Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, No. 3, May : 2023

AN ADAPTIVE SUNFLOWER OPTIMAZATION MODEL FOR EFFICEINT DIABETES PREDICTION USING MEDICAL DATA SET

K. Kanmani Research scholar, Dr. Ambedkar Govt Arts College, Vyasarpadi, Chennai- 600039, India, Assistant professor, Department of Computer Applications, College of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur- 603203. Chennai, TN, India
Dr. A. Murugan Associate Professor and Head, PG and Research Department of Computer Science, Dr. Ambedkar Govt. Arts College (Autonomous), Vyasarpadi, Chennai – 600039, TN, India

Abstract:

Objectives:The problem of diabetes prediction with medical data set is deeply analyzed . There are number of techniques in predicting the disease like Support vector machine (SVM), Genetic Algorithm, Ensemble learning, Artificial neural network, and so on. However, the methods struggling to achieve expected prediction accuracy. **Methods:** To handle this issue, an efficient Adaptive Sunflower Optimization based Diabetes Prediction Model (ASODPM) is presented towards diabetes prediction. The method start with preprocessing the diabetes data set given with the use of Value Based Normalization technique. The value based normalization technique identifies the noisy records and eliminates the noisy records. **Findings:** With the purated data set, the method extract the features and generates K number of sunflowers depending on the number of disease classes. Initially, the method uses the classfied labels to index the data points to the sunflower. At the training phase, the method iteratively applies the sunflower optimization algorithm with the objective function named Feature Centric Fitness (FCF) Estimation. **Novelty:**Using the objective function, the method computes the value of multi feature morality value (MFMV) for the sunflower considered. The method iteratively measure the MFMV value for different sunflowers of population and validates the clusters of sunflower. At the training phase, the method computes the value of Multi Feature Pollination Rate (MFPR) towards various group of sunflower. Based on the value of MFMV and MFPR, the method compute disease projection value (DPV) to identify the class of sample and performs diabetes prediction. The method hikes the prediction accuracy and reduces time complexity.

Keywords: Diabetes Prediction, Sunflower Optimzation, Medical Data, ASODPM, MFMV, MFPR.

1 Introduction

The changing lifestyle of human society introduces several challenges in their lifecycle. In this way, number of diseases are identified every year which target the health of human society. Some of them would pass like a cloud and disappear shortly but some of them becomes chronic and cannot be cured for their lifetime but managed. The diabetes is the one among them which is identified as a chronnic disease which cannot be cured but managed with set of drugs and other activities. However, predicting the upcoming disease would help the medical practitioner in managing and preventing the disease at the earliest. When you predict the diabetes for a person, then it can be prevented by modifying the lifestyle and diat of the person, the disease can be prevented.

Towards the scope, various schemes are available in literature. Ensemble based approaches are used in predicting the disease, which generates number of ensembles and maintains them. By measuring ensemble similarity with different ensemble set, the method would predict the disease. Similarly, support vector machine has been used in predicting the disease, which computes support value for the test sample in identifying the disease. The decision tree has been used in predicting diabetes which

works according to measuring the similarity among the features and values. Also, various techniques are discussed in literature available towards predicting diabetes for any person according to the symptoms and values given. The issue among the approaches are the lack of accuracy in predicting the disease. The methods consider only limited features and introduces poor performance in predicting the disease.

Machine learning algorithms have great influence on several medical problems. It has been used in the detection of various diseases. The machine learning algorithms are capable of finding hidden features and missing values which support the achievement of higher accuracy in any classification problem. The performance of disease prediction can be improved by adapting more number of features and more volume of data sets. The existing approaches consider only limited features and limited data sets which challenge them in achieving higher accuracy in disease prediction. By considering all these issues, an efficient Adaptive Sunflower Optimization Model is presented towards predicting diabetes. The proposed work uses sunflower optimization algorithm towards predicting the diabetes. The Sunflower algorithm is a hybrid algorithm which uses number of parameters and values to measure the morality of the sunflower and pollination of sunflower. The algorithm has been designed modified with the objective function Feature Centric Fitness (FCF) function which involve in measuring Multi Feature Pollination Rate (MFPR) to decide the inclusion or exclusion of sunflower from the set.

2. Related Works

There are number of approaches discussed in literature towards predicting diabetes and some of them are discussed in detail in this section.

An efficient glucose prediction model is prsented in [1], which uses historical data with local models. The method consider the glucose profiles of various sets and cluster them using fuzzy C-means. Also, the method uses Box-Jenkins methodology is used to detect the seasonal model and integrate them to produce prediciton. An Average Weighted Objective Disance (AWOD) model is presented in [2], towards predicting diabetes. The method consider various health factors of different individuals and finds set of factors with various priorities. Based on the factors identified, the method computes AWOD value to identify the possibility of disease.

A hybrid model is presented in [3], which combines different machine learning algorithm to perform classification. The method uses a weighted ensembling of various machine learning algorithms in predicting the disease according to their weights.

A machine learning based approach is presented in [4], to predict Gestational diabetes mellitus (GDM). The method uses PIMA (Polyisocyanurate Insulation Manufacturers Association) data set for performance evaluation. An probabilistic ensemble classification algorithm is presented in [5], which uses Local Median-based Gaussian Naive Bayes (LMeGNB) classification algorithm towards classification.

A hybrid fusion model is presented in [6], which combines Support Vector Machine (SVM) and Artificial Neural Network (ANN) models. The method analyze the tuples of data set and classify them. The SVM is used for future seleciton where the classificaiton is performed with ANN. The ANN applies classification based on the fuzzy logic generated from the selected features set.

An non parametric based approach with gaussian regression model is presented in [7], towards predicting diabetes. The method predicts type 1 diabetes according to the data set given. An multi layer perceptron model is presented in [8], towards diabetes prediction. The method applies Grey Wolf Optimization (GWO) and an Adaptive Particle Swam Optimization (APSO) for feature selection where MLP is used towards classification. A comprehensive classification model is presented in [9], which

consider number of approaches and analyze their performance according to the key metrics using the data set given.

A hybrid model for diabetes prediction is presented in [10], which analyze the performacne of random forest, decision tree and neural network towards predicting diabetes. The method performs cross validation on different models of diabetes prediction. The method reduces the dimensionality using principal component analysis (PCA) and minimum redundancy maximum relevance (mRMR). Machine learning based diabetes risk prediciton model is prsented in [11], for the prople of north kashmir india. The analysis conducted with six MLA (Member of Legislative Assembly) using Random Forest (RF), Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), Gradient Boost (GB), Decision Tree (DT), and Logistic Regression (LR).

An DNN (Deep Neural Network) based diabetes prediction model is presented in [12], which uses medical data set and its performance is measured. The method has produced higher accuracy in diabetes prediction. A logistic regression (LR) model is presented in [13], which adapted naïve Bayes (NB), decision tree (DT), Adaboost (AB), and random forest (RF) algorithms in predicting diabetes. Similarly, in [14], Decision Tree, SVM and Naive Bayes are used in predicting diabetes. An apriori and ANN based model is presented in [15], to support the prediction of diabetes. The method uses PCA and apriori in feature selection and classification is performed with ANN and random forest.

An ensemble-based framework named eDiaPredict is presented in [16], which uses ensemble modeling by combining XGBoost, Random Forest, Support Vector Machine, Neural Network, and Decision tree to predict diabetes status among patients.

A deep learning predicting diabetes (DLPD) model is presented in [17], which computes cross entropy to perform disease prediction.

3. Proposed Model

Adaptive Sunflower Optimization Based Diabetes Prediction Model (ASODPM):

The proposed adaptive sunflower optimization based diabetes prediction model uses the diabetes data set. The data set given has been preprocessed with value based normalization technique. With the normalized data set, the method extract the features and generates sunflower for each of the tuple present in the data set. Further, the method applies Adaptive Sunflower Optimization Clustering (ASOC) algorithm to group the data sets into number of clusters. At the test phase, the method estimates multi feature morality value (MFMV) and multi feature pollination rate (MFPR) to compute Disease Projection Value (DPV) to identify the class of tuple. The working of the method has been detailed in this section.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 5, No. 3, May : 2023

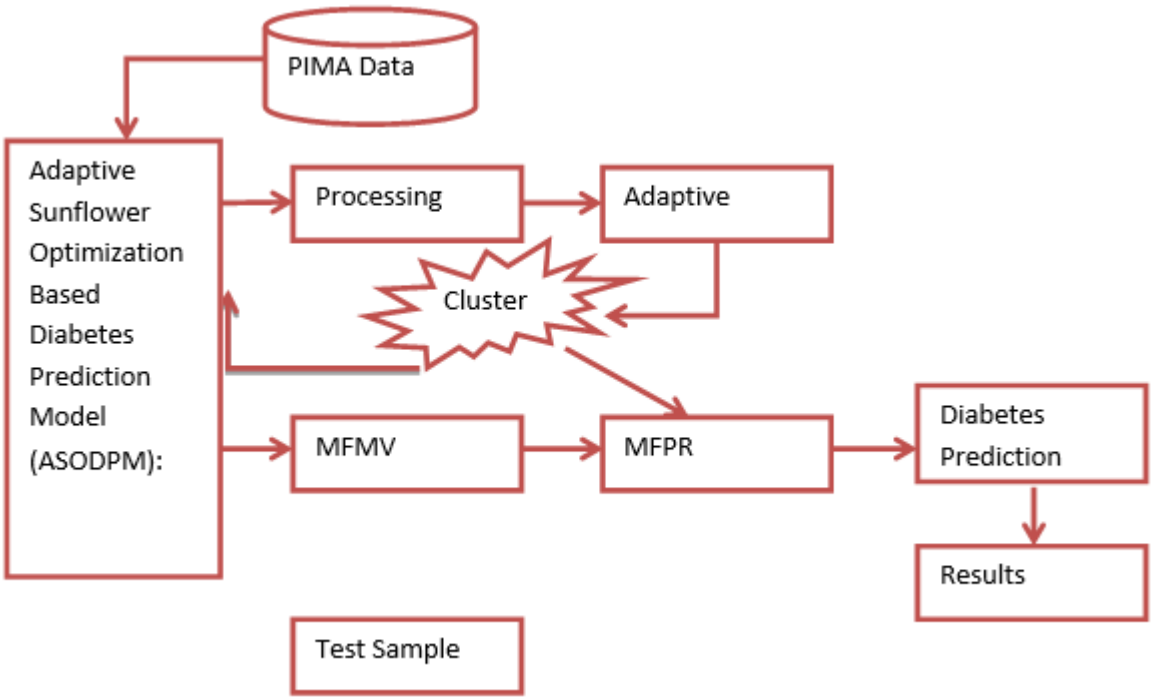

Fig 1: Architecture of proposed ASODPM model

The schematic model of ASODPM method is displayed in Fig1, which has been detailed in this section.

3.1 Value Based Normalization:

The value based normalization algorithm reads the given diabetic data set. The method finds the features available and their types as well. With the feature types, the normalization value is measured for the numeric types and and generates value sets for the non numeric values. Now, each tuple has been traversed and verified for the presence of all feature and values. For the missing values, the method adjust the value with the normalization value and value set according to the feature type.

Algorithm:

Given: Diabetic Data set Dds
Obtain: Preprocessed data set Pds
Start

Read Dds.

Identify feature set Fes = $Fes \cup ((\sum_{i=1}^{size(Dds)} Features \in Dds(i)) \ni Fes)$

For each feature f

Identify feature type Ftype = Fes(f).type
Add to type set Ts = Ts $\cup$ $Ftype$
If F.type==Numeric then

Compute Normalization value Nov = $\frac{\sum_{i=1}^{size(Dds)} DDs(i).f.value}{size(Dds)}$

Else

Identify value set Vs = $vs \cup (\sum_{i=1}^{size(Dds)} Dds(i).f.vallue \ni vs)$

End

End

For each tuple T

For each feature f

If f.type==value && t.f.value==null then

T(f).value=Random(f.vs)

Elseif f.type==numeric && t.f.value==null then

T(f).value= Nor(f)

End

End

Add tuple to preprocessed set Pds.

End

Stop

The value based normalization scheme preprocess the data set which identifies the features and types to normalize the data set. The method computes the value of normalization and value set to adjust the noisy records in the data set.

3.2 Adaptive Sunflower Optimization Clustering

The proposed Adaptive Sunflower Optimization Clustering algorithm initializes k number of clusters based on the number of disease classes considered. Also, the method initializes different parameters like Number Of Population nop, lower bound lb, upper bound ub, morality rate mr, pollinian rate pr, and maximum iteration Mi. Further the method generates random samples to each clusters. Then, for each tuple available a sunflower is generated and initialized with the features of the tuple. Third for each tuple available in the data set, the method computes Multi Feature Morality Value (MFMV) and Multi Feature Polination Rate (MFPR) towards different sunflowers present in different clusters. According to the morality value, the method identifies the group for a tuple and indexed. Fourth, the method generates initial population from each cluster of sunflowers, and computes MFMV and Multi Feature Polination Rate (MFPR) using the objective function towards various sunflower groups. According to the value of MFMV, the method assigns the populations to new clusters. This is iterated for all the populations. Further, the method computes MFMV and MFPR value towards each cluster of sunflower. Now, using both the value of MFMV and MFPR, the method removes the sunflower with poor MFMV value. This is iterated till there is no movement of sunflowers between the clusters.

Algorithm:

Given: Preprocesssed data set Pds

Obtain: Cluster set Cs

Start

Step 1: Read Pds.

Step 2: Initialize cluster set Cs.

Step 3: For each cluster c

Initialize parameters like Mr=8.0, Pr=7.5, maxiter=5, lb=3.5,ub8.5, nop=10, and ndim=11.

Generate random samples from data set and add to cluster c.

$$C = \sum_{i=1}^{size(nop)} random(1, size(pds))$$

End

Step 4: For each sample s

For each cluster c

$$\text{Compute MFMV} = \frac{\sum_{j=1}^{size(s)} Features(s(j) \in C(i)) / size(s)}{size(c)}$$

$$\text{Compute MFPR} = \frac{\sum_{j=1}^{size(s)} Dist((s(j), C(i)(j)) / size(s)}{size(c)}$$

Compute Cluster Attaction Score CAS = MFMV×MFPR

End

Choose the cluster with higher CAS value.

Index the sample to the selected cluster c.

End

Step 5: $\text{Population ps} = \sum_{i=1}^{size(nop)} random(1, size(pds))$

Step 6: For each popupulation p

Evaluate using Feature Centric Fitness function by computing MFMV.

Compute MFMV for population p.

Move the sunflower towards optimal cluster.

End

Step 7: For each instance I of each cluster c

Compute MFMV value.

If MFMV<Th then

$C = \sum Instances(C) \cap I$

end

End

Step 8: Iterate step 7 for maxiter times.

Step 9: Stop.

The above discussed algorithm estimates cluster attraction score (CAS) for various samples and identifies the group for each sample and evaluate the fitness of the samples according to the objective function designed. The method computes MFMV value for various samples to perform clustering.

3.3 Disease Prediction:

The proposed method performs disease prediction according to the features of the sample given. To perform this, the method first applies value based normalization technique and extract the features from the sample. With the features extracted and the sunflower cluster set available, the method computes MFMV, MFPR values. Using the value of MFMV and MFPR values, the method computes

the value of disease projection value (DPV) towards various cluster. Using the value of DPV, the method predict the disease as result.

Algorithm:

Given: Sunflower Cluster set Scs, Test sample Ts.

Obtain: Disease class Dc

Start

Read Scs and Ts.

For each cluster c

Compute MFMV= $\frac{\sum_{i=1}^{size(c)} \frac{\sum_{j=1}^{size(s)} Features(s(j) \in C(i)}{size(s)}}{size(c)}$

Compute MFPR = $\frac{\sum_{i=1}^{size(c)} \frac{\sum_{j=1}^{size(s)} Dist((s(j),C(i)(j))}{size(s)}}{size(c)}$

Compute DPV = MFMV×MFPR

End

Dc = $\underset{i=1}{\overset{size(Scs)}{Max(Scs(i).DPV).class}}$

Stop

The above discussed algorithm computes MFMV and MFPR values towards each cluster available. As per the value of MFMV and MFPR, the method computes the value of DPV to identify the disease class as result.

Results and Discussion:

The Adaptive Sunflower Optimization based diabetes prediction model (ASODPM) has been implemented using R language with the PIMA data set. The performance of the method has been evaluated under different conditions. In each case, the results are recorded and analysed.

Parameter	Value
Data Set	PIMA
Number of Features	11
Number of tuples	605
Tool used	R-language

Table 1: Evaluation Details

The constraints considered to analyse the efficiency of methods is displayed in Table 1.

Disease Prediction Accuracy %			
	1000 Tuples	3000 Tuples	5000 Tuples
AWOD	78	81	85
LMeGNB	81	84	89
DLPD	83	87	92
ASODPM	89	97	98

Table 2: Disease Prediction Accuracy %

The result of analysis conducted towards disease prediction accuracy has been measured for different methods at the presence of varying number of samples in the data set and presented Table 2. The HPMMM model achieves higher performance than the rest.

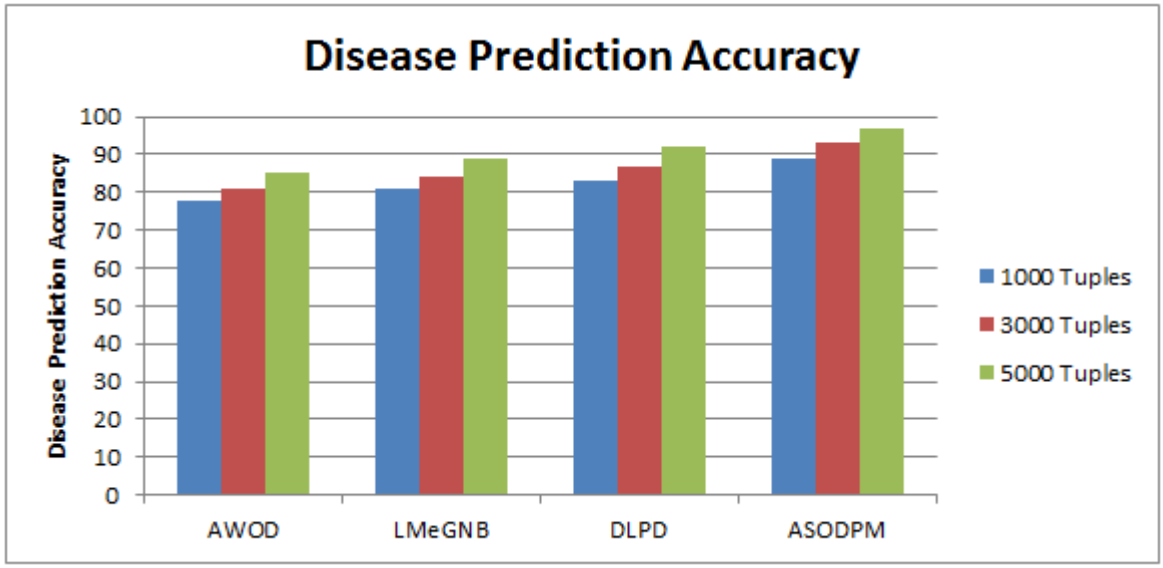

Fig 2: Analysis on Disease Prediction Accuracy

The result of analysis conducted towards disease prediction accuracy has been measured for different methods at the presence of varying number of samples in the data set and presented Fig 2. The ASODPM model achieves higher performance than the rest.

False Prediction Ratio %			
	1000 Tuples	3000 Tuples	5000 Tuples
AWOD	22	19	15
LMeGNB	19	16	11
DLPD	17	13	8
ASODPM	11	4	3

Table 3: False Prediction Ratio

The ratio of false prediction produced by different schemes with different tuples. The results are presented in Table 3. The ASODPM scheme shows poor false ratio than others.

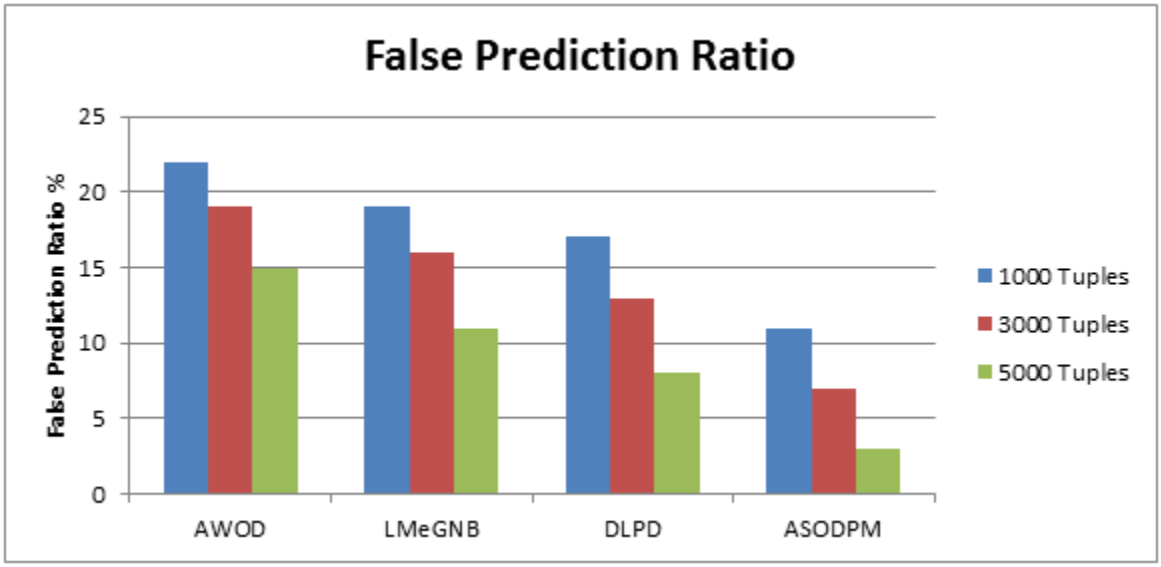

Fig 3: Analysis on False Classification Ratio

The false prediction generated by various models are sketched in Fig 3, where the proposed ASODPM model introduces less false ratio than other techniques in all the cases.

Time Complexity in Disease Prediction			
	1000 Tuples	3000 Tuples	5000 Tuples
AWOD	38	61	85
LMeGNB	35	58	79
DLPD	31	52	72
ASODPM	14	23	32

Table 5: Time Complexity in seconds

The prediction time complexity is measured for all the methods and plotted in Table 4. The ASODPM scheme shows less time complexity in all cases.

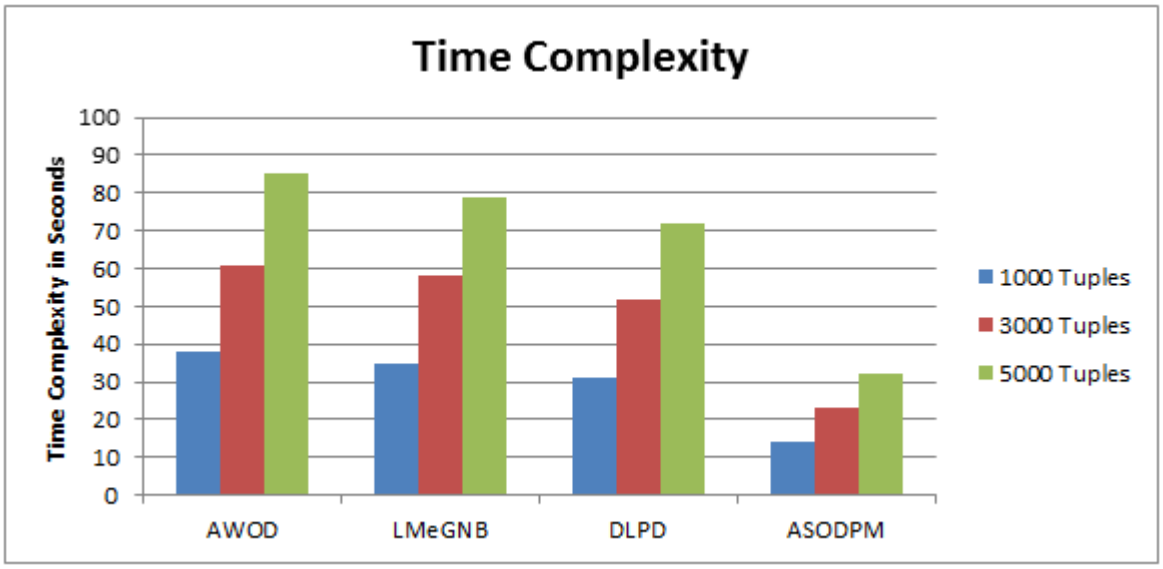

Fig 4: Analysis on Time Complexity

The value of time complexity generated by various schemes are plotted in Fig 4. The proposed ASODPM model has produced less time complexity compare to other methods.

4. Conclusion:

This paper presented a novel Adaptive Sunflower Optimization based Diabetes Prediction Model (ASODPM) towards predicting diabetes. The model applies value-based normalization technique to preprocess the data set given. Further, the method applies adaptive sunflower optimization clustering to group the tuples of the data set. Also, the method reads the test sample and computes disease projection value to identify the class of the sample. The proposed method improves the performance of disease prediction and reduces the false ratio. Considering that type 1 diabetes and type 2 diabetes have many differences in their treatment methods, this method will help to provide the right treatment for the patient. A comparison is also shown in each case. The highest accuracy obtained was around 97% for Dataset 1, after employing NAN removal and it was around 98.8 % for Dataset 2**,** after using the Sunflower Algorithm with pollination.

References:

1) E. Montaser, J. -L. Díez, P. Rossetti, M. Rashid, A. Cinar and J. Bondia, "Seasonal Local Models for Glucose Prediction in Type 1 Diabetes," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 7, pp. 2064-2072, July 2020, doi: 10.1109/JBHI.2019.2956704.
2) Nuankaew, S. Chaising and P. Temdee, "Average Weighted Objective Distance-Based Method for Type 2 Diabetes Prediction," in IEEE Access, vol. 9, pp. 137015-137028, 2021, doi: 10.1109/ACCESS.2021.3117269.
3) M. K. Hasan, M. A. Alam, D. Das, E. Hossain and M. Hasan, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers," in IEEE Access, vol. 8, pp. 76516-76531, 2020, doi: 10.1109/ACCESS.2020.2989857.
4)I. Gnanadass, "Prediction of Gestational Diabetes by Machine Learning Algorithms," in IEEE Potentials, vol. 39, no. 6, pp. 32-37, Nov.-Dec. 2020, doi: 10.1109/MPOT.2020.3015190.
5) L. Jia, Z. Wang, S. Lv and Z. Xu, "PE_DIM: An Efficient Probabilistic Ensemble Classification Algorithm for Diabetes Handling Class Imbalance Missing Values," in IEEE Access, vol. 10, pp. 107459-107476, 2022, doi: 10.1109/ACCESS.2022.3212067.
6) U. Ahmed et al., "Prediction of Diabetes Empowered With Fused Machine Learning," in IEEE Access, vol. 10, pp. 8529-8538, 2022, doi: 10.1109/ACCESS.2022.3142097.
7) F. Simone, F. Andrea, S. Giovanni, P. Gianluigi and D. F. Simone, "Linear Model Identification for Personalized Prediction and Control in Diabetes," in IEEE Transactions on Biomedical Engineering, vol. 69, no. 2, pp. 558-568, Feb. 2022, doi: 10.1109/TBME.2021.3101589.
8) T. M. Le, T. M. Vo, T. N. Pham and S. V. T. Dao, "A Novel Wrapper–Based Feature Selection for Early Diabetes Prediction Enhanced With a Metaheuristic," in IEEE Access, vol. 9, pp. 7869-7884, 2021, doi: 10.1109/ACCESS.2020.3047942.
9)F. A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab and S. A. C. Bukhari, "Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review," in IEEE Access, vol. 9, pp. 43711-43735, 2021, doi: 10.1109/ACCESS.2021.3059343.

10) Quan Zou, Kaiyang Predicting Diabetes Mellitus With Machine Learning Techniques, Front. Genet., Computational Genomics, Volume 9 – 2018, https://doi.org/10.3389/fgene.2018.00515.
11) Salliah Shafi Bhat,1Venkatesan Selvam,1Gufran Ahmad Ansari,2Mohd Dilshad Ansari,3and Md Habibur Rahman, Prevalence and Early Prediction of Diabetes Using Machine Learning in North Kashmir: A Case Study of District Bandipora, H5ndaw5, Advances in Medical Imaging Informatics with Artificial Intelligence and Big Data Analytics, Volume 2022 | Article ID 2789760 | https://doi.org/10.1155/2022/2789760.
12)F. A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab and S. A. C. Bukhari, "Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review," in IEEE Access, vol. 9, pp. 43711-43735, 2021, doi: 10.1109/ACCESS.2021.3059343.
13) Quan Zou, Kaiyang Predicting Diabetes Mellitus With Machine Learning Techniques, Front. Genet., Computational Genomics, Volume 9 – 2018, https://doi.org/10.3389/fgene.2018.00515.
14) Salliah Shafi Bhat,1Venkatesan Selvam,1Gufran Ahmad Ansari,2Mohd Dilshad Ansari,3and Md Habibur Rahman, Prevalence and Early Prediction of Diabetes Using Machine Learning in North Kashmir: A Case Study of District Bandipora, H5ndaw5, Advances in Medical Imaging Informatics with Artificial Intelligence and Big Data Analytics, Volume 2022 | Article ID 2789760 | https://doi.org/10.1155/2022/2789760.
15) Tawfik Beghriche,[1]Mohamed Djerioui,[2]Youcef Brik,[2]Bilal Attallah,[2]and Samir Brahim Belhaouari, An Efficient Prediction System for Diabetes Disease Based on Deep Neural Network, H5ndaw5, Volume 2021 | Article ID 6053824 | https://doi.org/10.1155/2021/6053824.
16) Md. Maniruzzaman, Md. Jahanur Rahman, Benojir Ahammed & Md. Menhazul Abedin, Classification and prediction of diabetes disease using machine learning paradigm, Springer, Health Information Science and Systems volume 8, Article number: 7 (2020).
17) Deepti Sisodiaa Dilip SinghSisodiab, Prediction of Diabetes using Classification Algorithms, Elsevier, Computer science, Volume 132, 2018, Pages 1578-1585.
18) Talha Mahboob Alam, Muhammad AtifIqbal, A model for early prediction of diabetes, Elsevier, Informatics in Medicine unlocked, vol. 16, 100204, 2019.
19) Ashima Singh, Arwinder Dhillon, DiaPredict: An Ensemble-based Framework for Diabetes Prediction, ACM Transactions on Multimedia Computing, Communications, and ApplicationsVolume 17, Issue 2, June 2021 Article No.: 66pp 1–26https://doi.org/10.1145/3415155.
20) Huaping Zhou, Raushan Myrzashova, Rui Zheng, Diabetes prediction model based on an enhanced deep neural network, EURASIP Journal on Wireless Communications and Networking, Volume 2020, issue 1, 2020, https://doi.org/10.1186/s13638-020-01765-7.