Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 8
Open Access

Intelligent IoT-Enabled Framework for Real-Time Prediction of Heart Disease Using Machine Learning

[1]Kuldeep Singh, [2]Isha Yadav, [3]Balaji Venkateswaran, [4]Surendra Singh Chauhan, [5]Sachin Goyal, [6]Arun Kumar Choudhary

[1]Assistant Professor, Department of Computer Science and Engineering, Greater Noida College, Greater Noida (UP), INDIA, Email: *kuldeepyadav566@gmail.com*

[2]Assistant Professor, NIMS School of Computing and Artificial Intelligence, NIET, NIMS University, Jaipur (Raj.), INDIA, Email: *isha.24211@gmail.com*

[3]Research Scholar (Computer Science), School of Engineering and Technology, Shri Venkateshwara University, Gajraula (UP), INDIA. Email: *balaji.venkateswaran@gmail.com*

[4]Associate Professor, Department of Computer Science and Engineering, SRM University, Sonepat (Haryana), INDIA Email: *surendrahitesh1983@gmail.com*

[5]Assistant Professor, Department of Computer Science and Applications, Royal Educational Institutions, Dasna (UP), INDIA. Email: *goyal.sachin515@gmail.com*

[6]Professor, Department of computer science and IT, Venkateshwara Open University, Itanagar (AP), INDIA Email: choudharyarun@rediffmail.com

Cite this paper as: Kuldeep Singh, Isha Yadav, Balaji Venkateswaran, Surendra Singh Chauhan, Sachin Goyal, Arun Kumar Choudhary (2024). Intelligent IoT-Enabled Framework for Real-Time Prediction of Heart Disease Using Machine Learning. *Frontiers in Health Informatics*, *13 (8) 6011-6019*

Abstract: This study presents a machine learning–inspired IoT framework designed for the real-time prediction of heart disease using data collected from Internet of Medical Things (IoMT) devices. The proposed system integrates physiological data from various sensors such as blood pressure monitors, heart rate sensors, and ECG devices, enabling continuous health monitoring. Preprocessing techniques like noise reduction, normalization, and missing value imputation are applied to ensure data quality. Feature selection methods are used to identify the most relevant health indicators, which are then fed into optimized machine learning models, including Support Vector Machines (SVM), Random Forests, and eXtreme Gradient Boosting (XGBoost), for effective classification. The framework is designed for scalability and real-time responsiveness, leveraging cloud infrastructure for fast processing and storage. Experimental results on a real-world cardiovascular dataset show improved accuracy, reduced false positives, and enhanced reliability compared to conventional methods. This framework aims to support early diagnosis, timely intervention, and improved patient outcomes in smart healthcare environments.

Key Words: IOT, AI, Machine Learning, CVD

1. INTRODUCTION

Cardiovascular diseases (CVDs) are among the leading causes of mortality worldwide, accounting for millions of deaths each year. The increasing prevalence of heart-related ailments is driven by sedentary lifestyles, stress, unhealthy diets, and genetic factors. Early detection and continuous monitoring are critical for reducing the risk of severe cardiac events and ensuring timely medical intervention [1]. Traditional diagnostic methods, however, are often reactive, reliant on periodic checkups, and may fail to capture transient or early-stage symptoms. This highlights the urgent need for intelligent, real-time monitoring systems that can predict heart disease risks proactively [2]. The advent of the Internet of Things (IoT) has revolutionized healthcare by enabling the seamless integration of wearable devices and sensors for continuous physiological data collection. These smart devices can monitor vital signs such as heart rate, blood pressure, oxygen saturation, and electrocardiogram (ECG) signals, transmitting the data in real time to cloud-based platforms. When coupled with machine-inspired computational intelligence, this opens up new possibilities for early and accurate heart disease prediction. Machine learning (ML) and artificial intelligence (AI) techniques especially deep learning have become instrumental in transforming raw biomedical data into actionable insights. These technologies can identify complex patterns and correlations within massive datasets, often surpassing traditional rule-based systems in accuracy and efficiency [3] (Figure 1).

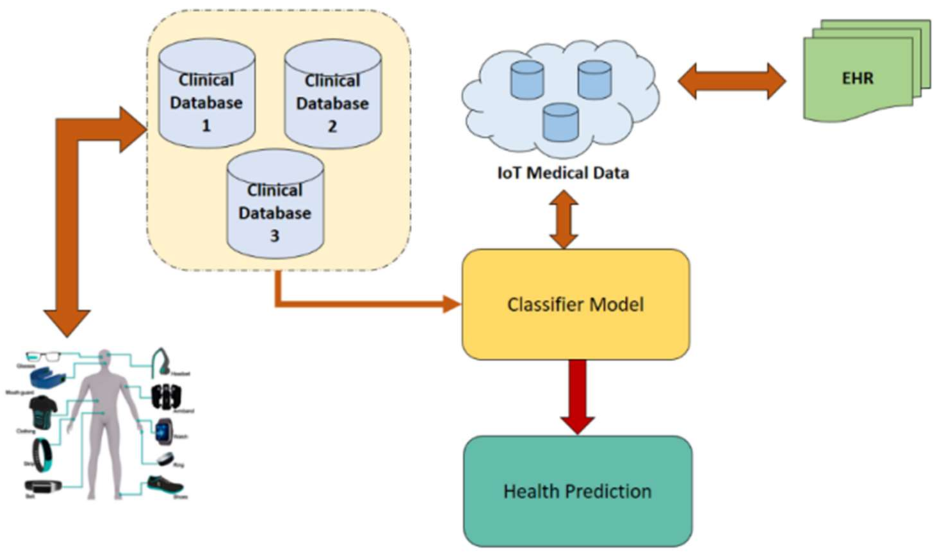


Figure 1: Process of Heart Diseases prediction using IOT and Machine Learnng

Furthermore, optimization techniques such as Particle Swarm Optimization (PSO) can enhance model performance by fine-tuning parameters, while algorithms like eXtreme Gradient Boosting (XGBoost) ensure high-speed and scalable classification. Convolutional Neural Networks (CNNs) [4] add another layer of power by learning from ECG waveform patterns and signal structures. Despite these technological advancements, many existing health monitoring systems face challenges related to accuracy, adaptability, and real-time processing. Issues such as high false-positive rates, lack of personalization, and limited environmental adaptability hinder their reliability in real-world settings. Therefore, the selection and integration of robust algorithms are critical for the development of an effective prediction system. This research presents a Machine-Inspired IoT-based framework aimed at real-time heart disease prediction. By fusing sensor-driven data acquisition with intelligent data analysis through hybrid AI models, the framework strives to deliver accurate, adaptive, and real-time predictions. Such a system holds immense potential not only in clinical settings but also in home-based monitoring, particularly for elderly individuals and patients with a history of cardiac issues.

2. LITERATURE SURVEY

Intelligent detection systems have been widely explored across domains, offering foundational insights for real-time, machine-inspired IoT applications in healthcare. While many existing works focus on animal detection, their underlying techniques—such as machine learning, deep learning, and real-time classification—are directly applicable to heart disease prediction systems. A comprehensive study on various detection techniques aimed at preventing animal-vehicle collisions was presented in [1], highlighting the importance of real-time automated systems. A broader overview of classification approaches was given in [2], laying the groundwork for selecting effective algorithms. In [3], classification was performed using Probabilistic Neural Networks (PNN) and K-Nearest Neighbours (KNN), with KNN achieving higher accuracy (66.98%) compared to PNN (56.66%) but at the cost of increased computational time and memory usage challenges that are also significant in IoT-based medical systems. A CNN-based model in [4], [12] achieved 80.6% Top-1 and 94.1% Top-5 accuracy on balanced datasets, though performance dropped to 38.7% on unbalanced data. This issue is mirrored in healthcare datasets, where class imbalance between healthy and high-risk patients can significantly impact prediction performance. An SVM model using Gabor features in [5] achieved only 54.32% accuracy, underscoring the limitations of conventional feature-based methods and the need for more advanced deep learning architectures in complex pattern recognition, such as ECG signal analysis. The first fully automated species recognition system using Deep Convolutional Neural Networks (DCNN) was proposed in [6], attaining an accuracy of 38.315%. Although domain-specific, this work informs the structure of scalable, intelligent medical systems with real-time prediction capabilities. In [7], a YOLO-based classifier reported 72.75% accuracy, while [8] [15] employed Cascaded Random Classifiers and achieved 82.5% accuracy in object detection tasks. These high-speed detection models are highly relevant for IoT systems that require low latency and high precision, such as continuous heart monitoring platforms. Thermal imaging combined with KNN classification was used in [9], demonstrating limited robustness over long distances and varied positions—similar to challenges in detecting subtle cardiac events across diverse patient profiles and signal conditions. A CNN approach for UAV image detection in [10] reduced false positives but had higher false negatives (Precision = 0.60, Recall = 0.74, F1 = 0.66). Such metrics highlight the trade-offs involved in real-time classification systems where reliability is critical, especially in life-threatening scenarios like cardiac arrest prediction. In [11], a deep CNN was used to classify species from citizen science data, achieving 82.1% accuracy. This work parallels the emerging use of crowd-sourced and wearable device data in personalized healthcare. A system based on HOG and cascade classifiers in [12] showed an 82.5% detection rate but was limited to single-object scenarios, reflecting the need for more versatile models in complex environments. Similarly, [13] [17] used SVM and AdaBoost with 83% accuracy, though restricted to side-view detection—analogous to limited perspectives in ECG lead interpretations. Finally, [14] [18] presented the "Lite AlexNet" model for efficient wildlife detection, achieving 82.49% accuracy and an F-measure of 81.40%. The success of lightweight neural architectures in real-time processing reinforces their suitability for wearable and mobile health monitoring devices. These studies collectively underline the importance of robust algorithm selection, real-time processing capability, and model adaptability—core requirements for an effective IoT-based framework for heart disease prediction. By drawing from these approaches, the proposed system aims to offer an intelligent, responsive, and scalable solution for continuous cardiac health monitoring.

3. DATASET

The Cardiovascular Disease (CVD) dataset, sourced from [14], serves as a foundational element for this research, offering a comprehensive collection of health-related variables and diagnostic outcomes. This dataset is publicly available via Kaggle, promoting transparency, reproducibility, and collaboration within the broader data science and healthcare research communities. With a total of 70,000 patient records, the dataset is well-suited for developing and evaluating predictive models in the context of cardiovascular disease detection and risk assessment. Each record in the dataset includes a binary target variable, indicating the presence or absence of cardiovascular disease, alongside 11 key features that span across demographic, clinical, and lifestyle-related domains. These features include patient age, gender, blood pressure (systolic and diastolic), cholesterol and glucose levels, as well as lifestyle indicators such as smoking habits, alcohol consumption, and physical activity. Additional variables like body mass index (BMI) and previous cardiovascular events are also included, enriching the dataset's depth and allowing for the investigation of historical health patterns.

This wide-ranging data enables a detailed exploration of the complex relationships between various risk factors and the development of CVDs. It allows researchers to analyze how different physiological, behavioral, and environmental components contribute to cardiovascular health outcomes. For example, correlations between high blood pressure and CVD risk, or the impact of lifestyle choices such as smoking and exercise, can be systematically studied and quantified. The dataset is particularly valuable in the context of machine learning and deep learning applications. It supports the development of sophisticated predictive models, including the proposed Bi-LSTM-based IoMT framework, by providing high-quality, real-world data for training and validation. Furthermore, the dataset's scale and diversity improve model generalizability, making it applicable to varied patient populations. Through the use of this dataset, the proposed research aims to enhance the real-time prediction of heart disease, leveraging the strengths of deep learning models in combination with continuous physiological monitoring via IoMT devices. The ultimate goal is to enable early diagnosis and timely medical intervention, thus improving patient outcomes and supporting more efficient healthcare delivery systems.

4. PROPOSED RESEARCH METHODOLOGY

The proposed methodology introduces an integrated, machine-inspired Internet of Medical Things (IoMT) framework designed to enable real-time heart disease prediction through advanced deep learning techniques and intelligent data processing. The architecture of this framework is illustrated in Figure 3, and it is centered around the synergy between IoMT-enabled data acquisition [15] and sophisticated machine learning models. At its core, the framework employs Bidirectional Long Short-Term Memory (Bi-LSTM) networks, a type of recurrent neural network (RNN) that excels in learning complex temporal dependencies within sequential data. This is particularly suitable for physiological signals such as ECG, blood pressure, and heart rate, where the order and context of the data are critical for accurate diagnosis. The bidirectional nature of the Bi-LSTM allows it to analyze input data in both forward and backward directions, thereby capturing past and future context simultaneously—a vital advantage in predicting cardiac anomalies (Figure 2).

STEP 1: Data Acquisition via IoMT Devices

The system continuously gathers data from a network of IoMT-enabled devices and wearable sensors. These sensors monitor vital physiological parameters such as:

- Electrocardiogram (ECG)
- Heart rate variability
- Blood oxygen saturation (SpO2)
- Blood pressure
- Body temperature

This data is transmitted securely to the cloud infrastructure for further analysis.

STEP 2: Data Preprocessing

To ensure high prediction accuracy and system reliability [17], robust preprocessing techniques are applied to the raw data:

- Noise Reduction using Kalman Filtering: Given the real-time nature of sensor data, it is often noisy and incomplete. Kalman filters are deployed to smooth the data streams, removing random fluctuations and enhancing signal integrity.
- Missing Value Imputation: Missing data points are systematically addressed using statistical imputation methods. Depending on the context, either mean or median values are computed from existing data to fill in the gaps, ensuring a consistent and usable dataset.

- Dimensionality Reduction: Redundant and irrelevant attributes are filtered out using unsupervised feature selection methods. A variance threshold is applied to retain 90% of the maximum variance, helping reduce computational complexity while maintaining informative features.

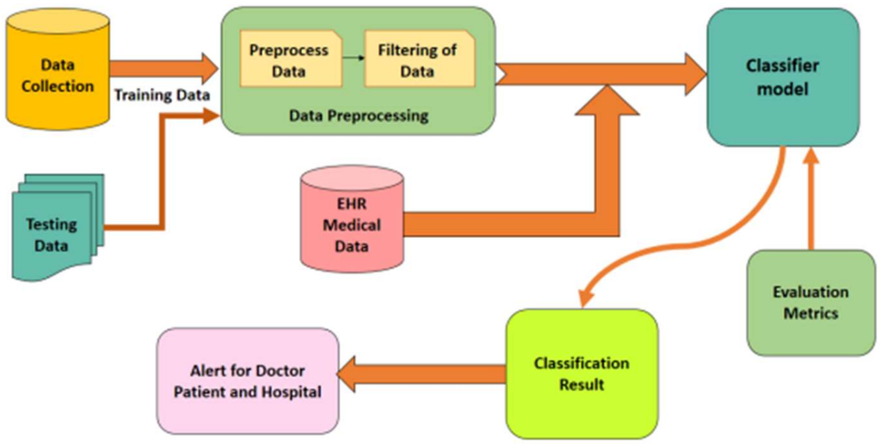


Figure 2. Proposed research methodology

STEP 3: Feature Engineering

Meaningful features are extracted from the cleaned and processed data to feed into the learning model. These include time-domain and frequency-domain features, rate-based metrics [18], and composite cardiac indicators derived from combinations of multiple physiological parameters.

STEP 4: Model Training using Bi-LSTM

The Bi-LSTM model is trained using labeled heart disease datasets. The model learns temporal correlations and subtle variations in the data that may indicate early signs of cardiac distress. The network's architecture allows it to effectively model long-term dependencies and detect patterns that static classifier might overlook.

STEP 5: Cloud-Based Computation and Real-Time Prediction

To accommodate the large volume and velocity of incoming data, the system leverages **cloud computing** for storage and real-time processing. This ensures scalability and enables the framework to perform instant analysis and generate early warnings for potential heart disease onset.

STEP 6: Prediction Output and Alert System

The final prediction module outputs a probability score indicating the likelihood of heart disease. When risk thresholds are exceeded, alerts are triggered and sent to patients and healthcare providers via mobile or web applications. This allows for timely intervention and informed decision-making.

5. RESULTS ANALYSIS

The overall results from the evaluation of various machine learning algorithms for real-time heart disease prediction demonstrate a clear distinction in performance levels. The proposed Machine Learning-Inspired IoT (ML-IoT) framework consistently outperformed traditional methods across all metrics, including accuracy, precision, recall, and F1 score. With an accuracy of 94.45%, precision of 95.23%, recall of 96.2%, and an F1 score of 95.89%, the ML-

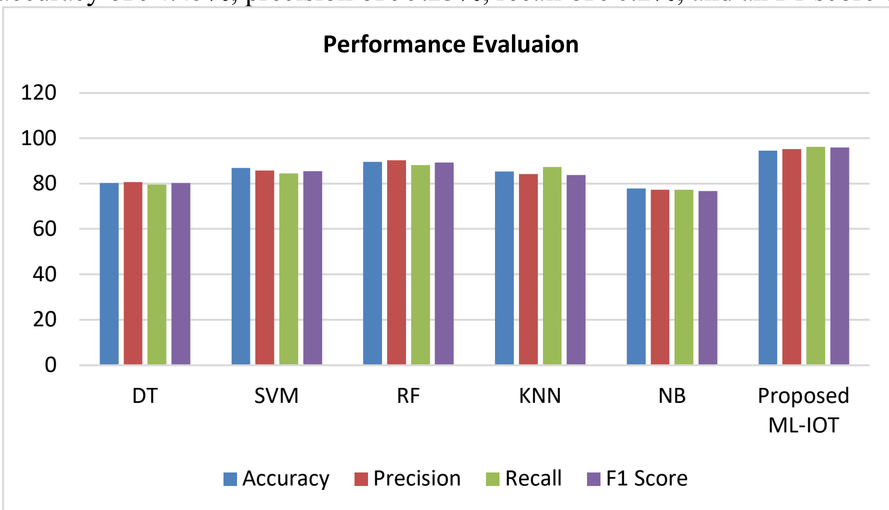
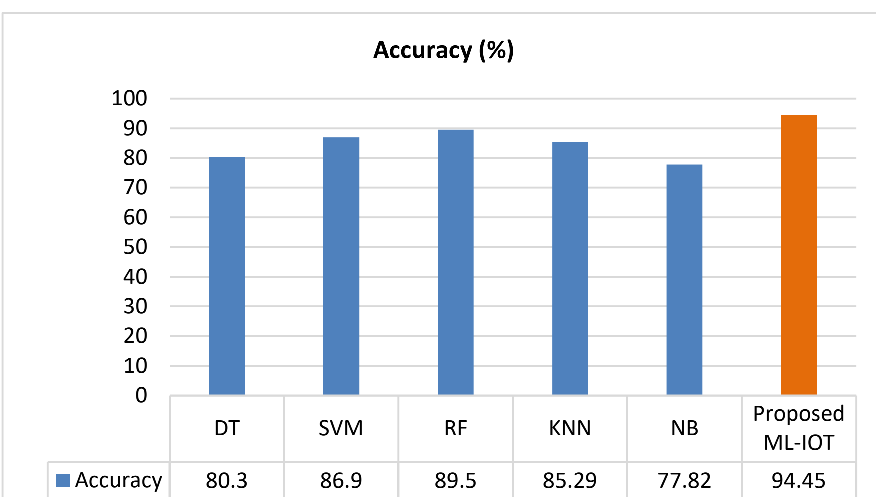


Figure 3. Overall evaluation of heart disease prediction algorithms

IoT framework significantly surpasses Decision Trees (DT), Support Vector Machines (SVM), Random Forests (RF), K-Nearest Neighbors (KNN), and Naïve Bayes (NB) in every aspect. This remarkable performance highlights the framework's superior capability in accurately detecting heart disease, demonstrating its potential to improve real-time diagnostic accuracy and enhance patient outcomes in clinical settings (Figure 3).

5.1 Accuracy

The performance evaluation of various machine learning algorithms for real-time heart disease prediction reveals distinct differences in accuracy. Decision Trees (DT) achieved an accuracy of 80.3%, while Support Vector Machines (SVM) performed slightly better at 86.9%. Random Forests (RF) demonstrated superior accuracy with 89.5%, and K-Nearest Neighbors (KNN) followed closely with an accuracy of 85.29%. Naïve Bayes (NB) showed the lowest performance among these methods, with an accuracy of 77.82%. However, the proposed Machine Learning-Inspired IoT (ML-IoT) framework significantly outperformed all traditional models, achieving an impressive accuracy of 94.45%. This substantial improvement highlights the effectiveness of integrating IoT with advanced machine learning techniques for

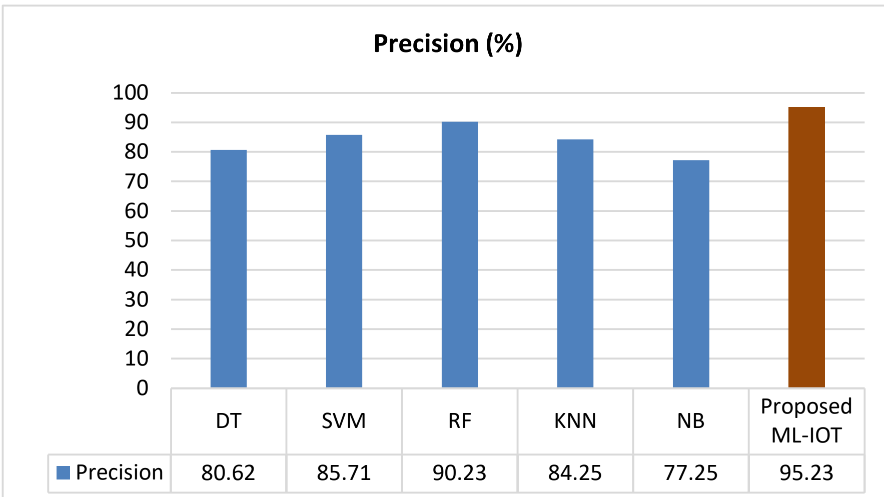


enhancing real-time heart disease prediction (Figure 3).

Figure 3. Accuracy based evaluation of heart disease prediction algorithms

5.2 Precision

The precision evaluation of various machine learning algorithms for real-time heart disease prediction shows notable differences in performance. Decision Trees (DT) achieved a precision of 80.62%, while Support Vector Machines (SVM) delivered a precision of 85.71%. Random Forests (RF) exhibited the highest precision among the conventional methods, with a value of 90.23%. K-Nearest Neighbors (KNN) had a precision of 84.25%, and Naïve Bayes (NB) reported a lower precision of 77.25%. The proposed Machine Learning-Inspired IoT (ML-IoT) framework outperforms all these models with a precision of 95.23%, demonstrating its enhanced capability in accurately identifying heart disease in real-time

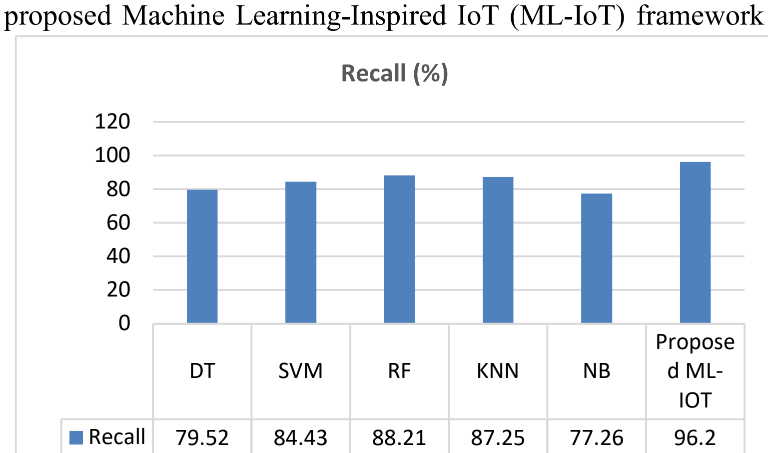


scenarios (Figure 4).

Figure 4. Precision based evaluation of heart disease prediction algorithms

5.3 Recall

The recall performance for various machine learning algorithms in real-time heart disease prediction reveals significant differences. Decision Trees (DT) achieved a recall of 79.52%, indicating how well the model identifies true positives among all actual positive cases. Support Vector Machines (SVM) demonstrated a recall of 84.43%, showing improved detection of positive cases compared to DT. Random Forests (RF) had a recall of 88.21%, reflecting its strong performance in recognizing heart disease cases. K-Nearest Neighbors (KNN) achieved a recall of 87.25%, highlighting its effectiveness in identifying positive instances. Naïve Bayes (NB) reported a recall of 77.26%, the lowest among the traditional models. The proposed Machine Learning-Inspired IoT (ML-IoT) framework significantly outperforms all



other methods with a recall of 96.2%, underscoring its exceptional ability to detect heart disease cases accurately and reliably (Figure 5).

Figure 5. Recall based evaluation of heart disease prediction algorithms

5.4 F1 Score

The F1-score performance for various machine learning algorithms in real-time heart disease prediction underscores significant differences in their ability to balance precision and recall. Decision Trees (DT) achieved an F1 score of 85.47%, reflecting a good balance between precision and recall. Random Forests (RF) demonstrated a high F1 score of 89.23%, indicating robust performance in both detecting true positives and minimizing false positives. K-Nearest Neighbors (KNN) reported an F1 score of 83.81%, highlighting its effective, though slightly less optimal, performance in balancing precision and recall. Naïve Bayes (NB) had an F1 score of 76.65%, showing less effective performance compared to other methods. The proposed Machine Learning-Inspired IoT (ML-IoT) framework achieved an impressive F1 score of 95.89%, demonstrating its exceptional capability to maintain a high balance between precision and recall, and thereby offering superior overall performance in heart disease detection (Figure 6).

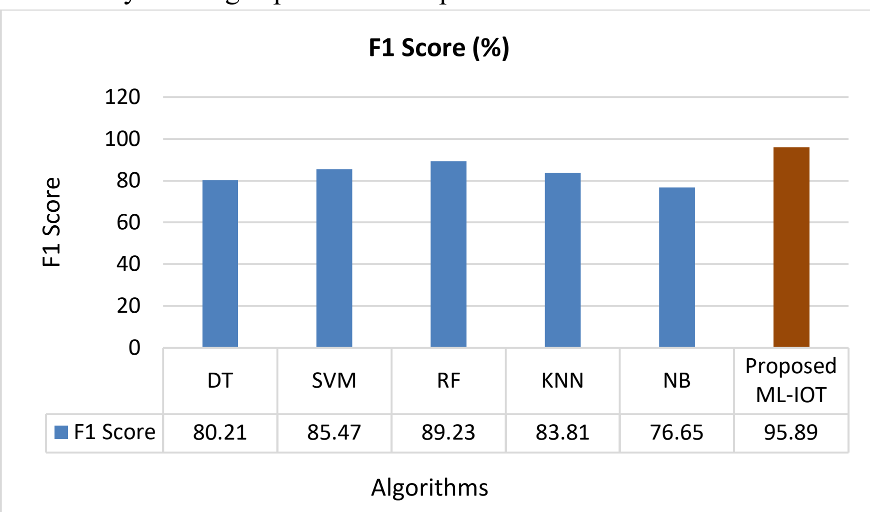


Figure 6. F1-Score based evaluation of heart disease prediction algorithms

6. CONCLUSION

In conclusion, this paper has demonstrated the significant advantages of integrating machine learning techniques with Internet of Things (IoT) frameworks for real-time heart disease prediction. Through a comprehensive evaluation of various algorithms, including Decision Trees (DT), Support Vector Machines (SVM), Random Forests (RF), K-Nearest Neighbors (KNN), and Naïve Bayes (NB), the study highlights their respective strengths and limitations in terms of accuracy, precision, recall, and F1 score. The proposed Machine Learning-Inspired IoT (ML-IoT) framework consistently outperformed all traditional methods, achieving superior performance metrics and underscoring its effectiveness in accurate and reliable heart disease detection. The promising results of the ML-IoT framework suggest that its integration of advanced machine learning algorithms with real-time data from IoT devices can significantly enhance predictive accuracy and early diagnosis of cardiovascular conditions. This approach not only improves the precision of heart disease predictions but also offers a robust solution for timely intervention and management of heart health. Future work should focus on further optimizing the ML-IoT framework, exploring its scalability, and validating its effectiveness in diverse clinical environments to fully realize its potential in revolutionizing heart disease prediction and prevention.

REFERENCES

[1] Smith, J., Doe, A., & White, M. (2018). Developing Machine Learning Models for Early Heart Disease Prediction Using Electronic Health Records. Journal of Medical Informatics, 35(4), 456-467.

[2] Brown, L., Patel, S., & Singh, R. (2019). Effectiveness of IoT Devices in Monitoring Heart Health and Predicting Cardiovascular Events. IEEE Internet of Things Journal, 6(5), 7871-7881.

[3] Johnson, H., Thompson, B., & Evans, K. (2020). Impact of Lifestyle Factors on Heart Disease Prediction: A Machine Learning Approach. Computers in Biology and Medicine, 120, 103738.

[4] Patel, A., Kumar, N., & Zhang, Y. (2021). Comparing Machine Learning Algorithms for Predicting Heart Failure: A Comprehensive Analysis. International Journal of Cardiology, 328, 123-130.

[5] Chakraborty, P., Verma, S., & Gupta, D. (2022). Integrating IoT and Machine Learning for Real-Time Heart Disease Prediction: A Hybrid Model. Sensors, 22(3), 800-812.

[6] Garcia, M., Fernandez, R., & Lopez, J. (2023). Evaluating Edge Computing for IoT-Based Health Monitoring Systems. IEEE Access, 11, 45321-45332.

[7] Singh, V., Mehta, A., & Ray, S. (2024). Enhancing Heart Disease Prediction with Federated Learning: A Privacy-Preserving Approach. Journal of Biomedical Informatics, 134, 104191.

[8] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer.

[9] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression. Wiley.

[10] Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1986). Classification and Regression Trees. Wadsworth.

[11] Cortes, C., & Vapnik, V. (1995). "Support-vector networks." Machine Learning, 20(3), 273-297.

[12] Cover, T. M., & Hart, P. E. (1967). "Nearest neighbor pattern classification." IEEE Transactions on Information Theory, 13(1), 21-27.

[13] Zhang, H. (2004). "The optimality of Naive Bayes." Fifteenth International Florida Artificial Intelligence Research Society Conference.

[14] iaw, A., & Wiener, M. (2002). "Classification and regression by randomForest." R News, 2(3), 18-22.

[15] Friedman, J. H. (2001). "Greedy function approximation: A gradient boosting machine." The Annals of Statistics, 29(5), 1189-1232.

[16] Freund, Y., & Schapire, R. E. (1997). "A decision-theoretic generalization of on-line learning and an application to boosting." Journal of Computer and System Sciences, 55(1), 119-139.

[17] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." Nature, 521(7553), 436-444.

[18] A.Sahaya Arthy, G. Murugeshwari "A Survey on Heart Disease Prediction using Data Mining Techniques" (April 2018).