International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 14, No.1 (Aug-2023)

http://dx.doi.org/10.12785/ijcds/140138

# Intellectual Gestational Diabetes Diagnosis System Using MLP-Whale Optimization Algorithm including Statistical Analysis

**Panduranga Vital Terlapu[1], JagadeeswaraRao G[1], Siva Prasad A[2], Ramesh Y[1], Ramesh Ch[3] and Yoshita B[1]**

[1]*Department of IT, Aditya Institute of Technology and Management, Tekkali, India*
[2]*Department of Computer Science, Govt. Degree College, Tekkali, India*
[3]*Department of CSE, Aditya Institute of Technology and Management, Tekkali, India*

*Received 08 Sep. 2022, Revised 07 May. 2023, Accepted 19 May. 2023, Published 01 Aug. 2023*

**Abstract:** Diabetes is a widely recognized medical condition and is generally mentioned as a silent killer in global healthcare. It is a metabolic and chronic disease that causes an increase in blood glucose (BGL) levels, leading to long-term damage to the blood vessels, heart, nerves, eyes, and kidneys, among other vital organs. The causes of diabetes include genetic factors, alcohol consumption, smoking, weight, the absence of actual work, unfortunate dietary propensities, and hypertension. Depending on the type and severity of diabetes, other organs in the body, such as the kidneys, heart, and eyes, are more susceptible to diseases. In this research investigation, we extend a model, the MLP-WOA, which uses a fine-tuned weight of the multi-layer perceptron (MLP) with the diabetes onset prediction using the whale (WOA) optimization algorithm. We used a benchmark dataset from the Kagle repository to train and test our model, and we evaluated its accuracy, precision, and recall. Support Vector Machines (SVM), K-nearest (KNN) neighbor, and other ML methods were evaluated against our model's outcomes, WOA-MLP, and decision trees (DTs). Our results showed that the MLP-WOA model performed superiorly to the other experimental models, achieving an accuracy of approximately 76%. Furthermore, we tested our MLP model using other existing optimizers and observed that the WOA optimizer gave better results. In conclusion, our study demonstrates that the MLP-WOA model prototype is a promising approach for predicting diabetes and that it outperforms other ML models. Patients with diabetes may see an improvement in their quality of life as a result of this strategy's ability to improve diagnosis accuracy.

**Keywords:** Diabetes, MLP, WOA, Machine Learning, Optimization

## 1. INTRODUCTION

In 2020, according to World Health Organisation (WHO) statistics, there were 34.2 million Americans with diabetes [1]. The United States had 77 million diabetes diagnoses in 2019, making it the country with the second-highest rate of diabetes worldwide [2]. Diabetes is generally classified into several types, including gestational diabetes, pre-diabetes, type 1 diabetes, and type 2 diabetes [3]. In India, diabetes affects over 30 million people, with many more at risk [4]. Therefore, early clinical diagnosis and treatment are necessary to predict and prevent diabetes and its related medical conditions. The chronic disease known as diabetes mellitus affects individuals worldwide and is commonly called a silent killer. It is primarily related to weight, age, physical inactivity, genetic predisposition, lifestyle choices, poor diet, hypertension, and other factors [5]. People with diabetes are at increased risk for other diseases, such as coronary heart disease, kidney disease, stroke, eye issues, nerve damage, and more [6]. Currently, medical clinics collect the necessary data for diabetes diagnosis through various tests, and established treatments are given for confirmed diagnoses. Gestational (GDM) diabetes mellitus is difficult while pregnant, affecting up to 15% of pregnant women worldwide. High blood (BGL) glucose levels, known as hyperglycemia, affect individuals with diabetes. Several factors can play a crucial role in hyperglycemia in individuals with diabetes, including skipping meals, insufficient insulin or other medications to lower blood glucose levels, and more. Table I provides a detailed overview of the types, symptoms, causes, and treatment of

*E-mail address: vital2927@gmail.com, jagadish4u.g@gmail.com, Siva.hod@gmail.com* 

diabetes. Hyperglycemia is a characteristic of diabetes, a metabolic disease characterized by malfunctions in either insulin action or insulin secretion or in both [7].

The ongoing hyperglycemia related to diabetes can cause long-term harm, brokenness, and disappointment in different organs, including the nerves, eyes, heart, kidneys, and veins. Figure 1 illustrates the processes of glucose inhibition and enhancement. The pancreas [12] is a tube-shaped, spongy, elongated organ situated in the abdomen behind the stomach and adjacent to the small intestine, resembling a fish. It measures around 6 inches in length. The central region of the pancreas is known as the body, while the tail is the tapered end that extends toward the left side of the body. The pancreas comprises two distinct functional parts: the endocrine function and the exocrine function. The exocrine function involves two integrated glands that produce digestive enzymes or compounds [13]. The cells of the exocrine system secrete their enzymes into ducts that merge to form the primary pancreatic duct. This essential pancreatic duct drains the digestive enzymes produced by the exocrine cells into the duodenum [14]. These enzymes include chymotrypsin and trypsin for protein digestion, amylase for carbohydrate absorption, and lipase for fat breakdown. The endocrine part of the pancreas comprises islet cells, known as the islets of Langerhans. Instead of releasing hormones into the pancreatic ducts, these endocrine cells do so directly into the body's circulation [15]. The essential pancreatic chemicals are glucagon, which raises BGLs, and insulin, which brings them down, supporting the control of glucose levels. The other hormones produced by the endocrine cells are pancreatic polypeptide and somatization. Maintaining proper glucose levels is crucial for the optimal functioning of the body, as glucose is essential for the body's energy needs. Instead of releasing hormones into the pancreatic ducts, these endocrine cells do so directly into the body's circulation [16]. The essential pancreatic chemicals are glucagon, which raises BGLs, and insulin, which brings them down BGLs, helps control glucose levels. Machine learning and soft computing concepts have found widespread use in the classification, prediction, and analysis of medical datasets, particularly in the identification and treatment of diabetes [17], [18]. Deep neural network models are extensively employed across various fields, including engineering, medicine, agriculture, and more. The human brain serves as inspiration for NNs, and they can be used successfully in both supervised and unsupervised learning applications. Neural networks can be categorized into six types. The feed-forward neural network, where input data travels in the same direction without back-propagation. CNN is a technique designed specifically for image recognition and processing that processes pixels through multiple convolutional and normalization layers [19], [20]. The modular neural network is made up of multiple neural network models that can act as modules, each solving a portion of the issue [21]. The data used in this study were gathered from the Pima Indian Diabetes Dataset repository in the Kagle repository, which included a total of 768 patients. The independent variables in the dataset were medical predictors, such as the number of pregnancies, BMI, insulin, etc., and the dependent variable was the target. Optimization involved finding the decision parameter values that maximize or minimize one or more specific objectives.

Researchers have proposed various categories of optimizers, as described in the literature [22], [23]. In the context of this paper, the combination of MLP and WOA has been implemented on the diabetes dataset to enhance its accuracy. MLP refers to a feed-forward NN with an input, an output, and a hidden layer. The I/P data layer processes signals, and the output layer categorizes them. Essentially, an MLP incorporates multiple hidden layers in neural networks. During training, back-propagation learning algorithms are used to adjust the neurons in the MLP. MLP is particularly suited for predicting an approximation of continuous functions and for solving non-linear problems.

The research work focuses as follows:

• Present the indications of diabetes and organ functionalization related to diabetes. • Describes Clinical Diabetes diagnosis and treatment, and prevention are exhaustive. • Detailed analysis of diabetes stages using measurable models. • Compare the existing ML models and proposal models in detail. • The proposed intelligent MLP-WOA technique is utilized for Diabetes classification and prediction.

The reaming contribution sections are as follows:

• Section 2 depicts the survey of the foundation's in-depth works. In this part, we evaluated many papers connected with this research work that introduced a portion of the considered research papers from reputed and believed journals like Elsevier, Springer, IEEE, and others. More researchers and analysts communicated in their studies about reasons for Diabetes, Diabetes identification, treatment and anticipations, and diabetes concerned with deep-learning, ML, and NNs with various benchmark and constant datasets. • Section 3 depicts the experimental and working arrangement and accomplishment of the proposed model in a detailed manner. This part introduced the statistics methods, ML approaches, and MLP-WOA-based models. Additionally, the mathematical representation of the proposal model MLP-WOA shows capacity and working process. Furthermore, it describes performance parameters and confusion matrix evolutions. • Section 4 projects the investigation with experimental results of 3 ML and MLP-WOA-based models statistical measurement values analysis on diabetes dataset. The proposal methodology performance is better than the remaining experimental ML models, and other research works related to diabetes. • Section 5 addresses the discussions and comparative analysis, which looks at the proposition of MLP-WOA model outcomes to other experimental ML model outcomes, and discussions about other early works with various datasets and models.

TABLE I. Types of Diabetes and their symptoms and treatments

| Diabetes Type | Symptoms | Caused by | Effected category | Treatment |
|---|---|---|---|---|
| Type 1 | Frequent urination, drowsiness, slow-healing wounds, weight loss [8] | Destruction of beta cells in the pancreas. | Children and young adults | Insulin therapy, regular exercise |
| Type 2 | Similar to Type 1 and discolored patches of skin on the neck and arms [9] | Inefficient use of insulin, or a decrease in insulin production. | Adults (age¿45) | Good lifestyle habits, Weight loss, Insulin therapy, and medication |
| Gestational diabetes | Abdominal pain, drowsiness, High blood pressure | Insulin is blocked by hormones produced during pregnancy. | Pregnant women | Insulin therapy may be needed, and a healthy diet [10] |
| Pre-diabetes | Same as type 1 and type 2 | High blood sugar | Adults | Good lifestyle habits, Managing weight, having a good diet, and physical [11] |

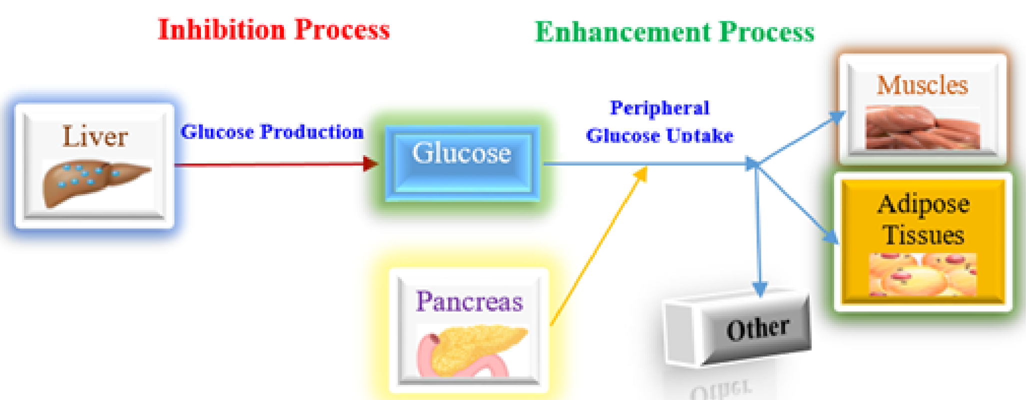



Figure 1. Inhibition process and Enhancement process of the glucose functionalities with Pancreas

• Section 6 demonstrates the conclusion and future related work. It depicts limitations and restrictions related to the Diabetes identification system.

## 2. LITERATURE REVIEW

In recent years, ML methodologies and tools have been utilized to examine and predict the probability of diabetes, and it has gained considerable interest. Researchers have analyzed and developed novel approaches and methodologies in this area. In this section, we present some significant works that have a significant impact on the diagnosis and detection of diabetes using ML models, soft computing and optimization analyzing models, neural network models, and hybrid models analysis. Mujumdar et al. (2019) proposed a prediction model for diabetes using external factors responsible for regular factors like BMI, insulin, age, and so on [24]. They further enforced diabetes prediction using a pipeline model intended to improve the accuracy of classification. Their proposed pipeline method for diabetes prediction is based on working on classification accuracy. Tigga et al. (2019) collected and analyzed data from 952 individuals using a questionnaire (online and offline) and applied ML models to this dataset as well as PID data [25]. As per their findings, the RF model is more accurate than other experimental algorithms. Diabetes is a metabolic disease characterized by a high level of sugar in the blood over a significant period. The severity and risk of diabetes can be decreased with an early forecast, which is possible with the use of ML techniques. Faruque et al. (2019) employed four ML procedures, including SVM, NB, KNN, and C4.5, to predict diabetes in adult population data [26]. Their observational results show that the C4.5 model achieved high accuracy compared to other ML experimental models. Sarwar et al. (2018) applied six different ML models to medical record datasets of diabetes patients, computed the performance of the applied experimental models, and compared them to each other [27]. Table 2 describes the contribution of diabetes with ML and other models.

As part of a project using MLP, several studies involving MLP and disease datasets are presented here. Nancy Masih et al. (2021) developed a deep neural network based on MLP, which predicted an accuracy of 92.10% in the early detection of coronary heart disease [37]. Verma et al. (2020) applied ensemble methods to predict skin diseases, selected features using hybrid methodologies and predicted 92% accuracy [38]. Djerioui et al. (2020) utilized MLP and LSTM models to predict ATr (accuracy training) and ATs (accuracy testing) at 95.73 and 89.18 percent, respectively [39]. Bhoyar et al. (2021) predicted 87.30% accuracy by applying MLP on the cardiovascular disease dataset and 85.71% accuracy using MLP, the UCI repository's datasetof heart disease [40]. Fuad Ali Mohammed Al-Yarimi et al. (2021) applied the FODW method, HRFLM (hybrid RF+LR model), and the HIFS model to a corpus of heart disease data, which predicted 92%, 90%, and 85% accuracy, respectively [41]. Mohapatra et al. (2019) obtained a classification accuracy of 77.5% by applying MLP on the PIMA Indians diabetes database [42]. Wlodarczak et al. (2015) aimed to determine a function that categorizes a set of correct outputs for a given set of inputs. The training repeats until the ANN (Artificial Neural Network) recognizes the person correctly in an image [43]. Mohan et al. (2019) produced 88.7% accuracy by applying HRFLM to the heart disease dataset. Sayan et al. (2020) used a hybrid-DL Model to predict the CV scores of Derm2Vec, DNN (Deep Neural Network), and Gradient Boosting(EGB) for the diagnosing Erythemato-Squamous Disease at 96.92 percent, 96.65%, and 95.80%, respectively [44]. Diabetes is a chronic disease characterized by hyperglycemia, which poses various complications. Due to the growing severity of the condition, it is predicted that by 2040, the number of people suffering from diabetes worldwide will reach 642 million, equating to one in every ten adult individuals. Zou et al. (2018) [45] conducted a study on physical medical examination data from China's Luzhou hospital, which consisted of a dataset with 14 attributes. The researchers employed 5-fold cross-validation to analyze the models and used min-max redundancy and relevance (mRMR) and PCA algorithms to reduce the dataset's dimensions. Nadeem et al. (2021) [46] conducted research on patients with diabetes using classification models such as ANN and SVM, with an accuracy rate of over 90%. The frequency of type 2 diabetes has significantly increased in recent years. With the advancements in AI applications in medical care, they are now being utilized for therapeutic, diagnostic, and predictive purposes, especially for type 2 diabetes. Abhari et al. (2019) [47] applied and revised AI models for identifying and examining different stages of type 2 diabetes. Sonar et al. (2019) [48] conducted research on predicting diabetes using ML models such as SVM, ANN, and DTs, with an accuracy rate of over 75%. Since diabetes is a non-curable disease, early detection, prevention, and protection are crucial. Khanam et al. (2021) [31] employed DM, neural network, and ML methods on the IPD diabetes dataset collected from the UCI ML repository. Silva et al. (2020) [49] examined twelve featured expected applicability methodologies for T2DM screening, and the meta-analysis delivered a decent pooled c-file (0.812). The issues relating to reporting and the quality of the observed methodologies were also analyzed in this study.

## 3. MATERIALS AND MODELS

In this section, we describe the proposed model, Dataset description, natural and mathematical model of the Wale Optimization (WOA) Algorithm, and others.

### A. *Proposal Model Description*

Figure 2 illustrates the Proposal model. The first step involves collecting data on pregnant diabetes and non-disease cases from the Kagle repository. Pre-processing and cleaning techniques are then applied to the data, and it is saved in CSV format in the database for classification and prediction purposes. The data is then split into two parts, training (80%) and testing (20%). ML models for instance SVM, DTs, and k-NN are applied, and MLP is optimized with WOA, Adam, Adamax, Ad-delta, and RMS optimizers. The models' performance is evaluated using performance parameters for instance AUC, ACC, recall, precision, and F1 values. In the final step, the models' performances are analyzed, and the best model for the Pregnant Diabetic dataset is selected. Unknown data is then fed into the prediction model, and the results are reported to the analysts for further analysis.

### B. *Diabetes Dataset Description*

The Pima Indian (PID) Diabetes dataset used in this study was obtained from the Kaggle repository. 768 patients' medical predictor (independent) variables are included in the dataset, while the outcome is the target (dependent) variable. Attributes like BMI, insulin, and the number of pregnancies are examples of independent variables. Table III provides a comprehensive explanation of the dataset, including attribute categories, attribute descriptions, and range values for each attribute. The 9th attribute in the dataset represents the class variable of each data point, indicating the outcome as 0 for non-diabetics and 1 for diabetics. To predict diabetes, a model was created, but the dataset was somewhat imbalanced, with around 500 records labeled as class 0 (negative) and 268 records labeled as class 1 (positive).

### C. *Whale Optimization Algorithm (WOA)*

The WOA is used to solve single-objective problems and imitates the humpback whales searching process in nature. When compared to existing optimization algorithms, WOA is a very aggressive algorithm designed to solve critical optimization problems. A nature-motivated inspired meta-heuristic algorithm imitates the hunting activity of humpback whales. In this algorithm, follow the strategy of bubble net hunting process like whales. Mainly, the whales like to chase the little fishes near the surface. It has

TABLE II. Descriptions of Authors Contributes about Diabetes Disease

| Ref | Author | Contribution | Dataset(s) Used | Year |
|---|---|---|---|---|
| [28] | Olisah et al˙ | Used supervised machine learning models, like RF, SVM, and twice-growth deep NN model for classification. | PIMA diabetes dataset. | 2022 |
| [29] | Suresh et al. | Proposed a novel stacking technique with MLP, SVM, and LR models. | PIMA Indian diabetes dataset. | 2021 |
| [30] | Kalagotla et al. | Applied Logistic regression and SVM algorithms | PIDDatasets | 2021 |
| [31] | Khanam et al. | Proposed a pipeline based deep learning technique | PID Datasets | 2021 |
| [32] | García-Ordás et al. | Proposed a Chaotic-Jaya hybridized ML model. | Pima Indian Diabetes dataset. | 2020 |
| [33] | Debata et al. | Proposed hybrid optimization technique by BGWO and integrating CSA optimizers. | Pima Indian Diabetes dataset. | 2021 |
| [34] | Mallika et al. | Used Neural Network (NN), SVM, AdaBoost, NB, RF and KNN, | Publically available dataset of in Sylhet, Sylhet-Diabetes (SDMC) Medical clinic , Bangladesh | 2021 |
| [35] | Chaves et al. | A new genetic (GA) algorithm, synthetic minority (SMO) oversampling, and DT based predictive model for a diabetes mellitus prediction system were proposed. | Pima Indian Diabetes dataset | 2020 |
| [36] | Azad et al. | Utilizing the Synthetic Minority (SMOTE) Over Sampling Technique, Outlier Detection utilizing DBSCAN-Based, and RF, a Hybrid Model was proposed for Hypertension, Diabetes Type 2. | Pima Indian Diabetes dataset | 2018 |

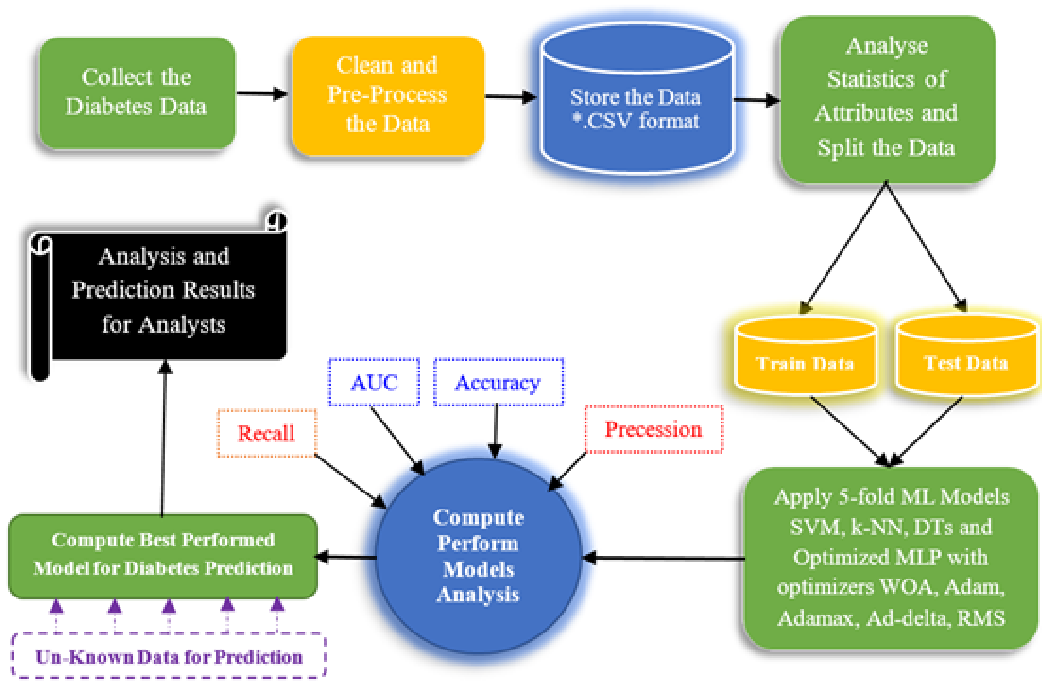



Figure 2. Detailed Proposed Model Diagram

TABLE III. Types of Diabetes and their symptoms and treatments

| Feature Attribute | Attribute Description | Category | Range (Min and Max) |
|---|---|---|---|
| X1 | Pregnancy (Number of Times) | Continuous | 0 to 17 |
| X2 | Glucose (Tolerance Test Of Oral Glucose) | Continuous | 0 to 199 |
| X3 | BloodPressure (BP (mm Hg)) | Continuous | 0 to 122 |
| X4 | SkinThickness (TricepsSkin-FoldThickness (mm)) | Continuous | 0 to 99 |
| X5 | Insulin (2-HourSerumInsulin (mu U/ml)) | Continuous | 0 to 846 |
| X6 | BMI (BodyMassIndex) | Continuous | 0 to 67.1 |
| X7 | DiabetesPedigreeFunction | Continuous | 0.078 to 2.42 |
| X8 | Age (years) | Continuous | 21 to 81 |
| C1 | Class (0 for Absence and 1 for Present) | Discrete | 0 (268) 1(500) |

been checked that this scrounging is finished by making unmistakable air pockets along a circle. Humpback whales find the target and dive 12 meters deep in the sea and create a bubble net in a '9' shape around the prey. The humpback swims in a spiral beneath the prey, making bubbles that trap them. Figure 3 shows the architecture of WOA.

*1) Mathematical model of WOA*

In the step of Encircling Prey, first, it finds the prey's location through a group of whales. The best agent will be selected based on the fitness value of each whale, other whales change their position with respect to the best search as shown in equation 1 and 2

$$\overrightarrow{D} = \left|\overrightarrow{C}.\overrightarrow{X^*}(t) - \overrightarrow{X}(t)\right| \quad (1)$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X^*}(t) - \overrightarrow{A}.\overrightarrow{D} \quad (2)$$

Where X* is the best solution's position vector, t is the current iteration, X is the vector of position, random vector r from 0 to 1, and a is a vector that will decrease linearly over the iterations.

$$\overrightarrow{A} = 2\overrightarrow{a}.\overrightarrow{r} - \overrightarrow{a} \quad (3)$$

$$\overrightarrow{C} = 2.\overrightarrow{r} \quad (4)$$

Equation 3 and 4 the "A" and "C" vectors are being used to update the values of the new position (X, Y) in the locality of the best agent (X*, Y*). WOA is likewise called the investigation stage, which includes two phases one is a contracting encompassing component and one more with a winding or spiral updating mechanism. Vector 'a' is the mathematical representation of the behavior of the shrinking encircling mechanism. Over the iterations 'a' value decreases from 2 to 0. Both these hunting behaviors occur simultaneously, and while shrinking their diameter these whales move upward in a spiral fashion. To do this WOA assumes a probability of 50% for position updating either first or second behavior in equation 7. The current whale's position is updated using equation 5

$$\overrightarrow{X}(t+1) = \overrightarrow{D'}.e^{bl}.cos(2\Pi l) + \overrightarrow{X^*}(t) \quad (5)$$

$$\overrightarrow{D'} = \left|\overrightarrow{X^*}(t) - \overrightarrow{X}(t)\right| \quad (6)$$

Where $\overrightarrow{D'}$ is calculated by equation 6 below and b is the constant that denotes the shape of a logarithmic spiral and l is a random number between -1 and 1.

$$\overrightarrow{X}(t+1) = \begin{cases} \overrightarrow{X^*}(t) - \overrightarrow{A}.\overrightarrow{D} & \text{if } p\leq 0.5 \\ \overrightarrow{D'}.e^{bl}.cos(2\Pi l) + \overrightarrow{X^*}(t) & \text{if } p>0.5 \end{cases} \quad (7)$$

The algorithm also introduces another mechanism called the exploration phase. In this phase, new search space is explored. Mathematically, this is done by using equation 9 by updating the position based on the vector 'A'. Updating vector $\overrightarrow{D}$ is done using equation 8

$$\overrightarrow{D} = \left|\overrightarrow{C}.\overrightarrow{X_{rand}}) - \overrightarrow{X}(t)\right| \quad (8)$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X_{rand}} - \overrightarrow{D}.\overrightarrow{A} \quad (9)$$

Where $\overrightarrow{X_{rand}}$ is a random whale position that is randomly taken from the population of whales. Figure 4 shows the detailed analysis of the whale optimization algorithm working process.

*D. Multilayer Perceptron (MLP)*

A feed-forward (FFN) NNs with three (Named) layers has been added to MLP: an I/P layer, an O/P layer, and one or more hidden (HLs) layers. The input layer processes the input signals. Categorization is performed by the output

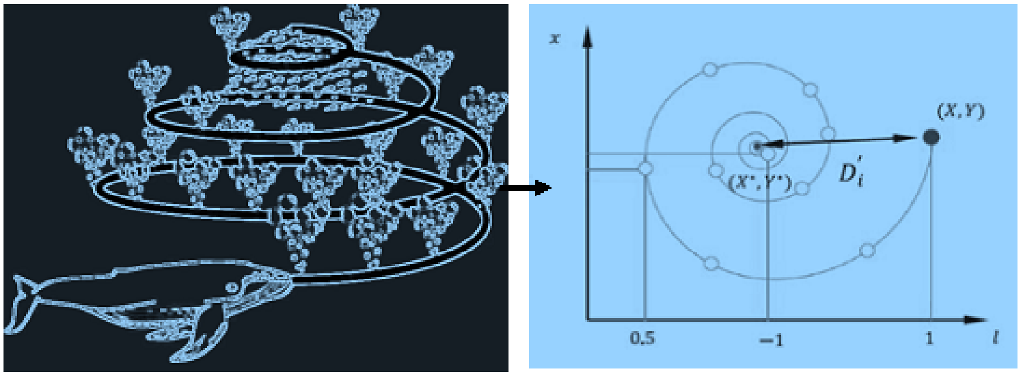



Figure 3. Humpback Whales Searching Process [50]

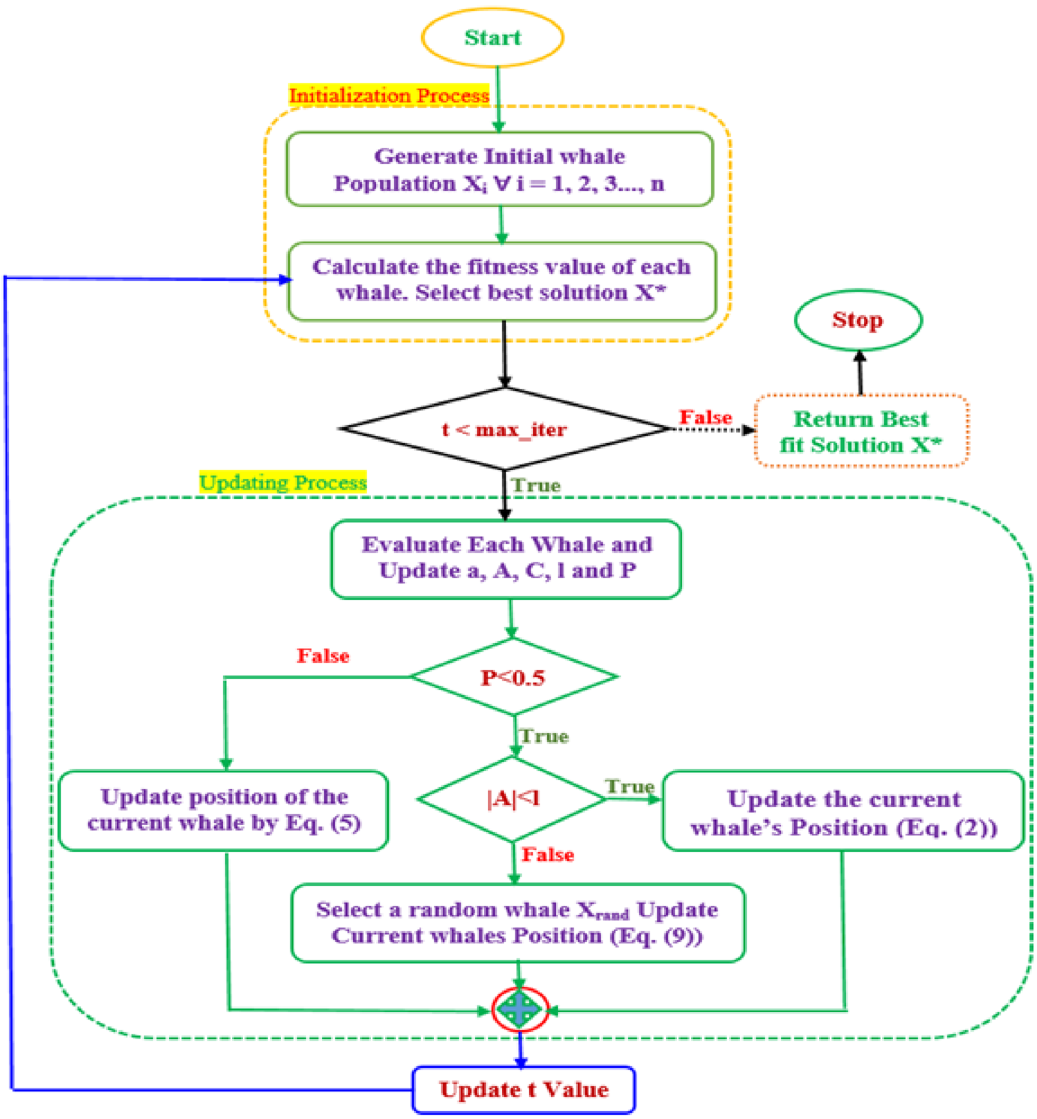



Figure 4. Whale Optimisation Algorithm Working Process

layer. In simple words, an MLP is having multiple hidden layers in neural networks. Figure 5 depicts the analysis of MLP with a two-hidden layer (H) process. One input layer (I) and output layer (O). The MLP solves non-linear problems and is mostly useful for classifying patterns, pattern recognition, estimation, and approximation. The feed-forward neural network may contain more than one hidden layer. Equation 10 shows the equation for processing the inputs by the hidden layer.

$$h_n = a \sum_{k=1}^{n} (i_k * w_{k,n} + b_n) \tag{10}$$

where $h_n$ is the number of hidden layers, a is the activation function applied to layers in MLP, $i_k$ is a number of inputs, $w_{k,n}$ are weights from one layer to another layer, $b_n$ is the bias used as a balancing factor to the layer.

*1) Activation function*

Activation functions are used in neural networks for making the decision and i.e., used to calculate the sum of the weights of its input and add a bias output of each node. The activation functions are applied only to the hidden layer and output layer. The selection and working of the activation function are shown in Figure 6.

*2) ReLU function*

ReLU stands for the rectified linear unit. In CNN, this is one of the most important activation functions. The equation for the ReLU function is shown in equation 11, range from 0 to infinite, including 0. Applied in the Hidden layer, the output layer for regression (positive outputs only).

$$\begin{aligned} f(z) &= 0, z < 0 \\ &= z, z \geq 0 \end{aligned} \tag{11}$$

*3) Sigmoid function*

This function is nonlinear. A variant of the logistic function is the sigmoid function. The O/P of the sigmoid neuron is either smooth or continuous. Range: All real numbers from 0 to 1. Used in: Hidden layer, output layer for classification. The equation for the sigmoid function is shown in equation **??**.

$$f(z) = \frac{1}{1 + e^{-z}} \tag{12}$$

*E. Multilayer Perceptron -Whale Optimization Algorithm (MLP-WOA)*

Figure 7 shows the general MLP-WOA functional architecture. The working process with a given dataset follows that in the first given data is pre-processed by the pre-processing algorithms like avoiding multiple records, filling with statistical values or removing the un-wanted and un-filled data, and so on. The pure data is classified with the MLP algorithm. The MLP weights are modified by the WOA (Whale optimization) algorithms the tune (smooth) the parametric values at the rich level. In the last, get the highly performed MLP-WOA model for the predictions of diabetes.

*F. Evaluation using Confusion Metrics*

In the model evaluation, we measure the predictive model's performance [51]. To determine how well a classifier performs, we use a confusion matrix. An instance of class A is classified as class B by counting the number of times it occurs. An actual class represents a row while a predicted class represents a column in a confusion matrix. The evaluation matrix gives the accuracy value, based on the prediction model. The main target user of the evaluated applications or objects is included in the evaluation matrix [52]. By using the evaluation matrix, you can attain the maximum level of accuracy or attention. Along with these the other metrics and their relevance for consideration are also given below. A confusion matrix structure is shown in Figure 8

Accuracy: Measuring an instrument's accuracy is defined as its ability to measure the exact value. A closeness to a standard or true. The value determines the accuracy of a measurement. Small readings can be used to determine accuracy [53]. Low readings can reduce calculation errors. It measures the proportion of correctly classified observations to all observations. Equation 13 shows the accuracy formula.

$$accuracy = \frac{(TP + TN)}{Total\_Observations} \tag{13}$$

Precision: The precision of measurement shows how close two or more measurements are to one another based on the information that is conveyed by its digits. However, accuracy is not included in its calculation. Equation 14 shows the precision formula.

$$Precision = \frac{TP}{(TP + FP)} \tag{14}$$

Recall: Also called sensitivity. Based on all the inspections in a class, the ratio of correctly predicted positives is called recall. Equation 15 shows the formula for recall.

$$Recall = \frac{TP}{(TP + FN)} \tag{15}$$

F1 rating score: a weighted average of precision and recall combined. Equation 16 shows the F1_score formula.

$$F1_score = \frac{TP}{(TP + FN)} \tag{16}$$

Specificity: Diagnostic tests are characterized by their specificity in identifying the number of true negative results. This gives us an indication of the accuracy of the test. Only true negative values are displayed, which were correctly classified. Equation 16 shows the formula for Specificity.

$$Specificity = \frac{TN}{(TN + FP)} \tag{17}$$

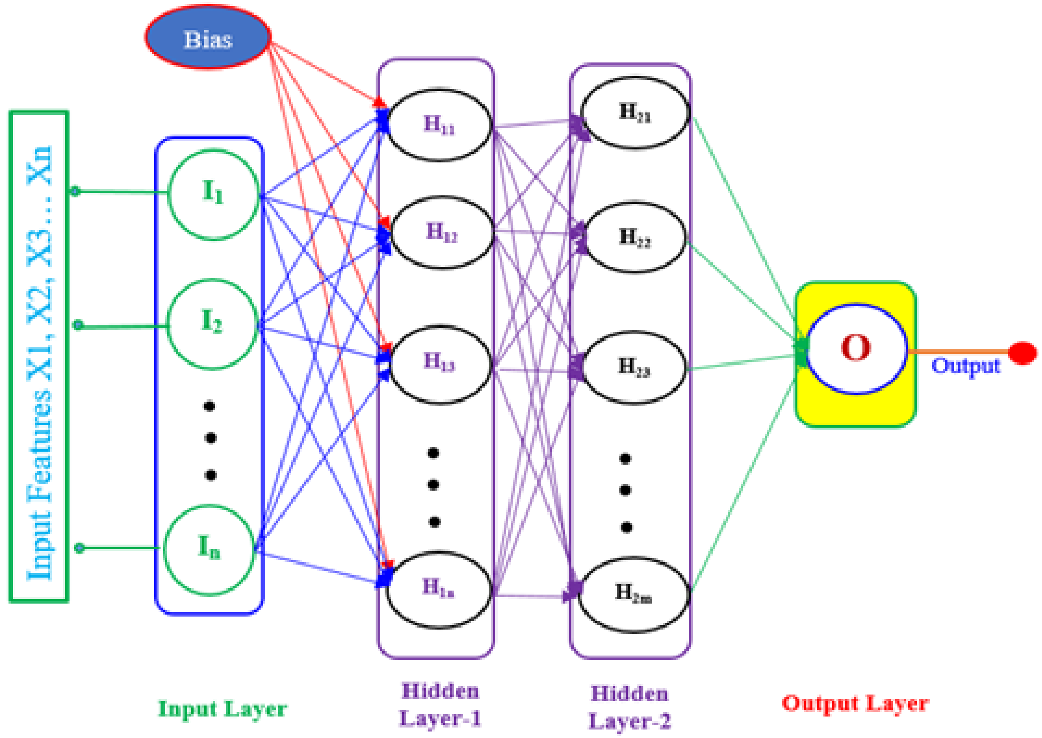



Figure 5. MLP Model

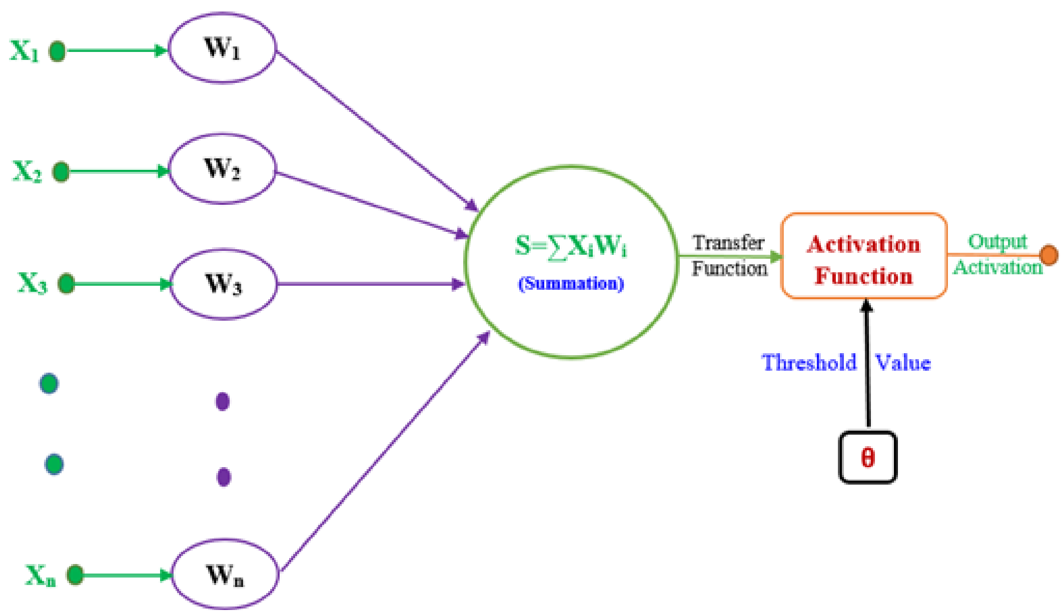



Figure 6. Activation Functions Computations in MLP

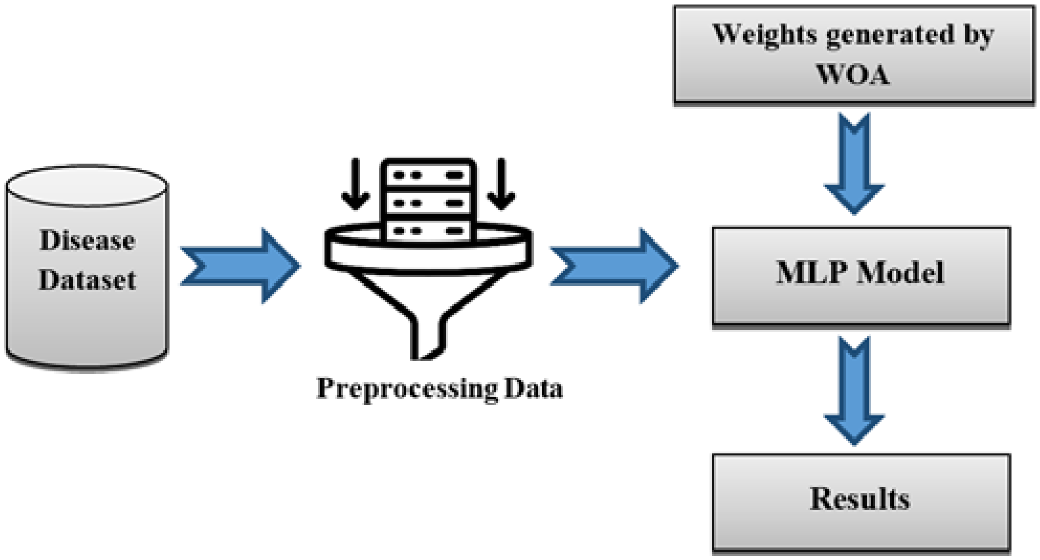



Figure 7. Multilayer Perceptron -Whale Optimization Algorithm (MLP-WOA)

| | | Actual values | |
|---|---|---|---|
| | | **Positive** | **Negative** |
| **Predicted values** | **Positive** | True Positive (TP) | False Positive (FP) |
| | **Negative** | False Negative (FN) | True Negative (TN) |

Figure 8. Confusion Matrix

ROC: The functioning measures that give a thorough analysis of the grouping model. It encapsulates the operation by adding the confusion matrices at all thresholds. AUC gives a numerical representation of a binary classifier's ROC curve. It takes the value between 0 and 1. AUC measures how accurately the model splits positive and negative values [54]. To determine how well a classifier performs, we use a confusion matrix. An instance of class A is classified as class B by counting the number of times it occurs. An actual class represents a row while a predicted class represents a column in a confusion matrix.

## 4. RESULTS

This section describes statistical analysis in detail on diabetes. Applied the ML models like SVM, DTs, and k-NN and their performance parameters. Furthermore, implementation and evaluating the performance parameters of model MLP with different optimizers like MLP-Adam, MLP-Ada-max, MLP-Ada-delta, MLP-RMS, and Proposal model MLP-WOA.

### A. *Statistical Analysis*

In this, we analyse the statistical values like Min, Max, Standard deviation, and Mean of each feature attribute (X1 to X8) in detailed diabetes. By the statistical analysis of diabetes, the class-0 (inactive) attribute values (standard deviation and Mean) are lesser than class-1 (active) and total dataset attribute values. Accordingly, observations, the marginal highest mean and standard deviation attributes are X2, X5, and somewhat X8 (highlighted in table). Table IV depicts a detailed statistical analysis of the Diabetes Dataset.

Figure 9 shows the line plot analysis. A line plot is a plot that shows the numerical information or digital data points in a series associated with straight-line fragments. It just works for numerical information or data. It is grouped with categorical attributes (class attributes). Figure 9(A) shows the Line plot with means per each class-0 and class-1. The red line indicates the means of class-1, and the blue line indicates the class-0 means. The black bars or lines indicate the error value or standard deviation value. The highest error indicator attributes are X2: Glucose, X5: Insulin, and X8: Age. Figure 10 shows the Line plot with individual data points and range ($10^{th}$ -$90^{th}$ percentile) grouping with class attributes (class-0 and class-1). The red lines indicate the data points of class-1, and the blue lines indicate the class-0 data points. The range shows between the 10th and 90th percentile and indicates a blurry light red color. Figure 11 shows the Diabetes variables correlation matrix with a heat map. The values are represented as correlation values between pairs of variables. The value 1 indicates the highly correlated attributes that are positive correlation. In this analysis, the age attribute is highly correlated with attribute pregnancies positively. High negatively correlated with 'skinThickness' attribute. This correlation matrix provides us with valuable information about the relationships

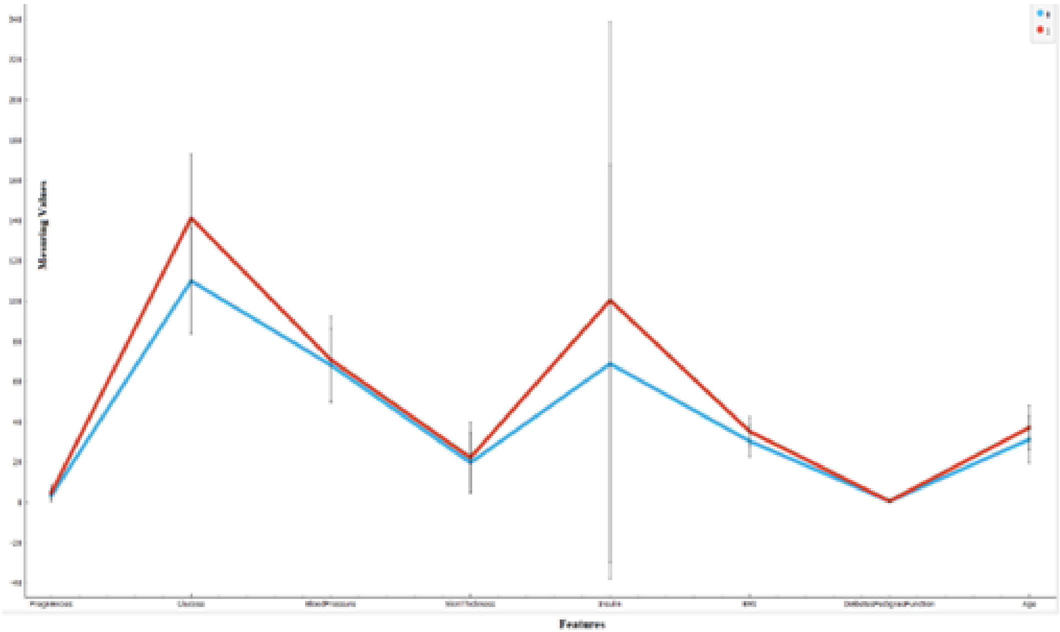



Figure 9. Line Plot with means and Std. Div. for each feature

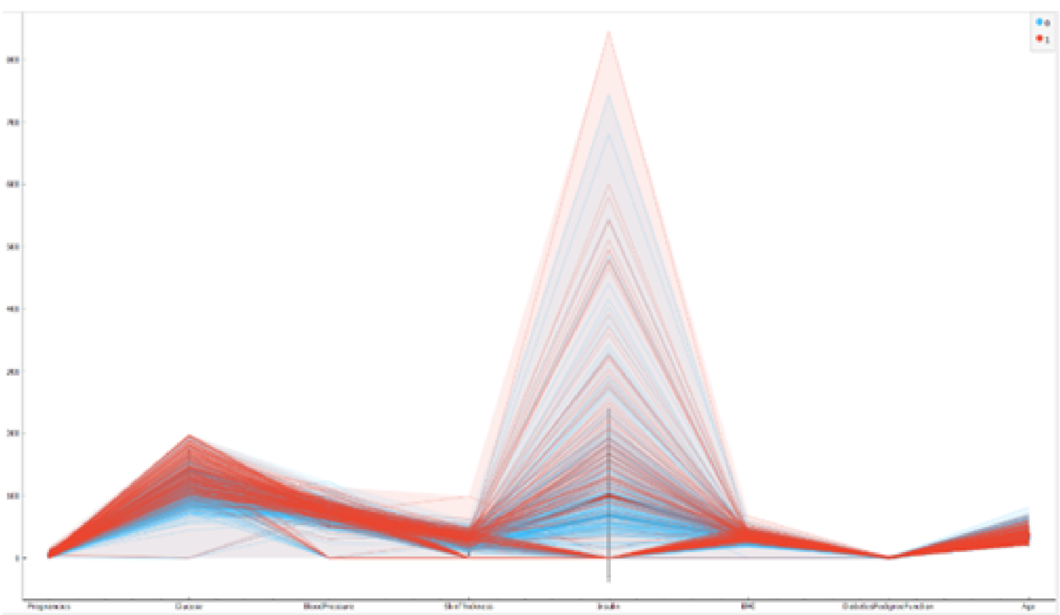


Figure 10. Line Plot with individual data points and range ($10^{th}$ -$90^{th}$ percentile)

TABLE IV. Statistical Analysis Class-0, Class-1 and Total Dataset of Diabetes

| Features | Class-0 (Count_268) | | | | Class-1(Count_500) | | | | Total set(Count_ 768) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Min | Max | Mean | Std.Dev. | Min | M ax | Mean | Std.Dev. | Min | Max |
| X1 | 3.298 | 3.017 | 0 | 13 | 4.865 | 3.75 | 0 | 17 | 3.845052 | 3.3696 | 0 | 17 |
| X2 | 109.98 | 26.141 | 0 | 197 | 141.25 | 31.941 | 0 | 199 | 120.8945 | 31.973 | 0 | 199 |
| X3 | 68.184 | 18.063 | 0 | 122 | 70.824 | 21.492 | 0 | 114 | 69.10547 | 19.356 | 0 | 122 |
| X4 | 19.664 | 14.889 | 0 | 60 | 22.164 | 17.679 | 0 | 99 | 20.53646 | 15.953 | 0 | 99 |
| X5 | 68.792 | 98.869 | 0 | 744 | 100.335 | 138.689 | 0 | 846 | 79.79948 | 115.245 | 0 | 846 |
| X6 | 30.304 | 7.6898 | 0 | 57.3 | 35.142 | 7.2629 | 0 | 67.1 | 31.99258 | 7.8842 | 0 | 67.1 |
| X7 | 0.4297 | 0.2990 | 0.078 | 2.329 | 0.5505 | 0.3722 | 0.088 | 2.42 | 0.471876 | 0.33133 | 0.078 | 2.42 |
| X8 | 31.19 | 11.667 | 21 | 81 | 37.067 | 10.968 | 21 | 70 | 33.24089 | 11.7603 | 21 | 81 |

Note* (Parameters or feature attribute Indicators) X1:Pregnancy (Number of Times), X2:Glucose(tolerance test of OralGlucose), X3:blood pressure (BP(mm Hg)), X4:Skin Thickness (TricepsSkinFoldThickness(mm)) X5:Insulin(2-HourSerumInsulin(mu U/ml),) X6:BMI(BodyMassIndex), X7:DiabetesPedigreeFunction, X8:Age (years)

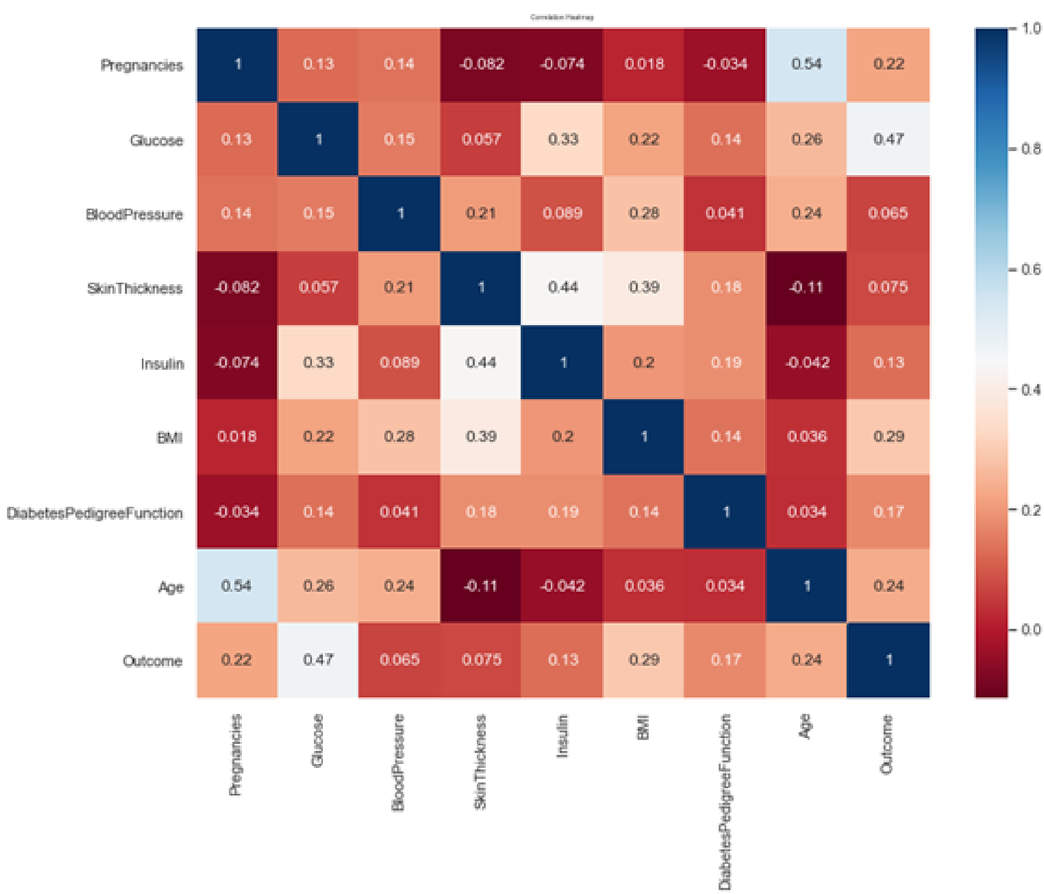



Figure 11. Diabetes Dataset Correlation Matrix(Heat Map)

between variables in the diabetes dataset. For example, we can see that glucose level is the strongest predictor of diabetes status, followed by BMI.

*B. Performances of ML Algorithms and Simulation Results*

In this sub-section, we analyse the data set with reputed ML models like SVM, k-NN, and DTs. The confusion matrices of all experimental ML models are depicted in Figure 12(A) (SVM), k-NN (Figure 12(B)), and DTs (Figure 12(C))). The accuracies of SVM are 0.997, k-NN is 0.698 and DTs are 0.710, and AUC values are 0.726, 0.733, and 0.690, respectively. The values of all experimental ML for the performance parameters Precision, accuracy, F1_score, Recall, and AUC are shown in Table V. models. As per comparison, the decision tree accuracy (0.710) is higher than other algorithms or models. One of the performance parameters of k-NN AUC is higher than other algorithms. The highest values are bolded in the table, and it gives a detailed analysis

*C. ROC Simulation Result Analysis*

Figure 13 shows the experimental ROC curves for ML models. The green curve shows the SVM, the blue cooled curve shows the DTs and the brown curve shows the k-NN ROC. SVM-0.726, k-NN-0.733, and Decision Tree-0.690 are the respective AUC values.

*D. Optimized MLP Simulation results*

The MLP architecture used in our work contains nine neurons in the input layer, eight neurons for input, and one for bias. The bias value is generated randomly in the range of 0 to 1. We have taken three hidden layers, which contain eight neurons in each layer and all are fully connected. We have used the stratified k-folding on the dataset, the k value is set to 5. The learning rate used is 0.89. In each fold, the MLP has trained for 100 epochs. The model hyper-parameter values are depicted in Table VI. In each epoch, the WOA algorithm was run 10 times to produce optimal weight. The objective (fitness) function chosen in our algorithm is the Rosen-brock function after testing the other nine types of fitness functions. The Rosen-rock function equation is given in Equation 18.

$$f(x) = \sum_{i=0}^{n-1} (100 * (x_i - x_{i-1}^2)^2 + (x_{i-1} - 1)^2) \tag{18}$$

Where n represents the total whale population and $x_i$ represents the $i^{th}$ whale.

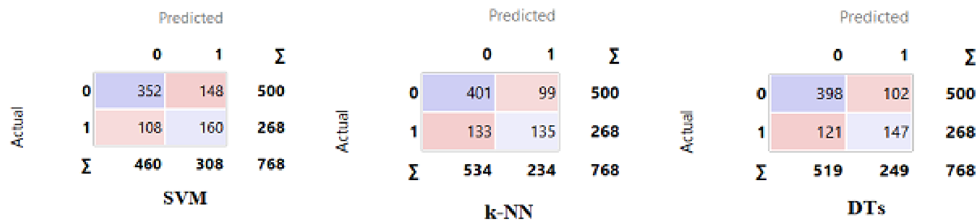
The ReLU and sigmoid activation functions were used at the input and hidden layers consequently. The accuracy is the mean of all the accuracies obtained in all stratified k-folding. The values set to different hyper-parameters in the MLP model are given in Table VI. The confusion matrices obtained after each fold are shown in Figure 14. The accuracy obtained was compared against MLP with WOA and shown in Table VIII.

The accuracy, precision, and recall values calculated at the end of each fold are shown in Figure 15. Figure 16 show the ROC curve for all five folds and plot between algorithms and their accuracy respectively. We also applied the existing optimizers such as adam, ada-max, ada-delta, and RMSprop on the same MLP network model. MLP model accuracy for different optimizers is shown in Table VIII.

## 5. Discussions and Comparative Analysis

The robust prediction of diabetes is difficult due to the narrow labeled dataset and the missing values and outliers present in the datasets of diabetes. Hasan et al. (2020) [55] researched diabetes prediction using ensembling ML models like AdaBoost, XGBoost, Naive Bayes, and so on. ML is a subset of AI, when joined with DM methodologies applies a promising functional part in the field of prediction. Birjais et al. (2019) [56] analyzed the diabetes dataset using ML models like LR, Gradient Boosting (GB), and NBs models for the diagnosis of diabetes. Gestational diabetes (GDM) Mellitus is a typical difficulty during pregnancy that affects tolerable to 15% of pregnant ladies all over the world. Wu et al. (2021) [57] analyzed and collected 16819 (training) and 14992 (testing) case data with 73 variables. For the prediction, they used a deep neural network model to achieve high accuracy of Gestational diabetes. The high-risk advanced pre-diabetes changes into diabetes might be targeted conveyance of interventional programs while preventing and treatment in those healthy. Rao et al. (2022) [58] presents a machine learning-based approach to the classification of diabetes. They used the DT approach and K-NN, two well-known machine learning techniques, to accomplish the same goal. Cahn et al. (2019) [59] concentrated on whether the utilization of the AI-ML model can work on the forecast of incident diabetes using patient information from electronic clinical records. As per simulation results of optimizer MLP with optimizers Adam, Ada-max, Ada-delta, RMS-Prop, and WOA accuracy values are 0.743, 0.734, 0.651, 0.739, and 0.759, respectively. Figure 17 shows the comparative analysis of MLP with optimizers like Adam, Ada-Max, Ada-delta, RMSprop, and WOA. This WOA optimizer with MLP very accurate than other optimization algorithms. The testing accuracy is 0.759. The second and third positions are MLP with Adam and MLP with Adam accuracies of 0.743 and 0.739, respectively. So, MLP-WOA is a high-performance model than other optimizers' MLPs. On the other hand, we compared MLP-WOA to existing ML algorithms SVM, k-NN, and DTs. In this analysis, MLP-WOA performed well with accuracy and error rates than other ML models. The table shows the detailed analysis of the Error rate and Accuracy values of each ML model, and highlighted values are the proposed model.

Figure 18 shows the accuracy of comparative analysis of existing ML and MLP-WOA algorithms. The proposal model MLP-WOA is in a higher position than other existing ML models, with near 76% performance accuracy. The DTs accuracy of 0.71 is the first position in ML models. So, we




(A) Support Vector Machine (B) k-Nearest Neighborhood (C) Decision Tree

Figure 12. Confution Matrices of SVM, k-NN and Decision Trees

TABLE V. Performance parameters of Experimental ML Models

| Model | AUC | Accuracy | F1_score | Precision | Recall |
|---|---|---|---|---|---|
| SVM | 0.726 | 0.667 | 0.667 | 0.669 | 0.667 |
| k-NN | 0.733 | 0.698 | 0.693 | 0.690 | 0.698 |
| Decision Tree | 0.690 | 0.710 | 0.707 | 0.705 | 0.710 |

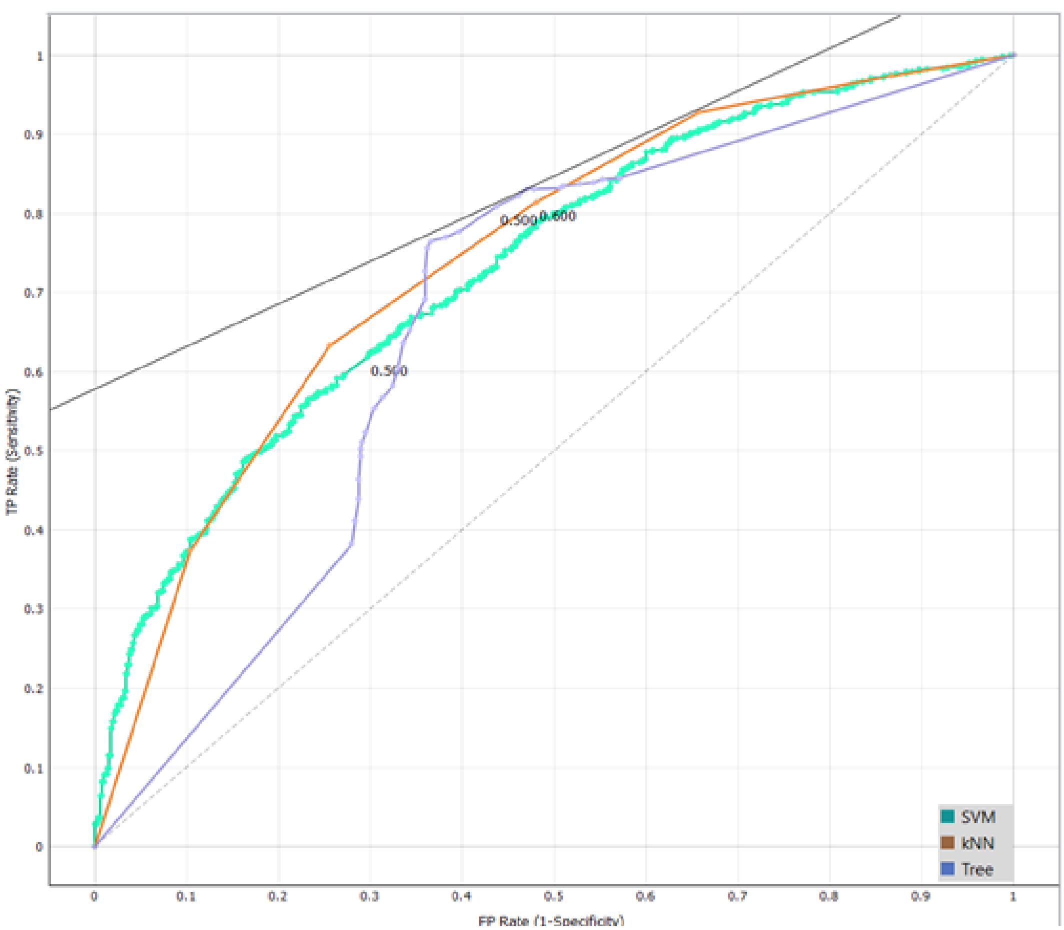



Figure 13. ROC Curves for Experimental ML Algorithms (SVM, k-NN and DTs)

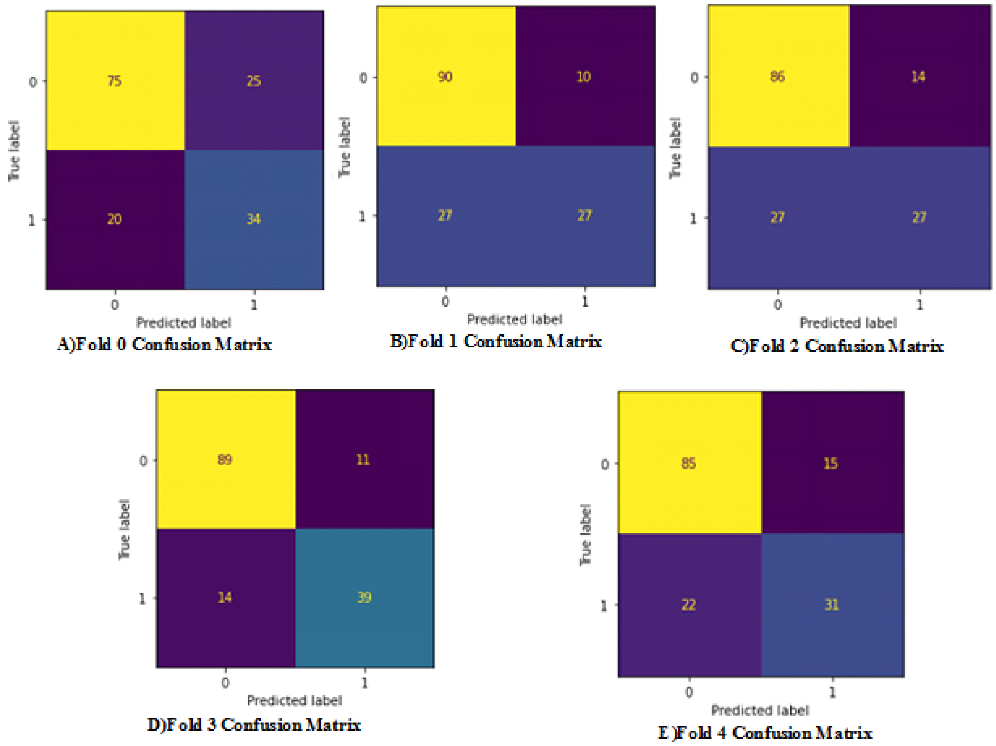



Figure 14. Testing Confusion Matrices for 5- Stratified Folding WOA-MLP

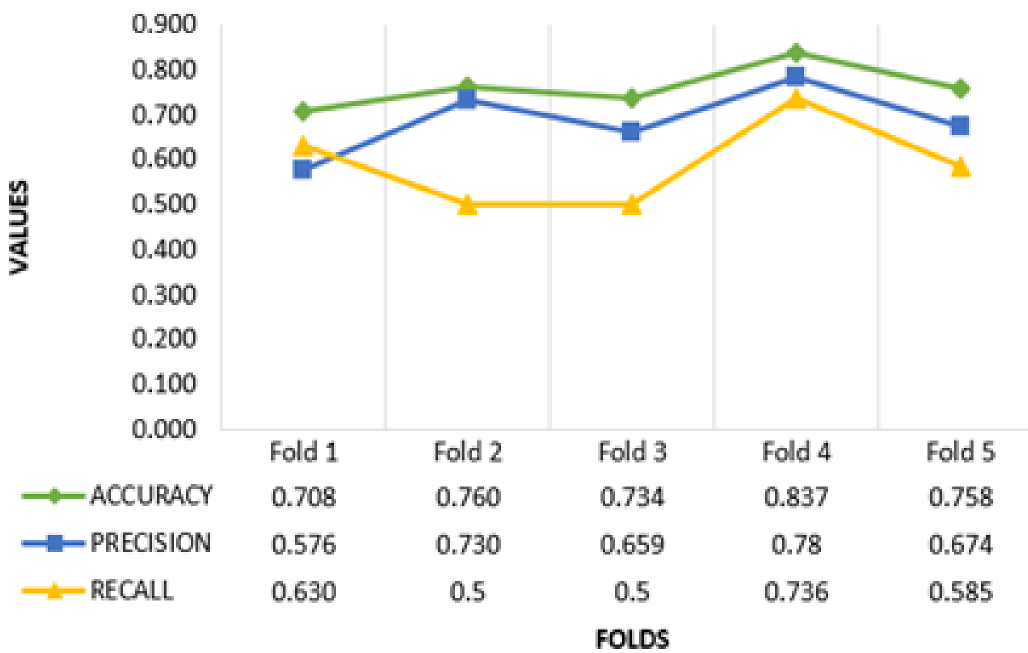



Figure 15. Accuracy, Precision, Recall in MLP_WOA

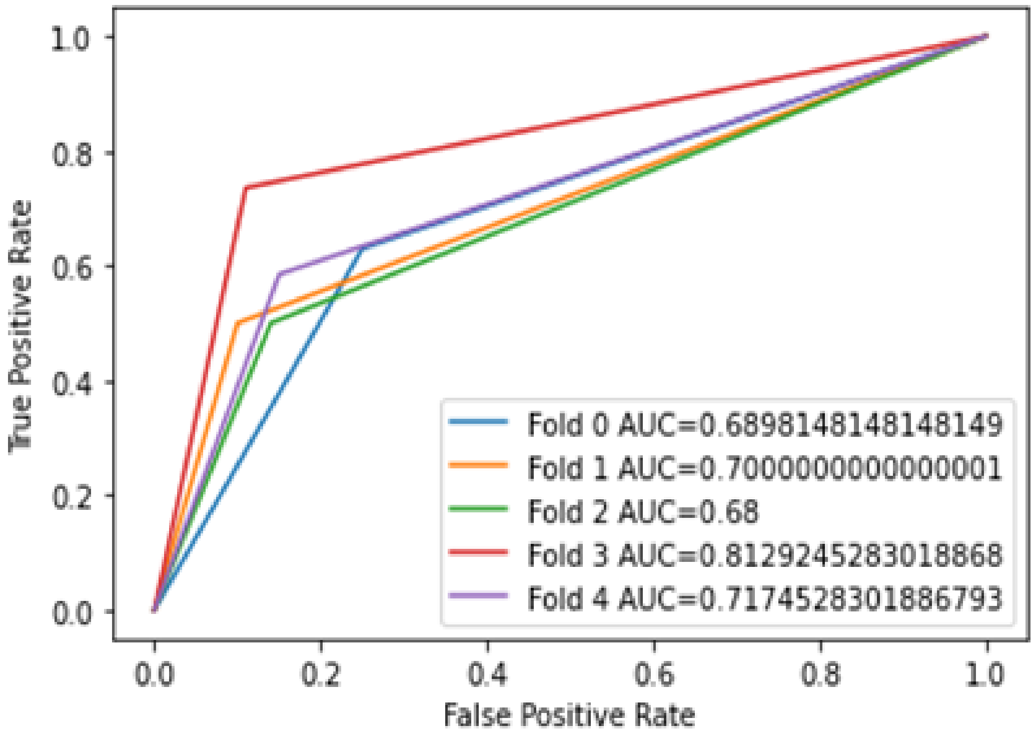



Figure 16. ROC curve for MLP_WOA

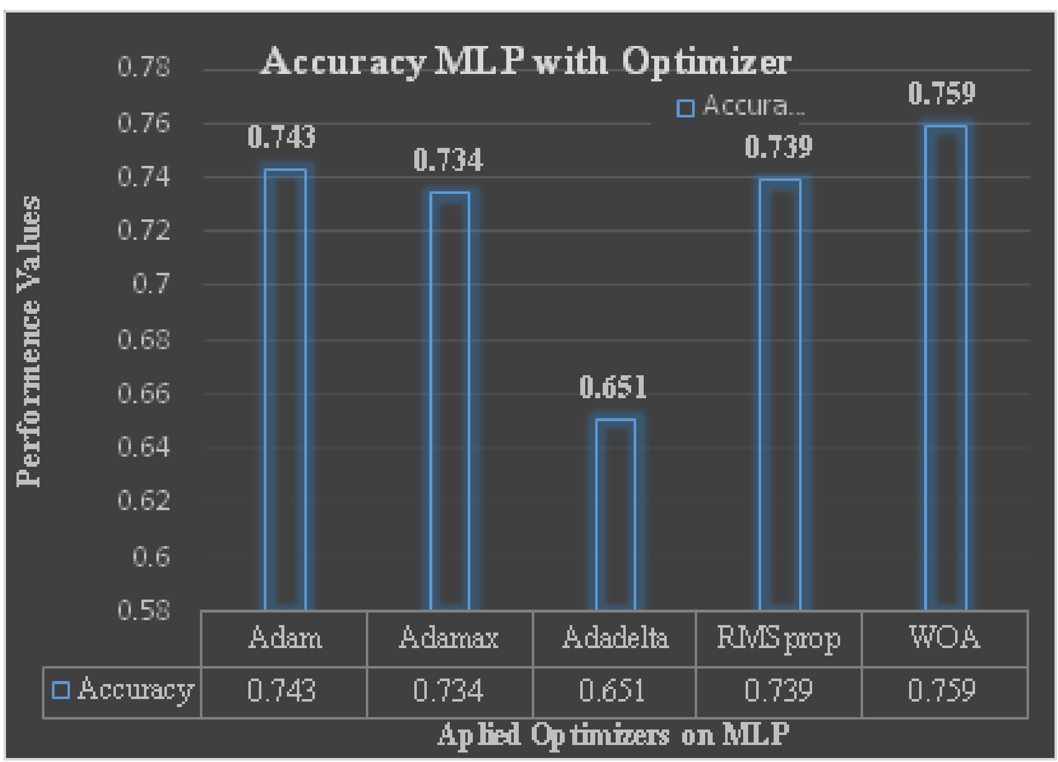



Figure 17. Comparative analysis MLP with optimizers

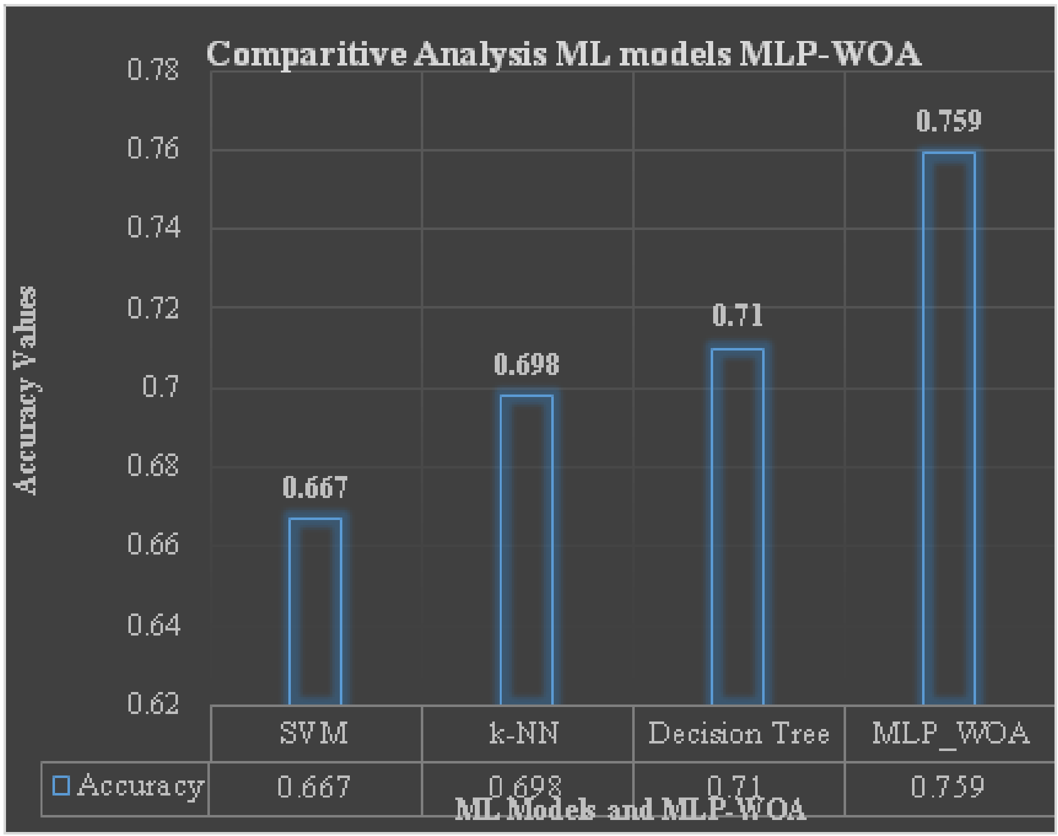



Figure 18. Accuracy comparisons experimental ML to MLP-WOA

TABLE VI. Hyper-Parameter values of our MLP model

| Hyper-Parameter | Values |
|---|---|
| Number of Layers | (1,2,1) |
| Neurons in each layer | 9, 8, 8, 1 |
| Epochs | 100 |
| Activation Functions | ReLU, Sigmoid |
| Optimizers | Adam, Adamax, Adadelta, RMSprop, WOA |
| Learning rate | 0.89 |
| Cross-validation | Five (stratified) |
| Train and Test ratio | 80:20 |
| Performance metrics | Classification accuracy (CA) and ROC |

conclude that the proposed model is the best methodology for predicting the diabetes data set.

## 6. Conclusion

Diabetes is a silent killer disease in the medical field. So many individuals are suffering from this all over the world. It is a tuner of the disease to other diseases like eye infection, heart, kidney, etc. The primary cause of this disease is metabolism, and another one is heredity. Early prediction of diabetes is essential for saving a patient's health. Prevention is another criterion in the medical field. We can analyse the preventions with statistical analysis methods. Prevent and protect against diabetes with good food habits and avoid bad habits like smoke, and drinking. So many studies tell that yoga, meditation, asana, and limited excursive and waking prevent diabetes. Some studies describe good and safe sexual life and peace of mind in preventing diabetes. For the prediction of diabetes, so many existing ML and hybrid optimization models are available. We proposed an MLP-WOA model for diabetes classification and prediction on a benchmark dataset like Pima Indian diabetes (Kagle repository). Compared the results against different ML algorithms. Moreover, the proposed MLP model was also tested with existing optimizers. From the results, it is evident that our MLP with WOA is outperforming compared with other algorithms. Optimized Further, the future scope of the results may be improved by using recent advances in heuristic and meta-heuristic optimization algorithms.

## References

[1] C. for Disease Control, Prevention *et al.*, "National diabetes statistics report: estimates of diabetes and its burden in the united states, 2014," *Atlanta, GA: US Department of Health and Human Services*, vol. 2014, 2014.

[2] P. Saeedi, I. Petersohn, P. Salpea, B. Malanda, S. Karuranga,

TABLE VII. Hyper-Parameter values of our MLP model

| MLP with Optimizers | AUC | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|---|
| MLP-Adam | 0.763 | 0.743 | 0.733 | 0.752 | 0.742 |
| MLP-Ada-max | 0.752 | 0.734 | 0.729 | 0.743 | 0.736 |
| MLP-Ada-delta | 0.680 | 0.651 | 0.658 | 0.647 | 0.652 |
| MLP-RMS-Prop | 0.755 | 0.739 | 0.727 | 0.742 | 0.734 |
| MLP-WOA | 0.803 | 0.759 | 0.760 | 0.743 | 0.751 |

TABLE VIII. Performance Parameters (ACC, ERR) values

| Algorithm | Accuracy | Error rate |
|---|---|---|
| SVM | 0.667 | 0.344 |
| k-NN | 0.698 | 0.312 |
| Decision Tree | 0.710 | 0.291 |
| MLP_WOA | 0.759 | 0.241 |

N. Unwin, S. Colagiuri, L. Guariguata, A. A. Motala, K. Ogurtsova *et al.*, "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the international diabetes federation diabetes atlas," *Diabetes research and clinical practice*, vol. 157, p. 107843, 2019.

[3] V. Jaiswal, A. Negi, and T. Pal, "A review on current advances in machine learning based diabetes prediction," *Primary Care Diabetes*, vol. 15, no. 3, pp. 435–443, 2021.

[4] M. Maniruzzaman, M. J. Rahman, B. Ahammed, and M. M. Abedin, "Classification and prediction of diabetes disease using machine learning paradigm," *Health information science and systems*, vol. 8, pp. 1–14, 2020.

[5] I. Hochberg, D. Daoud, N. Shehadeh, and E. Yom-Tov, "Can internet search engine queries be used to diagnose diabetes? analysis of archival search data," *Acta diabetologica*, vol. 56, pp. 1149–1154, 2019.

[6] T. Joshi, A. K. Singh, P. Haratipour, A. N. Sah, A. K. Pandey, R. Naseri, V. Juyal, and M. H. Farzaei, "Targeting ampk signaling pathway by natural products for treatment of diabetes mellitus and its complications," *Journal of Cellular Physiology*, vol. 234, no. 10, pp. 17 212–17 231, 2019.

[7] A. Berbudi, N. Rahmadika, A. I. Tjahjadi, and R. Ruslami, "Type 2 diabetes and its impact on the immune system," *Current diabetes reviews*, vol. 16, no. 5, p. 442, 2020.

[8] M. A. Atkinson, G. S. Eisenbarth, and A. W. Michels, "Type 1 diabetes," *The Lancet*, vol. 383, no. 9911, pp. 69–82, 2014.

[9] N. P. Tigga and S. Garg, "Predicting type 2 diabetes using logistic regression," in *Proceedings of the Fourth International Conference on Microelectronics, Computing and Communication Systems: MCCS 2019*. Springer, 2021, pp. 491–500.

[10] Y. Chen, J. Tang, Y. Zhang, J. Du, Y. Wang, H. Yu, and Y. He, "Astaxanthin alleviates gestational diabetes mellitus in mice through suppression of oxidative stress," *Naunyn-Schmiedeberg's Archives of Pharmacology*, vol. 393, pp. 2517–2527, 2020.

[11] N. W. Carris, R. R. Magness, and A. J. Labovitz, "Prevention of diabetes mellitus in patients with prediabetes," *The American journal of cardiology*, vol. 123, no. 3, pp. 507–512, 2019.

[12] X. Dai, Z.-c. Luo, L. Zhai, W.-p. Zhao, and F. Huang, "Artificial pancreas as an effective and safe alternative in patients with type 1 diabetes mellitus: a systematic review and meta-analysis," *Diabetes Therapy*, vol. 9, pp. 1269–1277, 2018.

[13] E. A. Al-Suhaimi, *Emerging Concepts in Endocrine Structure and Functions*. Springer, 2022.

[14] T. P. Vital, "Empirical study on uddanam chronic kidney diseases (uckd) with statistical and machine learning analysis including probabilistic neural networks," in *Handbook of Computational Intelligence in Biomedical Engineering and Healthcare*. Elsevier, 2021, pp. 283–314.

[15] P. V. Terlapu, R. P. R. Sadi, R. K. Pondreti, and C. R. Tippana, "Intelligent identification of liver diseases based on incremental hidden layer neurons ann model," *International Journal of Computing and Digital System*, 2021.

[16] P. V. Terlapu, S. B. Gedela, V. K. Gangu, and R. Pemula, "Intelligent diagnosis system of hepatitis c virus: A probabilistic neural network based approach," *International Journal of Imaging Systems and Technology*, vol. 32, no. 6, pp. 2107–2136, 2022.

[17] Y. Xu, X. Liu, X. Cao, C. Huang, E. Liu, S. Qian, X. Liu, Y. Wu, F. Dong, C.-W. Qiu *et al.*, "Artificial intelligence: A powerful paradigm for scientific research," *The Innovation*, vol. 2, no. 4, p. 100179, 2021.

[18] A. Mints, "Analysis of the stability factors of ukrainian banks during the 2014–2017 systemic crisis using the kohonen self-organizing neural networks," *Banks and Bank Systems*, vol. 14, no. 3, p. 86, 2019.

[19] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020.

[20] V. A. Sindagi and V. M. Patel, "A survey of recent advances in cnn-based single image crowd counting and density estimation," *Pattern Recognition Letters*, vol. 107, pp. 3–16, 2018.

[21] J. W. Smith, J. E. Everhart, W. Dickson, W. C. Knowler, and R. S. Johannes, "Using the adap learning algorithm to forecast the onset of diabetes mellitus," in *Proceedings of the annual symposium on computer application in medical care*. American Medical Informatics Association, 1988, p. 261.

[22] N. Panda and S. K. Majhi, "Oppositional salp swarm algorithm with mutation operator for global optimization and application in training

higher order neural networks," *Multimedia Tools and Applications*, vol. 80, no. 28-29, pp. 35 415–35 439, 2021.

[23] ——, "Improved salp swarm algorithm with space transformation search for training neural network," *Arabian Journal for Science and Engineering*, vol. 45, no. 4, pp. 2743–2761, 2020.

[24] A. Mujumdar and V. Vaidehi, "Diabetes prediction using machine learning algorithms," *Procedia Computer Science*, vol. 165, pp. 292–299, 2019.

[25] N. P. Tigga and S. Garg, "Prediction of type 2 diabetes using machine learning classification methods," *Procedia Computer Science*, vol. 167, pp. 706–716, 2020.

[26] M. F. Faruque, I. H. Sarker *et al.*, "Performance analysis of machine learning techniques to predict diabetes mellitus," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*. IEEE, 2019, pp. 1–4.

[27] M. A. Sarwar, N. Kamal, W. Hamid, and M. A. Shah, "Prediction of diabetes using machine learning algorithms in healthcare," in *2018 24th international conference on automation and computing (ICAC)*. IEEE, 2018, pp. 1–6.

[28] C. C. Olisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective," *Computer Methods and Programs in Biomedicine*, vol. 220, p. 106773, 2022.

[29] T. Suresh, Z. Brijet, and T. B. Sheeba, "Cmvhho-dkmlc: A chaotic multi verse harris hawks optimization (cmv-hho) algorithm based deep kernel optimized machine learning classifier for medical diagnosis," *Biomedical Signal Processing and Control*, vol. 70, p. 103034, 2021.

[30] S. K. Kalagotla, S. V. Gangashetty, and K. Giridhar, "A novel stacking technique for prediction of diabetes," *Computers in Biology and Medicine*, vol. 135, p. 104554, 2021.

[31] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, 2021.

[32] M. T. García-Ordás, C. Benavides, J. A. Benítez-Andrades, H. Alaiz-Moretón, and I. García-Rodríguez, "Diabetes detection using deep learning techniques with oversampling and feature augmentation," *Computer Methods and Programs in Biomedicine*, vol. 202, p. 105968, 2021.

[33] P. P. Debata and P. Mohapatra, "Diagnosis of diabetes in pregnant woman using a chaotic-jaya hybridized extreme learning machine model," *Journal of Integrative Bioinformatics*, vol. 18, no. 1, pp. 81–99, 2020.

[34] C. Mallika and S. Selvamuthukumaran, "A hybrid crow search and grey wolf optimization technique for enhanced medical data classification in diabetes diagnosis system," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, p. 157, 2021.

[35] L. Chaves and G. Marques, "Data mining techniques for early diagnosis of diabetes: a comparative study," *Applied Sciences*, vol. 11, no. 5, p. 2218, 2021.

[36] C. Azad, B. Bhushan, R. Sharma, A. Shankar, K. K. Singh, and A. Khamparia, "Prediction model using smote, genetic algorithm and decision tree (pmsgd) for classification of diabetes mellitus," *Multimedia Systems*, pp. 1–19, 2021.

[37] N. Masih, H. Naz, and S. Ahuja, "Multilayer perceptron based deep neural network for early detection of coronary heart disease," *Health and Technology*, vol. 11, pp. 127–138, 2021.

[38] A. K. Verma, S. Pal, and B. Tiwari, "Skin disease prediction using ensemble methods and a new hybrid feature selection technique," *Iran Journal of Computer Science*, vol. 3, pp. 207–216, 2020.

[39] M. Djerioui, Y. Brik, M. Ladjal, and B. Attallah, "Heart disease prediction using mlp and lstm models," in *2020 International Conference on Electrical Engineering (ICEE)*. IEEE, 2020, pp. 1–5.

[40] S. Bhoyar, N. Wagholikar, K. Bakshi, and S. Chaudhari, "Real-time heart disease prediction system using multilayer perceptron," in *2021 2nd International Conference for Emerging Technology (INCET)*. IEEE, 2021, pp. 1–4.

[41] F. A. M. Al-Yarimi, N. M. A. Munassar, M. H. M. Bamashmos, and M. Y. S. Ali, "Feature optimization by discrete weights for heart disease prediction using supervised learning," *Soft Computing*, vol. 25, pp. 1821–1831, 2021.

[42] S. K. Mohapatra, J. K. Swain, and M. N. Mohanty, "Detection of diabetes using multilayer perceptron," in *International Conference on Intelligent Computing and Applications: Proceedings of ICICA 2018*. Springer, 2019, pp. 109–116.

[43] P. Wlodarczak, J. Soar, and M. Ally, "Multimedia data mining using deep learning," in *2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC)*. IEEE, 2015, pp. 190–196.

[44] S. Putatunda, "A hybrid deep learning approach for diagnosis of the erythemato-squamous disease," in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. IEEE, 2020, pp. 1–6.

[45] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting diabetes mellitus with machine learning techniques," *Frontiers in genetics*, vol. 9, p. 515, 2018.

[46] M. W. Nadeem, H. G. Goh, V. Ponnusamy, I. Andonovic, M. A. Khan, and M. Hussain, "A fusion-based machine learning approach for the prediction of the onset of diabetes," in *Healthcare*, vol. 9, no. 10. MDPI, 2021, p. 1393.

[47] S. Abhari, S. R. N. Kalhori, M. Ebrahimi, H. Hasannejadasl, and A. Garavand, "Artificial intelligence applications in type 2 diabetes mellitus care: focus on machine learning methods," *Healthcare informatics research*, vol. 25, no. 4, pp. 248–261, 2019.

[48] P. Sonar and K. JayaMalini, "Diabetes prediction using different machine learning approaches," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 2019, pp. 367–371.

[49] K. De Silva, W. K. Lee, A. Forbes, R. T. Demmer, C. Barton, and J. Enticott, "Use and performance of machine learning models for type 2 diabetes prediction in community settings: A systematic review and meta-analysis," *International journal of medical informatics*, vol. 143, p. 104268, 2020.

[50] A. Brodzicki, M. Piekarski, and J. Jaworek-Korjakowska, "The

whale optimization algorithm approach for deep neural networks," *Sensors*, vol. 21, no. 23, p. 8003, 2021.

[51] T. PanduRanga Vital, M. Murali Krishna, G. Narayana, P. Suneel, and P. Ramarao, "Empirical analysis on cancer dataset with machine learning algorithms," in *Soft Computing in Data Analytics: Proceedings of International Conference on SCDA 2018*. Springer, 2019, pp. 789–801.

[52] P. V. Terlapu, M. Yugandhar, B. Ramesh, B. V. Kumar, and R. Pemula, "Student cognitive learning capability (sclc) prediction system using pca-ann based model," in *2022 International Conference on Computing, Communication and Power Technology (IC3P)*. IEEE, 2022, pp. 11–18.

[53] V. Balasankar, S. Penumatsa, and P. Terlapu, "Intelligent socio-economic status prediction system using machine learning models on rajahmundry ap, ses dataset," *Indian Journal of Science and Technology*, vol. 13, no. 37, pp. 3820–3842, 2020.

[54] T. P. R. Vital, J. Nayak, B. Naik, and D. Jayaram, "Probabilistic neural network-based model for identification of parkinson's disease by using voice profile and personal data," *Arabian Journal for Science and Engineering*, vol. 46, no. 4, pp. 3383–3407, 2021.

[55] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, "Diabetes prediction using ensembling of different machine learning classifiers," *IEEE Access*, vol. 8, pp. 76 516–76 531, 2020.

[56] R. Birjais, A. K. Mourya, R. Chauhan, and H. Kaur, "Prediction and diagnosis of future diabetes risk: a machine learning approach," *SN Applied Sciences*, vol. 1, pp. 1–8, 2019.

[57] Y.-T. Wu, C.-J. Zhang, B. W. Mol, A. Kawai, C. Li, L. Chen, Y. Wang, J.-Z. Sheng, J.-X. Fan, Y. Shi *et al.*, "Early prediction of gestational diabetes mellitus in the chinese population via advanced machine learning," *The Journal of Clinical Endocrinology & Metabolism*, vol. 106, no. 3, pp. e1191–e1205, 2021.

[58] G. Jagadeeswara Rao, A. Siva Prasad, S. Sai Srinivas, K. Sivaparvathi, and N. Panda, "Data classification by ensemble methods in machine learning," in *Advances in Intelligent Computing and Communication: Proceedings of ICAC 2021*. Springer, 2022, pp. 127–135.

[59] A. Cahn, A. Shoshan, T. Sagiv, R. Yesharim, R. Goshen, V. Shalev, and I. Raz, "Prediction of progression from pre-diabetes to diabetes: Development and validation of a machine learning model," *Diabetes/metabolism research and reviews*, vol. 36, no. 2, p. e3252, 2020.

**Dr.PanduRanga Vital Terlapu** obtained his Bachelor of Science in Computer Science from Andhra University in A.P, India in 1995, and completed his Master of Computer Application from Andhra University in 1998. He further pursued his M. Tech in Computer Science and Engineering from Acharya Nagarjuna University in A.P., India, and completed his Ph.D. in Computer Science and Engineering from GITAM University in A.P., India. With a total of 23 years of teaching and 18 years of research experience, he currently holds the position of Associate Professor in the Department of Information Technology at Aditya Institute of Technology and Management (AITAM), India. Dr. Terlapu is an esteemed member of the Association for Computing Machinery (ACM) and holds Lifetime Memberships from the International Computer Science and Engineering Society (ICSES), USA, and the Indian Society for Technical Education (ISTE), New Delhi, India, and has 5 reputed memberships with the close association. Dr. Terlapu has contributed to the field of computer science with over 44 research papers published in reputed international journals, including SCI, SCOPUS-indexed journals, and conferences published by Springer, Elsevier, and available online. He is also a reviewer for reputable journals from Springer, Elsevier, and IEEE databases. His primary research focus is on Machine Learning, Image Processing and Deep Learning, Data Mining, Data and Big Data Analytics, IoT and Computational Intelligence, Voice Analysis and Voice Processing, signal processing, and Bioinformatics.

**JagadeeswaraRao G** is currently pursuing Ph.D. in Computer science and systems engineering department, Andhra University, Visakhapatnam. He completed his M. Tech in Information Technology from NCET, Vijayawada in 2010. With a total of 12 years of teaching and 9 years of research experience, he currently holds the position of Assistant Professor in the Department of Information Technology at Aditya Institute of Technology and Management (AITAM), India. His areas of interest are Machine Learning, Deep Learning, Data Mining, and Bioinformatics.

**Dr. A Siva Prasad** currently working as a professor and Head of the computer science department in Government PG College, Tekkali, Andhra Pradesh. He received his doctorate in 2009 from Andhra University, Visakhapatnam. He is having years of experience in both R&D and academics helped in being and growing as a scientist in Computer Science Engineering, and also in showing significant leadership and teaching skills. I have generated a substantial body of original research and played a significant role and achieved sustained and productive research output of International papers and review articles in peer-reviewed journals of significant impact and many others in the pipeline. He is widely known for my energy and enthusiasm for research. He participated in organizing conferences, Workshops, and symposia organized at national and international levels.

**Dr. Yegireddi Ramesh** has been working as a Professor & HOD in the Department of IT at AITAM, Tekkali, Srikakulam, Andhra Pradesh, India. He completed his Bachelor of Science (B.Sc) from Andhra University. He completed a Master of Computer Applications (MCA) from Osmania University, a Master of Technology (M.Tech - CSE) from JNTUH Hyderabad, and Ph. D (CSE) from JNTUK Kakinada. He has more than 22 years of experience in teaching and research and also has good knowledge of Information Security, Networks, Future Internet Architecture & Technologies, Block-chain, and the Internet of Things along with academic subjects, etc. Currently, he is an active member of ISTE, CSI, and ACM. He is a good reviewer for some eminent journals and conferences. He also wrote one Textbook entitled A Methodical Observation of the Analysis of Cryptographic Algorithms. He published 23 Research papers in various reputed (UGC-approved, Scopus, SCI, Web of Science) International and National Journals, magazines, and conferences.

**Dr. Chappa Ramesh** Presently working as a professor, Dept of Computer Science & Engg and Dean, IQAC, Aditya Institute of Technology and Management (AITAM), Tekkali, Andhra Pradesh. He completed his Ph.D. in CSE with a specialization of image processing from Jawaharlal Nehru Technological University, Hyderabad in the year 2015. He did his M.Tech (CSE) from Jawaharlal Nehru Technological University, Hyderabad in 2004 & M.Tech (Remote Sensing) from Andhra University, Vishakhapatnam in 1997, and B.Tech from V.R.Siddhardha Engg College, Vijayawada in 1992. He is having an experience of more than 27 years including 3 years of industrial and 24 years of academic and teaching experience. He has been holding various academic and administrative positions during his career. He is having vast experience teaching students of B.Tech, M.Tech, and MCA. He is currently the chairman of the Institute of Engineers (I.E), Srikakulam local center. His research areas include image processing, data mining, and machine learning. He has published more than 60 papers in reputed international journals and conference proceedings. He has successfully completed two major research projects sponsored by the Natural Resources Data Management System (NRDMS), Dept of Science and Technology (DST). He organized four faculty development programs sponsored by DST (NRDMS), and DST (STUTI). He is the accreditation ambassador for the UGC-para marsh scheme. Presently 5 Research scholars are pursuing Ph.D. under his supervision. He has been a member of professional societies like the Indian Society for Technical Education (ISTE) -Life Member, the Computer Society of India (CSI) - Life Member, Institute of Engineers (IE)-Life Member.

**Yoshita Bellala** is currently working as Associate Professional Software Engineer at EIT Services India P. Ltd of DXC Technology, Hyderabad. She completed her B. Tech in Information Technology from AITAM, Tekkali in 2022. Her areas of interest are Machine Learning and Cloud Computing.