transfer learning and federated learning could enhance the generalizability of models across diverse populations and datasets. As the healthcare landscape continues to evolve, ongoing collaboration between data scientists and medical practitioners will be essential to harness the full potential of AI in improving patient outcomes.

REFERENCES

[1] S. J. Al'Aref, K. Anchouche, G. Singh, P. J. Slomka, K. K. Kolli, A. Kumar, J. K. Min *et al.*, "Machine learning in cardiovascular medicine: A review and future perspectives," *Journal of the American College of Cardiology*, vol. 73, no. 14, pp. 1707–1716, 2019.

[2] S. Dey, P. Nandy, and S. Biswas, "A comprehensive review on machine learning techniques for heart disease prediction," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 4, pp. 1577–1596, 2020.

[3] H. Nguyen, Q. Nguyen, and P. Le, "A hybrid deep learning approach for the early prediction of cardiovascular disease," *Computer Methods and Programs in Biomedicine*, vol. 200, p. 105949, 2021.

[4] A. Bresó, A. Monzón, and R. Fernández-Álamo, "Deep learning-based heart disease prediction using electronic health records," *Journal of Medical Systems*, vol. 47, no. 3, p. 23, 2023.

[5] Y. Zhang and X. Zhang, "Enhancing heart disease diagnosis through machine learning and optimization techniques," *Artificial Intelligence in Medicine*, vol. 127, p. 102226, 2022.

[6] V. M. Patel, S. Gaurav, and S. Chaudhary, "Hybrid approach for heart disease prediction using data mining and optimization techniques," *Expert Systems with Applications*, vol. 178, p. 114989, 2021.

[7] S. Taheri, A. Arabameri, and S. Samadi, "Optimized support vector machine model for heart disease prediction based on genetic algorithms," *BMC Medical Informatics and Decision Making*, vol. 22, no. 1, p. 231, 2022.

[8] A. Singh, S. Thakur, and M. Sharma, "Leveraging ensemble learning for the prediction of heart disease using optimized machine learning models," *Journal of King Saud University – Computer and Information Sciences*, 2022.

[9] P. Gupta and R. Walia, "A hybrid machine learning model for cardiovascular disease prediction using feature selection and ensemble methods," *Journal of Biomedical Informatics*, vol. 139, p. 104198, 2023.

[10] B. Kolukisa and B. Bakir-Gungor, "Ensemble feature selection and classification methods for machine learning-based coronary artery disease diagnosis," *Computer Standards & Interfaces*, vol. 84, p. 103706, 2023.

[11] A. A. Ahmad and H. Polat, "Prediction of heart disease based on machine learning using jellyfish optimization algorithm," *Diagnostics*, vol. 13, no. 2392, 2023.

[12] S. A. Almutairi, "An optimized feature selection and hyperparameter tuning framework for automated heart disease diagnosis," *Computer Systems Science and Engineering*, vol. 47, no. 2, pp. 2599–2624, 2023.

[13] A. Gupta and A. Singh, "EDL-NSGA-II: Ensemble deep learning framework with NSGA-II feature selection for heart disease prediction," *Expert Systems*, vol. 40, no. 7, p. e13254, 2023.

[14] J. Sekar, P. Aruchamy, H. S. L. Abdul, A. S. Mohammed, and S. Khamuruddeen, "An efficient clinical support system for heart disease prediction using TANFIS classifier," *Computational Intelligence*, vol. 38, no. 2, pp. 610–640, 2021.

[15] A. Purnomo, M. A. Barata, M. A. Soeleman, and F. Alzami, "Adding feature selection on naive bayes to increase accuracy on classification heart attack disease," *Journal of Physics: Conference Series*, vol. 1511, p. 012001, 2020.

[16] A. Winursito, F. Arifin, A. Nasuha, A. S. Priambodo, and Muslikhin, "Design of robust heart abnormality detection system based on wavelet denoising algorithm," *Journal of Physics: Conference Series*, vol. 2111, p. 012048, 2021.

[17] E. J. Benjamin, P. Muntner, A. Alonso, M. S. Bittencourt, C. W. Callaway, A. P. Carson, S. S. Virani *et al.*, "Heart disease and stroke statistics—2019 update: A report from the american heart association," *Circulation*, vol. 139, no. 10, pp. e56–e528, 2019.

[18] S. S. Virani, A. Alonso, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, C. W. Tsao *et al.*, "Heart disease and stroke statistics—2021 update: A report from the american heart association," *Circulation*, vol. 143, no. 8, pp. e254–e743, 2021.

[19] G. A. Roth, G. A. Mensah, C. O. Johnson, G. Addolorato, E. Ammirati, L. M. Baddour, C. J. L. Murray *et al.*, "Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the GBD 2019 study," *Journal of the American College of Cardiology*, vol. 76, no. 25, pp. 2982–3021, 2020.

[20] R. Arora and M. S. Baghini, "Application of support vector machine in classification of biomedical data," *IEEE Access*, vol. 8, pp. 151 464–151 472, 2020.

[21] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AI Communications*, vol. 33, no. 3, pp. 169–190, 2020.

[22] M. Rashidi and M. Rahman, "Applications of logistic regression in data analysis and predictive modeling," *Advances in Intelligent Systems and Computing*, vol. 1173, pp. 1013–1018, 2021.

[23] M. Pal and D. Santanu, "A comprehensive review on random forest techniques in big data environment," *Big Data Research*, vol. 23, p. 100183, 2020.

[24] A. M. Rajasekaran and N. Sharma, "Enhanced naive bayes classifier for text classification," *Journal of Big Data*, vol. 7, no. 1, p. 42, 2020.

[25] R. Walia and R. Kakkar, "Decision trees: A comprehensive review and future research directions," *Applied Artificial Intelligence*, vol. 35, no. 9, pp. 661–692, 2021.

[26] X. Pan and T. Wu, "Exploring the interpretability of XGBoost in mortality prediction for patients with COVID-19," *Journal of Biomedical Informatics*, vol. 103, p. 103380, 2020.

[27] Y. Zhang and W. Chen, "A review of convolutional neural networks in image and video analysis," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 3, pp. 611–636, 2021.

[28] S. Park and Y. Kim, "Recent advances in autoencoder applications for unsupervised anomaly detection," *Electronics*, vol. 9, no. 12, p. 2095, 2020.

[29] J. Liu and Z. Deng, "An improved TSK fuzzy inference system for predicting multi-variable time series," *IEEE Access*, vol. 8, pp. 141 659–141 669, 2020.

[30] Y. Cheng and S. Chen, "The application of artificial neural networks in intelligent systems," *Neural Processing Letters*, vol. 53, no. 3, pp. 1689–1707, 2021.