Predicting Microvascular Complications in Diabetic Mellitus Using Improved Enhanced Coati Optimizer

[*1]Ms.Mayuri Kulkarni, [2]Dr. Shailesh Deore

[*1a] Research Scholar,
Department of Computer Engineering,
SSVPs's B.S. Deore College of Engineering, Dhule
K.B.C. North Maharashtra University, India
b Assistant Professor, Department of Computer Engineering,
SVKM's Institute of Technology, Dhule, Maharashtra 424001, India.
[*]Email: mayurikulkarni@gmail.com
[2]Research Guide & Associate Professor,
Department of Computer Engineering,
SSVPs's B.S. Deore College of Engineering, Dhule
K.B.C. North Maharashtra University, India.
Email: shaileshdeore@gmail.com

Abstract:

Objective:** Diabetes complications are classified as Macro and Microvascular Diseases. Microvascular complications in type 2 Diabetic patients commonly occur as diabetic retinopathy, diabetic neuropathy, and diabetic nephropathy. Therefore detecting these microvascular complications from the clinical dataset is very important. Method: In this paper, a machine learning model is proposed for predicting and detecting microvascular diseases in type 2 diabetic Patients. The initial stage is preprocessing where data processing . After the preprocessing operation is performed feature selection process is carried out using the Improved Enhanced Coati algorithm. The optimal features from the Improved Enhanced Coati Optimization algorithm are applied to various classification algorithms. The reason behind applying this feature selection algorithm to various models is to check the performance with the traditional classifiers. Hence model performance is compared with XGB, KNN, SVM, RF, AdaBoost, Tree, and ANN algorithms. Findings: For the classification of diabetic retinopathy, the selected features are age, sex, BMI, BP, FPS, Family History, and Medical Adherence. Similarly, the features selected to classify Diabetic Nephropathy as Sex, SP, FPS, Family History, Onset Age, and HbA1C and FPS used to classify Diabetic Neuropathy. On optimal selection of features various ML classification algorithms are applied. The results are compared with algorithms as XGB, KNN, SVM, RF, AdaBoost, Tree, and ANN. The results are measured by considering parameters for training and testing accuracy and Random Forest Classifier results are optimal for the AdaBoost estimator for type 2 diabetic patients for the diabetic retinopathy is 99.9% and 94.78%, diabetic nephropathy, and diabetic neuropathy is 99.8% and 95.44%. Novelty: In the proposed methodology the feature selecting fitness function is selected based upon the received optimal accuracy from the feature selecting estimator as AdaBoost. In Coati Optimizer the feature selection process is carried out by selecting a fitness function that provides the minimum error.

***Keywords:** Enhanced Coati Optimizer, Feature Selection, Microvascular Complications, Machine learning classification algorithms, Bio-inspired algorithms.*

Introduction:

Diabetes is a long-term condition that continues to be an expanding and international concern as it affects the health of all people worldwide[55]. It is a common disease caused by to metabolic condition that causes elevated glucose levels. Basically, in type 2 diabetes pancreas makes insulin but cells are inefficient to resist as they should. Hence glucose can't get into cells and raises the blood sugar level [1]. This long-term prevalence of diabetes affects human organs and is responsible for microvascular and macrovascular complications in patients. Due to macro and microvascular complications in type 2 diabetic patients various health issues may occur. Hence the statically understanding of the count of the diabetic patient is very essential. As per the IDF Diabetes Atlas 10^{th} edition, in the age group 20-79 overall world the diabetic patent count is predicted to be 643 million by 2030. And the specified count of mortality due to diabetic complications is 6.7 million which is a serious issue [3].The possible complications in diabetic type 2 patients are categorized into macro and microvascular complications. These complications are occurred due to the long-term presence of diabetes in type 2 patients [39].

Macrovascular challenges originate from damage in the enormously blood arteries of the circulatory system, brain, and legs [4]. This causes chronic coronary artery disease, peripheral arterial disease, and neurological illness. [39]. Similarly, in diabetic patients, if the complication affects the small blood arteries in organs such as the retina, kidney, or nerves, it creates microvascular complications such as diabetic retinopathy, nephropathy, and neuropathy[2]. These diabetic complications leading impact morbidity and mortality.

The paper focuses on microvascular diabetic complications which are very commonly occurring in type 2 patients. Therefore early prediction of diabetic retinopathy, diabetic nephropathy, and diabetic neuropathy is very essential. For this medicinal practitioners regularly conduct the health checkup. In that checkup process patients need to appear for specified clinical tests. Nowadays this data is stored in the form of digital forms called medical records of the patients, hospital records, and clinical reports of the patients. By using these digital records analysis of the patient's health data is possible. Similarly, this record is analyzed by the doctors. Based on doctors' experience and knowledge diagnosis is done which can be inaccurate as it's a manual decision. In some cases, unnoticed hidden patterns affect the decision-making of doctors. Therefore machine learning algorithms help doctors to understand the hidden patterns and help in making decisions accurately [5, 6].

The clinical parameters that are analyzed for predicting diabetic microvascular complications are the duration of diabetes, age, sex, Hb1AC, Blood pressure, Eating habits, stress, family History, etc. Due to these factors, it is very easy to predict the probable occurring complications in patients using machine learning systems. Data-driven approaches utilizing supervised machine learning techniques for identifying patients with these conditions.

Several researchers have attempted to develop a precise diabetes model for prediction over the years. However, this area still faces substantial open research challenges due to a lack of relevant knowledge sets and prediction tools, pushing researchers to use analytics- and machine learning (ML)-based methodologies. Using four different machine learning algorithms, the inquiry seeks to solve problems and examine healthcare prediction analytics[55]. Machine learning perdition algorithms have been developing and reliable in recent years. These algorithms are used to preprocess and select relevant features from the datasets and automate the prediction. Machine learning models allow us to identify the risk of microvascular and microvascular complications [7]. Statistical tests such tests are ANOVA, and descriptive analysis techniques are used to predict diabetic complications [8].ML predictive models use the standard deviation and mean for the performance tree, and for validation use c-statistics [9].

These Machine learning techniques significantly influence the model performance and shape new horizons for prediction, and outcome definition [9].
The paper discusses a novel approach suggested for the feature selection technique in microvascular complications. Later these optimal features are provided to the various classification algorithms and results are measured on various measuring parameters.

Literature survey:

Various research uses the machine learning algorithm for predicting complications in the healthcare sectors. The researcher developed various learning models for predicting the risk of diseases. In type 2 diabetic patients various researchers are developing machine learning models for predicting diabetic retinopathy, diabetic neuropathy, and diabetic nephropathy. Related work concerning microvascular complications is discussed in this section.

Cichosz et al suggested the predictive model for long-term chronic diseases caused due to diabetes based on the predictive models. Therefore the suggested model is based upon the data pattern as a multiple logistic model which can find the depended complications based on the independent parameter [10].

Sarah Kanbour et al discussed machine learning models for predicting diabetic retinopathy, diabetic kidney, and diabetic neuropathy in type 2 diabetic patients. These microvascular complications are predicted from 256 features and [9]

Dagliati et al proposed a data mining and predictive model for the type 2 microvascular complication prediction model with center profiling for assessing the features and predictive model constructed using RF imputation algorithm as miss Forest. RF imputation uses 100 trees and a maximum of 100 iterations using RMSE and RMSEN on missing values. To validate the results, the Leave-one-out strategy is used [11]. The prevalence of microvascular complications is predicted using log-binomial and Poisson regression. The model suggests prevalence due to different pathophysiological mechanisms [12].

Vamsi et al proposed a machine learning model RF with a Decision tree that provides the optimal results for classifying microvascular and macrovascular complications. The network for diabetic complications identifies the relationship between the patients and the common health complications.[13]

In Smart Framework researchers suggested a machine learning model to predict the prevalence of. Microvascular diseases in patients based on the decision theory. In a smart framework, the prediction parameters are auto-stunned, and hence the uncertainty and Errors exist in the selection model. The decision theory consists of three elements plausible features, alternative decisions, and objectives. The uncertainties are associated with the decision-making using the likelihood prediction technique for classifying multiple replications. Hence this framework works upon data-driven model selection [14].

Nicolucci A. et al, suggested a supervised machine learning model as XGBoost for predicting six groups of diabetic complications retinopathy, neuropathy, nephropathy, cardiovascular diseases, cerebrovascular, and peripheral vascular disease. This model helps in finding the greater risk of diabetic complications using XGBoost. This improves the quality of diabetic care [15].

The researchers Branimir L et al, suggested RNN LSTM and RNN GRU are designed to predict microvascular complications and results are compared with random forest and multilayer perceptron for diagnosing and designing are compared with the random forest and multiple perception [16].

The risk engine suggested by Hui Shao et al is an analytical tool that collects a large amount of the population which allows the simulation to indicate the progression of diabetes using the BRAVO model using the ACCORD trial database to predict the complications due to diabetes [17].

Weiss et al suggested a machine learning model for predicting microvascular complications from electronic health data by calculating risk scores. Risk calculation using a regularized Cox model and wavelet reconstruction network. The c-statistics test is used for predicting the risk of microvascular complications [18].

Rashid et al used demographic, clinical, and laboratory datasets for predicting microvascular complications. For feature selection applied chi-square test. After selecting features from the chi-square test these features are applied to various machine learning models and compared to the performance of the model. The diagnosed complications as CAN, DPN, NEP, and RET [40].

Dagliati et al, suggested data mining pipelined model comprises 5 stages central profiling, predicting model targeting, predicting model construction, and model validation. The central profiling stage focuses on derivative prediction. In the second stage of pipelining predictive model construction works on the dataset collected from the physician and set threshold values for microvascular complications as 3,5, and 7 years. In predictive model construction data imputation is performed using mean, mode, median, and miss forest. Prediction model validation uses LOO, Sensitivity, specificity, accuracy, PPV, NPV and ROC, and MCC [11].

Vamsi et al designed a classification model for micro and macrovascular disease prediction. The data preprocessing step executes by deleting rows and columns where more than 70% of data is missing. Feature selection is performed using a chi-square test and applied to a machine-learning model where a random forest with a decision tree provides more accuracy [13].

XGBoost algorithm is used for classifying diabetic complications as micro and macrovascular diseases. The work predicts the probability of any micro and macrovascular complications in type 2 diabetic patients throughout 3, or 5 years and model evaluation uses performance measures such as ROC and AUC [42].

Missing values are calculated using various classifiers and more than 60% of missing values found in such data are deleted from the database. Data imputation uses three approaches replacing missing values by mean, k-NN model, and missForest. Later RMSE was calculated for the dataset and missForest results are more reasonable. Balancing of the dataset is carried out by using SMOTE analysis. The feature selection process is carried out using Logistic Regression and SVM linear classifier and classification results are best for RF, AdaBoost, and XGBoost classifiers [41].

Haque et al used LMAV and NSV methods for feature extraction for EMG in the time domain.GRF feature extraction is carried out using Discrete Wavelet decomposition. Feature selection is performed using the chi-squared test. PCA is used for dimensionality reduction purposes. Later machine learning algorithms such as DAC, ECM, KCM, KNN, LCM, NBC, SVM, and BDC classification algorithms are used[42].

Jelinek et al, suggested automated detection of diabetic neuropathy using machine learning. ECG signals are applied to multi-scale Allen factors to determine heart rate variability. This uses a based learning system for diagnosis purposes [43].

Author	***Feature selection algorithm***	***Diagnosed Microvascular complication***

Rashid et al [40]	Chi-square test	CAN, DPN, NEP, and RET
Dagliati et al[11]	RF approach	Nephropathy, neuropathy and retinopathy
Vamsi et al[13]	Chi-square test	NEP, NEU, RET, CVD, PVD
Jian et al [41]	LR and Linear SVM	Metabolic Syndrome, dyslipidemia, NEP, NEU, diabetic foot, hypertension, obesity, and RET.
Haque et al[42]	Chi-squared	Diabetic Neuropathy
Jelinek et al [43]	Multi-scale Allen Factor	Diabetic Neuropathy

Feature selection is the process of selecting subsets of features from the available features based on certain selection criteria. Generally, bio-inspired algorithms are inspired by physical properties. These selection algorithms are used for dimension reduction. These algorithms are Genetic algorithms, Evolutionary Strategies, and Differential Evolution. Swarm Intelligence is an emerging paradigm that uses adaptive systems and is based on a genetic adaptation of organisms and the social behavior of organisms.

Some suggested feature selection algorithms are Artificial Bee Colony Algorithm, Ant colony ,Fish Swarm Algorithm, Artificial Immune system algorithm, Firefly algorithm, Group Search optimizer, Shuffled Frog Leaping Algorithm, PS20, Intelligent Water Drops algorithm, and many more[23]. These algorithms are discussed below:

1. Particle Swarm Optimization: A population-based algorithm modeled by the social movements of birds or fish schools. Birds' social behavior is to flock or roost. The individual particles iteratively modify the entire solution. Each particle searches the space with an identifiable velocity in the same direction and has its impact by its own best location found so far, the best solution, and the global solution[36].

2. Ant colony algorithm:It is a phase algorithm in which the solutions are searched by the ants to construct an Ant Solution and these solutions are improved through local search in apply local search phase and finally updated in update pheromones[27-28]. Ant Colony Optimization uses two approaches pheromone and decision making. The pheromone variable is associated with each edge and can be read and updated by ants. The pheromone value is associated with the solution component. The ACO optimization algorithms' most successful variants are the MAX-MIN ant system and Ant colony system [32].

3. Glow-worm Swarm Optimization : This algorithm suggests multiple optima of multimodal function which makes it different from the PSO and ABC algorithms. This calculates the fitness of the current location and accordingly calculates the objective function into luciferin value which is broadcasted to neighbors. Identification of neighbor and movement computation exploited using adaptive neighborhood by sensor range [37]. This algorithm solves the issue by locating a global optimum solution for instances that suffer from the issue as low accuracy, local optimum, convergence of success rate, and reduced speed [38].

4. Artificial Bee Colony Algorithm:The Artificial Bee Colony Algorithm is based on the particular intelligent behavior of the honeybee. In this algorithm, three groups of bees exist employed bees, onlookers, and scouts. The employed bee and onlookers bees search the food and scout group present at the hive. This algorithm converts the problem of finding to best parameter which minimizes an objective function [33].

5. Zebra Optimization Algorithms:The name suggests that it is inspired by the Zebra Optimization algorithm. The performance of the algorithm is evaluated on 68 benchmark functions. These functions are unimodal, high, fixed dimensional multimodal, according to Congress on Evolutionary Computation Standards 2015 and 2017. In ZOA two natural behaviors of Zebras in wild animals are considered foraging and defense strategies against predators. In the foraging phase pioneer zebra is the best member and leads the population towards the search space and in Defence strategies against predator attacks update the position of the population in the search space. In this problem, each zebra is a candidate solution for the problem which searches space for the problem. While considering the predator attack there are two possibilities attack by a lion or other than a lion. In case of a lion attack zebra chooses to escape and for others, it selects to offensive strategy [26,56].

6. Bacterial Foraging Optimization:Bacterial foraging is inspired by the theory of foraging means animals search for nutrients and obtain nutrients to maximize energy intake. This uses E.coli bacteria and models chemotaxis, tumbling, and swimming behavior to navigate search space and to find an optimal solution [31].BCF combines chemotaxis with cell-to-cell which improves the speed of bacterial colony. The area concentrates on search and maintains diversity in the search process. BCF models are a self-adaptive foraging strategy that dynamically balances exploration and exploitation behavior and enables information sharing among ant colonies [34,57].

7. Cuckoo Search Algorithm (CSA):The Cuckoo Search Algorithm is based on obligated brood parasite habits of several cuckoo species as well as the Levy fight behavior of some birds and fruit flies. Parasitic cuckoos select a nest where the host bird just laid down its eggs which increases the cuckoo's chick's share of food provided by its host bird. The cuckoo search follows three idealized rules as [44,58]

1. Laying one egg at a time and dumping it to a randomly chosen nest.
2. The best nest for high-quality eggs carries over to the next generation
3. Several available host nests are fixed and the egg laid by the cuckoo is discovered by the host with probability 0 or 1.

8. Cuttlefish optimization algorithm:The method replicates the cuttlefish's color-changing behavior mechanism, it is subsequently used to solve numerical global optimization problems. This algorithm is associated with two processes reflection and visibility. These processes are achieved by stacking Chromophores, Iridophores, and Leucophores. The Cuttlefish Optimization algorithm initializes the population with random solutions and calculates and keeps the best solution and average values of the best solution point. The interaction among chromatophores and iridophores cells in global search cases produces the reflection and visibility of the entire search space and escapes local optima. The iridophores cell calculates the new solution based on reflected light from the best solution and visibility matches the local search. The leucophores in the Enhanced local search case are responsible for producing a new solution. The leucophores operate in mimic cases and are responsible for a random solution by reflecting incoming light called global search [52].

9. Salp Swarm Algorithm:The Salp Swarm algorithm has characteristics such as competency, flexibility, and simplicity. SSA is a stochastic algorithm that has a considerable number of random components which improves the performance of the metaheuristic algorithm. In the Salp swarm algorithm population randomly searches the space corresponding to the dimensions of the problem. The leader salp updates the position as per the best objective function value in the current iteration. Leader Salp searches for the optimal solution using mathematical functions by considering the current and random values. As per the leader Salp's position followers, Salp updated their position and position chain and tried to be closer to a better Solution. The objective function values of salps are evaluated after each update. The iteration continuous until the stopping criteria is not met. This stopping criterion is a maximum number of iterations or achieving a desirable level of accuracy [51].

10. Fish Swarm Algorithm:The Fish Swarm algorithm is a metaheuristic optimization technique inspired by fish schools from the behavior of collective movement and social behaviors. The steps in the fish swarm algorithm are initializing the search space corresponding to the problem dimension. This assigns fitness value to the objective function. In the movement step preying, swarming, and following such steps exist. In the preying stage, the position of fitness value is updated based on the current position. In the swarming stage, it avoids overcrowding and in the following step, fish with lower fitness tend to be close to higher fitness value. Later evaluation of each fitness is carried out using an objective function [53].

11. Intelligent Water Drops Algorithm :Inspired by the drop of water flowing in the river. This drop of water flows with lots of twists and turns in the river with its two main properties as velocity and soil. IWD has two kinds of parameters to remain constant during the lifetime called static parameters and another parameter as dynamic which reinitializes each iteration. The basic principle of the intelligent water drops algorithm is to populate the "water drops".The water drops are used by the algorithm as virtual agents in a virtual setting.
The "soil" that each drop of water brings, and each drop represents a possible way to solve the issue at hand. Water droplets represent various options or states in the problem as they flow across a network of nodes. They change the quantity of "soil" they carry and may even improve their solutions as they travel and interact with the surroundings and other water droplets[54].

Sr. No.	Algorithm	Computational Complexity
1	Ant Colony[30]	$O(n^2\ 2*m)$ where n: number of nodes in the graph m: number of ants in colony[28].
2	Artificial Bee Colony Algorithm[29]	Initialization O(S*N) where S: Colony Size and N: Number of Variables. Employed Bee Phase: O(S*N) Onlooker Bee Phase: $O(S^2)$ Scout Bee Phase: O(S) Worst Case: O(S*N)*iterations
3	Zebra Optimization Algorithm[26]	Initialization O(N.m) where N: Number of Zebras and m: number of problem variables Update Process Complexity: O(2*N*m*T) Total Computational Complexity: O(N*m*(1+2*T))
4	Bacterial Foraging Optimization[35]	θ(n) successive iterations θ (t) The bacterium lives in continuous time and at t-th instant its position

These algorithms are used in healthcare for feature selection and classification. The role of Bio-inspired feature selection and classification algorithms is discussed in below:

S. Murugesan et al used a CAD system to diagnose whether the prevalence of the disease had been developed or not. For feature selection purposes used algorithms are cat swarm optimization, kill herd, and bacterial foraging. After selecting optimal features from these to classify the disease used classification algorithm a Support Vector Machine algorithm is used [24].

Sakri et al used PSO for predicting features. The PSO algorithm can explore optimal solutions as the particle can explore different parts of solution space. This stores the feature selection in the memory and knowledge of solution based on particle fly within the problem space. PSO performance is unaffected by the dimension of the problem. PSO selects best-fit features that are applied to the classification model [25].

The researchers provided various models which are lacking due to optimal feature selection. The measure parameters don't provide the optimal solution for microvascular prediction. Hence the paper suggests a bio-inspired feature selection model Enhanced Coati Optimization model, which selects solutions based on the fitness function. These fitness functions are calculated from global and local solutions. These three feature selection algorithms add three back propagation neural networks.

Proposed Methodology:

Through the identification and selection of the most pertinent and useful features from a dataset, feature selection algorithms significantly contribute to the performance and comprehension of predictive models in the context of healthcare labeling difficulties. Hence the researchers use feature selection as the primary and important step in predicting microvascular complications in diabetic type 2 patients.

Correlation-based feature selection works upon heuristics search strategy by evaluating the appropriate correlation measures. This finds the similarity measures among the two features in between the +1 and -1 correlation coefficients [45]. This feature selection techniques are used for various evaluation measures. These evaluation measures are Information theory and consistency-based measures. Correlation-based feature selection selects the highest correlation with the target variable and the lowest correlation with each other.

Genetic algorithms are used for feature selection which selects the best features by simplifying mathematical models [46]. The genetic algorithm solves constrained and unconstrained optimization problems. These genetic algorithms are population-based and heuristic methods that are inspired by man [47]. Genetic algorithms use three important steps selection, reproduction, and termination. In a genetic algorithm, Individuals within the population compete for resources and mates. Individuals who are successful (fittest) mate to produce more children than others. Genes from the "fittest" parent are passed down across generations, which means that parents may produce offspring that outperform either parent. Thus, each subsequent generation is more matched to their surroundings [48].

To achieve a Pareto front of non-dominated solutions with both low cost and high classification performance, a PSO-based multi-objective feature selection strategy is used for the cost-based situation. We examine HPPSOFS, a multi-objective approach to PSO with hybrid mutation, to reach this objective [49].

Discrete issues with partial information and poor computational power are referred to as optimization difficulties. These difficulties can be overcome using Meta-heuristic algorithms. Basically, Meta-heuristic algorithms take into account three elements:
Three factors influence an objective function[50]:

(a) it is maximized or minimized;
(b) a set of unknowns or variables impacts the objective function; and

(c) a set of restrictions, on which optimization problems are focused, allows the unknown to accept some values and exclude others.

The proposed system is a bio-inspired feature selection model. This is metaheuristic bio-inspired by the attacking and hunting behavior of iguanas called exploration and escape from predators called exploitation [19]. The fitness function selects the optimal features that are based on the classification algorithm which provides greater accuracy. Figure 1 shows the workflow of the proposed system. The Enhanced Coati Optimizer [20] finds a robust and efficient algorithm. It uses the balanced exploration and exploitation called Levy flight search, Adaptive learning rate, and information sharing.

The proposed method uses a Feature selection algorithm as Enhanced Coati Optimized. To this, the applied feature estimators' algorithms are KNN, SVM, AdaBoost, and Tree. The optimal results from the estimator are later applied to the Classification algorithm to classify the microvascular disease

In the proposed system, the levy flight search uses the estimator functions as KNN, SVM, AdaBoost, and Tree. These algorithms search the efficient patterns. Later stage apply adaptive learning which adjusts the iterations based upon the selection of the estimators which improves the exploitation. Information sharing among the coatis enhances the collaboration. The details of the Improved Enhanced Coati algorithms are explained below:

The algorithm for the Improved Enhanced Coati Algorithm is discussed below:

Algorithm: Proposed System

Step 1: Select the data set and divide it for training and testing

Step 2: Apply Feature Estimator for the Enhanced Coati Optimizer

a. Apply estimator as KNN, SVM, AdaBoost and Tree

b. Compare the Accuracy of the each estimator

c. Select the Optimal accuracy providing estimator.

In this algorithm, each coati stores the information about its current position called local solution and fitness quality Global solution. The sharing of the information among each other is carried out during each iteration and hence position storage and fitness value storage is done. While performing these processes in the improved Enhanced Coati algorithm based on the current understanding called local solution the information is carried forward to explore the new promising area. Later the Local solution, current solution, and global solution are compared to refine the Global solution.

These information-sharing probabilities are based on fitness-based sharing which is the better solution and provides the higher quality of solution.

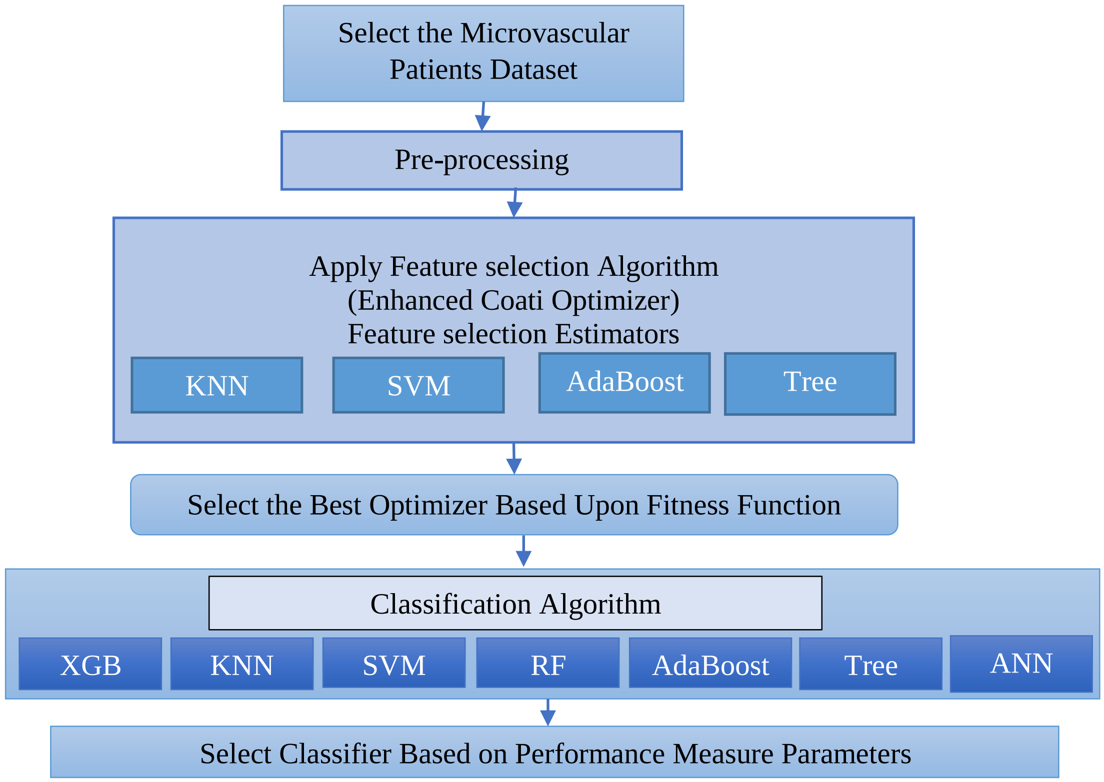

Figure 1 Proposed Approach for Predicting Microvascular Disease

Results and Discussions:

The proposed feature selection algorithm provides optimal results on the Adaboost estimator. Adaboost classifier based on the Meta-learning method. It is an ensemble classification algorithm that uses multiple weak learners. These weak learners calculate the weights and greater weights are given to continue to train the model until a smaller error is returned. This provides more flexibility for the feature set [21].

The design of the AdaBoost estimator for estimating features is determined by the prior knowledge. This a priori knowledge is available in the domain learning problem for decision stumps. These decision stumps are defined on the index of the features that cut the threshold and sign of the decision [22].

Positive Stump decision expressed below in equation 1-

$$h_{j,\theta+}(x) \triangleq 2I_{\{x^{(j)} \geq \theta\}}\ {}_{-1=}\left\{\begin{matrix} 1 \\ -1 \end{matrix}\right.\ if\ x^{(j)} \geq \theta \ldots\ldots\ldots eq(1)$$

The negative Stump decision is expressed below in Equation 2-

$$h_{j,\theta-}(x) \triangleq -h_{j,\theta+(x)=}\ 2I_{\{x^{(j)} < \theta\}}\ {}_{-1=}\left\{\begin{matrix} 1 \\ -1 \end{matrix}\right.\ if\ x^{(j)} < \theta \ldots\ldots\ldots eq(2)$$

In the coati algorithm, the optimal solution is found based upon predefined search space not concerning the most relevant features. The main purpose of the coati optimizer is to select the best solution within the predefined search space.

The Adaboost estimator is an ensemble model that builds strong classifiers from several weak classifiers. For that, it uses the weak classifiers in the series and builds the training model first. Another model works upon the error correction which is present in the first model. This error-correcting and generating results from weak classifiers is continuous and the model is added until the complete training set is not predicting correctly. In short, it adds the maximum number of models are added. Due to this approach, the model accuracy improves also the model doesn't suffer from overfitting issues. Apart from its accuracy and overfitting issues, it deals with the problem of imbalanced data due to the boosting method. This estimator increases the interpretability of the model by dividing the model for the decision-making process into multiple processes.

In the Adaboost estimator, estimation works on the stage-wise gathering of multiple weak learners together for the formation of a strong classifier. The value of the second model in the Adaboost ensemble learning model is indirectly proportional to the error of the weak learner.

The mathematical model for the Adaboost estimator is given below in Equation 3:

$$S(x) = sign\left(\sum_{\mu=1}^{M} \lambda_\mu \emptyset_\mu(x)\right) \ldots\ldots\ldots eq(3)$$

Where λ_μ: weights for learning, when $\lambda_\mu = 0$ corresponding classifier $\emptyset_\mu(x)$ is not selected.

The weak classifiers' data $\emptyset_\mu$(x) divides the data into two parts correctly classified or wrongly classified. The mathematical expression for correctly classified weak classifiers is expressed in equations 4 and 5 respectively.

$$W_\mu^+ = \left\{i : y^i \emptyset_\mu\left(x^i\right) = 1\right\} \ldots\ldots\ldots eq(4)$$

$$W_\mu^- = \left\{i : y^i \emptyset_\mu\left(x^i\right) = -1\right\} \ldots\ldots\ldots eq(5)$$

The weights of training data are expressed in Equation 6:

$$D_i^t = \frac{e^{-y^i} \sum_{\mu=1}^{M} y^i \emptyset_\mu\left(x^i\right)}{\sum_{i=1}^{N} e^{-y^i} \sum_{\mu=1}^{M} y^i \emptyset_\mu(x^i)} \ldots\ldots\ldots eq(6)$$

The search space is represented by various parameters depending upon the features associated with the microvascular complications. Hence the output of the coati algorithm is the combination of minimum cost with maximum efficiency.

Class	Feature Selection Estimator	Features	Classification Algorithm	Training			Testing		
				Accuracy	Precision	Recall	Accuracy	Precision	Recall
Diabetic Retinopathy	AdaBoost	Age, Sex, BMI, BP,	XGB	81.56	81.68	81.56	70.58	71.61	70.58
			KNN	81.46	81.53	81.46	62.76	63.54	62.76

		FPS, Family History, Medical Adherence	SVM	52.31	61.41	52.31	46.36	36.94	46.36
			RF	99.90	99.90	99.90	94.78	94.78	94.78
			AdaBoost	65.16	65.16	65.61	56.14	56.93	56.14
			Tree	99.91	99.91	99.91	94.78	94.79	94.79
			ANN	51.98	59.01	51.98	56.79	60.79	56.79
Diabetic Nephropathy	AdaBoost	Sex, SP, FPS, Family History, Onset Age	XGB	78.85	79.16	78.85	69.71	70.35	69.71
			KNN	82.72	82.72	82.72	62.87	63.05	62.87
			SVM	53.04	51.63	53	50.16	50.09	50.16
			RF	99.89	99.89	99.89	95.44	95.45	95.44
			AdaBoost	64.84	64.97	64.84	59.72	60.43	59.72
			Tree	99.89	99.89	99.89	95.98	96.02	95.98
			ANN	54.82	65.42	54.82	51.14	61.01	51.14
Diabetic Neuropathy	AdaBoost	HbA1C, FPS	XGB	72.24	72.31	72.24	62.32	62.61	62.32
			KNN	82.21	82.32	82.21	65.36	65.63	65.36
			SVM	53.52	53.33	53.52	49.29	49.2	49.29
			RF	99.81	99.81	99.81	94.03	94.03	94.03
			AdaBoost	60.74	60.93	60.74	56.46	56.52	56.46
			Tree	99.81	99.81	99.81	93.59	93.64	93.59
			ANN	52.73	53.55	52.73	52.01	52.18	52.01

- **Diabetic Retinopathy:**

The selected features for retinopathy as age, sex, BMI, BP, FPS, Family History, and Medical Adherence. The classification results are found 99.9% and 94.78% for the Random Forest algorithm for training and testing accuracy. The comparison of various classification algorithms is shown in the figure 2.

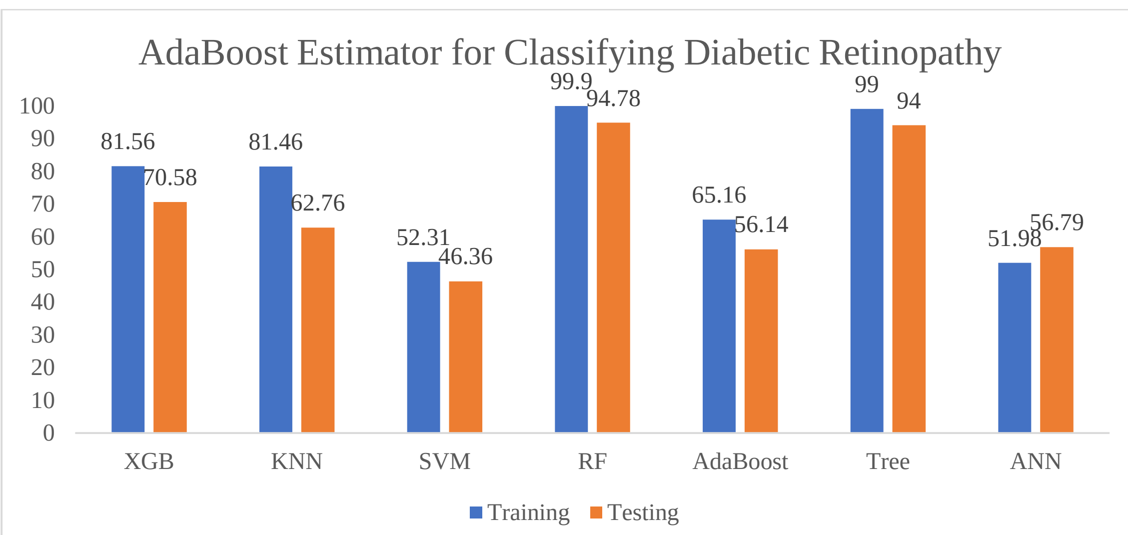

Figure 2 Comparison of Diabetic Retinopathy with XGB, KNN, SVM, RF, AdaBoost, Tree, and ANN classifier

- **Diabetic Nephropathy:**

The selected features for Nephropathy as Sex, SP, FPS, Family History, and Onset Age. The classification results are found 99.8% and 95.44% for the Random Forest algorithm for training and testing accuracy. The comparison of various classification algorithms is shown in Figure 3.

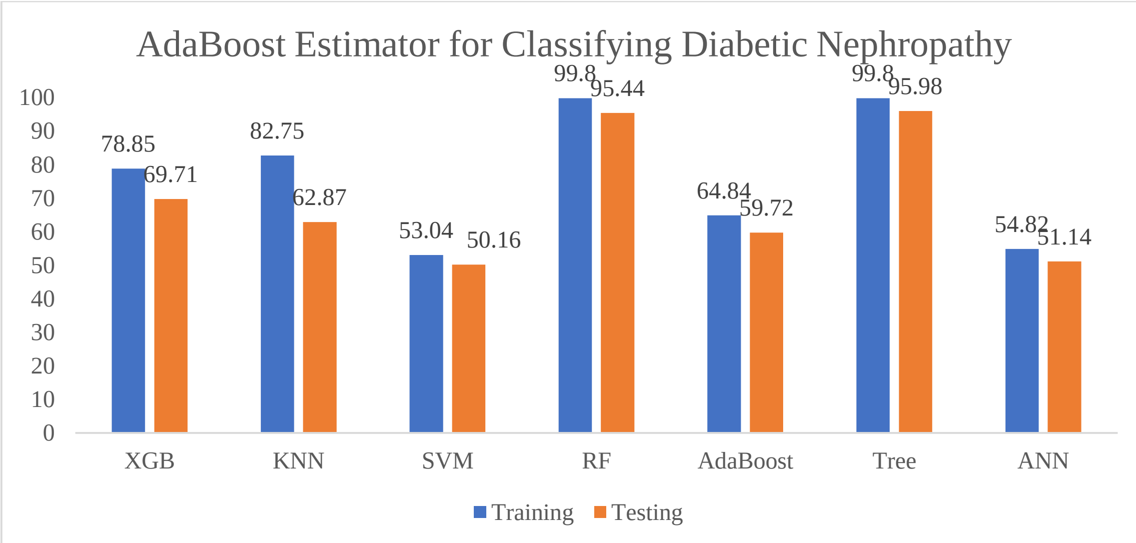

Figure 3 Comparison for Diabetic Nephropathy with XGB, KNN, SVM, RF, AdaBoost, Tree, and ANN classifier

- **Diabetic Neuropathy:**

The selected features for Neuropathy are HbA1C and FPS. The classification results are found 99.8% and 95.44% for the Random Forest algorithm for training and testing accuracy. The comparison of various classification algorithms is shown in the figure 3.

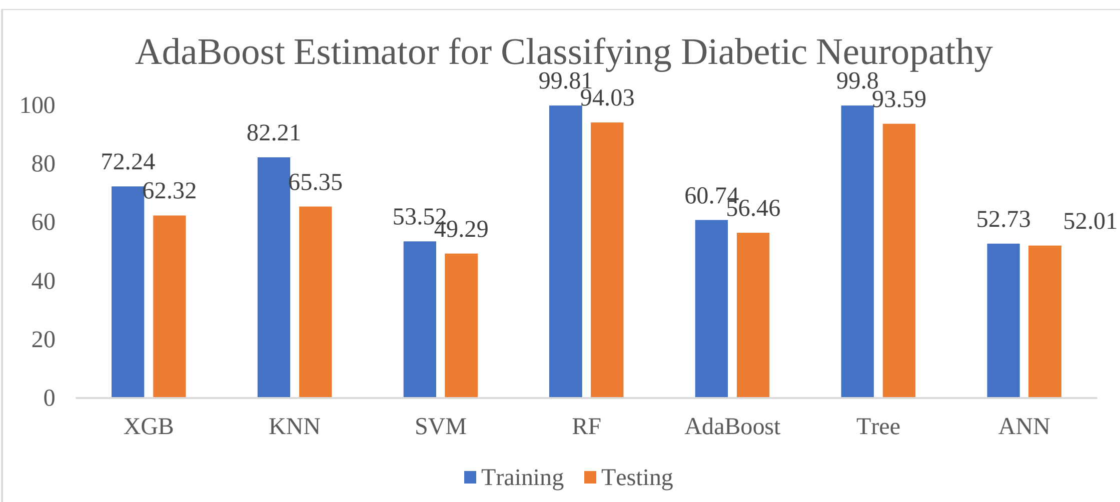

Figure 4 Comparison of Diabetic Neuropathy with XGB, KNN, SVM, RF, AdaBoost, Tree, and ANN classifier

Conclusion:

Microvascular complications in type 2 Diabetic patients commonly occur as diabetic retinopathy, neuropathy, and nephropathy. This complication depends upon the various parameters. In the proposed system the essential parameters are extracted using AdaBoost Estimator. This Improved Enhanced Coati feature extraction selects the optimal feature selector estimator which provides the optimal result based upon the fitness function. The selected features are age, sex, BMI, BP, FPS, Family History, Medical Adherence for diabetic retinopathy, Sex, SP, FPS, Family History, Onset Age for Nephropathy, and HbA1C and FPS for Neuropathy. The proposed feature selection algorithm results best for the Adaboost feature selection algorithm and provides suitable results for training and testing accuracy for the Random Forest classifier

Conflict-of-Interest: No

Research involving human participants and/or animals: No.

Informed Consent: Yes

Funding : No

Reference:

[1] M. L. Gavin, Ed., "What is type 2 diabetes? (for teens) - nemours kidshealth," KidsHealth, https://kidshealth.org/en/teens/type2.html#:~:text=Insulin%20helps%20open%20cells%20throughout,the%20blood%20sugar%20level%20rises. (accessed Feb. 17, 2024).

[2] Diabetes mellitus: Management of microvascular and macrovascular complications, https://www.clevelandclinicmeded.com/medicalpubs/diseasemanagement/endocrinology/diabetes-mellitus/#:~:text=Definitions,Nonproliferative%20retinopathy%20and%20proliferative%20retinopathy. (accessed Feb. 17, 2024).

[3] "IDF diabetes atlas," IDF Diabetes Atlas, https://diabetesatlas.org/ (accessed Dec. 7, 2023).

[4] Department of Health & Human Services, "Diabetes - long-term effects," Better Health Channel, https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/diabetes-long-term-effects (accessed Feb. 17, 2024).

[5] J. Chaki, S. Thillai Ganesh, S.K. Cidham, S. Ananda Theertan, Machine learning and artificial intelligence-based diabetes mellitus detection and self-management: Asystematic review, J. King Saud Univ.-Comput. Inf. Sci. (2020).

[6] I. Contreras, J.Vehi,Artificial intelligence for diabetesmanagement anddecision support:Literature review, J.Med. InternetRes. 20 (5) (2018) e10775.

[7] Schallmoser S, Zueger T, Kraus M, Saar-Tsechansky M, Stettler C, Feuerriegel S. Machine Learning for Predicting Micro- and Macrovascular Complications in Individuals With Prediabetes or Diabetes: Retrospective Cohort Study. J Med Internet Res. 2023 Feb 27;25:e42181. doi: 10.2196/42181. PMID: 36848190; PMCID: PMC10012007.

[8] Sambyal, N., Saini, P. & Syal, R. Microvascular Complications in Type-2 Diabetes: A Review of Statistical Techniques and Machine Learning Models. Wireless Pers Commun 115, 1–26 (2020). https://doi.org/10.1007/s11277-020-07552-3

[9] Kanbour S, Harris C, Lalani B, et al. Machine Learning Models for Prediction of Diabetic Microvascular Complications. Journal of Diabetes Science and Technology. 2024;0(0). doi:10.1177/19322968231223726

[10] Cichosz, S. L., Johansen, M. D., & Hejlesen, O. (2015). Toward Big Data Analytics. *Journal of Diabetes Science and Technology*. https://doi.org/10.1177/1932296815611680

[11] Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, De Cata P, Chiovato L, Bellazzi R. Machine Learning Methods to Predict Diabetes Complications. J Diabetes Sci Technol. 2018 Mar;12(2):295-302. doi: 10.1177/1932296817706375. Epub 2017 May 12. PMID: 28494618; PMCID: PMC5851210.

[12] Anne Gedebjerg, Thomas Peter Almdal, Klara Berencsi, Jørgen Rungby, Jens Steen Nielsen, Daniel R. Witte, Søren Friborg, Ivan Brandslund, Allan Vaag, Henning Beck-Nielsen, Henrik Toft Sørensen, Reimar Wernich Thomsen, "Prevalence of micro- and macrovascular diabetes complications at time of type 2 diabetes diagnosis and associated clinical characteristics: A cross-sectional baseline study of 6958 patients in the Danish DD2 cohort,Journal of Diabetes and its Complications",Volume 32, Issue 1,2018,Pages 34-40,ISSN 1056-8727,https://doi.org/10.1016/j.jdiacomp.2017.09.010.(https://www.sciencedirect.com/science/article/pii/S1056872717308991)

[13] Vamsi, B., Bataineh, A. A., & Doppala, B. P. (2022). Prediction of micro vascular and macro vascular complications in type-2 diabetic patients using machine learning techniques. International Journal of Advanced Computer Science and Applications, 13(11) doi:https://doi.org/10.14569/IJACSA.2022.0131103

[14] Swan BP, Mayorga ME, Ivy JS. The SMART Framework: Selection of Machine Learning Algorithms With ReplicaTions-A Case Study on the Microvascular Complications of Diabetes. IEEE J Biomed Health Inform. 2022;26(2):809-817. doi:10.1109/JBHI.2021.3094777

[15] Antonio Nicolucci, Luca Romeo, Michele Bernardini, Marco Vespasiani, Maria Chiara Rossi, Massimiliano Petrelli, Antonio Ceriello, Paolo Di Bartolo, Emanuele Frontoni, Giacomo Vespasiani,Prediction of complications of type 2 Diabetes: A Machine learning approach,Diabetes Research and Clinical Practice,Volume 190,2022,110013,ISSN 0168-8227,https://doi.org/10.1016/j.diabres.2022.110013. (https://www.sciencedirect.com/science/article/pii/S0168822722008270)

[16] Branimir Ljubic, Ameen Abdel Hai, Marija Stanojevic, Wilson Diaz, Daniel Polimac, Martin Pavlovski, Zoran Obradovic, Predicting complications of diabetes mellitus using advanced machine learning algorithms, Journal of the American Medical Informatics Association, Volume 27, Issue 9, September 2020, Pages 1343–1351, https://doi.org/10.1093/jamia/ocaa120

[17] Shao, H., Shi, L., Lin, Y., & Fonseca, V. (2022). Using modern risk engines and machine learning/artificial intelligence to predict diabetes complications: A focus on the BRAVO model. *Journal of Diabetes and its Complications*, *36*(11), 108316. https://doi.org/10.1016/j.jdiacomp.2022.108316

[18] Weiss, Jeremy C. "On microvascular complications of diabetes risk: development of a machine learning and electronic health records risk score." (2019).

[19] Dehghani, Mohammad & Montazeri, Zeinab & Trojovska, Eva & Trojovsky, Pavel. (2023). Coati Optimization Algorithm: A new bio-inspired metaheuristic algorithm for solving optimization problems. Knowledge-Based Systems. 259. 110011. 10.1016/j.knosys.2022.110011.

[20] Hany M. Hasanien, Ibrahim Alsaleh, Abdullah Alassaf, Ayoob Alateeq, Enhanced coati optimization algorithm-based optimal power flow including renewable energy uncertainties and electric vehicles,Energy,Volume 283,2023,129069,ISSN 0360-5442,https://doi.org/10.1016/j.energy.2023.129069.

[21] Ying CAO, Qi-Guang MIAO, Jia-Chen LIU, Lin GAO,Advance and Prospects of AdaBoost Algorithm,Acta Automatica Sinica,Volume 39, Issue 6,2013,Pages 745-758,ISSN 1874-1029,https://doi.org/10.1016/S1874-1029(13)60052-X.(https://www.sciencedirect.com/science/article/pii/S187410291360052X)

[22] Kégl, Balázs. "Introduction to adaboost." (2009): 011-014.

[23] Binitha, S., and S. Siva Sathya. "A survey of bio inspired optimization algorithms." *International journal of soft computing and engineering* 2.2 (2012): 137-151.

[24] S. Murugesan, R. S. Bhuvaneswaran, H. Khanna Nehemiah, S. Keerthana Sankari, Y. Nancy Jane, "Feature Selection and Classification of Clinical Datasets Using Bioinspired Algorithms and Super Learner", *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 6662420, 18 pages, 2021. https://doi.org/10.1155/2021/6662420

[25] S. B. Sakri, N. B. Abdul Rashid and Z. Muhammad Zain, "Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction," in *IEEE Access*, vol. 6, pp. 29637-29647, 2018, doi: 10.1109/ACCESS.2018.2843443.

[26] E. Trojovská, M. Dehghani and P. Trojovský, "Zebra Optimization Algorithm: A New Bio-Inspired Optimization Algorithm for Solving Optimization Algorithm," in IEEE Access, vol. 10, pp. 49445-49473, 2022, doi: 10.1109/ACCESS.2022.3172789.

[27] Dorigo, Marco & Birattari, Mauro & Stützle, Thomas. (2006). Ant Colony Optimization: Artificial Ants as a Computational Intelligence Technique. IEEE Computational Intelligence Magazine. 1. 28-39. 10.1109/CI-M.2006.248054.

[28] M. Dorigo, V. Maniezzo and A. Colorni, "Ant system: optimization by a colony of cooperating agents," in *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 26, no. 1, pp. 29-41, Feb. 1996, doi: 10.1109/3477.484436.

[29] Karaboga, Dervis. (2010). Artificial bee colony algorithm. Scholarpedia. 5. 6915. 10.4249/scholarpedia.6915.

[30] Dirk Sudholt, Christian Thyssen, “Running time analysis of Ant Colony Optimization for shortest path problems”,Journal of Discrete Algorithms,Volume 10,2012,Pages 165-180,ISSN 1570-8667,https://doi.org/10.1016/j.jda.2011.06.002.(https://www.sciencedirect.com/science/article/pii/S1570866711000682)

[31] K. M. Passino, "Biomimicry of bacterial foraging for distributed optimization and control," in *IEEE Control Systems Magazine*, vol. 22, no. 3, pp. 52-67, June 2002, doi: 10.1109/MCS.2002.1004010.

[32] M. Dorigo, M. Birattari and T. Stutzle, "Ant colony optimization," in *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28-39, Nov. 2006, doi: 10.1109/MCI.2006.329691.

[33] http://www.scholarpedia.org/article/Artificial_bee_colony_algorithm

[34] Hanning Chen, Ben Niu, Lianbo Ma, Weixing Su, Yunlong Zhu,Bacterial colony foraging optimization, Neurocomputing, Volume 137, 2014, Pages 268-284, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2013.04.054.(https://www.sciencedirect.com/science/article/pii/S0925231214002409)

[35] Das, Swagatam & Biswas, Arijit & Dasgupta, Sambarta & Abraham, Ajith. (2009). Bacterial Foraging Optimization Algorithm: Theoretical Foundations, Analysis, and Applications. 10.1007/978-3-642-01085-9_2.

[36] Kennedy, James, and Russell Eberhart. "Particle swarm optimization." *Proceedings of ICNN'95-international conference on neural networks*. Vol. 4. IEEE, 1995.

[37] Kaipa, Krishnanand & Ghose, Debasish. (2008). Glowworm swarm optimization for simultaneous capture of multiple local optima of multimodal functions. Swarm Intelligence. 3. 87-124. 10.1007/s11721-008-0021-5.

[38] Zainal, Nurezayana & Zain, Azlan & Radzi, Nor & Udin, Amirmudin. (2013). Glowworm Swarm Optimization (GSO) Algorithm for Optimization Problems: A State-of-the-Art Review. Applied Mechanics and Materials. 421. 507-511. 10.4028/www.scientific.net/AMM.421.507.

[39] *Cade WT. Diabetes-related microvascular and macrovascular diseases in the physical therapy setting. Phys Ther. 2008 Nov;88(11):1322-35. doi: 10.2522/ptj.20080008. Epub 2008 Sep 18. PMID: 18801863; PMCID: PMC2579903.*

[40] Rashid M, Alkhodari M, Mukit A, Ahmed KIU, Mostafa R, Parveen S, Khandoker AH. Machine Learning for Screening Microvascular Complications in Type 2 Diabetic Patients Using Demographic, Clinical, and Laboratory Profiles. *Journal of Clinical Medicine*. 2022; 11(4):903. https://doi.org/10.3390/jcm11040903

[41] Jian, Y.; Pasquier, M.; Sagahyroon, A.; Aloul, F. A Machine Learning Approach to Predicting Diabetes Complications. *Healthcare* **2021**, *9*, 1712. https://doi.org/10.3390/healthcare9121712

[42] Haque, F.; Reaz, M.B.I.; Chowdhury, M.E.H.; Ezeddin, M.; Kiranyaz, S.; Alhatou, M.; Ali, S.H.M.; Bakar, A.A.A.; Srivastava, G. Machine Learning-Based Diabetic Neuropathy and Previous Foot Ulceration Patients Detection Using Electromyography and Ground Reaction Forces during Gait. *Sensors* **2022**, *22*, 3507. https://doi.org/10.3390/s22093507

[43] Jelinek, H. F., D. J. Cornforth, and A. V. Kelarev. "Machine learning methods for automated detection of severe diabetic neuropathy." *J. Diabet. Complicat. Med* 1.02 (2016): 1-7.

[44] X.-S. Yang; S. Deb (December 2009). Cuckoo search via Lévy flights. World Congress on Nature & Biologically Inspired Computing (NaBIC 2009). IEEE Publications. pp. 210–214. arXiv:1003.1594v1.

[45] Blessie, E. Chandra and E. Karthikeyan. "Sigmis: A Feature Selection Algorithm Using Correlation Based Method." *Journal of Algorithms and Computational Technology* 6 (2012): 385 - 394.

[46] R. Leardi,3 - Genetic Algorithms in Feature Selection,Editor(s): James Devillers,In Principles of QSAR and Drug Design,Genetic Algorithms in Molecular Modeling,Academic Press,1996,Pages 67-86,ISBN 9780122138102,https://doi.org/10.1016/B978-012213810-2/50004-9. (https://www.sciencedirect.com/science/article/pii/B9780122138102500049)

[47]Babatunde, Oluleye & Armstrong, Leisa & Leng, J. & Diepeveen, Dean. (2014). A Genetic Algorithm-Based Feature Selection. International Journal of Electronics Communication and Computer Engineering. 5. 889-905.

[48] Mathew, Tom V. "Genetic algorithm." *Report submitted at IIT Bombay* (2012): 53.

[49] Y. Zhang, D. -w. Gong and J. Cheng, "Multi-Objective Particle Swarm Optimization Approach for Cost-Based Feature Selection in Classification," in *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 14, no. 1, pp. 64-75, 1 Jan.-Feb. 2017, doi: 10.1109/TCBB.2015.2476796.

[50] S. Sharma and H. M. Pandey, "Genetic Algorithm, Particle Swarm Optimization and Harmony Search: A quick comparison," *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, Noida, India, 2016, pp. 40-44, doi: 10.1109/CONFLUENCE.2016.7508044.

[51] *Sobhi Ahmed, Majdi Mafarja, Hossam Faris, and Ibrahim Aljarah. 2018. Feature Selection Using Salp Swarm Algorithm with Chaos. In Proceedings of the 2nd International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence (ISMSI '18). Association for Computing Machinery, New York, NY, USA, 65–69. https://doi.org/10.1145/3206185.3206198*

[52] Ahmed, Sobhi, et al. "Feature selection using salp swarm algorithm with chaos." *Proceedings of the 2nd international conference on intelligent systems, metaheuristics & swarm intelligence*. 2018.

[53] Xiao, Li-min. "An Optimizing Method Based on Autonomous Animats: Fish-swarm Algorithm." *Systems Engineering - Theory & Practice* (2002): n. pag.

[54] Hamed Shah-Hosseini. 2009. The intelligent water drops algorithm: a nature-inspired swarm-based optimization algorithm. Int. J. Bio-Inspired Comput. 1, 1/2 (January 2009), 71–79. https://doi.org/10.1504/IJBIC.2009.022775

[55] Krishnamoorthi R, Joshi S, Almarzouki HZ, Shukla PK, Rizwan A, Kalpana C, Tiwari B. A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques. J Healthc Eng. 2022 Jan 11;2022:1684017. doi: 10.1155/2022/1684017. Retraction in: J Healthc Eng. 2023 May 24;2023:9872970. PMID: 35070225; PMCID: PMC8767376.

[56] E. Trojovská, M. Dehghani and P. Trojovský, "Zebra Optimization Algorithm: A New Bio-Inspired Optimization Algorithm for Solving Optimization Algorithm," in *IEEE Access*, vol. 10, pp. 49445-49473, 2022, doi: 10.1109/ACCESS.2022.3172789.

[57] Hanning Chen, Ben Niu, Lianbo Ma, Weixing Su, Yunlong Zhu, Bacterial colony foraging optimization,Neurocomputing, Volume 137,2014,Pages 268-284,ISSN 0925-2312,https://doi.org/10.1016/j.neucom.2013.04.054.(https://www.sciencedirect.com/science/article/pii/S0925231214002409)

[58] Scaria, Abyson & George, Kiran & Sebastian, Jobin. (2016). An Artificial Bee Colony Approach for Multi-objective Job Shop Scheduling. Procedia Technology. 25. 1030-1037. 10.1016/j.protcy.2016.08.203.