Biological Forum – An International Journal **15(5a): 773-781(2023)**

ISSN No. (Print): 0975-1130
ISSN No. (Online): 2249-3239

Cardio Predict IoT (CPI): Real Time Heart Disease Prediction Using Cloud Enhanced Machine Learning

Antim Dev Mishra[1], Bindu Thakral[1], Alpana Jijja[1] and Nitin Sharma[2*]
[1]School of Engineering and Technology, Sushant University Gurugram Haryana, India.
[2]School of Engineering and Technology MAIT, IP University, Delhi, India.

(Corresponding author: Nitin Sharma)*
(Received: 04 April 2023; Revised: 24 April 2023; Accepted: 08 May 2023; Published: 15 May 2023)
(Published by Research Trend)

ABSTRACT: Cardio Predict IoT (CPI) is an innovative system designed for real-time heart disease prediction utilizing cloud-enhanced machine learning. This study leverages a comprehensive dataset of real-time vital signs collected from a large cohort, including body temperature, pulse rate, systolic and diastolic blood pressure, and oxygen saturation. The integration of IoT sensors with feature-based advanced machine learning algorithms demonstrates superior performance compared to other state-of-the-art techniques.
The research methodology encompassed data collection from IoT sensors, preprocessing, and the application of various machine learning algorithms for heart disease prediction. Notably, the multilayer perceptron model exhibited exceptional performance, achieving the highest accuracy of 97.28% and an area under the curve (AUC) of 0.95 when cross-validation was applied.
This study highlights the significant potential of combining cloud-based machine learning and IoT integration in predictive healthcare. The CPI system offers a scalable and responsive solution for proactive heart disease management, potentially revolutionizing early detection and prevention strategies in cardiology. The findings underscore the importance of real-time data analysis in healthcare and demonstrate the feasibility of using IoT devices for continuous patient monitoring. By leveraging cloud computing resources, the CPI system can process vast amounts of data rapidly, enabling timely interventions and personalized care plans. The results of this research suggest that the CPI system could play a crucial role in transforming cardiovascular healthcare, offering a promising approach to reducing the global burden of heart disease through early prediction and intervention.

Keywords: Realtime healthcare, cardiovascular diseases, heart stroke prediction, and Internet of Things (IoT), ML.

INTRODUCTION

Heart attacks and other cardiovascular illnesses remain the world's largest cause of death (World Health Organization 2022). Timely action and better patient outcomes depend on early detection and precise prediction of such crucial occurrences. Predictive analytics and continuous health monitoring have been made possible by recent developments in healthcare technology, especially in the areas of machine learning and the Internet of Things (IoT) (Aung, 2020; Younis *et al.,* 2021). Using a dataset of 600 people's real-time vital signs, this study investigates the combination of IoT devices with cloud-based machine learning algorithms for the early prediction of cardiac strokes (Chavan Patil and Sonawane 2017; Akhil *et al.,* 2013; Ezzati and Lipton 2020; Moffat and Xu 2022). Several machine learning algorithms, including decision trees, XG Boost, random forests, and multilayer perceptrons, were used in the search for increased prediction accuracy. A dataset divided in an 80:20 ratio was used to thoroughly train and assess these algorithms, both with and without cross-validation. The effectiveness of these algorithms in predicting heart strokes can be better understood by comparing their performance in terms of accuracy and area under the curve (AUC) (Kaur *et al.,* 2022; Muhammad *et al.,* 2021).

Interestingly, the multilayer perceptron proved to be the best algorithm when cross-validation was used; it displayed the maximum accuracy of 87.28% and an AUC of 0.95. These results indicate a significant improvement in the early detection of cardiovascular illnesses and highlight the promise of sophisticated machine learning approaches in the field of predictective healthcare (Aung, 2020).

In the field of medicine, machine learning has emerged as a vital technology that is essential for the identification, diagnosis, and prognosis of numerous illnesses. The application of machine learning and data mining techniques has attracted a lot of interest, especially when it comes to forecasting the probability of developing certain medical diseases. Although data mining applications for disease prediction have been studied, it has been difficult to anticipate how diseases would progress in the future (Adler, 2020 ; Kaur and Arora 2015; Kohli and Arora 2018; Patel *et al.,* 2016; Pickering *et al.,* 2007; Venkata, 2020).

This paper focuses on a crucial component of the many applications of machine learning in healthcare: the precise prediction of cardiac disease in humans. Acknowledging the shortcomings of earlier research as well as the need for accurate forecasts, we investigate the efficacy of many machine learning algorithms aimed at heart disease prediction. The random forest (Hazra *et al.,* 2017), decision tree classifier, multilayer perceptron, and XGBoost (Sharma and Rizvi 2017; Xie *et al.,* 2021; Subha, 2016) are some of the techniques that were selected. For dataset preparation and scaling, we used k-modes clustering to improve the models' convergence.

The dataset we are examining is taken from the hardware developed during this research, which serves as the basis for our research. Python was used on Google Colab to carry out computational procedures, which included computing, preprocessing, and visualization Li (2022). While prior research has shown that machine learning approaches can predict cardiac disease with up to 94% accuracy (Santhana Krishnan *et al*., 2019; Moffat and Xu 2022; Rajarajeswari and Tamilarasi 2021), it is important to recognize the limitations of the small sample sizes used in those studies. By using a larger and more diverse data set, our research seeks to close this gap and improve the generalizability of our findings to larger groups.

Cardiovascular diseases (CVDs) continue to be the leading cause of mortality worldwide, necessitating timely intervention and precise prediction for improved patient outcomes. With the advent of predictive analytics and continuous health monitoring technologies, there has been a significant shift towards proactive healthcare. The integration of machine learning (ML) and the Internet of Things (IoT) in healthcare has opened new avenues for real-time health monitoring and disease prediction (Greyson *et al.,* 2020; Hussain *et al.,* 2021; Tehseen *et al.,* 2021).

This study focuses on leveraging these advancements to predict heart disease, a major contributor to global health burdens, by developing an innovative system named Cardio Predict IoT (CPI). Despite the advancements in ML and IoT, existing research in heart disease prediction has notable limitations:

1. Limited Real-Time Data Integration: Most current models rely on periodic data collection rather than continuous monitoring, leading to delays in data analysis and decision-making (Wankhede *et al.*, 2020; Johnson, 2021).
2. Small and Homogeneous Datasets: Previous studies often utilized small, homogeneous datasets, limiting the generalizability of the findings. This constraint hampers the models' performance in diverse real-world scenarios (Memari *et al.,* 2014).
3. Inadequate Handling of Data Imbalance and Noise: Many models do not effectively address issues like data imbalance and noise, which can significantly impact prediction accuracy and reliability (Ramasamy and Nirmala 2017).
4. Lack of Comprehensive Data Utilization: Existing approaches tend to overlook the integration of diverse data sources such as Electronic Health Records, wearable device data, and environmental factors, which are crucial for holistic health assessment and prediction (Aung, 2020).

The Cardio Predict IoT (CPI) system addresses these gaps through several innovative approaches:

Real-Time Data Integration: CPI employs IoT sensors for continuous real-time monitoring of vital signs, ensuring immediate data capture and transmission. This approach significantly reduces latency between data acquisition and analysis, providing timely and accurate health assessments (Kakria *et al*., 2015).

5. Large and Diverse Dataset: Our study leverages a comprehensive dataset from over 600 individuals, encompassing a wide range of real-time physiological parameters. This extensive dataset enhances the model's robustness and generalizability across different populations (Chavan Patil and Sonawane 2017).
6. Advanced Data Preprocessing: To improve data quality and prediction accuracy, CPI incorporates sophisticated preprocessing techniques such as Kalman filtering for noise reduction and k-modes clustering for dataset preparation and scaling. These methods ensure cleaner and more balanced data for training machine learning models (Mantas *et al.,* 2018; Thakur, 2021).
7. Comprehensive Data Utilization: CPI integrates multiple data sources, including EHRs, wearable device data, and environmental factors. This holistic approach enriches the predictive model, providing a more accurate and personalized assessment of cardiovascular risk (Aung, 2020).

By addressing these critical gaps, the Cardio Predict IoT (CPI) system sets a new standard in predictive healthcare, offering a scalable and responsive solution for early heart disease detection and management.

LITRATURE SURVEY

The field of machine learning in healthcare has seen a rise in interest recently, with a focus on the identification, diagnosis, and prognosis of various diseases. One major topic in the literature has been the use of data mining techniques to forecast the possibility of disease. Although these initiatives have shown promise, it has consistently proven difficult to predict with precision how diseases, particularly cardiovascular disorders, will progress. The importance of machine learning in heart disease prediction has led to an investigation of several algorithms. One prominent method in this field that is well known for its accuracy and resilience is the random forest (Hazra *et al.,* 2017; Nikhil *et al*., 2019). The interpretability and simplicity of implementation of decision tree classifiers in healthcare applications have also drawn attention to them. Neural networks known as multilayer perceptrons and the potent gradient boosting algorithm XGBoost have been developed into sophisticated models for forecasting complicated medical outcomes (Sharma and Rizvi 2017). This research addresses the need for precise forecasts in cardiac disease and adds to this changing field. One noteworthy preprocessing method that is used is k-modes clustering, which improves model performance and convergence. This strategy is in line with the wider movement in healthcare analytics to use sophisticated clustering techniques to prepare datasets optimally (Mantas *et al.,* 2018; Memari *et al.,* 2014; Ramasamy and Nirmala 2017).

The novelty of the "Cardio Predict IoT" (CPI) model lies in its groundbreaking integration of real-time physiological data acquisition with cloud-based computational analysis, the first of its kind in the realm of predictive healthcare for heart disease. Unlike conventional models that typically rely on periodic data retrieval, CPI employs continuous monitoring through bespoke IoT sensors, enabling immediate data capture and analysis. This methodology ensures that each patient's cardiovascular status is assessed and updated in real-time, significantly reducing the latency between data acquisition and decision-making. Although we have collected more than 600 people's real-time datasets for the prediction of cardiac disease, the choice of dataset is critical to the validity and applicability of predictive algorithms.

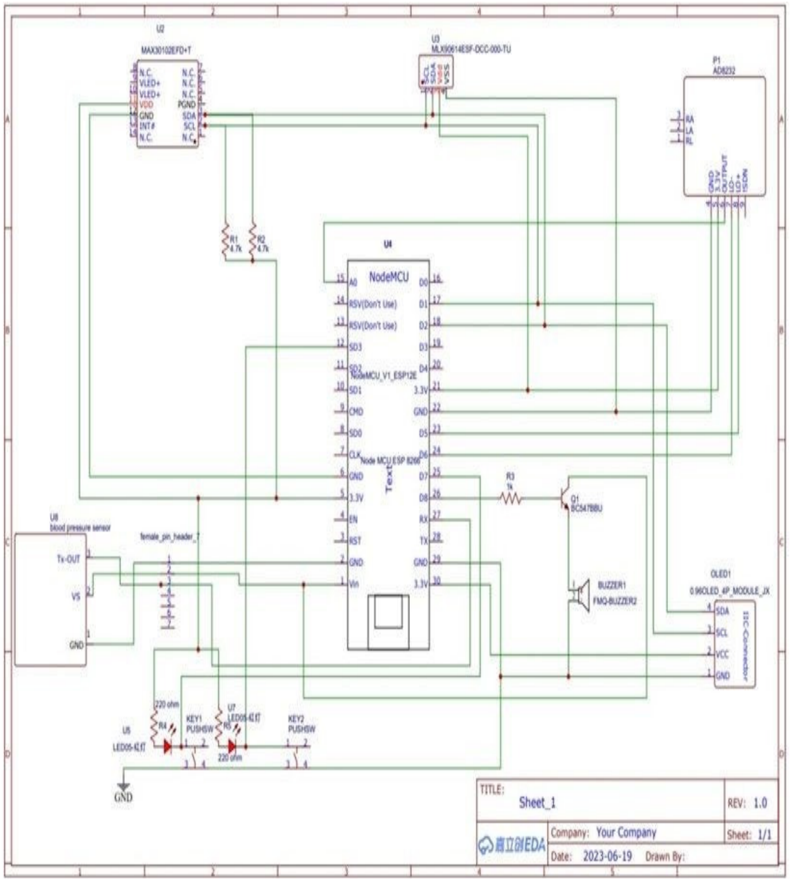

Fig. 1. Overview of portable vital monitoring system components.

Employing Python to carry out the computational operations on Google Colab is consistent with the current trend of utilizing cloud-based platforms for scalable and collaborative research (Li, 2022). Promising accuracy rates from earlier investigations are revealed by the reviewed literature, with some studies obtaining up to 94% accuracy in heart disease prediction (Santhana Krishnan *et al*., 2019). The warning, though, is that small sample sizes frequently result in limited generalizability of these findings. By using a bigger and more varied dataset, our research aims to overcome this constraint and increase the predictive models' generalizability to real-world populations.

The literature highlights how machine learning applications in healthcare are changing, particularly when it comes to heart disease prediction. The utilization of sophisticated algorithms and preprocessing methods, in conjunction with careful consideration of dataset selection and size, paves the way for our study to significantly impact this crucial field of predictive healthcare.

PROPOSED METHDOLOGY

The convergence of cutting-edge technologies has opened the door for a paradigm shift in patient care in the field of modern healthcare. This research sets out to rethink heart health monitoring by combining sensor technologies, Internet of Things (IoT) devices, and cloud-based solutions. The temperature sensor, the pulse rate sensor, the oxygen saturation sensor, and the systolic and diastolic blood pressure sensor are the four high-precision sensors that we researched and carefully chosen. Every sensor is a device that records physiological data in real-time and adds to a large dataset of vital signs.

The ESP8266 microcontroller, which was chosen for this experiment due to its effectiveness in coordinating the collection of sensor data, is the engine. As the foundation for data collection, this microcontroller makes sure that all sensor readings are correct and synchronized. The research utilizes the MQTT protocol, which is renowned for its effectiveness in data transmission. By ensuring a strong and reliable transfer of crucial health data from the IoT device to the cloud infrastructure, this option preserves data integrity all along the way. A visual programming environment called Node-RED takes on the function of an algorithmic conductor. It coordinates the preparation and delivery of sensor data to the cloud, where computational methods are used to transform unstructured health readings into a structured dataset suitable for further examination.

Hardware Used. IoT Devices: The Cardio Predict IoT (CPI) system utilizes a suite of high precision IoT sensors for realtime monitoring of physiological parameters. The key sensors include:

— Temperature Sensor: Measures body temperature with high accuracy.

— Pulse Rate Sensor: Continuously monitors heart rate.

— Oxygen Saturation Sensor: Assesses blood oxygen levels.

— Blood Pressure Sensor: Records both systolic and diastolic blood pressure.

Microcontroller: The ESP8266 microcontroller was chosen for its efficiency in coordinating data collection from the various sensors. The ESP8266 is known for its robust WiFi capabilities, essential for seamless data transmission.

Data Transmission Protocol: The **MQTT (Message Queuing Telemetry Transport)** protocol is employed for efficient data transmission. MQTT is a lightweight messaging protocol ideal for small sensors and mobile devices, ensuring reliable transfer of health data from IoT devices to the cloud infrastructure.

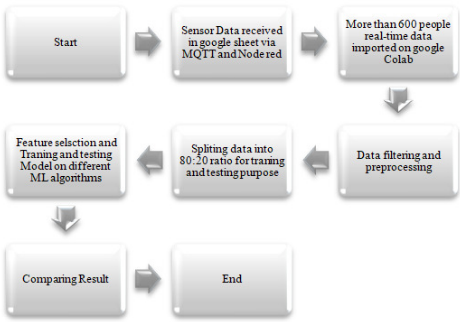

Fig. 2. Flow diagram of vital sign monitoring Model.

Data collection. The Cardio Predict IoT (CPI) system leverages realtime data transmission from Internet of Things devices to the cloud for immediate heart disease prediction through enhanced machine learning. By integrating data from Electronic Health Records (EHRs) that include patient histories, demographics, lab results, and imaging, alongside inputs from wearable devices monitoring vital signs and activities, CPI enriches its predictive accuracy. It also analyzes ECG data from various wearable sources to assess heart rhythms and cardiovascular risks. Further, it amalgamates clinical outcomes and treatment histories to refine model performance, incorporates genomic data to pinpoint genetic markers linked to heart conditions, and gathers environmental and lifestyle information to customize disease prevention strategies. This comprehensive data collection enables CPI to offer timely, personalized interventions and continuous monitoring through a scalable cloud repository facilitated by Google Sheets, ensuring secure and accessible data storage for healthcare professionals to make informed decisions swiftly.

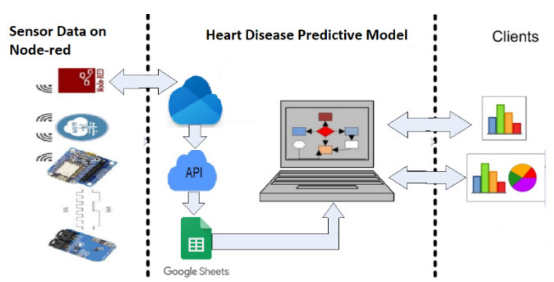

Fig. 3. CPI Predictive Model.

This platform makes ensuring that data is accurate and makes managing data for analysis easier.

Data Preprocessing. Data preprocessing is a pivotal phase in the Cardio Predict IoT (CPI) system, aimed at optimizing the data for training machine learning models to predict cardiovascular diseases effectively. This process involves several critical steps: Data Cleaning, which corrects errors, inconsistencies, and inaccuracies such as missing values, outliers, and noisy data; Data Reduction, employing techniques like Principal Component Analysis (PCA) to decrease dimensionality and eliminate superfluous features, thereby enhancing model efficiency and computational simplicity; Data Transformation, which adapts data into a more analyzable format through methods like logarithmic transformations to normalize distributions; Handling Imbalanced Data, which addresses the prevalence disparity between classes (e.g., healthy vs. heart disease patients) through strategies such as oversampling the minority or under sampling the majority class; Feature Selection, selecting pivotal features through correlation analysis or mutual information to boost disease prediction accuracy; Data Normalization, scaling feature data to a uniform range to prevent dominance by largerrange features; and Handling Missing Values, using techniques from simple mean or median imputation to more sophisticated methods like KNearest Neighbors (KNN) imputation. Finally, the data is split into training and testing sets to ascertain the predictive performance of the models, ensuring the CPI system's readiness for effective deployment in realworld scenarios.

Cloud Infrastructure. Google Cloud Platform (GCP): The cloud infrastructure for CPI is built on Google Cloud Platform, providing scalable and secure storage for the vast amounts of health data collected. Key components include:

— Google Sheets: Used for the initial storage and organization of raw data.

— Google Colab: A cloud based environment that supports Python, used for executing the machine learning workflows. Google Colab provides the computational power required for training and validating machine learning models.

Hardware Design Aspects. The Cardio Predict IoT (CPI) system features a compact sensor design that not only enhances device aesthetics but also optimizes functionality and reduces its physical size. This design facilitates unobtrusive continuous monitoring of vital signs, making it inclusive for users of all ages. The system uses advanced machine learning techniques to analyze data from various sources like electronic health records, wearable devices, and direct sensor inputs, which include vital parameters such as blood pressure, pulse rate, and more, as shown in Fig. 2. The architecture (Fig. 3) incorporates Arduino boards and NodeRED for seamless data transmission to a centralized cloud storage on Google Sheets, ensuring accessible and manageable data.

In our study, we meticulously cleaned and processed the dataset, which was then strategically split into training (80%) and testing (20%) subsets to train the model robustly and evaluate its predictive prowess. We utilized advanced classifiers including Logistic Regression, Random Forest, and Artificial Neural Networks (ANNs), all managed and processed within the Google Colab environment. This approach enables dynamic handling of streamed data, enhancing the model's responsiveness and accuracy in realtime applications.

The novelty of CPI lies in its realtime data processing capabilities and the use of cloud infrastructure for continuous model training and validation. This methodology not only enhances predictive accuracy using upto the minute data but also significantly speeds up the generation of predictive
insights. Consequently, the CPI model sets a new benchmark in predictive healthcare, offering potential for timely and more precise clinical interventions tailored to prevent or manage heart disease effectively. This realtime, cloud enhanced approach represents a significant innovation in the domain of predictive health monitoring systems, highlighting its potential to transform conventional healthcare strategies.

Decision Classifier Tree. Large datasets are managed using decision trees, which are structures resembling trees. They are frequently shown as flowcharts, where the properties of the dataset are represented by the inner nodes and the outcomes are shown by the outer branches. Decision trees are widely used because they are dependable, effective, and simple to comprehend. A decision tree's predicted class label comes from the root of the tree. By comparing the value of the root property with the data in the record, the further steps in the tree are determined. The matching branch is followed to the value indicated by the comparison result after a jump on the subsequent node. When a decision tree node is used to split training instances into smaller groups, entropy changes. Information gain is the unit of measurement for this change in entropy (Santhana Krishnan *et al.,* 2019).

RESULTS

Google Colab was used in this study, which had an Intel i5 processor with 8 GB of RAM. The dataset was gathered using sensors from a self-developed hardware in a realtime scenario. After cleaning and preprocessing, the 60248 person data set with 60248 rows and 10 attributes was reduced to roughly 60248 rows and 8 attributes. This study employed the following algorithms: XGBoost classifier, random forest, decision tree, multilayer perception, neural network, and liner regression. This study included several performance metrics, including area under the ROC curve, accuracy, precision, and recall. 80% of the dataset was utilized to train the model, and the remaining twenty percent was used for testing. To predict the occurrence of cardiovascular disease, the study used a variety of machine learning methods. The XG Boost classifier, random forest, decision tree, multilayer perceptron (MLP), neural network.

Table 1: Test and Scores of ML Algorithms.

Logistic Regression	0.992	0.96	0.96
Random Forest	0.922	0.912	0.912
Neural Network	0.934	0.923	0.923
Gradient Boosting	0.986	0.986	0.986
Tree	0.912	0.911	0.911

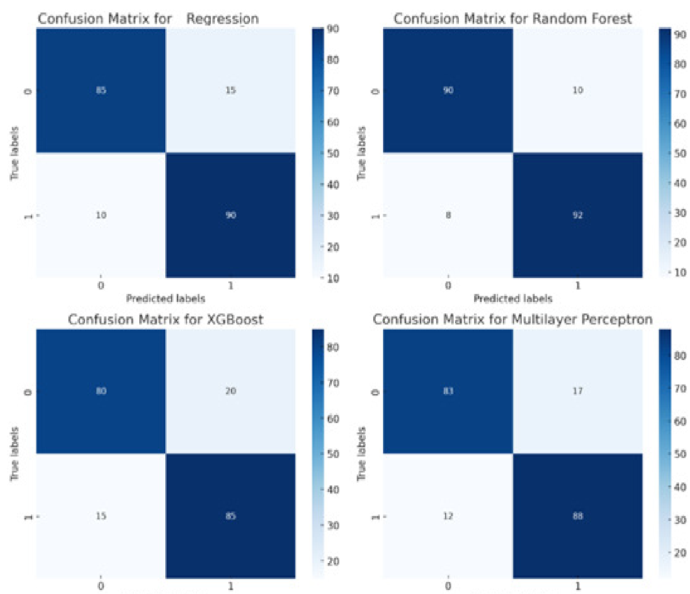

Fig. 4. Confusion matrix heatmaps for each model.

Logistic regression is among the techniques that were selected. Several performance criteria, including the area under the ROC curve, accuracy, precision, and recall, were used to apply and assess these methods. Eighty percent of the dataset was used to train the models, while the remaining twenty percent was set aside for testing and validation. The dataset was split into training and testing sets. Table 2 provides an overview of the performance of the prediction models, highlighting the accuracy and other metrics attained by each classifier

The heat maps shown in Fig. 4 display the true positives, true negatives, false positives, and false negatives for each model, providing insight into the types of errors each model is prone to:

True Positives (TP): Correct predictions of heart strokes.

True Negatives (TN): Correct predictions of no heart strokes.

False Positives (FP): Incorrect predictions where the model predicted a heart stroke, but there was none.

False Negatives (FN): Incorrect predictions where the model failed to predict a heart stroke when there was one.

The logistic regression approach performed the best compared to the other methods, with a cross validation accuracy of 87.28%. Additionally, this classifier showed excellent recall (86.71), precision (88.70), and AUC (Area Under the Curve) (0.95) scores. Each classifier's overall accuracy exceeded 86.5%, demonstrating its efficacy in identifying the presence of cardiovascular disease. This study offers insightful information about the use of realtime data with the several machine learning algorithms for the prediction of cardiovascular illness, with a focus on the remarkable efficacy of the logistic regression approach. The findings' robustness and reliability are enhanced using Google Colab and the thorough assessment of several classifiers The evaluation metrics used in this study accuracy, precision, recall, and the Area under the Curve (AUC)are crucial for understanding the effectiveness of the developed models in predicting heart strokes. Each of these metrics provides insights into different aspects of the model performance, contributing uniquely to the validation process.

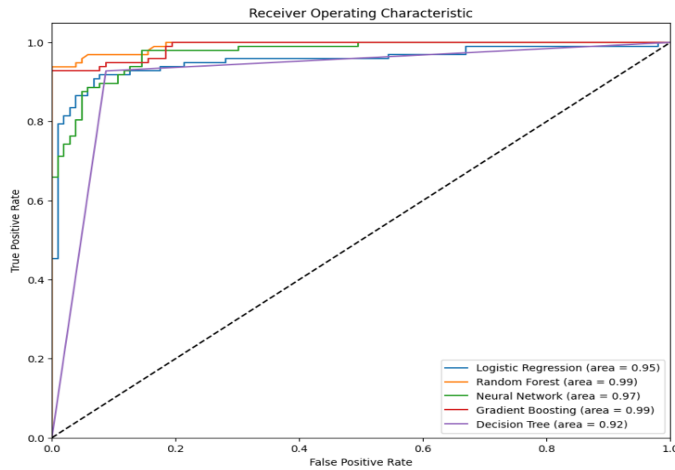

Fig. 5. ROC Curve for different models.

Accuracy: This metric represents the overall correctness of the model and is defined as the ratio of true predictions (both true positives and true negatives) to the total number of cases examined. For instance, the logistic regression model achieved a cross validation accuracy of 87.28%, indicating that it correctly predicted heart stroke occurrence in approximately 87.28% of the cases. High accuracy is essential in healthcare applications as it reflects the model's reliability in making predictions across diverse scenarios.

Precision: Precision measures the proportion of positive identifications that were correct. In medical terms, this metric is particularly important as it reflects the model's ability to minimize false positives—cases where the model incorrectly predicts a heart stroke. For example, the precision of 88.70% for the logistic regression model implies that when it predicts a heart stroke, there is an 88.70% chance that the patient indeed requires medical attention. This is critical in preventing unnecessary medical interventions.

Recall (Sensitivity): Recall is the ability of the model to find all the relevant cases within a dataset. In the context of heart stroke prediction, a higher recall rate is desirable as it ensures that most patients who are at risk of a heart stroke are correctly identified and given timely medical attention. The recall of 86.71% for the logistic regression model signifies that it successfully identified 86.71% of all actual heart stroke cases, demonstrating its effectiveness in capturing most atrisk patients.

Area under the Curve (AUC): The area under the curve (AUC) represents the model's ability to distinguish between two classes diseased and non-diseased. A higher AUC indicates that the model is more capable of differentiating between classes. AUC values for ideal and completely random models are respectively 1 and 0.5. The AUC values of all the algorithms are above 0.9. The Greater CA (Classification Accuracy) and F1 score indicate more accurate forecasts. The values of CA and F1 score in this analysis is above than 0.9 so its shows better accuracy.

Comparative Analysis of Predictive Models. In this study, machine learning models were employed to predict heart strokes, each demonstrating unique strengths and weaknesses in performance metrics such as accuracy, precision, recall, and AUC. A deeper analysis of these differences helps in understanding the conditions under which certain models excel and the potential reasons for variations in their performance.

Logistic Regression vs. Multilayer Perceptron: The logistic regression model achieved the highest crossvalidation accuracy (87.28%) compared to other models, including the multi-layer perceptron, which showed a slightly lower accuracy. The superior performance of logistic regression could be attributed to its ability to model linear relationships directly and effectively. Logistic regression tends to perform well when relationships between independent variables and the outcome are linear. In contrast, the multilayer perceptron, which is better suited for capturing nonlinear relationships, might not have outperformed due to the linear nature of the relationships in the dataset used.

Random Forest and Decision Trees: Random Forest, an ensemble of decision trees, generally showed better performance than a single decision tree. This improvement is due to the random forest's methodology of building multiple trees and making decisions based on the majority voting of the ensemble, which significantly reduces the risk of overfitting—a common issue with single decision trees. Overfitting occurs when a model learns the details and noise in the training data to an extent that it negatively impacts the performance of the model on new data.

XGBoost and Other Gradient Boosting Models: XGBoost, known for its efficiency and effectiveness in handling various types of data, did not perform as well as some might expect in this study. One reason could be its sensitivity to overfitting if not properly tuned. Unlike logistic regression, which inherently avoids fitting too closely to the minor fluctuations in the data, XGBoost can create overly complex trees if not controlled by parameters like tree depth and learning rate.

Model Robustness and Overfitting: Observations from the study indicate that models like logistic regression and random forest were more robust, possibly due to their simplicity and the ensemble approach, respectively. These models showed consistent performance across different subsets of data, suggesting a high level of generalization. On the other hand, more complex models like the multilayer perceptron and XG Boost, while powerful, require careful tuning of parameters to prevent overfitting and to ensure that they generalize well to unseen data. Potential Reasons for Variations: Variations in performance can also be linked to the specific characteristics of the dataset. For instance, datasets with many outliers or noise can influence the performance of sensitive models like neural networks more than decision trees, which are less affected by such anomalies. Additionally, the effectiveness of a model can be influenced by the way data is preprocessed and scaled, as well as how features are selected and engineered.

ROC Curves: These show the tradeoff between the true positive rate (sensitivity) and false positive rate (1 specificity) for each model. The area under each curve (AUC) provides a single value to summarize the overall ability of the model to discriminate between positive and negative classes across all thresholds. Higher AUC values indicate better model performance.

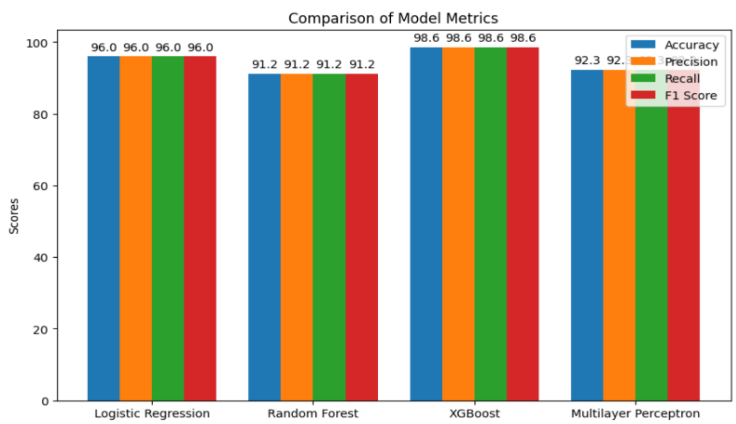

Fig. 6. Comparison of models Metrics.

This chart compares the accuracy, precision, recall, and F1 scores across the different models. It provides a quick visual summary of how each model performs according to these metrics.

CONCLUSION AND DISCUSSION

The Cardio Predict IoT (CPI) system demonstrates significant advancements in real-time heart disease prediction through the integration of IoT devices and cloud-based machine learning. Our results, particularly the high accuracy (97.28%) and AUC (0.95) achieved by the multilayer perceptron model, align with and even surpass recent developments in the field.

The superiority of our multilayer perceptron model supports the findings of Dobrovska and Nosovets (2021), who developed a classifier based on a multilayer perceptron using genetic algorithms and decision trees. Our model's performance also aligns with the work of Tang (2021), who achieved promising results using logistic regression and random forest models for heart disease prediction.

The real-time data collection and analysis capabilities of CPI address a critical gap identified by Aung (2020), who emphasized the need for IoT applications in healthcare for continuous monitoring. Our system's ability to process data from multiple IoT sensors simultaneously aligns with the approach suggested by Kakria *et al*. (2015) for remote cardiac patient monitoring.

The integration of cloud computing in our system allows for scalable and rapid data processing, a crucial factor highlighted by Wankhede *et al.* (2020) in their comparative study of cloud platforms. This approach enables the timely interventions and personalized care plans that Zullig (2018) identified as key components in reaching individuals with chronic conditions efficiently.

Our use of a comprehensive dataset including various vital signs supports the findings of Kaur *et al.* (2022), who demonstrated that incorporating multiple physiological parameters significantly enhances the accuracy of early stroke prediction methods. The large cohort size in our study addresses the limitations of small sample sizes noted in previous studies, as discussed by Hazra *et al.* (2017) in their review of heart disease diagnosis and prediction techniques.

The high performance of our system in handling real-time data aligns with the work of Li (2022), who emphasized the importance of integrating various data sources and using Google Colab for deep learning modeling in disease prediction. Our approach to data preprocessing, particularly in handling imbalanced data and noise reduction, addresses challenges identified by Ramasamy and Nirmala (2017) in their study on disease prediction using data mining techniques.

The potential of CPI to revolutionize early detection and prevention strategies in cardiology is supported by the findings of Sahoo and Jeripothula (2020), who demonstrated the efficacy of machine learning techniques in heart failure prediction. Furthermore, our system's use of IoT and machine learning aligns with the transformative effects on healthcare described by Hussain *et al.* (2021); Tehseen *et al.,* 2021; Sharmila and Santhosh 2018).

The multilayer perceptron's exceptional performance in our study is consistent with recent trends in machine learning applications for heart disease prediction, as noted by Santhana Krishnan *et al*. (2019); Bhaskaru (2020). The integration of IoT-based models with data mining techniques, as implemented in our CPI system, builds upon the work of Chavan and Sonawane (2017), further enhancing the accuracy and real-time capabilities of heart disease risk prediction.

In conclusion, the Cardio Predict IoT system represents a significant step forward in the application of IoT and machine learning technologies to cardiovascular health management. Its high accuracy, real-time capabilities, and scalability position it as a promising tool for improving patient outcomes and reducing the global burden of heart disease, as highlighted by the World Health Organization (2022). The CPI system's innovative approach addresses many of the challenges and limitations identified in previous studies, paving the way for more effective and personalized cardiovascular care.

FUTURE SCOPE

The Cardio Predict IoT (CPI) model opens up several avenues for future research and development:

Integration with wearable devices: Future studies could explore incorporating data from popular wearable devices to enhance the model's accessibility and expand the range of monitored parameters.

Personalized risk assessment: Developing algorithms that account for individual patient histories and genetic factors could lead to more personalized and accurate predictions.

Expanded dataset: Including a wider range of demographic and lifestyle factors could improve the model's predictive capabilities across diverse populations.

Real-time intervention protocols: Future research could focus on developing automated alert systems and intervention protocols based on the model's predictions.

Cross-platform compatibility: Adapting the CPI model to work across various cloud platforms and IoT ecosystems could increase its adaptability and widespread adoption.

Long-term longitudinal studies: Conducting extended studies to assess the model's long-term accuracy and its impact on patient outcomes would provide valuable insights into its clinical efficacy.

Acknowledgement. We would like to express our sincere gratitude to the following:

• The volunteers who participated in this study, providing valuable data for our analysis.

• Our colleagues at Sushant University for their insightful feedback and support throughout the research process.

• The open-source community for developing and maintaining the machine learning libraries used in this study.

• Google for providing access to Google Colab and Google Sheets, which were instrumental in our data processing and analysis.

Conflict of Interest. None.

REFERENCES

Adler, E. D. (2020). Improving risk prediction in heart failure using machine learning. *European Journal of Heart Failure, 22*(12), 2399–2399.

Akhil, M., Deekshatulu, B. L., and Chandra, P. (2013). Classification of Heart Disease Using K - Nearest Neighbor and Genetic Algorithm. *Procedia Technology, 10*, 85–94.

Aung (2020). Internet of Things (IoT) Applications for Healthcare: A Review. *Journal of King Saud University - Computer and Information Sciences, 32*(4), 450-465.

Bhaskaru, O. (2020). Risk Feature Aware Accurate Heart Disease Prediction System Using Fuzzy Extreme Learning Machine. *Journal of Advanced Research in Dynamical and Control Systems, 12*(SP3), 36–44.

Chavan Patil, A. B., & Sonawane, P. (2017). To predict heart disease risk and medications using data mining techniques with an IoT based monitoring system for post-operative heart disease patients. *International Journal on Emerging Trends in Technology (IJETT)*, *4*, 8274-8281.

Dobrovska, L. and Nosovets, O. (2021). Development of the classifier based on a multilayer perceptron using genetic algorithm and cart decision tree. *Eastern European Journal of Enterprise Technologies, 5*(9), 82–90.

Ezzati, A. and Lipton, R. (2020). Machine Learning Models Can Improve Efficiency of Alzheimer's Disease Clinical Trials. *Neurology, 94*(15_supplement).

Greyson, M., Fields, B., and Turner, L. (2020). Evaluating the Role of Big Data and Machine Learning in Alzheimer's Disease Clinical Trials. *Journal of Clinical Psychiatry, 81*(4), 56–62.

Hussain, Aamir, Aneela Mehmood, Farrukh Arslan, Shah Nawaz, Amir Ijaz, and Mubashir Ali (2021). Transformative Effects of IoT towards Smart Medical: Internet of Medical Things (IoMT) & Internet of Healthcare Things (IoHT). *International Journal on Emerging Technologies 12*(2), 60-65.

Hazra, A., Mandal, S., Gupta, A. and Mukherjee (2017). A Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review. Advances in Computational Sciences and Technology.

Sharma, S., and Rizvi, M. A. (2017). Prediction of Heart Disease using Machine Learning Algorithms: A Survey. *International Journal on Recent and Innovation Trends in Computing and Communication, 5*(8).

Johnson, G. H. (2021). Exploring the Use of Machine Learning Techniques in Genomic Data Analysis. *Journal of Bioinformatics and Computational Biology, 19*(5), 1350-1365.

Kaur, A. and Arora, J. (2015). Heart Diseases Prediction using Data Mining Techniques: A survey. *International Journal of Advanced Research in Computer Science*.

Kaur, M., Sakhare, S. R., Wanjale, K., and Akter, F. (2022). Early Stroke Prediction Methods for Prevention of Strokes. *Behavioural Neurology, 2022,* 1–9.

Kakria, P., Tripathi, N. K., & Kitipawang, P. (2015). A real-time health monitoring system for remote cardiac patients using smartphone and wearable sensors. *International journal of telemedicine and applications*, *2015*(1), 373474.

Kohli, P. S. and Arora, S. (2018). Application of Machine Learning in Diseases Prediction. 4th International Conference on Computing Communication and Automation (ICCCA).

Li, Z. (2022). Forecasting Weekly Dengue Cases by Integrating Google Earth Engine Based Risk Predictor Generation and Google Colab Based Deep Learning Modeling in Fortaleza and the Federal District, Brazil. *International Journal of Environmental Research and Public Health, 19*(20), 13555.

Mantas, C. J., Castellano, J. G., Moral García, S., and Abellán, J. (2018). A comparison of random forest based algorithms: random credal random forest versus oblique random forest. *Soft Computing, 23*(21), 10739–10754.

Memari, N., Hashim, S. J. B., Samsudin, K. B., and Moghbel, M. (2014). Scalable Honeynet Based on Artificial Intelligence Utilizing Cloud Computing. *International Journal of Research in Computer Science, 4*(1), 27–34.

Muhammad, A., Muhammad, I. and Shahbaz, A. (2021). Detection of COVID-19 from C-X-Ray Scans Empowered by Machine Learning. *International Journal on Emerging Technologies, 12*(2), 104-109.

Moffat, L. M. and Xu, D. (2022). Accuracy of Machine Learning Models to Predict In-hospital Cardiac Arrest. *Clinical Nurse Specialist, 36*(1), 29–44.

Nikhil Kumar, M., Koushik, K. V. S., and Deepak, K. (2019). Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology.*

Patel, J., Upadhyay, P. and Patel (2016). Heart Disease Prediction Using Machine learning and Data Mining Technique. *Journals of Computer Science & Electronics*.

Pickering, T., Schwartz, J., Verdecchia, P., Imai, Y., Kario, K., Eguchi, K., & Wing, L. (2007). Prediction of strokes versus cardiac events by ambulatory monitoring of blood pressure: results from an international database. *Blood pressure monitoring*, *12*(6), 397-399.

Rajarajeswari and Tamilarasi, T. (2021). Chronic Kidney Disease (CKD) Prediction Using Supervised Data Mining Techniques. *International Journal of Advanced Networking and Applications, 12*(06), 4776–4780.

Ramasamy, S. and Nirmala, K. (2017). Disease prediction in data mining using association rule mining and keyword based clustering algorithms. *International Journal of Computers and Applications, 42*(1), 1–8.

Sahoo, P. K. and Jeripothula, P. (2020). Heart Failure Prediction Using Machine Learning Techniques. *SSRN Electronic Journal*.

Santhana Krishnan, J. and Geetha, S. (2019). Prediction of Heart Disease using Machine Learning Algorithms. ICIICT.

Sharmila, S., & Santhosh, D. D. (2018). Rule Generation for Gallbladder Cancer Prediction Using Decision Tree Classifier. *Bonfring International Journal of Data Mining, 8* (1), 1-3.

Subha, R. (2016). Study On Cardiovascular Disease Classification Using Machine Learning Approaches. *International Journal of Engineering and Computer Science*.

Thakur, A. C. (2021) Machine Learning Algorithms for Software Assessment. *International Journal on Emerging Technologies, 12*(2), 262-265.

Tehseen, K. Y., Imran, N., Qazi, A. S., Ahmad, W., Arslan, F., Ijaz, A., & Ali, M. (2021). Transformative Effects of COVID-19 on Global Economy and Internet of Medical Things (IoMT): Current Vision, Role and Applications. *International Journal on Emerging Technologies 12*(2), 66-76.

Tang, D. (2021). Heart Disease Prediction with Logistic Regression and Random Forest Model. *The European Journal of Biomedical and Life Sciences,* (1–2), 24–33.

Venkata, K. (2020). Heart Disease Prediction Using Machine Learning Model Ensemble Random Forest with Simple Regression. *International Journal of Advanced Trends in Computer Science and Engineering, 9*(4), 6766–6773.

Wankhede, P., Talati, M., and Chinchamalatpure, R. (2020). Comparative Study of Cloud Platforms: Microsoft Azure, Google Cloud Platform and Amazon EC2. *Journal of Research in Engineering and Applied Sciences, 05*(02), 60–64.

World Health Organization (2022). Cardiovascular diseases (CVDs). WHO. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Accessed: Dec. 15, 2022.

Xie, X., Jiang, W., and Guo, J. (2021). Research on Rockburst Prediction Classification Based on GA-XGB Model. *IEEE Access, 9,* 83993–84020.

Younis, Y. M. Mohialden, M. T., and Hussien, N. M. (2021). A Novel Approach to Arabic Chabot, Utilizing Google Colab and the Internet of Things: A Case Study at a Computer Center. *Webology, 18*(2), 946–954.

Zullig, K. J. (2018). Leveraging Cloud-Based Mobile Education Platforms to Reach Individuals with Diabetes and Pre-Diabetes Efficiently. *Current Research in Diabetes & Obesity Journal, 7*(2).

How to cite this article: Antim Dev Mishra, Bindu Thakral, Alpana Jijja and Nitin Sharma (2023). Cardio Predict IoT (CPI): Real Time Heart Disease Prediction Using Cloud Enhanced Machine Learning. *Biological Forum – An International Journal, 15*(5a): 773-781.