**Diabetes Prediction Using Dolphin Swarm Optimization based on Radial Basis Neural Network**

**1. Introduction**

Diabetes is a prolonged health problem with injurious, yet preventable disease. It is caused by high glucose levels in blood due to faults in production of insulin [1]. Type 1 diabetes results from beta cell destruction and this usually results in a complete insulin deficiency in the body [2]. This type is usually identified in kids and teenagers. Type 2, the most common method of diabetes, is a more progressive kind of the disease that is typically diagnosed in adults and is considered by an insulin secretory defect [3]. In 2000, the number of type 2 diabetes patients are 15.1 million and in 2030, it is expected to rise about 36.6 million. There are 12 million men and 11.5 women with diabetes. In adult patients, non-Hispanic White are 6.6%, non-Hispanic Black are 11.8%, Hispanic are 10.4%, and Asian are 7.5% [5].There is no treatment for diabetes as of yet, nevertheless, an early diagnosis of this disease [6], followed by a suitable medication, a balanced diet, and regular physical activity go a long way in controlling glucose levels in blood and decreasing the danger of rising complications [7]. Controlling the blood glucose level of diabetic patients and keeping it within the normal range (70 mg/dL -120 mg/dL) is therefore the focal goal of physicians [8]. However, the main challenge in blood glucose control is the need to keep its level as close to the normal range as possible, while keeping the number of hypoglycemia to a minimum [9]. Hypoglycemia is a situation that happens when blood glucose falls dangerously low (below 60 mg/dL).

A predictive analysis is a technique which includes numerous machine learning algorithms (MLA), statistical methods, and data mining methods which utilize present as well as the past data for predicting the future. These analytics are performed using both machine learning and regression technique. Machine learning is an important artificial intelligence feature that supports the upgrading in the computer system to achieve knowledge from the past experience. This technique is crucial in current circumstances instead to except human hard work by supporting automation. Many metaheuristic algorithms and MLA are widely used in the detection of diabetes. The Support Vector Machine (SVM) [10], Artificial Neural Network (ANN), Convolution Neural Network (CNN), Gravitational Search Algorithm [11], and also Optimization techniques like Ant Colony Optimization (ACO) [12] are utilized by most of the researchers. Also researchers utilize many machine learning algorithms like Decision tree, J48, Naïve Bayes, Random forest [13], etc. for diagnosing various diseases. These methods possess integral limitations. Few problems in this are high computational time, slow convergence, and stuck in to local optima. To overcome these issues we proposed Diabetic Prediction using Dolphin Swarm Optimization (DSO) based on RBNN.

**2. Literature Review**

Fiarni et al. [14] introduced a data mining (DM) method for forecasting diabetics. The medical information of diabetics are divided into 4 sets, they are neuropathy, retinopathy, nephropathy and others. The DM consist of K-mean clustering (KMC), Naïve bayes (NB), J48 decision tree (DT). For constructing rule-based model for forecasting determination, they evaluate the presentation from clustering and classification method. The clustering method consist of KMC, the classification method consist of NB & DT. Then compare with these 2 methods. From the analysis of DM, it classify and sub features into three main micro vascular diabetes difficulty diseases. This method gives diabetes complication diseases, however the current accuracy level have to improve.

Yuvaraj et al. [15] presented an algorithm of machine learning (ML) on account of Handoop cluster (HC) for prediction of diabetes disease. The forecasting system consist of 2 parts, diabetes collection module and diabetes prediction module. To hold the large information set the R tool is fixed in HC. Then the extracted features are divided using different ML, they are NB, DT and random forest algorithm (RFA) for finding diabetes. Then by analyzing these parameters it is clear that RFA gives accuracy better than others. This methods gives high efficiency, but less number of nodes in HC.

Devi *et al.* [16] proposed a method by integrating Farthest First (FF) and a classifier procedure to diagnose patients with Diabetes mellitus (DM). The FF algorithm clusters the information into numerous groups. FF comprises of two steps: (i) centroid selection (ii) cluster assignment. At the first step, a random data position is chosen as the cluster center and at the next step, an alternate cluster center is decided. In this way, the information ideas are clustered to the close centroid. After clustering, the data is given to SOM for classification process. The SOM helps to select the optimal values for the multipliers. This hybrid technique has attained higher accuracy than other existing schemes.

Zhu *et al.* [17] proposed a technique by integrating k-means and PCA to predict diabetics. Their approach includes pre-processing and classification stages. The irrelevant features are initially filtered by PCA technique which helps to reduce the training time. It also enhances the performance by reducing the dimensionality without loss of sensitive data. At the next step, clustering is performed by k-means technique which is said to be simple and efficient classifier. Finally, logistic regression is applied to classify data into several categories. Here, 0 denotes the patients who are tested negative for diabetics and 1 denotes the patients tested positive for diabetics. The accuracy of this method is higher than several other existing techniques.

Jayashree J, Kumar SA. [18] Proposed evolutionary correlated gravitational search algorithm (ECGS). Here, they utilized this protocol for choosing the optimal features. Then these features are given to Hopfield NN (HNN) for processing. This work is categorized in to two parts: part 1 and part 2. Initial one is to collect the details about the patient and to decrease the measurement of the data using feature selection method ECGS. The fitness value in this method is estimated on account of fuzzy membership function. In final they utilize HNN for the prediction of diabetes. The diabetic data is collected from the Pima Indian Diabetic Dataset. The accuracy of this system is better. Even though, the efficiency is quite low.

Kumar NK *et al* [19] proposed hybrid optimized random forest classifier (GA-ORF) along with the Genetic Algorithm (GA). Here GA has been utilized by the random forest classifier in which each chromosomes resembles different type of trees and also the different length of chromosomes. The preprocessed data is categorized in to two sets they are training set and validation set. They send the training set to the GA in which it decompose in two classes positive and negative. Later the values ntrees and mtrees are decomposed and given to the random forest classifier. Here they use this obtained classifier for predicting the diabetes. The performance of this system is better. Moreover, the accuracy is quite low.

Lukmanto et al. [20] introduced F-score feature selection (FS) and fuzzy SVM. In this method the dataset is used for predicting diabetes. Initially the data is pre-processed. Second the FS is utilized to recognize the valuable features in dataset. Then this data is given to SVM, for create fuzzy rules, SVM train the dataset. At last fuzzy inference is utilized to categorize the output. This method gives simple classification strategy, however, the accuracy should be enhanced in future.

**3. Problem definition**

From the literature review it is concluded that the existing methods used in the classification of patients with and without diabetes is inefficient in predicting. In classification based on neural network, the network structure of conventional methods are complex with multiple layers. Some drawbacks of prediction of diabetes is low convergence speed, inaccuracy, and low significant performance. To maximize the convergence speed and accuracy Dolphin Swarm Optimization Algorithm based on Radial Based Neural Network is proposed.

**4. Objective**

The key objective of the research wok is as follows:

* To maximize the significant performance and accuracy in the classification of patients with diabetes or without diabetes.
* To discover the optimal cluster center value for better classification a Dolphin Swarm Optimization Algorithm is newly proposed.
* To enrich the training and classification of patients with diabetic and without diabetic, a Dolphin Swarm optimization (DSO) algorithm based on Radial Basis Neural Network (RBNN) is proposed.
* To show the dominance of the proposed approach by comparing it with the existing methods.

**5. Proposed Methodology**

The proposed system of the Diabetes prediction is given in the Figure 1. In our proposed method, the first step is to preprocess input data. This is for selecting the accurate attributes for additional processing. Then to achieve an accurate prediction, the dimensionality of the data is reduced using the Fast Correlation Based Feature Selection (FCBS). The next step is the clustering process. Here we adopt Mean shift clustering which is a circular sliding window. The final step is classification, here the prediction of patients with diabetes is done by Dolphin Swarm Optimization algorithm based RBNN. The hidden nodes of RBNN uses the radial basis function and the optimal center of the cluster is evaluated using DSO. By utilizing this method we can find the classification of patients with diabetes and without diabetes.



Figure 1 Proposed Method for Diabetes Prediction

**6. Process flow**

1. ***Preprocessing***

This phase handles the data set to achieve an accurate data. Therefore, we assign values for some carefully chosen attributes such as BMI, glucose level, Age, Blood Pressure because the values of this attributes cannot be zero. This converts the medical value into diagnosis value.

1. ***Feature Selection***

This is a dimension reduction step which removes the unresolved data to reduce the size and to improve the learning accuracy. Here we adopt Fast Correlation Based Feature selection (FCBF).

1. ***Clustering***

In clustering phase random data position is chosen as the cluster center and at the next step, an alternate cluster center is decided. We adopt Mean Shift Clustering (MSC) which allots the data points to the clusters by shifting points.

1. ***Classification***

Finally in Classification, a DSO based RBNN classifies the patients with or without Diabetes.

**7. Expected Outcomes**

The proposed RBNN based on DSO approach for predicting patients with diabetes or without diabetes will be tested under MATLAB/Simulink using PIMA Indian Diabetes Dataset and its performance based on accuracy is to be compared with the existing Diabetes Prediction techniques.

**References**

1. Sivaparthipan CB, Karthikeyan N, Karthik S. Designing statistical assessment healthcare information system for diabetics analysis using big data. Multimedia Tools and Applications. 2020 Apr;79(13):8431-44.
2. DiMeglio LA, Evans-Molina C, Oram RA. Type 1 diabetes. The Lancet. 2018 Jun 16;391(10138):2449-62.
3. Navarro-González D, Sánchez-Íñigo L, Pastrana-Delgado J, Fernández-Montero A, Martinez JA. Triglyceride–glucose index (TyG index) in comparison with fasting plasma glucose improved diabetes prediction in patients with normal fasting glucose: The Vascular-Metabolic CUN cohort. Preventive medicine. 2016 May 1;86:99-105.
4. Zhang T, Li L, Shang Q, Lv C, Wang C, Su B. Circulating miR-126 is a potential biomarker to predict the onset of type 2 diabetes mellitus in susceptible individuals. Biochemical and biophysical research communications. 2015 Jul 17;463(1-2):60-3.
5. Seo DC, Choe S, Torabi MR. Is waist circumference≥ 102/88 cm better than body mass index≥ 30 to predict hypertension and diabetes development regardless of gender, age group, and race/ethnicity? Meta-analysis. Preventive medicine. 2017 Apr 1;97:100-8.
6. Sebastiani G, Nigi L, Grieco GE, Mancarella F, Ventriglia G, Dotta F. Circulating microRNAs and diabetes mellitus: a novel tool for disease prediction, diagnosis, and staging?. Journal of endocrinological investigation. 2017 Jun 1;40(6):591-610.
7. Goodenough CJ, Liang MK, Nguyen MT, Nguyen DH, Holihan JL, Alawadi ZM, Roth JS, Wray CJ, Ko TC, Kao LS. Preoperative glycosylated hemoglobin and postoperative glucose together predict major complications after abdominal surgery. Journal of the American College of Surgeons. 2015 Oct 1;221(4):854-61.
8. Sheen YJ, Sheu WH. Association between hypoglycemia and dementia in patients with type 2 diabetes. Diabetes research and clinical practice. 2016 Jun 1;116:279-87.
9. Martyn-Nemeth P, Farabi SS, Mihailescu D, Nemeth J, Quinn L. Fear of hypoglycemia in adults with type 1 diabetes: impact of therapeutic advances and strategies for prevention-a review. Journal of Diabetes and its Complications. 2016 Jan 1;30(1):167-77.
10. Maniruzzaman M, Kumar N, Abedin MM, Islam MS, Suri HS, El-Baz AS, Suri JS. Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. Computer methods and programs in biomedicine. 2017 Dec 1;152:23-34.
11. Adem K. Exudate detection for diabetic retinopathy with circular Hough transformation and convolutional neural networks. Expert Systems with Applications. 2018 Dec 30;114:289-95.
12. Chen W, Chen S, Zhang H, Wu T. A hybrid prediction model for type 2 diabetes using K-means and decision tree. In2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS) 2017 Nov 24 (pp. 386-390). IEEE.
13. Das H, Naik B, Behera HS. Classification of diabetes mellitus disease (DMD): a data mining (DM) approach. InProgress in computing, analytics and networking 2018 (pp. 539-549). Springer, Singapore.
14. Fiarni C, Sipayung EM, Maemunah S. Analysis and prediction of diabetes complication disease using data mining algorithm. Procedia Computer Science. 2019 Jan 1;161:449-57.
15. Yuvaraj N, SriPreethaa KR. Diabetes prediction in healthcare systems using machine learning algorithms on Hadoop cluster. Cluster Computing. 2019 Jan 16;22(1):1-9.
16. Devi RD, Bai A, Nagarajan N. A novel hybrid approach for diagnosing diabetes mellitus using farthest first and support vector machine algorithms. Obesity Medicine. 2020 Mar 1;17:100152.
17. Zhu C, Idemudia CU, Feng W. Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. Informatics in Medicine Unlocked. 2019 Jan 1;17:100179.
18. Jayashree J, Kumar SA. Evolutionary correlated gravitational search algorithm (ECGS) with genetic optimized hopfield neural network (GHNN)–a hybrid expert system for diagnosis of diabetes. Measurement. 2019 Oct 1;145:551-8.
19. Kumar NK, Vigneswari D, Krishna MV, Reddy GP. An optimized random forest classifier for diabetes mellitus. InEmerging Technologies in Data Mining and Information Security 2019 (pp. 765-773). Springer, Singapore.
20. Lukmanto RB, Nugroho A, Akbar H. Early detection of diabetes mellitus using feature selection and fuzzy support vector machine. Procedia Computer Science. 2019 Jan 1;157:46-54.