# Smart Health Monitoring for Predicting Heart Disease using IoT-Fog-Cloud Computing Model

**Hafsat Jalo Suleiman**
Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Johor Malaysia | School of Engineering Technology, Department of Computer Engineering Technology, Gombe State Polytechnic Bajoga, Nigeria
hafsatsuleimanjalo@gspb.edu.ng (corresponding author)

**Isredza Rahmi A. Hamid**
Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia
rahmi@uthm.edu.my

**Oyebayo Ridwan Olaniran**
Department of Statistics, Faculty of Physical Sciences, University of Ilorin, Nigeria
olaniran.or@unilorin.edu.ng

*Received: 27 December 2024 | Revised: 2 February 2025 | Accepted: 5 February 2025*

 *DOI: https://doi.org/10.48084/etasr.10048*

## ABSTRACT

**Cloud computing enables access to various resources online, supporting services across numerous sectors. However, meeting real-time demands in IoT-based computing is challenging due to high latency issues. This is particularly problematic for low-latency applications, such as health monitoring and traffic surveillance, which require fast processing of large datasets. Performance drop occurs when data moves between central databases and cloud data centers. Edge and fog computing have emerged as new solutions to address this. These models place computing resources closer to users, significantly reducing latency and energy consumption while improving data processing efficiency. This paper presents a prediction system utilizing a fog-cloud framework, combining machine learning and deep learning with wearable IoT devices for real-time cardiovascular disease prediction. The system is trained using cardiovascular data from Gombe State, Nigeria, and evaluated based on energy consumption, precision, accuracy, recall, F1 score, and AUC. The proposed Optimized Naïve Bayes Random Forest (ONBRF) model offers a reliable and energy efficient approach to predicting heart disease.**

***Keywords-IoT; heart disease; fog computing; cloud computing; accuracy; energy consumption***

## I. INTRODUCTION

The rapid expansion of the Internet of Things (IoT) is particularly noticeable in the medical field [1]. The global IoT market is experiencing rapid growth, with forecasts indicating a compound annual growth rate of 7.4% from 2020 to 2025, leading to a projected market value of USD 110.6 billion [2]. The healthcare sector, in particular, is witnessing significant advances in IoT adoption. Telemonitoring within the home of patients is expected to be a key application area for IoT [3]. There is great promise in improving healthcare quality and lowering costs through remote health monitoring based on IoT, allowing early disease detection and preventive care [4]. Health monitoring, chronic disease management in real-time, early diagnosis, and emergency healthcare are some applications that IoT-capable health networks can support. Moreover, IoT devices lead to health records and medical servers, a gateway to providing on-demand real-time healthcare services [4].

The health sector is gradually transforming due to advances in the IoT, aiming to streamline the patient process of submitting health data and receiving prescriptions from physicians online. Given the prevailing trend toward centralization, cloud computing represents an ideal network framework for IoT applications. Services such as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) play crucial roles in this cloud computing ecosystem. Although it has the potential to support efficient data analysis, processing speed, and reliability in IoT healthcare, cloud computing cannot fully address the requirements of connecting healthcare systems due to the demands of the vast amount of medical information generated by these devices [5]. This overload results in problems such as high latency, network congestion, and high energy consumption [6]. Consequently, the performance of IoT healthcare systems is reduced due to the incorporation of cloud computing [7]. Cloud-based IoT applications have been successfully implemented in several areas of medical testing. However, a significant response time is observed when using current cloud-based Internet of Healthcare Things (IoHT) systems, along with scalability issues. In [8], a framework for IoHT was proposed, which integrated deep learning and fog computing to address these specific challenges, ensuring timely diagnosis support and improved analytics for cardiovascular patients. Simulation results showed that this model improved energy consumption, bandwidth utilization, access time, and processing speed for different fog IoHT scenarios.

Fog computing serves as a solution to some of the limitations posed by cloud computing, particularly in healthcare applications that require real-time processing. By offering cloud-like services at the network edge, fog computing improves mobility, privacy, and energy efficiency while effectively addressing latency issues that challenge traditional cloud infrastructures [9-10]. Integrating IoT-fog-cloud computing into healthcare has been claimed to mitigate problems such as network congestion and long response times, which is particularly beneficial for detecting and managing critical diseases such as cardiovascular disease [10]. Despite its advantages, fog computing has limitations, such as lower storage capacities, necessitating connection to cloud services for long-term data storage and more computationally intensive tasks.

Cardiovascular diseases are the leading cause of mortality worldwide, with early identification crucial for effective intervention [11]. Advanced machine learning techniques, including support vector machines and random forest models, have shown promise in predicting heart disease by analyzing extensive medical datasets [12-14]. Recent studies have emphasized the importance of integrating machine learning with IoT and fog computing to enhance real-time monitoring systems, thus improving diagnostic accuracy and reducing energy consumption [1,8,15]. Innovations in wearable technologies and IoT-based monitoring systems reflect ongoing efforts to create efficient cardiovascular prediction models, culminating in the proposed Smart Health Monitoring for Heart Disease (SHMHD) model that leverages these advanced technologies for real-time prediction of cardiovascular issues.

## II. MATERIAL AND METHODS

This paper introduces the SHMHD framework, which consists of three fundamental components: (i) the IoT layer, (ii) the fog layer, and (iii) the cloud layer.

### *A. Internet of Things Layer (IoT Layer)*

The proposed smart healthcare system integrates multiple hardware devices with software tools, allowing smooth, organized, end-to-end connectivity for faster and more precise results. For further analysis and monitoring, patient data is transmitted to connected gateway devices. The proposed predictive healthcare model has two main sources of data: physiological information and clinical observations. The collected information includes metrics such as blood pressure, blood sugar levels, respiration rate, heart rate, blood oxygen saturation, and cholesterol levels, as well as electrocardiograms, electroencephalograms, and electromyograms. Additionally, a comprehensive medical history, which serves as a crucial indicator to predict diseases, is derived from electronic clinical data that includes lab results, clinical notes, prescribed medications, and allergy records, all securely stored in a cloud-based system. These data are transmitted to relevant nodes within the fog layer, such as gateway devices, for further analysis through wireless technologies such as Bluetooth and Zigbee, as outlined in [16], allowing connectivity between IoT devices.

### *B. Dataset*

This study collected the heart disease dataset from the Federal Teaching Hospital, Gombe, Gombe State, Nigeria. It is a prospective study with clinical and physiological data collected electronically from 106 patients aged 13 to 89. Table I summarizes detailed descriptions of the clinical and physiological features collected.

TABLE I. DATASET DESCRIPTION

| Feature | Description | Type |
|---|---|---|
| ID | A six-digit patient's unique hospital or unit number | Numeric |
| AGE | Diagnosed patients' age (in years) | Numeric |
| SEX | Gender of diagnosed patients with two values: Value 1 indicates male, Value 0 indicates female | Nominal |
| BP | Upon hospital admission, the patient's blood pressure is at rest (measured in mmHg). | Numeric |
| PR | Pulse rate or heartbeats of patients (in bpm on admission to the hospital) | Numeric |
| ECG | Patient's electrocardiogram status with two values: Value 1 indicates abnormal ECG scan, Value 0 indicates normal ECG scan. | Nominal |
| ECHO | Patient's echocardiogram status with two values: Value 1 indicates abnormal ECHO status, and Value 0: indicates normal ECHO status. | Nominal |
| TCL | Patient's total cholesterol (in mmol/L) | Numeric |
| HDL | High-density lipoprotein or good cholesterol (in mmol/L) | Numeric |
| LDL | Low-density lipoprotein or bad cholesterol (in mmol/L) | Numeric |
| TGL | Triglycerides (in mmol/L) | Numeric |
| FBS | Fasting blood sugar of patients (measured in mmol/L) | Numeric |
| NUM | Diagnosis of CHD contains two values: Value 1 (CHD): the presence of CHD complications; Value 2 (NON-CHD): absence of CHD complications | Nominal |

This research adhered to the regulations set by the National Library of Nigeria depository law and the Helsinki Declaration regarding safeguarding personal information. It received approval from the ethics committee at the Federal Teaching Hospital in Gombe, Gombe State, Nigeria, on May 19, 2021, under reference number NHREC/25/10/2013.

### *C. Fog Layer*

IoT devices act as fog devices by collecting data from various sensors and transmitting it to fog nodes for processing within the fog bus system. This system includes a data manager that handles incoming requests and a resource manager that organizes patient data according to urgency for the cloud integrator, facilitating data flow between the fog and cloud layers. In addition, the resource monitor regularly evaluates the status of resources and manages data processing priorities.

### D. Deep Learning Module

The deep learning module utilizes Convolutional Neural Networks (CNNs) to handle data filtering and processing. It autonomously makes decisions based on the processed information and ongoing training data. The Convolutional Layer is crucial for extracting features. Each layer has multiple convolutional kernels. This layer convolves the input matrix by using these kernels. Consider the input matrix as $X$, where $X=\{x_{i,j}|i=1,2,\dots,I,\ j=1,2,\dots,J\}$. When using this process for heart prediction, $I$ represents the number of patients, and $J$ represents the number of features. The convolution kernel is represented as $W$, where $W=\{w_{m,n}|m=0,1,\dots,F-1,\ n=0,1,\dots,\ F-1\}$. $F$ represents the size, which is the same for both the convolutional kernel's width and height. Equation (1) describes the convolutional layer.

$$a_{i,j}=f\left(\sum_{m=0}^{F-1}\sum_{n=0}^{F-1}w_{m,n}x_{i+m,j+n}+b\right)$$
$$i=1,2,\dots,I,\quad j=1,2,\dots,J \qquad (1)$$

### E. Machine Learning Module

#### 1) Naïve Bayes (NB) Model

NB is a probabilistic classifier that relies on Bayes' theorem and operates under the assumption that the features are independent. Laplace smoothing (also known as additive smoothing) is used to handle the model's zero probabilities. Laplace smoothing ensures that it does not get a zero probability when a particular feature-category combination does not appear in the training set. The Bayes theorem states that [17-21]:

$$P(C|X)=\frac{P(X|C)P(C)}{P(X)}$$

where $P(C|X)$ represents the probability of class $C$ occurring given the feature vector $X$, $P(X|C)$ indicates the probability of observing the feature vector $X$ if class $C$ is true, $P(C)$ denotes the probability of class $C$ before considering the feature vector $X$, and $P(X)$ is the overall probability of the feature vector $X$.

The NB classifier simplifies the computation by assuming that features are conditionally independent given the class. The classification rule is given by:

$$\hat{C}=\text{argmax}_C\, P(C)\prod_{i=1}^{n}P(x_i|C)$$

where $x_i$ are the features in the feature vector $X$. Laplace smoothing addresses the issue of zero probabilities in NB classification. If a specific feature-category pair is absent from the training data, its probability estimates default to zero, which can disrupt the accuracy of the overall probability calculations. Laplace smoothing adds a small constant $\alpha$ to each count to ensure that the probability is never zero. The formula for calculating the smoothed probabilities is as follows:

$$P(x_i|C)=\frac{n_{x_i,C}+\alpha}{n_C+\alpha|V|}$$

where $n_{x_i,C}$ is the count of feature $x_i$ in class $C$, $n_C$ is the total count of all features in class $C$, $\alpha$ is the smoothing parameter (usually 1), and $|V|$ is the number of possible feature values.

#### 2) Random Forest (RF) Model

RF is an ensemble classification algorithm that is based on the aggregation of several Decision Trees (DT) [18]. In DT, classification is achieved by partitioning the entire feature space $X$ into regions using a top-down procedure. Final DT prediction is achieved using:

$$\text{argmax}_C\, P\left(C|R_{x_i}\right)$$

where $R_{x_i}$ is the region rule for a specific feature $x_i$ and $P\left(C|R_{x_i}\right)$ is the probability of class $C$ in the region $R_{x_i}$. However, due to the tree terminal splitting procedure in DT, different fitting iterations often yield different final predictions, thus making the prediction of DT unstable. To address this, RF was introduced so that a random subset of feature space $X$, denoted by $mtry$, is selected and then fitted on $B$ DTs. The final prediction from RF is estimated by:

$$\hat{C}_{RF}=\text{argmax}_C(\{DT\}_1,\{DT\}_2,\dots,\{DT\}_B)$$

where

$$\{DT\}_i=\text{argmax}_C\, P\left(C|R_{x_i};x_i\in mtry\right)$$

is the DT model based on a random subset of features. RF's predictive performance strongly relies on an optimal $mtry$ and a sufficient number of trees $B$ that make the forest.

#### 3) Particle Swarm Optimization (PSO)

PSO is a computational technique modelled after the collective behavior observed in bird flocks and fish schools. This method is employed to find optimal solutions for various functions. The process starts by setting up a group of particles, each symbolizing a potential solution, with designated positions and velocities. Initial parameters such as the number of particles, inertia weight $w$, cognitive coefficient $c_1$, and social coefficient $c_2$, are set. Each particle's fitness is evaluated using the objective function to determine how close its position is to the optimal solution. The particles subsequently adjust their velocities and locations by referencing their individual best positions ($pBest$) and the best position found by any particle ($gBest$). This involves using random numbers ($r_1$ and $r_2$) to simulate social and cognitive behaviors. The fitness of the particles is re-evaluated at their new positions, and the personal and global bests are updated accordingly. This process is repeated until a stopping criterion is met, such as reaching a maximum number of iterations, achieving a satisfactory fitness level, or observing negligible improvement over iterations.

#### 4) Naïve Bayes Random Forest (NBRF)

The proposed NBRF is a two-stage algorithm. In the first stage, predictions are made using NB, and in the second stage, predictions are made using RF. The predictions and features from the NB stage are used as input for the RF stage.

#### 5) Optimized Naïve Bayes Random Forest (ONBRF)

The machine learning module is finalized by performing the analytical prediction with Optimized NBRF (ONBRF) for heart disease problems using PSO. The first step involves optimizing the parameters of NB and RF. The NB parameter to be optimized is the Laplace smoothing $\alpha$. It is critical to tune the Laplace smoothing parameter $\alpha$ because it helps to balance the bias-variance trade-off of the NB model. The parameters for optimizing the RF model include the number of trees ($B$), and the subsample size ($mtry$). The fitness functions to be optimized correspond to the prediction loss $f_1(\Omega)$ and latency $f_2(\Omega)$ given by:

$$f_1(\Omega) = -\frac{1}{N}\sum_{i=1}^{N}[C_i \log(p_i) + (1 - C_i)\log(1 - p_i)]$$

and:

$$f_2(\Omega) = T_{stop} - T_{start}$$

where $\Omega = \{\alpha, B, mtry\}$, $C_i$ is the actual class for an instance $i$, $p_i$ is the fitted probability of NBRF, and $T_{stop}$ and $T_{start}$ are the start and stop times for the NBRF algorithm. The multiobjective minimization function is given by:

$$\min_{\Omega} \mathrm{F} = [f_1(\Omega), f_2(\Omega)]$$

subjected to: $\alpha \geq 1$ , $100 \leq B \leq 500$ , $1 \leq mtry \leq \lfloor\sqrt{k}\rfloor$ , $\alpha, B, mtry > 0$ , and $\alpha, B, mtry \in \mathbb{Z}$ . The optimization constraints are based on the minimum threshold for the Laplace smoothing and the recommended number of trees and $mtry$. The parameter $k$ in the upper bound of $mtry$ is the number of features in the feature space $X$.

The optimized parameters $\Omega_o$ = ($\alpha_o$, $B_o$, and $mtry_o$) are supplied to the SHMHD framework for diagnosing cardiovascular disease patients. These parameters specifically help reducw final prediction loss and processing time, which eventually leads to lower energy consumption using PSO:

$$\begin{aligned} v_{id}(\Omega + 1) = {} & w \times v_{id}(\Omega) + \\ & c_1 \times r_1() \times [pbest_{id}(\Omega) - \mathbf{F}_{id}(\Omega)] + \\ & c_2 \times r_2() \times [gbest_{id}(\Omega) - \mathbf{F}_{id}(\Omega)] \end{aligned} \quad (2)$$

### *F. Cloud Layer*

The third element is the cloud data center, designed to improve data processing, provide effective storage solutions, and transmit information to doctors, hospitals, and patients. The proposed model enhances resource utilization within the cloud layer to manage input data sizes that exceed the usual limits. In the final stage of the SHMHD framework, the proposed health monitoring model is assessed and compared against current IoT-fog-cloud models for cardiovascular diseases.

### *G. Evaluation Metrics*

The evaluation metrics are divided into two major parts namely predictive classification metrics and energy consumption metrics. In addition, the model performance evaluation was carried out through a 10-fold cross-validation of the dataset. The performance evaluation metrics results were obtained from the one-tenth holdout test dataset.

#### *1) Predictive Classification Metrics*

The predictive classification metrics consist of the following:

- Accuracy assesses how well a classifier performs. It is determined by dividing the number of correct cardiovascular disease predictions by the overall number of predictions made by the classifier:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad (3)$$

- Recall is used to evaluate the performance of the IoT-Fog-Cloud health monitoring model. It assesses how well the model can accurately recognize all positive cases in the dataset:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

True Positives (TP) refers to the count of actual positive cases the model has accurately identified, and False Positives (FP) represents the count of actual negative cases the model mistakenly classified as positive.

- The F1 score is a widely used measure for assessing the effectiveness of a binary classification model. It merges the model's precision and recall into a unified metric that represents both dimensions of its effectiveness. F1 can be calculated as the harmonic mean of precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

- Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is a summary statistic that indicates the overall effectiveness of a classification model. It involves the ROC curve, which illustrates the True Positive Rate (TPR) versus the False Positive Rate (FPR) across various threshold levels. TPR measures the fraction of actual positives accurately recognized by the model, whereas the FPR represents the fraction of negatives incorrectly classified as positives:

$$AUC_{ROC} = \int_0^1 TPR(FPR)\, dFPR \quad (6)$$

where TPR(FPR) is the TPR at a given FPR value, and the integral is calculated over the range of FPR values from 0 to 1.

#### *2) Energy Consumption Metrics*

The following evaluation metrics measure energy consumption:

- Latency ($L$) denotes the delay experienced in delivering crucial services. For scenarios where uninterrupted or immediate service is critical, real-time stream processing must be implemented through fog computing to reduce delays. This is particularly important for latency-sensitive applications such as complex event processing or stream mining.

$$L = Response_{time} - Request_{time} \quad (7)$$

- In a fog computing setup, Energy Consumption ($EC$) occurs as resources provide essential services or forward requests to the cloud for further data processing:

$$EC = Power \times Processing_{time} \quad (8)$$

## III. IMPLEMENTATION OF SHMHD

Algorithm 1 summarizes the flow process of SHMHD across the three layers. The procedure starts with Patient Clinical/physiological Data (PCD) collection using a Wearable Device *(*WD*)*, stored in an IoT Layer Device (ITLD) through a Bluetooth Gateway Device (BGD). The data from the PCD is subsequently sent to the Fog Computing Layer (FCL) through a secure internet network protocol. Here, a local Master Computing System (MCS) was used as the FCL. Within the FCL, the patients' attributes, such as age, gender, systolic blood pressure, etc., are pre-processed and transformed using CNN before being sent to the machine learning segment where prediction occurs. In the machine learning segment of FCL, the ONBRF procedure is applied to train the transformed PCD. The final prediction function is transferred from the FCL using R Studio v.2.9.2 as the cloud integrator to the R Shiny Web Cloud Layer (RSWCL). RSWCL stores the ONBRF procedure such that when a patient or physician requests the heart disease status, SHMHD outputs a prediction and health advice based on the result. The web application can be accessed at [22].

```
Algorithm 1: SHMHD Process Flow
1: Start: Define the parameters: PCD, WD,
   ITLD, BGD, FCL, MCS, CNN, ONBRF,
   and RSWCL.
2: PCD Collection:
   Collect PCD using a WD.
   Transfer and Store PCD in ITLD via BGD
3: Data Transfer to FCL:
   Transfer PCD from ITLD to FCL using a
   secure internet networking protocol.
   Use local MCS as FCL.
4. Data Pre-processing and Transformation
   in FCL:
   Pre-process and transform patient
   attributes (e.g., age, gender, systolic
   blood pressure) using CNN.
5: Machine Learning in FCL:
   Apply ONBRF procedure to train
   transformed PCD.
   Transfer final prediction function from
   FCL to RSWCL using R Studio v.2.9.2
   (cloud integrator).
6: Storage and Prediction in RSWCL Cloud
   layer:
   Store ONBRF procedure in RSWCL.
   On request from the patient or
   physician, retrieve heart disease
   status prediction based on supplied
   attributes.
   Output prediction and health advice
   based on the result.
7: End
```

## IV. RESULTS

Table II shows the performance metrics for predicting cardiovascular heart disease using different models, indicating that the proposed ONBRF model achieved the highest overall accuracy at 92.3%, followed closely by the RF model at 90.4%. The NB model had the lowest accuracy at 64.9%. In terms of precision, the proposed ONBRF and RF also led with 95.3% and 93.1%, respectively. NB achieved a perfect precision of 100%, although with a significantly lower recall (55.8%), suggesting that it is highly specific but not sensitive in detecting the presence of cardiovascular heart disease.

TABLE II. MEAN METRICS FOR PREDICTING CARDIOVASCULAR HEART DISEASE (STANDARD DEVIATION), BASED ON TEST DATA FROM 10-FOLD CROSS-VALIDATION

| Metrics | Methods | | | | | |
|---|---|---|---|---|---|---|
| | **DT** | **RF** | **SVM** | **NB** | **NN** | **ONBRF** |
| Accuracy | 82.0% (6.4%) | 90.4% (6.3%) | 84.7% (8.5%) | 64.9% (8.3%) | 84.9% (4.1%) | 92.3% (4.6%) |
| Precision | 96.1% (3.6%) | 93.1% (4.9%) | 93.0% (5.4%) | 100.0% (0.0%) | 89.7% (4.5%) | 95.3% (2.7%) |
| Recall | 80.8% (10.2%) | 95.2% (5.2%) | 87.9% (12.5%) | 55.8% (9.9%) | 91.7% (3.2%) | 95.2% (5.2%) |
| F1 | 87.4% (5.1%) | 94.0% (3.9%) | 89.8% (6.5%) | 71.2% (8.3%) | 90.6% (2.5%) | 95.1% (3.0%) |
| Specificity | 87.0% (12.0%) | 73.0% (19.2%) | 74.0% (21.6%) | 100.0% (0.0%) | 59.0% (20.4%) | 82.0% (10.4%) |
| AUC | 85.3% (6.5%) | 95.1% (4.2%) | 87.0% (9.8%) | 91.7% (6.8%) | 84.8% (10.2%) | 95.8% (3.8%) |

RF has the highest recall (95.2%), followed by the proposed ONBRF (95.2%), indicating that these models are better at correctly identifying the presence of heart disease. The F1 score, which balances precision and recall, is again highest for ONBRF (95.1%), showing its overall robustness. Specificity is another critical measure where NB excels at 100%, meaning that it effectively identifies TN, but its overall performance is compromised by lower accuracy and recall. The AUC values further corroborate these findings, as the proposed ONBRF has the highest AUC at 95.8%, showing an excellent discriminative ability. RF follows closely with an AUC of 95.1%, while NB falls behind despite its high specificity and precision.

In addition, Table II presents stability metrics (standard error) to reflect the variability in performance across different trials (10-fold cross-validation). The proposed ONBRF and NN exhibited the lowest accuracy variability at 4.6% and 4.1%, respectively, suggesting that they consistently perform well across different datasets. NB showed significant stability in precision with a variability of 0.0%, but this is less meaningful given its lower recall stability (9.9%) and specificity variability (0.0%). Stability in recall and the F1 score is crucial for reliable predictions. Here, ONBRF demonstrates strong stability with low variability in recall (5.2%) and F1 score (3.0%). RF and SVM (SVM) also show commendable stability in these metrics, indicating their reliable performance.

Table III shows the energy consumption, which is a critical factor for fog computing applications. Latency and energy consumption metrics show that DT had the lowest latency at 13.6 ms and energy consumption at 4.1 Wms, making it the most efficient in computational resource consumption. NN exhibited the highest latency (1195.6 ms) and energy consumption (358.7 Wms), which could be a significant drawback in real-time applications. The proposed ONBRF, while highly accurate and stable, has moderate latency (301.4

ms) and energy consumption (90.4 Wms), positioning it as a middle ground between efficiency and performance.

TABLE III. MEAN METRICS FOR ENERGY CONSUMPTION (STANDARD ERROR) ACROSS VARIOUS MODELS, BASED ON THE TEST DATA FROM 10-FOLD CROSS-VALIDATION

| Metrics | Methods | | | | | |
|---|---|---|---|---|---|---|
| | **DT** | **RF** | **SVM** | **NB** | **NN** | **ONBRF** |
| Latency (ms) | 13.6 (2.96) | 129.9 (17.06) | 25.1 (3.32) | 24.3 (3.44) | 1195.6 (121.76) | 301.4 (52.66) |
| Energy (Wms) | 4.1 (0.89) | 39.0 (5.12) | 7.5 (1.00) | 7.3 (1.03) | 358.7 (36.53) | 90.4 (15.80) |

Stability in energy consumption is essential for consistent performance in fog environments. DT again leads in stability with the lowest variability in latency (2.96 ms) and energy consumption (0.89 Wms). The proposed ONBRF shows moderate stability in both latency (52.66 ms) and energy (15.80 Wms), suggesting reliability compared to NN, which exhibits high variability despite its high energy and latency, indicating less predictability in energy consumption that could be problematic for sustained fog deployment.

## V. COMPARATIVE ANALYSIS

To evaluate the external validity of the SHMHD method, its predictive accuracy was compared with that of leading existing methods using publicly available datasets. Four published heart disease datasets from the UCI machine learning repository [23] were used to achieve this, which are summarized in Table IV.

TABLE IV. COMPARATIVE ANALYSIS DATASET SUMMARY

| Dataset | Number of instances | Number of features | Heart disease rate |
|---|---|---|---|
| Cleveland | 304 | 14 | 45.7% |
| Hungary | 293 | 14 | 36.2% |
| Switzerland | 123 | 14 | 93.5% |
| VA Long Beach | 200 | 14 | 74.5% |
| All | 920 | 14 | 55.3% |

The accuracy comparison in Table V shows the superior performance of the proposed SHMHD method, utilizing PSO, NB, and RF techniques across multiple datasets compared to existing methods. In the Cleveland dataset, SHMHD demonstrated a remarkable accuracy of 98.4%, surpassing other techniques. For example, in [24], a 90.8% accuracy was reported using K-NN, while in [28], 93.4% accuracy was achieved with a combination of RF, KNN, LR, NB, GB, AB, and SVE classifiers.

In the Hungarian dataset, SHMHD reached a perfect accuracy of 100%, surpassing the model with Chisquare feature selection (CHI), Principal Component Analysis (PCA), and RF in [26] and the one with Harris Hawk Optimization (HHO) feature selection and Long Short-term Memory (LSTM) in [27]. For the Switzerland dataset, SHMHD again achieved 100%, outperforming the ANN in [28]. In the VA Long Beach dataset, SHMHD obtained 96.2%, exceeding the ANN in [28]. When considering all datasets combined, SHMHD's accuracy of 99.3% was higher than that of the stacked approach in [29] and the Extreme Gradient Boosting (XGB) in [30]. These results consistently demonstrate the robustness and effectiveness of the proposed SHMHD method in achieving higher accuracy than existing competing methods.

TABLE V. ACCURACY COMPARISON WITH SOME COMPETING METHODS

| Dataset | Study | Method | Accuracy |
|---|---|---|---|
| Cleveland | [24] | K-NN | 90.8% |
| | [25] | RF, KNN, LR, NB, GB, AB, SVE classifier | 93.4% |
| | [31] | DT, RF | 88.0% |
| | [32] | BO, SVM | 93.3% |
| | [33] | SVM, NB, ConvSGLV, and ensemble methods | 93.0% |
| | [34] | SVM | 85.0% |
| | SHMHD | ONBRF (PSO, NB, RF) | 98.4% |
| Hungarian | [26] | CHI, PCA, RF | 99.0% |
| | [27] | HHO, LSTM | 98.0% |
| | [35] | PSO, SVMGS | 97.0% |
| | [36] | LR, NB | 92.7% |
| | SHMHD | ONBRF (PSO, NB, RF) | 100.0% |
| Switzerland | [28] | ANN | 94.4% |
| | SHMHD | ONBRF (PSO, NB, RF) | 100.0% |
| VA Long Beach | [29] | ANN | 94.0% |
| | SHMHD | ONBRF (PSO, NB, RF) | 96.2% |
| All | [16] | Friend (DNN, ML, and EL approaches) | 94.3% |
| | [30] | Stacked [ETC, RF, XGB] | 92.3% |
| | [30] | XGB | 98.1% |
| | SHMHD | ONBRF (PSO, NB, RF) | 99.3% |

K-NN (K Nearest Neigbour), BO (Bayesian Optimization), ConvSGLV (Convolution Stochastic Gradient Logistic Vector), GB (Gradient Boosting), AB (Adaboost), SVE (Soft Voting Ensemble), CHI (Chisquare feature selection), PCA (Principal Component Analysis), HHO (Harris Hawk Optimization), LSTM (Long Short Term Memory), ETC (ExtraTrees Classifier), XGB (Extreme Gradient Boosting), DNN (Deep Neural Network), ANN (Artificial Neural Network), ML (Machine Learning), EL (Ensemble Learning).

## VI. CONCLUSIONS

Accurate and reliable prediction of cardiovascular disease is crucial for early diagnosis and effective intervention. While existing machine learning models such as RF, NN, and SVM offer promising results, they often struggle with imbalanced datasets, computational efficiency, and performance stability. This study proposes the ONBRF model within the SHMHD framework to address these limitations. The ONBRF model integrates the strengths of NB and RF through an optimized feature selection and ensemble learning approach. The results show that ONBRF achieved the highest accuracy (92.3%) and precision (95.3%) among competing models, ensuring minimal false positives while effectively identifying true cases (95.2% recall). Its F1 (95.1%) and AUC (95.8%) scores highlight its superior performance in handling imbalanced datasets. Furthermore, ONBRF exhibits strong computational efficiency, with a lower energy consumption (90.4 Wms) compared to NN (358.7 Wms), making it suitable for real-time applications in fog computing environments. External validation on multiple datasets confirms its robustness, achieving accuracy of up to 100% on the Hungarian and Switzerland datasets.

Collectively, the ONBRF model provides a novel, high-accuracy, and computationally efficient solution for the prediction of cardiovascular disease. By overcoming the limitations of existing models, it sets a new benchmark in predictive analytics, demonstrating its potential for integration into real-time healthcare systems. These findings position ONBRF as a cutting-edge advancement in machine learning applications for disease classification.

## ACKNOWLEDGEMENT

We would like to extend our gratitude to Yahaya Lamido of the Department of Computer Science at Modibbo Adama University of Technology in Yola, Adamawa State, Nigeria, for his assistance in gathering and providing the heart disease dataset from the Federal Teaching Hospital in Gombe State, Nigeria

## REFERENCES

[1] E. Moghadas, J. Rezazadeh, and R. Farahbakhsh, "An IoT patient monitoring based on fog computing and data mining: Cardiac arrhythmia usecase," *Internet of Things*, vol. 11, Sep. 2020, Art. no. 100251, https://doi.org/10.1016/j.iot.2020.100251.

[2] A. Rejeb, K. Rejeb, S. Simske, H. Treiblmaier, and S. Zailani, "The big picture on the internet of things and the smart city: a review of what we know and what we need to know," *Internet of Things*, vol. 19, Aug. 2022, Art. no. 100565, https://doi.org/10.1016/j.iot.2022.100565.

[3] N. Y. Philip, J. J. P. C. Rodrigues, H. Wang, S. J. Fong, and J. Chen, "Internet of Things for In-Home Health Monitoring Systems: Current Advances, Challenges and Future Directions," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 2, pp. 300–310, Oct. 2021, https://doi.org/10.1109/JSAC.2020.3042421.

[4] Md. Asif-Ur-Rahman *et al.*, "Toward a Heterogeneous Mist, Fog, and Cloud-Based Framework for the Internet of Healthcare Things," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4049–4062, Jun. 2019, https://doi.org/10.1109/JIOT.2018.2876088.

[5] K. S. Awaisi, S. Hussain, M. Ahmed, A. A. Khan, and G. Ahmed, "Leveraging IoT and Fog Computing in Healthcare Systems," *IEEE Internet of Things Magazine*, vol. 3, no. 2, pp. 52–56, Jun. 2020, https://doi.org/10.1109/IOTM.0001.1900096.

[6] S. K. Sood and I. Mahajan, "IoT-Fog-Based Healthcare Framework to Identify and Control Hypertension Attack," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1920–1927, Apr. 2019, https://doi.org/10.1109/JIOT.2018.2871630.

[7] S. Rastegar, H. Gholam Hosseini, and A. Lowe, "Hybrid CNN-SVR Blood Pressure Estimation Model Using ECG and PPG Signals," *Sensors*, vol. 23, no. 3, Jan. 2023, Art. no. 1259, https://doi.org/10.3390/s23031259.

[8] S. Tuli *et al.*, "HealthFog: An ensemble deep learning based Smart Healthcare System for Automatic Diagnosis of Heart Diseases in integrated IoT and fog computing environments," *Future Generation Computer Systems*, vol. 104, pp. 187–200, Mar. 2020, https://doi.org/10.1016/j.future.2019.10.043.

[9] H. F. Atlam, R. J. Walters, and G. B. Wills, "Fog Computing and the Internet of Things: A Review," *Big Data and Cognitive Computing*, vol. 2, no. 2, Jun. 2018, Art. no. 10, https://doi.org/10.3390/bdcc2020010.

[10] A. A. Mutlag, M. K. Abd Ghani, N. Arunkumar, M. A. Mohammed, and O. Mohd, "Enabling technologies for fog computing in healthcare IoT systems," *Future Generation Computer Systems*, vol. 90, pp. 62–78, Jan. 2019, https://doi.org/10.1016/j.future.2018.07.049.

[11] A. Jain, M. Ahirwar, and R. Pandey, "A Review on Intutive Prediction of Heart Disease Using Data Mining Techniques," *International Journal of Computer Sciences and Engineering*, 2019, https://doi.org/10.26438/ijcse/v7i7.109113.

[12] N. Absar *et al.*, "The Efficacy of Machine-Learning-Supported Smart System for Heart Disease Prediction," *Healthcare*, vol. 10, no. 6, Jun. 2022, Art. no. 1137, https://doi.org/10.3390/healthcare10061137.

[13] S. Guruprasad, V. L. Mathias, and W. Dcunha, "Heart Disease Prediction Using Machine Learning Techniques," in 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), Mysuru, India, Dec. 2021, pp. 762–766, https://doi.org/10.1109/ICEECCOT52851.2021.9707966.

[14] G. Choudhary and S. Narayan Singh, "Prediction of Heart Disease using Machine Learning Algorithms," in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, Bengaluru, India, Oct. 2020, pp. 197–202, https://doi.org/10.1109/ICSTCEE49637.2020.9276802.

[15] S. Kumari, M. Bhatia, and G. Stea, "Fog-Computing Based Healthcare Framework for Predicting Encephalitis Outbreak," *Big Data Research*, vol. 29, Aug. 2022, Art. no. 100330, https://doi.org/10.1016/j.bdr.2022.100330.

[16] A. Pati, M. Parhi, M. Alnabhan, B. K. Pattanayak, A. K. Habboush, and M. K. Al Nawayseh, "An IoT-Fog-Cloud Integrated Framework for Real-Time Remote Cardiovascular Disease Diagnosis," *Informatics*, vol. 10, no. 1, Mar. 2023, Art. no. 21, https://doi.org/10.3390/informatics10010021.

[17] O. R. Olaniran and M. A. A. Abdullah, "Bayesian weighted random forest for classification of high-dimensional genomics data," *Kuwait Journal of Science*, vol. 50, no. 4, pp. 477–484, Oct. 2023, https://doi.org/10.1016/j.kjs.2023.06.008.

[18] O. R. Olaniran and A. R. R. Alzahrani, "On the Oracle Properties of Bayesian Random Forest for Sparse High-Dimensional Gaussian Regression," *Mathematics*, vol. 11, no. 24, Jan. 2023, Art. no. 4957, https://doi.org/10.3390/math11244957.

[19] A. Arjmand and N. Giannakeas, "Fat Quantitation in Liver Biopsies Using a Pretrained Classification Based System," *Engineering, Technology & Applied Science Research*, vol. 8, no. 6, pp. 3550–3555, Dec. 2018, https://doi.org/10.48084/etasr.2274.

[20] A. Satty, M. M. Y. Salih, A. A. Hassaballa, E. A. E. Gumma, A. Abdallah, and G. S. M. Khamis, "Comparative Analysis of Machine Learning Algorithms for Investigating Myocardial Infarction Complications," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12775–12779, Feb. 2024, https://doi.org/10.48084/etasr.6691.

[21] S. M. Alanazi and G. S. M. Khamis, "Optimizing Machine Learning Classifiers for Enhanced Cardiovascular Disease Prediction," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12911–12917, Feb. 2024, https://doi.org/10.48084/etasr.6684.

[22] "Optimized NaiveBayes Random Forest Heart Disease Prediction." https://rid4stat.shinyapps.io/FOGCHD/.

[23] A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "Heart Disease." UCI Machine Learning Repository, 1989, https://doi.org/10.24432/C52P4X.

[24] S. A. A. Shah, A. H. Saleh, M. Ebrahimian, and R. Kashef, "Early Detection of Heart Disease Using Advances of Machine Learning for Large-Scale Patient Datasets," in *2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, Halifax, Canada, Sep. 2022, pp. 274–280, https://doi.org/10.1109/CCECE49351.2022.9918215.

[25] N. Chandrasekhar and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," *Processes*, vol. 11, no. 4, Apr. 2023, Art. no. 1210, https://doi.org/10.3390/pr11041210.

[26] A. K. Gárate-Escamila, A. Hajjam El Hassani, and E. Andrès, "Classification models for heart disease prediction using feature selection and PCA," *Informatics in Medicine Unlocked*, vol. 19, Jan. 2020, Art. no. 100330, https://doi.org/10.1016/j.imu.2020.100330.

[27] G. Rajkumar, T. Gayathri Devi, and A. Srinivasan, "Heart disease prediction using IoT based framework and improved deep learning approach: Medical application," *Medical Engineering & Physics*, vol. 111, Jan. 2023, Art. no. 103937, https://doi.org/10.1016/j.medengphy.2022.103937.

[28] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, "Atherosclerosis disease prediction using Supervised Machine Learning Techniques," in *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco, Apr. 2020, pp. 1–5, https://doi.org/10.1109/IRASET48871.2020.9092082.

[29] A. Tiwari, A. Chugh, and A. Sharma, "Ensemble framework for cardiovascular disease prediction," *Computers in Biology and Medicine*, vol. 146, Jul. 2022, Art. no. 105624, https://doi.org/10.1016/j.compbiomed.2022.105624.

[30] G. N. Ahmad, H. Fatima, S. Ullah, A. Salah Saidi, and Imdadullah, "Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV," *IEEE Access*, vol. 10, pp. 80151–80173, 2022, https://doi.org/10.1109/ACCESS.2022.3165792.

[31] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, Jan. 2021, pp. 1329–1333, https://doi.org/10.1109/ICICT50816.2021.9358597.

[32] S. P. Patro, G. S. Nayak, and N. Padhy, "Heart disease prediction by using novel optimization algorithm: A supervised learning prospective," *Informatics in Medicine Unlocked*, vol. 26, Jan. 2021, Art. no. 100696, https://doi.org/10.1016/j.imu.2021.100696.

[33] F. Rustam, A. Ishaq, K. Munir, M. Almutairi, N. Aslam, and I. Ashraf, "Incorporating CNN Features for Optimizing Performance of Ensemble Classifier for Cardiovascular Disease Prediction," *Diagnostics*, vol. 12, no. 6, Jun. 2022, Art. no. 1474, https://doi.org/10.3390/diagnostics12061474.

[34] E. A. Ogundepo and W. B. Yahya, "Performance analysis of supervised classification models on heart disease prediction," *Innovations in Systems and Software Engineering*, vol. 19, no. 1, pp. 129–144, Mar. 2023, https://doi.org/10.1007/s11334-022-00524-9.

[35] A. K. Dubey, A. K. Sinhal, and R. Sharma, "An Improved Auto Categorical PSO with ML for Heart Disease Prediction," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8567–8573, Jun. 2022, https://doi.org/10.48084/etasr.4854.

[36] R. Rajendran and A. Karthi, "Heart disease prediction using entropy based feature engineering and ensembling of machine learning classifiers," *Expert Systems with Applications*, vol. 207, Nov. 2022, Art. no. 117882, https://doi.org/10.1016/j.eswa.2022.117882.