Journal of Theoretical and Applied Information Technology
28th February 2023. Vol.101. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REVIVED ANT COLONY OPTIMIZATION-BASED ADABOOST ALGORITHM FOR HEART DISEASE AND DIABETES (HDD) PREDICTION

S.USHA[1], Dr.S.KANCHANA[2]

[1]Research Scholar, Department Of Computer Science,College of Science and Humanitites, SRM Institute of Science and Humanities, Kattankulathur

[2]Assistant Professor, Department of Computer Science,College of Science and Humanitites SRM Institute of Science and Humanities, Kattankulathur

us3648@srmist.edu.in,kanchans@srmist.edu.in

ABSTRACT

Hidden information in the massive medical data collected by the healthcare industries assists in making productive decisions. Using cutting-edge data mining tools, we can get accurate findings and make informed decisions based on that information. Affordable quality treatment is a big concern for healthcare institutions like hospitals and clinics. The world's leading killers are heart disease and diabetes (HDD), which are often misdiagnosed. Earlier diagnosing of HDD saves the victim's life and the costs associated with treating the disease. Additionally, most existing machine learning algorithms tend to specialize in forecasting certain diseases only. It's possible that a classifier that can reliably forecast the frequency of several. In this paper, Revived Ant Colony Optimization-based Adaboost Algorithm (RACOAA) for HDD Prediction. The Adaboost algorithm is enhanced by fixing the threshold value for predicting the HDD, and it will assist in enhancing the classification accuracy. To predict more accurately, this research applies the enhanced version of ant colony, RACO. The deposit of pheromones is optimized for better classification. This study uses the Cardiovascular Disease Dataset and the PIMA Indian Diabetes Dataset to assess the efficacy of the proposed classifier. The assessment results show that the suggested classifier achieves higher classification accuracy than the state-of-the-art classifiers.
Keywords: *Ant Colony, Adaboost, Diabetes, Heart Disease, Optimization*

1. INTRODUCTION

The expansion of data-generation processes in healthcare has brought renewed interest in machine learning (ML), an old idea. One study found that 86 percent of healthcare companies employ ML solutions, and 80 percent of healthcare company CEOs have an AI strategy [1]. Machine learning is a crucial subfield when discussing the greater area of artificial intelligence. ML can be described as the capability of a computer to replicate intelligent human behavior. Algorithms may be taught to recognize patterns in data and make choices based on the presence or absence of particular traits or variables in newly collected data. Data science, which is needed to apply different ML models, is essentially the meeting point of several fields, including but not limited to computer science, statistics, and mathematics [2]. Humans' weak short-term memory makes them fallible despite their superior intelligence. As the volume and variety of available data allow for the development of more sophisticated machine learning algorithms, people can benefit from using these tools to make judgments that consider a wider range of relevant factors. Recently, ML has found applications in several areas of medicine, including managing heart failure, clinical decision assistance in medical imaging, and clinical medicine [3].

While conventional statistical analysis is motivated by hypotheses, machine learning concerns how well a model can predict future outcomes. Epidemiologic methodologies have generally driven the development of healthcare data [4]. The absence of efficient connectivity between different disciplines is a severe barrier, primarily when data analysts and public health professionals work together to answer questions. ML approaches provide new ways to address these concerns. Even though there are many diverse terminologies in these fields, it has been suggested that many terms

mean the same thing [5]. Statisticians and epidemiologists have used these notions for a long time in hypothesis-driven research, so it shouldn't take much education to help them comprehend how ML uses such principles with a standard structure in mind. Clinicians, public health doctors, pathologists, and radiologists, among others, are more accustomed to hypothesis-driven research than ML algorithms. Because of this, it's essential to show how the two different but related fields of thought are similar [6].

Research on techniques for resolving optimization issues is still an ongoing process in the current world. The determinism of a new optimization method is not set in milestones. In other words, the output of deterministic algorithms is always the same as the input [7]. However, stochastic algorithms use pseudo-randomness to deal with the unknowns inherent in the ML process. While applied in optimization, the issues present in deterministic algorithms are ineffective as problem size gradually increases because they demand too much timing and computing power [8]. Therefore, in contrast to deterministic algorithms, the computational efficiency of bio-inspired-based ML algorithms becomes a major factor in their adoption. By modeling natural processes as restricted optimization, bio-inspired algorithms provide a novel heuristic method [9], [10], [19], [20], [11]–[18].

1.1 Problem Statement

Using Machine Learning (ML) algorithms for diagnosing Heart Disease and Diabetes (HDD) presents several obstacles. ML models cannot be built without access to high-quality data that is statistically and demographically representative of the target audience. Therefore, if ML models is included in health care, it becomes crucial to develop efficient data-handling methods on all levels. The processes present in ML algorithms can convert raw data into usable datasets. The relevant parties must develop a solid data governance policy to make the most of the data being produced. Another significant difficulty is that ML-based forecasts seldom come with justifications for their healthcare expert conclusions. Legal processes are not optimized in case of a possible mistake while using an ML model to forecast a health consequence.

Detection is the primary difficulty with HDD. Predictive tools for HDD exist, but they are either prohibitively costly or inefficient when applied to estimating an individual's risk of developing heart disease. Death rates and other consequences associated with HDD may be reduced with early diagnosis. However, it is not feasible to monitor patients every day correctly, and 24-hour consultation with a doctor is not accessible due to the increased sagacity, time, expense, and experience required. The current world has access to enhanced quality data, and it can be analyzed using various ML techniques to uncover previously unseen patterns. The underlying patterns in medical data can be utilized to identify illnesses accurately.

1.2 Objective

The primary goal of this research work is to propose Revived Ant Colony Optimization-Based Adaboost Algorithm (RACOAA), a bio-inspired optimization-based classifier that can be used for the prediction of both heart disease and diabetes (HDD).

2. LITERATURE REVIEW

This section breaks the relevant literature into two groups: (1) heart disease classification-based algorithms and (2) diabetes disease classification-based algorithms.

2.1 Heart Disease Classifiers

"IoT-based heart disease prediction and diagnosis model" [21] is proposed to provide elevated service via online healthcare. It makes use of sensors for predicting heart disease among the public. IoT-HDPDM is trained by utilizing the standard dataset. In the testing phase, data gathered from patients are used to find the presence of heart disease. While applying the J48 classifier, IoT-HDPDM has given better accuracy. "Medical Decision Support System" [22] is proposed to predict atherosclerosis, which falls under the category of heart disease. K-means and K-medoids are applied for performing the classification in MDSS. K-Nearest Neighbor and Artificial Neural Network are applied to perform the prediction. MDSS evaluated using Matthews Correlation Coefficient. "ML-based Heart Disease Prediction" [23] is proposed to analyze

heart disease datasets with many features. Missing data lead to complexity, and complete and partial traits handle it. Feature minimization controls the classifier performance. Ada-boost and Decision trees are further used for prediction. "Intelligent Diagnosis" [24] is proposed to minimize the cost of diagnosing the disease. It allows full access to initiate the non-invasive method, treated as an early diagnosis method. IoT embedded with sensor devices makes communication with other devices with a low level of human intervention. Machine learning concepts are used to make intelligent communication with other devices. "Clinical Event Prediction Comparison" [25] studied machine learning strategies to predict heart diseases. Decision trees, SVM, random forest, logistic regression, and neural networks have been studied to perform an adequate comparison. Area-Under-Curve (AUC) is used to manipulate the data to test the algorithm's effectiveness.

"Machine Learning Prediction Model" [26] is proposed to predict chronic liver disease related to heart disease. Significant parameters in the dataset are identified using regression tree concepts and linear regression methods, which assist in classifying hepatocellular carcinoma. AUC is used to analyze the performance of hepatocellular carcinoma classification. "Exploratory Data Analysis" [27] is proposed to detect the presence of mistakes and seek appropriate data to predict heart disease. It decides between the correlation and explanatory variables. It avoids statistical modeling and inferences. Analytics is done to predict the hidden patterns related to heart diseases. "Smart Heart Disease Prediction" [28] is proposed to predict heart disease risk factors. Navies Bayesian is applied for classification. Dataset is split into two sets, i.e., 80% and 20%, where one is used for training and another for testing. Advanced Encryption Standard has been applied to provide security to medical data during transmission. "Analytical Study" [29] has been conducted to identify the best classification algorithm to predict heart disease. Analyzing different classification algorithms, it was found that the Apriori algorithm provides better classification accuracy than ANN, SVM, and KNN. "Hidden Naïve Bayes Classifier" [30] is proposed as a prediction system for heart disease. HNBC is proposed to avoid the demerits present in the traditional Naïve Bayes strategy. HNBC aims to predict heart disease with enhanced accuracy than TNB.

"Boosting Support Vector Machine (BSVM)" [31] was proposed to predict heart disease risks with enhanced accuracy. The method was tested on data from the Cleveland dataset. The listwise method was used for data cleansing to remove the six absent entries. Due to the small sample size and lack of biases introduced by the random selection, this method had no impact on the experiment. The boosting method chooses the most relevant characteristics to save time and increase precision. The data is then separated into two sets, one for training and another for testing, using the train/test split methodology. Then, SVM is utilized for training and testing the records in the dataset. In this case, a linear kernel with a C value of 0.05 is used. "Swarm-Artificial Neural Network (S-ANN)" [32] is proposed to forecast heart diseases accurately. It is an ANN-based model that uses all the observed healthcare data on heart disease as input. The neural networks (NN) with the predetermined population size are implemented by assigning a random weight. There are three distinct phases of data processing involved in S-ANN. First, the method is used to build an ANN using data from the sample population. The S-ANN generates a three-layer feedforward ANN with random weights at the beginning of each iteration. Each randomly generated S-ANN is fed as a single pattern, and their weights are adjusted in the second phase using a back-propagation approach. After that, the swarm employs a stochastic weight adjustment method created in the third stage. An ANN population winner is then selected using the evaluation of its performance as the last step. A stochastic function determines which neurons present in the population will get the winning neuron's acquired weight and bias.

2.2 Diabetes Classifiers

"Personalized Prediction" [33] is a sequential modeling method at the patient level that uses sequential dependencies to provide personalized calculation of the prescription efficiency of diabetic patients. This prediction model was executed with a Recurrent Neural Network, which utilizes the progression of

every of the previous record fields as an input for predicting the efficiency using the prescription of every patient. These historical records of each patient are used effectively for making the prediction. "Fuzzy Based Rule Set Model" [34] is proposed for predicting people with diabetes by applying the Grey Wolf Optimization technique. The perception of artificial intelligence was incorporated into the model to study the fuzzy rule. "Diabetic Predicting Technique" [35] is proposed for classifying the patients according to the feature sets developed to evaluate the real-world data with a machine learning algorithm called the Hoeffding Tree algorithm. It also provides a way to analyze diabetic pathology. "Deep Neural Network Framework" [36] is proposed for classifying the data towards predicting people with type 2 diabetes with stacked encoders. The features are extorted and classified using the SoftMax layer, and the Back-propagation method is used for fine-tuning. Evaluation metrics are used to evaluate the framework performance of classifying the Type 2 diabetic data. "Peripheral Neuropathy Abnormalities Prediction" [37] is proposed to predict Type 2 Diabetics. It predicts cardiac autonomic neuropathy with the clinical features of the tests. The test which shows the frequency of the disease are: heart rate response, sustained handgrip, blood pressure response for standing position, and the response of the blood pressure. Based on the test, it could predict CAN disease and diabetes.

"Estimation of Peripheral Vascular Occlusion (PVOc)" [38] for food in diabetics has been calculated by utilization of a Support Vector Machine (SVM) classifier along with the Wolf optimization algorithm. The increasing blood sugar intensity of Type 2 diabetics was also evaluated using the PVOc method. Synchronizing Chaotification is also performed to measure the asymmetry-of-PPG signals, bilateral similarity, and the level of PVOc was estimated. The proposed technique also finds three butterfly motion patterns to represent its severities. The "Integrated Electronic Health Record" [39] model was built to predict undiagnosed diabetic individuals with dental backgrounds. The feature selection was performed on the dataset retrieved from the Marshfield Clinic Health System by applying predictive model validation. Classifiers, namely; Multilayer Perceptron, SVM, Random Forest, and Logistic Regression, were also used for measuring its performance.

"Elementary Analysis" [40] is proposed to classify type 2 diabetes using toenails. Forty-six machine learning structure models were compared with the training and testing data set. The Random Forest model is trained by taking information on age, gender, and smoking history into account. The receiver Operating Characteristic is used to measure the elemental analysis. "Novel Prototype" [41] is proposed to measure the electrical activity distribution in the human skin LED matrix of 20×10. It was used to capture the signals using sensors that encompass 200 in number. Light intensities from the LED matrix were evaluated, and diabetic and normal individuals' intensities were calculated. The group of individuals was used for evaluation, and based on intensity level, the model makes the recognition. "Prediction of Type 2 Diabetic Mellitus" [42] is developed to make the model predictive for more than one dataset. The k-means and logistic regression algorithms are used for preprocessing, and the Knowledge Analysis tool kit was used to analyze. Two diabetic datasets were used to analyze the model's efficiency and for predicting and managing the diabetic.

"Modified Support Vector Machine (MSVM)" [43] is presented for diabetes prediction. MSVM uses a modified principal component analysis (MPCA) technique to extract features from the input medical record for preliminary processing. The data is projected onto a new subspace with the exact dimensions or less by calculating the covariance matrix's eigenvectors using MPCA. To choose the most relevant and valuable characteristics for subsequent categorization, MSVM is used to extract them. The predictions and classifications are based on the best characteristics that have been retrieved. The search for the ideal hyperplane satisfies the classification requirement and employs a tailored method to get the best possible margin of separation. This classification is used for analyzing data that is both linear and nonlinear, and it helps to lessen the experimental categorization error. The input vector is mapped into a higher-dimensional space when

the greatest separation hyperplane is determined. The diabetes prediction system "Deep Learning for Predicting Diabetes (DLPD)" [44] is constructed using deep neural networks' hidden layers. Overfitting is avoided using dropout regularization. DLPD makes use of a binary cross-entropy loss function with finely tuned parameters. Normalization layers help the model retain its training results and adapt to novel inputs. Dropout randomly resets a certain number of units to zero with each training update. The hyperparameters are scaled to a fixed value to accommodate the accumulation of previous gradients. Through a recursion process, we may define the total of the gradients as the average attenuation of all previous square gradients. Classification times are independent of variables other than the current average and gradient.

3. REVIVED ANT COLONY OPTIMIZATION BASED ADABOOST ALGORITHM

3.1 Adaboost Algorithm

AdaBoost is a powerful ensemble learning method that makes the most of a finite set of training instances by dynamically redistributing their relative importance. The core steps of AdaBoost is provided in Algorithm 1.

Algorithm 1: Pseudocode of Core Adaboost Algorithm

1. **Input:**

(a) An algorithm that learns weakly

(b) Size of ensembling F

(c) Dataset used for training
$E = \{(P_1, Q_1), (P_2, Q_2), \dots\dots, (P_c, Q_c)\}$.

2. **Identification of variable data for training:**

Distribute weights equally while starting the training set: $\pi_s^1 : \pi_s^1 = \frac{1}{c}, s = 1,2, \dots, c.$

3. **Foreach**
$f = (1,2,3, \dots\dots\dots, F-2, F-1, F)$, perform

(a). Create a naive classifier by making utilization of:

- ✓ The weighted distribution strategy.
- ✓ The process of underlying weak learning classifier.

(b). Compute the training error present in the weight σ_f: $\sigma_f = \sum_{s=1}^{c} \pi_s^f, Q_s \neq l_f(P_s)$;

(c). Allocate a value for l_t and it will have an effect in δ_f: $\delta_f = \frac{1}{2} \ln\left(\frac{1-\sigma_f}{\sigma_f}\right)$;

(d). Fine-tune the training set's weights $\left(\pi_s^{f+1} = \frac{\pi_s^f . ex[(-\delta_f . q_s . l_f(P_s))}{l_g}\right)$, using the $l_g = \sum_{s=1}^{c} \pi_s^f . ex[\left(-\delta_f . q_s . l_f(P_s)\right)$ which refers to standardization.

4. **Output:**

Ensemble-based classifier: $g(P) = sign\left(\sum_{f=1}^{F} \delta_f . l_f(P)\right)$

Step 2 (i.e., Identification of variable data for training) establishes a consistent weight distribution across all samples with the whole dataset $E, (P_s, Q_s)$ serving as the sth sample with $P_s \, \omega \, B^y$. By displaying the characteristics, $Q_s \, \omega \{-1,1\}$ operate labeling i.e., $(s = 1,2,3, \dots, c-2, c-1, c)$. After that, the AdaBoost algorithm iteratively calls the Weakly Learning algorithm for the classification, as demonstrated in Step 3. In the fth iteration, the classifier refines the l_f i.e., learning of samples, and the resampling of data π^f based on the outcomes that are obtained from the previous iteration's predictions on the training data. Step 3(d) involves assigning lower weights to "uncomplicated" samples that are successfully identified by l_f and greater weights to "complicated" samples that are incorrectly labeled.

To account for the possibility of a different distribution π^{f+1}, a new classification algorithm l_f will be introduced in the subsequent cycle. That's why AdaBoost prioritizes complicated samples. AdaBoost linearly integrates every component of the classifier to generate a single-end hypothesis g, and it does this for a total of F iterations.

3.2. Threshold Value-based Classification

AdaBoost iteratively trains learners by raising the weights of error-classifying

instances, narrowing their attention to the misclassified "complicated" samples. Multiple distinct classifiers are pooled to get a consensus using a weighted majority selection process. AdaBoost uses the accessible Weakly Learning strategy rather than simply building an outstanding algorithm for classification. Most samples with larger values of a certain attribute fall into the same group, whereas the samples with smaller values fall into the other different group. AdaBoost determines the optimal threshold from all aspects of the classification algorithm in each iteration, and it offers a rudimentary cum doable learning algorithm. Algorithm 2 provides a pseudocode of Threshold Value-based Classification (TVC).

Algorithm 2: Pseudocode of TVC

1. **Input:**

(a).Training Dataset:

$E = \{(P_1, Q_1), (P_2, Q_2), (P_3, Q_3), \dots, (P_{c-1}, Q_{c-1}), (P_c, Q_c)\}$;

(b).Distribution of weight π.

2. Foreach $w = 1,2,3, \dots, y-2, y-1, y$

(a).Using the size of wth attribute, order the c training samples.

(b).Using the size of l_w attribute, order the β_w training samples.

3. **Output:**

(a).Perform a comparison among the weighted errors $\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_{y-2}, \sigma_{y-1}, \sigma_y$

(b).Identify the error that has the minimum weight, i.e., $\sigma = min\{\sigma_w\}(w = 1,2, \dots, y)$.

(c).Adaboost classification that is based on Threshold value is $l = l_w$ with the error $\sigma_w = \sigma$.

In the dataset E, (P_s, Q_s) is the sth sample, $P_s = (p_{s1}, p_{s2}, \dots . p_{sw})$ describes the sth sample $(w = 1,2, \dots, y)$, and $Q_s \omega \{-1,1\}$ does the label $(s = 1,2,3, \dots . c-2, c-1, c)$. So, the wth ascribe the sth instance, i.e., p_{sw}, and each sample has y characteristics, and its label will describe the same. This research work identifies the threshold β_w as a classifier from $\{p_{1w}, p_{2w}, \dots . p_{cw}\}$ which are the wth characteristics of all the instances, to get the lowest weighed error $\sigma_w = \sum_{s=1}^{c} \pi_s, Q_s \neq l_f(P_s)$. When discriminating the cases whose wth characteristics are bigger than β_w and those that are less than β_w will get the error that has minimum weight. When the majority of instances are either more positive than β_w or more negative than β_w in terms of their wth attributes, then Eq.(1) is applied:

$$l_w(P_w) = \begin{cases} 1, & p_{sw} \geq \beta_w \\ -1, & p_{sw} < \beta_w \end{cases} \quad (1)$$

whenever the majority of instances are either excellent or negative, then the majority of the wth qualities are smaller than β_w or bigger than β_w: $l_w(P_w) = \begin{cases} 1, & p_{sw} \geq \beta_w \\ -1, & p_{sw} < \beta_w \end{cases}$ at step 2. Finally, the threshold classification-based weak classifier l has an accuracy of $1 - \sigma$.

3.3 Multi-Threshold Value-based Classification

Rather than building a complex association among all labels and attributes, the TVC prioritizes the more superficial association between labels and the most significant attribute. However, the classification may result in average or below average accuracy with deprived generalization due to the following factors. The two problems that persist with TVC are:

(a).It still has a propensity for overfitting, where it concentrates on training the samples that are misclassified by just one criterion in every iteration.
(b).It seeks the ideal cutoff from all the characteristics as the classified rule, wasting future benefits in other attributes.

There is a risk of low accuracy and poor generalization when using TVC if the label tends to have a specific degree of linear associations with more than one characteristic, such as the link correlation between product review comments and its star rating. The multi-Threshold Value-based classification Algorithm (MTVC) employs not just one but three critical threshold values for performing the classification to enhance the TVC. The MTVC begins by sorting the samples into sequences based on their attributions, then selects thresholds from each sequence. The three different stages of MTVC are: (i) the

selection of the top three threshold values with the lowest error rates, (ii) Assigning random weights, and (iii) Utilizing the weighted majority procedure to select and combine the results. Algorithm 3 provides the pseudocode of MTVC.

Algorithm 3: Pseudocode of MTVC

1. **Input:**

(a). Training Dataset:

$E = \{(P_1, Q_1), (P_2, Q_2), (P_3, Q_3), \dots, (P_{c-1}, Q_{c-1}), (P_c, Q_c)\}$;

(b). Distribution of weight π.

2. **Foreach** $w = 1,2,3, \dots, y-2, y-1, y$

(a). Using the wth attribute's size, order the c training examples.

(b). For the wth characteristic, determine the threshold value β_w which results l_w to get the minimum error with the weight $\sigma_w = \sum_{s=1}^{c} \pi_s, Q_s \neq l_f(P_s)$.

3. **Output:**

(a). Identification of three classifiers l', l'', l''' with the highest minimum errors $\sigma', \sigma'', \sigma'''$, where $0 \leq \sigma' \leq \sigma'' \leq \sigma''' \leq 1$.

(b). Weighted errors $\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_{y-2}, \sigma_{y-1}, \sigma_y$ are compared for $l_1, l_2, l_3, \dots, l_{y-2}, l_{y-1}, l_y$ classifiers.

(c). $l(P) = sign\{(1-\sigma')l'(P) + (1-\sigma'')l''(P) + (1-\sigma''')l'''(P)\}$ predicts the results with the error $l = \sum_{s=1}^{c} \pi_s^f, Q_s \neq l(P_s)$.

A classifier's error value of ρ indicates that it will properly categorize the samples with a likelihood of $1-\sigma$. Hence, in Algorithm 3, the poor classifiers l with various Detection limits will reliably categorize the instances over the next 5 scenarios:

1. l', l'', l''' classifiers perform exact classification $l'(P) = Q, l''(P) = Q, l'''(P) = Q$ with the probability $l(P) = Q$ if and only if Eq.(1) holds the same.

$$m_1 = (1-\sigma')(1-\sigma'')(1-\sigma''') \quad (2)$$

2. l' and l'' classifiers perform the exact classification where l''' does not perform better classification $l'(P) = Q, l''(P) = Q, l'''(P) \neq Q, l(P) = Q$ with the condition $\sigma' \leq \sigma'' \leq \sigma'''$. Eq.(3) provides the probability value of the same.

$$m_2 = (1-\sigma')(1-\sigma'')\sigma''' \quad (3)$$

3. l'' and l''' perform exact classification, but l'' not. Eq.(4) provides the probability value of the same.

$$m_3 = (1-\sigma')\sigma''(1-\sigma''') \quad (4)$$

4. l'' and l''' perform exact classification, but l' doesn't perform the exact classification where the aggregated accuracy of $l''$and $l'''$is more than that of $l', (1-\sigma'') + (1-\sigma''') \geq (1-\sigma')$, or $\sigma'' + \sigma''' - \sigma' \leq 1$. Eq.(5) provides the probability value of the same with the condition of $\sigma'(1-\sigma'')(1-\sigma''') \leq 1$.

$$m_4 = \sigma'(1-\sigma'')(1-\sigma''').m \quad (5)$$

5. l' performs exact classification, but l'' and l'''doesn't perform the exact classification where the aggregated accuracy of $l''$and $l'''$is less than that of $l', (1-\sigma'') + (1-\sigma''') \leq (1-\sigma')$, or $\sigma'' + \sigma''' - \sigma' \geq 1$. Eq.(6) provides the probability value of the same.

$$m_5 = (1-\sigma')\sigma''(1-\sigma''').(1-m) \quad (6)$$

Eq(2) to Eq.(6) assist in predicting the classification accuracy $l(P)$.

$$l(P) = m_1 + m_2 + m_3 + m_4 + m_5 \quad (7)$$

Let's pretend the errors have a normal distribution $\sigma' \sim o(0,1), \sigma'' \sim o(\sigma', 1), \& \sigma''' \sim o(\sigma'', 1)$. Eq.(8) provides the joint density function of the same.

$$g(p,q,i)|\sigma', \sigma'', \sigma''' = \begin{cases} \frac{1}{(1-p)(1-q)}, 1 \\ 0, \quad el \end{cases} \quad (8)$$

$\sigma'' + \sigma''' - \sigma' \geq 1 \ \& \ 0 \leq \sigma' \leq \sigma'' \leq \sigma''' \leq 1$ has the maximum mean value $\{1 + \sigma' - \sigma'', \sigma''\} \leq \sigma''' \leq 1$, where $\sigma' \leq \sigma'' 1.1 + \sigma' - \sigma''$ consequently performs comparison $\begin{cases} 1 + \sigma' - \sigma'' > \sigma'', as \sigma'' < \frac{1+\sigma'}{2} \\ 1 + \sigma' - \sigma'' \leq \sigma'', as \sigma'' \geq \frac{1+\sigma'}{2} \end{cases}$ in addition to Eq.(8). The $l(P)$ in E.(7) is determined by calculating $l(P)$ as $\theta = \int_0^1 yp \int_p^{\frac{1+p}{2}} yq \int_{1+p-q}^1 \frac{1}{(1-p)(1-q)} yi + \int_0^1 yp \int_{\frac{1+p}{2}}^1 yq \int_q^1 \frac{1}{(1-p)(1-q)} yi = in\ 2$
.

TVC and MTVC have almost similar classification accuracy with a minor difference. The probability of $1 - \sigma'$ is 52.3%, if and only if the probability of $\sigma' \sim o(0,1)$ is 47.7% for σ'. MTVC, presented in Algorithm 4, has higher prediction accuracy when compared with TVC.

Algorithm 4: Pseudocode of Adaboost-based MTVC

1. **Input:**

(a). Size of ensembling F

(b). Dataset used for training $E = \{(P_1, Q_1), (P_2, Q_2), \ldots, (P_c, Q_c)\}$.

2. **Initializing the variables for training:**

As Algorithm 1

3. **Foreach** $f = 1,2,3, \ldots, F-2, F-1, F$

(a). Results in the generation of poor classifier $l_f(P)$, As like Algorithm 3.

(b). $\sigma_f = max\{\sigma''', h\}$ favors assigning a higher weight to $l_f(P)$ by selecting a more significant error σ_f between σ''' and h. This is because the more significant the error σ_f of training samples $l_f(P)$, the lower the weight of $l_f(P)$. $\delta_f = \frac{1}{2} \mathrm{In}\left(\frac{1-\sigma_f}{\sigma_f}\right)$ is applied for calculating the weights of $l_f(P)$.

(c). Alter the training data weight: $\pi_{f+1}(s) = \frac{\pi_f(s).f^{-\delta_f.q_s.l_f(P_s)}}{I_f}$, where $I_f = \sum_{s=1}^{c} \pi_f(s).f^{-\delta_f.q_s.l_f(P_s)}$ is the assimilation.

4. **Output:**

Ensembled-based classifier: $g(P) = sign\{\sum_{f=1}^{F} \pi_f.l_f(P)\}$

3.4 Revived Ant Colony Optimization

The above-discussed AA (i.e., Adaboost Algorithm) lacks better classification accuracy. Hence, this research work attempts to enhance the above AA using the bio-inspired optimization-based strategy, namely revived ant colony optimization ($RACO$), which is inspired from the natural characteristics of ants towards foraging its food.

3.4.1 Inspiration

One method cannot provide optimal performance across the board because of the varied nature of the challenges. Quality improvement improvements or deterioration in searching efficiency can be attributed to the pheromone update approach. While a pheromone updating method functions slowly but steadily, pheromone increase can produce good results in single-modal situations. Still, it won't guarantee the algorithm's effectiveness when dealing with multiple modalities. The problem's characteristic is typically unknown at the outset of the search process. In addition, it may exhibit various behaviors across the various optimization stages. Here, the Revived ACO ($RACO$) needs to be dynamic and flexible during its search process. Due to the varied nature of the challenges, no single solution will yield optimal results across the board. The pheromone update technique is crucial for effective searches. The algorithm's performance cannot be guaranteed in the face of multi-modal difficulties using a pheromone update technique that involves a slow pheromone increase to get adequate results for single-modal problems. In actuality, the problem's characteristic is usually unknown up until the point where the search begins. Additionally, it may exhibit varying behaviors during the various optimization stages.

$RACO$ is an interpersonal evolutionary information-driven pheromone update technique to solve the optimal local problem and ensure the Pareto-front has reached continuously. As the optimization progresses, the adaptive pheromone updating technique

modifies the method to tailor the search to specific issues and enhance the convergence rate. Specifics of the two methods are discussed below.

3.4.2 Pheromone Updating Strategy

Instead of updating the pheromone in an inter manner, $RACO$ updates the pheromone in an intra-manner, which refers to updating the pathways created by the same group of ants. The two primary designs in this strategy are intra-update of pheromone and avoidance of local optima.

(a) Intra-Update of Pheromone

The pheromone updating technique helps the algorithm converge more quickly by influencing the ants' search patterns. As part of each iteration, the pheromone matrix gets updated and will deposit more pheromones to the recently visited edges to increase their influence. The ants in $RACO$ all come from distinct colonies. Pheromone matrices are group-specific. The algorithm used to update pheromone matrices must determine how the ant colonies should be credited for exploring new routes. The newly visited edges may be recognized as a h_t if the set of non-dominated solutions (NDS) at the $f-1$ th iteration is H_{mv}, and the set of NDS with freshly created non-dominated Pareto remedies in the fth iteration is H_{mt},. Algorithmically comparing the vertices in H_{mv} and H_{mt} yields a h_t. An edge is h_t, if it is part of H_{mt} but not H_{mv}. Next, the $RACO$ compiles all h_t into a database of H_t. The pheromone associated with the w group on the $(a,z)\beta_{a,z}^{w}$ edge is updated in the following way, according to Eq. (4):

$$\beta_{a,z}^{w} = \varphi\beta_{a,z}^{w} + \sum_{p\omega SS}\frac{1}{j(p|\exists^{w})} + \Delta\beta \quad (4)$$

The pace at which the preceding pheromone is preserved and it is controlled by the parameter φ, where $j(p|\exists^{w})$ is the wth sub-problem. The best solution set for this group is the integer Φ. The value of $\Delta\beta$ is determined by Eq. (5):

$$\Delta\beta = \begin{cases} \frac{1}{j(p|\exists^{w})} & if\ (a,z)\ belongs\ to\ H_i \\ 0, & otherwise \end{cases} \quad (5)$$

Consequently, more pheromone is distributed around the boundaries that have recently been visited. The ants are likelier to continue exploring the solution space by exploring recently-visited edges.

The approach constrains the maximum and minimum pheromone concentrations along an edge to prevent it from settling into a suboptimal state. At the beginning of each loop, the maximum allowable concentration of pheromone is increased by β_{max} which is provided in Eq.(6).

$$\beta_{max} = \frac{V+1}{(1-\varphi)j_{min}} \quad (6)$$

wherein V is the current iteration's count of NDS produced by group a, and j_{min} does an ant colony achieving process with the lowest objective value. In addition, Eq. (7) will be applied to determine the smallest pheromone, β_{min}:

$$\beta_{min} = \in \beta_{max} \quad (7)$$

wherein $0 \prec \in \prec 1$is a tuning knob. If the value of $\beta_{a,z}^{w}$, is lower than β_{min}, then set it to β_{min}; otherwise, set it to $\beta_{a,z}^{w} = \beta_{max}$; and vice versa for $\beta_{a,z}^{w} \succ \beta_{max}$.

(b) Avoidance of Local Optima

Traditional ACO gets stuck in local optima and prevents finding the best possible solution. In this study, we suggest a strategy for avoiding this problem by trying to avoid reaching a local optimum. By resetting the pheromone matrices when the ants in a given colony reach an optimum level, this technique can increase the variety of pathways explored by the colony. In $RACO$, a stagnant condition is defined as a situation when there is no discernible change in the appearance of edges along the pathways. That is to say, the newly discovered edges contain few, if any, previously undiscovered edges. Using Eq.(8), the suggested method modifies the pheromone whenever a steady state is reached, avoiding the pitfalls of reaching a local optimum.

$$\beta_{a,z}^{w} = \varphi.\left(\beta_{max} - \beta_{a,z}^{w}\right) \quad (8)$$

3.4.3 Constant Update of Pheromone

With the adaptive pheromone update technique based on Learning Automata

Journal of Theoretical and Applied Information Technology
28th February 2023. Vol.101. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

($LGAT$), the issue's nuances are dissected at various optimization stages. The $LGAT$ modifies the techniques for updating the pheromones used by each colony of ants based on the feedback.

(a) Strategy descriptions for potential pheromone updates

After optimum solutions for sub-problems have been selected, the pheromone updating procedures are applied. Each cluster's pheromone matrix is updated using the newly identified solutions leading to the modification of attraction between the ants. The first stratagem o_1 is a pheromone-updating plan that considers intragroup evolutionary information (i.e., Eq (4)).

Eq.(9) describes the second stratagem o_2.

$$\beta_{a,z}^{w} = \varphi\beta_{a,z}^{w} + \sum_{p\omega ss}\frac{1}{f(p|\exists^{w})} \quad (9)$$

In contrast to o_1 and o_2, o_2 does not care if the edge has been searched in earlier rounds. Edges that have recently been accessed (i.e., traveled) by the ants won't gain any pheromone boosts. Eq. (10) describes the demonstration of the third stratagem o_3:

$$\beta_{a,z}^{w} = \varphi\beta_{a,z}^{w} + \sum_{p\omega ss}\frac{1}{\exists.q^{F}} \quad (10)$$

where the target vector of the p solution is $q = g(p) = (g_1(p), g_2(p), \ldots . g_c(p))$. Meanwhile, the weight vector for the p solution is $\exists$. The maximum and minimum limits of the pheromones (i.e., (a, z)) are similasrly constrained by o_1, o_2 , and o_3 as shown in Eq.(6) and Eq.(7) . The proposed optimization employs $LGAT$ to regulate the update as mentioned above techniques. The $LGAT$ picks a suitable strategy from the candidate mentioned above to improve the searching efficiency at various stages of the optimization process.

4. ABOUT DATASET AND PERFORMANCE METRICS

4.1 Dataset

The Cardiovascular Disease (CD) dataset and the PIMA Indians Diabetes (PID) dataset, found on the Kaggle website and downloaded for free, are used to evaluate the suggested classifier's performance. The cardiovascular Disease dataset holds 70000 records, and the PIMA Indians Diabetes dataset holds 768 records. Descriptions of both datasets are provided in Table 1 and Table 2.

Table 1. CD Dataset

Feature	Feature Type	Description
Age	Objective	Age of Patient
Height	Objective	Height of Patient in centimeters (cm)
Weight	Objective	Weight of Patient in kilogram (kg)
Gender	Objective	Gender of Patient. 1 indicates Male, 2 indicates Female
Systolic Blood Pressure	Examination	The sudden influx of blood brings on pressure into the arteries. It is measured in millimeters of mercury (mmHg).
Diastolic Blood Pressure	Examination	Arterial blood pressure refers to the force exerted by the blood on the walls of the arteries in between heartbeats. It is measured in millimeters of mercury (mmHg).
Cholesterol	Examination	Cholesterol present in patient blood is indicated in 3 different levels, which are: 1 indicating normal level, 2 indicating above the normal level, 3 indicating well above the normal level
Glucose	Examination	Glucose present in patient blood is indicated in 3 different levels, which are: 1

		indicating normal level, 2 indicating above the normal level, 3 indicating well above the normal level
Smoking	Subjective	Whether the patient has the habit of smoking or not.
Alcohol Intake	Subjective	Whether the patient consumes alcohol or not.
Physical Activity	Subjective	Whether the patient do any physical activity or not.
Cardiovascular Disease Presence/Absence	Subjective	Whether the patient has the disease of Cardiovascular or not

Table 2. PID Dataset

Feature	Feature Type	Description
Preg	Subjective	Indicates the number of times the patient got pregnant.
Gluc	Examination	Glucose present in the patient blood.
Diastolic Blood Pressure	Examination	Arterial blood pressure refers to the force exerted by the blood on the walls of the arteries in between heartbeats.
Triceps Skin Fold Thickness	Objective	Indicates triceps thickness of skin fold.
Insulin	Examination	Insulin serum for 2 hours.
Body Mass Index	Examination	Medical screening tool that compares the height and weight of the patient to identify the presence of fat.
Diabetes Pedigree Function	Subjective	Determines whether or not a patient has a high risk of developing diabetes depending on their family.
Age	Objective	Age of Patient.
Diabetes Presence/Absence	Subjective	Whether the patient has the disease of diabetes or not

4.2 PERFORMANCE METRICS

- **Classification Accuracy (CA):** It indicates how many predictions were accurate from a total of input samples.
- **F-Measure (FM):** The method enables the integration of recall and accuracy into a single metric that accounts for both features.
- **Matthews Correlation Coefficient (MCC):** It evaluates the discrepancies between observed and expected values.
- **Fowlkes-Mallows Index (FMI):** It is a measure to evaluate the similarity between clustering or classification quantitatively. In short, it acts as a geometric means to recall and precision.

The four variables involved in the metrics mentioned above calculation, which are True Positive ($TrPst$), False Positive ($FlPst$), True Negative ($TrNgt$), False Negative ($FlNgt$):

- $TrPst$ = Count of medical cases precisely diagnosed as the patient.
- $FlPst$ = Count of medical cases inexactly diagnosed as the patient.
- $TrNgt$ = Count of medical cases precisely diagnosed as healthy.
- $FlNgt$ = Count of medical cases inexactly diagnosed as healthy.

5. RESULTS AND DISCUSSION

5.1 CA and FM Analysis

Figure 1 compares CA and FM between the existing heart disease classification algorithms (BSVM and S-ANN) and the proposed algorithm (RACOAA). Figure 2 compares CA and FM between the existing diabetes disease classification algorithms (MSVM and DLPD). It is obvious that the RACOAA has attained better CA and FM than the existing algorithms in both the considered datasets. RACOAA is optimized to classify heart disease and diabetes disease.

Journal of Theoretical and Applied Information Technology
28th February 2023. Vol.101. No 4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

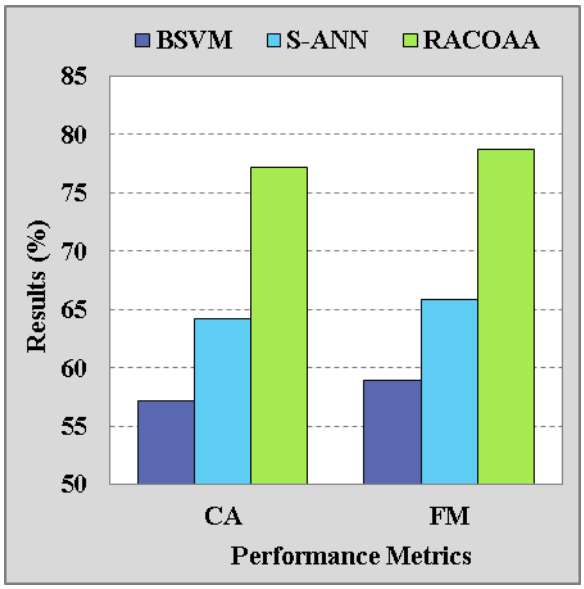

Fig 1. CA and FM Analysis on CD Dataset

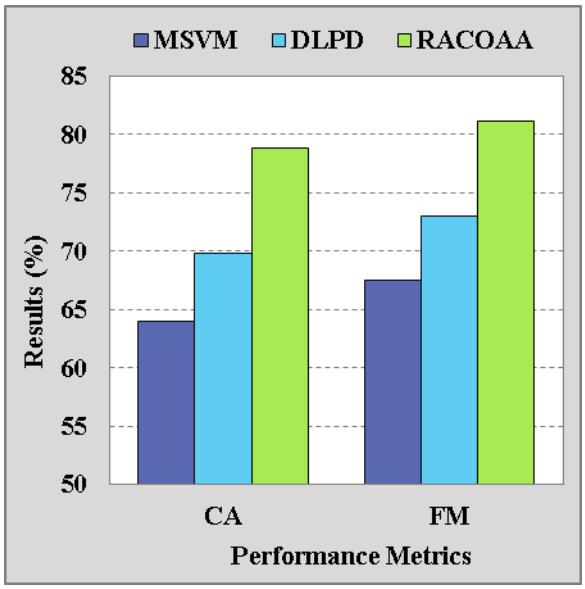

Fig 2. CA and FM Analysis on PID Dataset

Table 3. Result Values CA and FM Analysis on CD Dataset

Metrics → / Algorithms ↓	CA	FM
BSVM	57.134	58.920
S-ANN	64.200	65.856
RACOAA	77.120	78.645

Table 4. Result Values CA and FM Analysis on PID Dataset

Metrics → / Algorithms ↓	CA	FM
MSVM	63.932	67.526
DLPD	69.792	72.960
RACOAA	78.776	81.069

5.2 FMI and MCC Analysis

Figure 3 compares FMI and MCC between the existing heart disease classification algorithms (BSVM and S-ANN) and the proposed algorithm (RACOAA). Figure 4 compares FMI and MCC between the existing diabetes disease classification algorithms (MSVM and DLPD). It is apparent that the RACOAA has attained better FMI and MCC than the existing algorithms in both the considered datasets. RACOAA is optimized to classify heart disease and diabetes disease. Threshold value-based classification in RACOAA leads to better results than the existing algorithms.

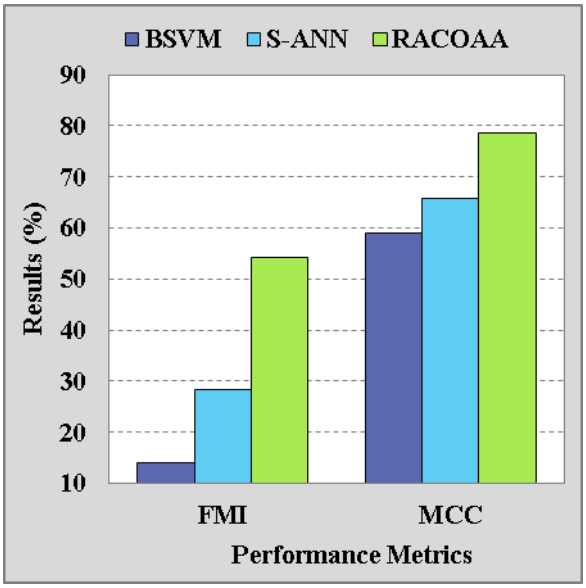

Fig 3. FMI and MCC Analysis on CD Dataset

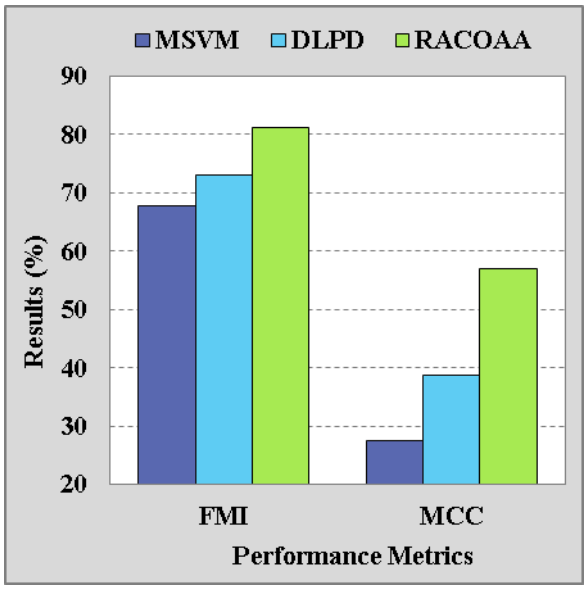

Fig 4. ***FMI*** and ***MCC*** Analysis on PID Dataset

Table 3. Result Values ***FMI*** *and* ***MCC*** *Analysis on CD Dataset*

Metrics → / Algorithms ↓	CA	FM
BSVM	14.124	58.924
S-ANN	28.247	65.860
RACOAA	54.129	78.672

Table 4. Result Values ***FMI*** *and* ***MCC*** *Analysis on PID Dataset*

Metrics → / Algorithms ↓	CA	FM
MSVM	67.657	27.588
DLPD	72.964	38.757
RACOAA	81.081	56.982

6. CONCLUSION

Both heart disease and diabetes (HDD) are significant causes of death worldwide. Understanding the symptoms and signs of these diseases in patients is the first step in stopping their progression. This research work has proposed a bio-inspired optimization-based classification algorithm, Revived Ant Colony Optimization-based Adaboost Algorithm (RACOAA). The RACOAA is influenced by the ants' inherent tendencies to forage. For better classification accuracy, the adaboost algorithm is improved in RACOAA by establishing a threshold value. Drawbacks in threshold setting towards the classification of HDD are overcome by bio-inspired optimization, i.e., Revived Ant Colony Optimization. RACOAA is evaluated against the current classification algorithm using two benchmark datasets with standard machine learning performance metrics. Results indicate that RACOAA has better performance than the current classification algorithms in terms of classification accuracy.

REFERENCES:

[1] M. C. S. Geetha, I. E. Shanthi, and N. S. Sehnaz, "Analyzing the suitability of relevant classification techniques on medical data set for better prediction," in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2017, pp. 665–670. doi: 10.1109/I-SMAC.2017.8058262.

[2] P. Ramprakash, R. Sarumathi, R. Mowriya, and S. Nithyavishnupriya, "Heart Disease Prediction Using Deep Neural Network," in *2020 International Conference on Inventive Computation Technologies (ICICT)*, 2020, pp. 666–670. doi: 10.1109/ICICT48043.2020.9112443.

[3] D. Deepika and N. Balaji, "Effective heart disease prediction using novel MLP-EBMDA approach," *Biomed. Signal Process. Control*, vol. 72, p. 103318, 2022, doi: https://doi.org/10.1016/j.bspc.2021.103318

[4] D. Verma and N. Mishra, "Analysis and prediction of breast cancer and diabetes disease datasets using data mining classification techniques," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 2017, pp. 533–538. doi: 10.1109/ISS1.2017.8389229.

[5] L. Lu, R. R. Sun, M. Liu, Y. Zheng, and P. Zhang, "The Inflammatory Heart Diseases: Causes, Symptoms, and Treatments," *Cell Biochem. Biophys.*, vol. 72, no. 3, pp. 851–855, Jul. 2015, doi: 10.1007/s12013-015-0550-7.

[6] S. Dhar, K. Roy, T. Dey, P. Datta, and A. Biswas, "A Hybrid Machine Learning Approach for Prediction of Heart Diseases," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, 2018, pp. 1–6. doi: 10.1109/CCAA.2018.8777531.

[7] L. Ali *et al.*, "An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure," *IEEE Access*, vol. 7, pp. 54007–54014, 2019, doi: 10.1109/ACCESS.2019.2909969.

[8] P. Sharma, K. Choudhary, K. Gupta, R. Chawla, D. Gupta, and A. Sharma, "Artificial plant optimization algorithm to detect heart rate & presence of heart disease using machine learning," *Artif. Intell. Med.*, vol. 102, p. 101752, 2020, doi: https://doi.org/10.1016/j.artmed.2019.101752.

[9] R. Jaganathan and R. Vadivel, "Intelligent Fish Swarm Inspired Protocol (IFSIP) for Dynamic Ideal Routing in Cognitive Radio Ad-Hoc Networks," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 1063–1074, 2021, doi: 10.12785/ijcds/100196.

[10] J. Ramkumar and R. Vadivel, "Improved Wolf prey inspired protocol for routing in cognitive radio Ad Hoc networks," *Int. J. Comput. Networks Appl.*, vol. 7, no. 5, pp. 126–136, 2020, doi: 10.22247/ijcna/2020/202977.

[11] J. Ramkumar and R. Vadivel, "Whale optimization routing protocol for minimizing energy consumption in cognitive radio wireless sensor network," *Int. J. Comput. Networks Appl.*, vol. 8, no. 4, pp. 455–464, 2021, doi: 10.22247/ijcna/2021/209711.

[12] J. Ramkumar and R. Vadivel, "Multi-Adaptive Routing Protocol for Internet of Things based Ad-hoc Networks," *Wirel. Pers. Commun.*, vol. 120, no. 2, pp. 887–909, Apr. 2021, doi: 10.1007/s11277-021-08495-z.

[13] J. Ramkumar and R. Vadivel, "Performance Modeling of Bio-Inspired Routing Protocols in Cognitive Radio Ad Hoc Network to Reduce End-to-End Delay," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, pp. 221–231, 2019, doi: 10.22266/ijies2019.0228.22.

[14] P. Menakadevi and J. Ramkumar, "Robust Optimization Based Extreme Learning Machine for Sentiment Analysis in Big Data," *2022 Int. Conf. Adv. Comput. Technol. Appl. ICACTA 2022*, pp. 1–5, Mar. 2022, doi: 10.1109/ICACTA54488.2022.9753203.

[15] J. Ramkumar, C. Kumuthini, B. Narasimhan, and S. Boopalan, "Energy Consumption Minimization in Cognitive Radio Mobile Ad-Hoc Networks using Enriched Ad-hoc On-demand Distance Vector Protocol," *2022 Int. Conf. Adv. Comput. Technol. Appl. ICACTA 2022*, pp. 1–6, Mar. 2022, doi: 10.1109/ICACTA54488.2022.9752899.

[16] J. Ramkumar and R. Vadivel, "Meticulous Elephant Herding Optimization based Protocol for Detecting Intrusions in Cognitive Radio Ad Hoc Networks," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 8, pp. 4548–4554, 2020, doi: 10.30534/ijeter/2020/82882020.

[17] J. Ramkumar, "Bee inspired secured protocol for routing in cognitive radio ad hoc networks," *Indian J. Sci. Technol.*, vol. 13, no. 30, pp. 2159–2169, 2020, doi: 10.17485/ijst/v13i30.1152.

[18] J. Ramkumar and R. Vadivel, "Improved frog leap inspired protocol (IFLIP) – for routing in cognitive radio ad hoc networks (CRAHN)," *World J. Eng.*, vol. 15, no. 2, pp. 306–311, 2018, doi: 10.1108/WJE-08-2017-0260.

[19] J. Ramkumar, R. Vadivel, and B. Narasimhan, "Constrained Cuckoo Search Optimization Based Protocol for Routing in Cloud Network," *Int. J. Comput. Networks Appl.*, vol. 8, no. 6, pp. 795–803, 2021, doi: 10.22247/ijcna/2021/210727.

[20] J. Ramkumar and R. Vadivel, "CSIP—cuckoo search inspired protocol for routing in cognitive radio ad hoc networks," in *Advances in Intelligent Systems and Computing*, 2017, vol. 556, pp. 145–153. doi: 10.1007/978-981-10-3874-7_14.

[21] M. Ganesan and N. Sivakumar, "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019, pp. 1–5. doi: 10.1109/ICSCAN.2019.8878850.

[22] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, "Classification and Prediction of atherosclerosis diseases using machine learning algorithms," in *2019 5th International Conference on Optimization and Applications (ICOA)*, 2019, pp. 1–5. doi: 10.1109/ICOA.2019.8727688.

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

[24] Y. Zhou, Y. Lu, and Z. Pei, "Intelligent diagnosis of Alzheimer's disease based on internet of things monitoring system and deep learning classification method," *Microprocess. Microsyst.*, vol. 83, p. 104007, 2021, doi: https://doi.org/10.1016/j.micpro.2021.104007.

[25] J.-J. Beunza *et al.*, "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)," *J. Biomed. Inform.*, vol. 97, p. 103257, 2019, doi: https://doi.org/10.1016/j.jbi.2019.103257.

[26] S. Hashem *et al.*, "Machine Learning Prediction Models for Diagnosing Hepatocellular Carcinoma with HCV-related Chronic Liver Disease," *Comput. Methods Programs Biomed.*, vol. 196, p. 105551, 2020, doi: https://doi.org/10.1016/j.cmpb.2020.105551.

[27] R. Indrakumari, T. Poongodi, and S. R. Jena, "Heart Disease Prediction using Exploratory Data Analysis," *Procedia Comput. Sci.*, vol. 173, pp. 130–139, 2020, doi: https://doi.org/10.1016/j.procs.2020.06.017.

[28] A. N. Repaka, S. D. Ravikanti, and R. G. Franklin, "Design And Implementing Heart Disease Prediction Using Naives Bayesian," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 292–297. doi: 10.1109/ICOEI.2019.8862604.

[29] C. Sowmiya and P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques," in *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, 2017, pp. 1–5. doi: 10.1109/ITCOSP.2017.8303115.

[30] M. A. Jabbar and S. Samreen, "Heart disease prediction system based on hidden naïve bayes classifier," in *2016 International Conference on Circuits, Controls, Communications and Computing (I4C)*, 2016, pp. 1–5. doi: 10.1109/CIMCA.2016.8053261.

[31] E. Owusu, P. Boakye-Sekyerehene, J. K. Appati, and J. Y. Ludu, "Computer-Aided Diagnostics of Heart Disease Risk Prediction Using Boosting Support Vector Machine," *Comput. Intell. Neurosci.*, vol. 2021, p. 3152618, 2021, doi: 10.1155/2021/3152618.

[32] S. Nandy, M. Adhikari, V. Balasubramanian, V. G. Menon, X. Li, and M. Zakarya, "An intelligent heart disease prediction system based on swarm-artificial neural network," *Neural Comput. Appl.*, 2021, doi: 10.1007/s00521-021-06124-1.

[33] S. Kang, "Personalized prediction of drug efficacy for diabetes treatment via patient-level sequential modeling with neural networks," *Artif. Intell. Med.*, vol. 85, pp. 1–6, Apr. 2018, doi: 10.1016/J.ARTMED.2018.02.004.

[34] S. S. G. and M. K., "Diagnosis of diabetes diseases using optimized fuzzy rule set by grey wolf optimization," *Pattern Recognit. Lett.*, vol. 125, pp. 432–438, 2019, doi: https://doi.org/10.1016/j.patrec.2019.06.005.

[35] F. Mercaldo, V. Nardone, and A. Santone, "Diabetes Mellitus Affected Patients Classification and Diagnosis through Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 112, pp. 2519–2528, 2017, doi: https://doi.org/10.1016/j.procs.2017.08.193.

[36] K. Kannadasan, D. R. Edla, and V. Kuppili, "Type 2 diabetes data classification using stacked autoencoders in deep neural networks," *Clin. Epidemiol. Glob. Heal.*, vol. 7, no. 4, pp. 530–535, 2019, doi: https://doi.org/10.1016/j.cegh.2018.12.004.

[37] T. Haq, T. Ahmed, Z. A. Latif, M. A. Sayeed, and S. M. Ashrafuzzaman, "Cardiac autonomic neuropathy in patients with type 2 diabetes mellitus having peripheral neuropathy: A cross-sectional study," *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 13, no. 2, pp. 1523–1528, Mar. 2019, doi: 10.1016/J.DSX.2019.03.015.

[38] C. M. Li *et al.*, "Synchronizing chaotification with support vector machine and wolf pack search algorithm for estimation of peripheral vascular occlusion in diabetes mellitus," *Biomed. Signal Process. Control*, vol. 9, no. 1, pp. 45–55, Jan. 2014, doi: 10.1016/J.BSPC.2013.10.001.

[39] H. Hegde, N. Shimpi, A. Panny, I. Glurich, P. Christie, and A. Acharya, "Development of non-invasive diabetes risk prediction models as decision support tools designed for application in the dental clinical environment," *Informatics Med. unlocked*, vol. 17, Jan. 2019, doi: 10.1016/J.IMU.2019.100254.

[40] J. A. Carter, C. S. Long, B. P. Smith, T. L. Smith, and G. L. Donati, "Combining elemental analysis of toenails and machine learning techniques as a non-invasive diagnostic tool for the robust classification of type-2 diabetes," *Expert Syst. Appl.*, vol. 115, pp. 245–255, Jan. 2019, doi: 10.1016/J.ESWA.2018.08.002.

[41] C. Ionescu-Tirgoviste, P. A. Gagniuc, and E. Gagniuc, "The electrical activity map of the human skin indicates strong differences between normal and diabetic individuals: A gateway to onset prevention," *Biosens. Bioelectron.*, vol. 120, pp. 188–194, Nov. 2018, doi: 10.1016/J.BIOS.2018.08.057.

[42] H. Wu, S. Yang, Z. Huang, J. He, and X. Wang, "Type 2 diabetes mellitus prediction model based on data mining," *Informatics Med. Unlocked*, vol. 10, pp. 100–107, Jan. 2018, doi: 10.1016/J.IMU.2017.12.006.

[43] S. Thenappan, M. Valan Rajkumar, and P. S. Manoharan, "Predicting Diabetes Mellitus Using Modified Support Vector Machine with Cloud Security," *IETE J. Res.*, pp. 1–11, Jul. 2020, doi: 10.1080/03772063.2020.1782781.

[44] H. Zhou, R. Myrzashova, and R. Zheng, "Diabetes prediction model based on an enhanced deep neural network," *EURASIP J. Wirel. Commun. Netw.*, vol. 2020, no. 1, p. 148, 2020, doi: 10.1186/s13638-020-01765-7.