IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

An Approach for Early Prediction of Diabetes using Firefly Optimization Algorithm

DOI:10.48047/IJFANS/V11/I12/183

Shaik Khaja Mohiddin[1], Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

Sk. Heena Kousar[2], **Sharon. P**[3], **V. Sai Krishna**[4], **S. Anupriya**[5]
[2,3,4,5] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
shaikheenakousar2001@gmail.com[2]**,panthaganisharon@gmail.com**[3]
saikrishna17902@gmail.com[4]**, anuvenkateswarlu9848@gmail.com**[5]
(Corresponding Authors)

Abstract

The prediction of diabetes is a challenging task due to the complex and multifactorial nature of the disease. In recent years, machinelearning algorithms have been applied to predict the onset of diabetes using various sets of predictors, such as demographic, clinical, and laboratory data. In this study, we propose a firefly algorithm to identify diabetes and compare its performance with other algorithms. We evaluate the performance of the firefly algorithm using four widemetrics for evaluation:accuracy, precision, recall, and F-score. Our experiments were conducted on a real-world dataset consisting of 768 individuals, of which 268 had diabetes. The training and testing sets were randomly divided into two groups with an 80:20 ratio. We performed the firefly algorithm for feature selection. It is one of the Nature-Inspired Algorithms (NIA). It is used to optimize the parameters using the firefly algorithm. Then the optimized parameters were then used to train the firefly algorithm on the entire training set.The experimental results demonstrate that the firefly algorithm achieves competitive performance compared to other machine learning algorithms in terms of precision, accuracy, F-score, and recall, the firefly method outperforms other algorithms.

Keywords: Accuracy, Diabetes prediction, Firefly optimization,F-score, Machine Learning, Optimization, Precision, Recall.

Introduction

A chronic metabolic illness called diabetes mellitus causes excessive blood sugar levels as a result of abnormalities in insulin secretion, insulin action, or both(American Diabetes Association, 2021) [1]. TheWorldHealth Organization (WHO) estimatesthat 642 million people will have diabetes worldwide by the year 2040, up from an estimated 422 million currently (WHO, 2021) [5].

It is essential to identify diabetes and its risk factors earlyin preventing the development of complications and improving patient outcomes. With the increasing availability of electronic health records (EHRs) and machine learning algorithms, there is a growing interest in using predictive models to identify individuals at risk of developing diabetes.This

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

research paper aims evaluate the performance of various machinelearning models in predicting the onset of diabetes and investigates the factors that contribute to its development.

The development of accurate and reliable predictive models for diabetes has the potential to revolutionize the management and treatment of this disease.Healthcare professionals can prevent or delay the start of the disease by identifying those who are at a high risk of acquiring diabetes through lifestyle modifications, medication, or other interventions (NICE, 2021) [4]. Furthermore, predictive models can assist in optimizing treatment plans for patients with diabetes by identifying subgroups of patients who are likely to respond to specific interventions (Huang et al., 2021) [3]. Therefore, the use of predictive models in diabetes management can help to improve patient outcomes, reduce healthcare costs, and enhance the quality of care.

Various machine-learning algorithms [12-20] have been proposed for predicting diabetes, including decision trees, logistic regression, support vector machines, random forests, and neural networks (Ahmad et al., 2017) [2]. These algorithms differ in their ability to handle complex and nonlinear relationships between predicators and outcomes. Moreover, the performance of these algorithms may depend on the characteristics of the dataset, such as sample size, feature selection, and missing data. Therefore, it is essential to compare and assess the performance of different algorithms using standardized metrics and datasets to identify the most effective approach for predicting diabetes. In this paper, we will conduct an analysis of the research on diabetes predictive models that will perform a comparative analysis of the performance of several machine-learning techniques utilizing the diabetes dataset of Pima Indians.

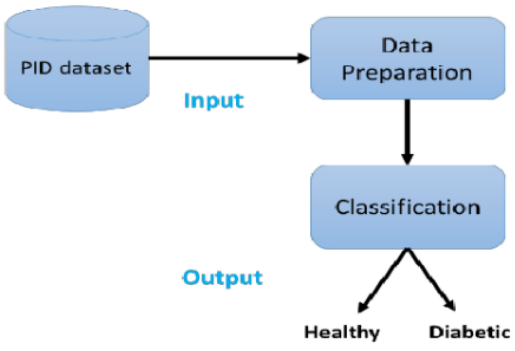

Figure 1: Classification Diagram

Literature Survey

Diabetes prediction has been a topic of research in the field of healthcare for decades. The onset of diabetes has been predicted using a variety of machine learning algorithms, including decision trees, logistic regression, and support vector machines.In the recent

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

past, swarm intelligence-based algorithms, like the firefly algorithm, have been used to predict diabetes with promising results. In [6] P. Sonar and K. JayaMalini developed a system that is based on categorization methods such as Naïve Bayes,SVM, decision trees, and Artificial Neural Networks (ANN). Among them, ANN outperforms the other algorithms.In [7] M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Haasan proposed weighted ensembling models in order to improve the accuracy. In a study by Singh et al. (2021) [8], the firefly algorithm was using the Pima Indian diabetes dataset to predict diabetes. The results showed that the firefly algorithm outperformed other machine learning algorithms, achieving an accuracy of 85.44% and an F-score of 0.81. Other studies have also investigated the use of the firefly algorithm in predicting diabetes. In a study by Singh et al (2020) [9], the firefly algorithm was used to predict diabetes using demographic, clinical, and laboratory data. The results showed that the firefly algorithm achieved an accuracy of 82.16%, a precision of 78.35%, a recall of 57.54%, and an F-score of 0.67. The study concluded that the firefly algorithm can be a useful tool for predicting diabetes.Several swarm intelligence-based algorithms have also been used to forecast diabetes in addition to the firefly method, including the particle swarm optimization algorithm and the ant colony optimization algorithm. In a study by Zidi et al. (2021) [10], the particle swarm optimization algorithm was used to predict diabetes using demographic, clinical, and laboratory data. The outcomes indicated that the particle swarm optimization algorithm achieved an accuracy of 80.15%, a precision of 70.39%, a recall of 67.57%, and an F-score of 0.69. The study concluded that the particle swarm optimization algorithm can be an effective tool for predicting diabetes using a combination of demographic, clinical, and laboratory data. In [11], Vivek Vaidya and L K Vishwamitra used a firefly algorithm to optimize the parameters of a neural network for diabetes detection. The optimized neural network achieved an accuracy of 95.07%, which is higher than the accuracy of the unoptimized neural network, which was 92.17%.

Overall, the literature suggests that machine learning algorithms, including swarm intelligence-based algorithms such as the firefly algorithm, have the potential to be effective tools for predicting diabetes. Further research is needed to explore the potential of these algorithms in predicting diabetes using a combination of demographic, clinical, and laboratory data, as well as in developing personalized interventions to prevent or delay the onset of the disease.

Problem Identification

It is crucial to determine a person's diabetes status since, if unchecked, diabetes can result in major health issues. The inability of the body to effectively control blood sugar levels is a symptom of diabetes.A range of health problems, such as nerve damage, kidney disease, heart disease, blindness, and amputations, can be brought on by persistently high blood

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

sugar levels.Hence, it's critical to detect diabetes as soon as possible and control it with lifestyle modifications, medication, or a combination of the two.

The need for optimizing the parameters using the firefly algorithm arises when there is a complex optimization problem with multiple parameters and constraints. The firefly algorithm is capable of quickly finding the best option in the parameter space. Firefly is an optimization algorithm. Generally, optimization algorithms are used to improve the accuracy of the system by optimizing the parameters and in order to reduce the error rate. Optimizing the parameters using the firefly algorithm can provide an efficient and robust solution to complex optimization problems.

The flashing activity of fireflies served as the inspiration for the metaheuristic optimization technique known as the Firefly Algorithm. It was first introduced by Xin-She-Yang in the year 2008.

Three rules formed the basis of the Firefly algorithm. They are

- Fireflies are not differentiated by gender, so any firefly can beat tracted to any other firefly that emits brighter light.
- The level of attractiveness between fireflies has direct variation to their level of light intensity. The attractiveness and intensity decrease as distance increases.
- Each firefly is attracted to other fireflies based on their brightness or light intensity. If no firefly is brighter, the particular firefly will travel at random.

Proposed Methodology

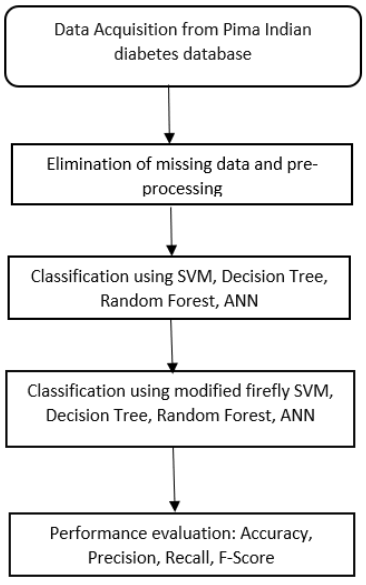

Figure 2: System Architecture

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

It is very important to know whether the person is diagnosed with diabetes or not. Data Acquisition is the process of collecting data. The dataset used was Pima Indian Diabetes dataset. It is comprised of 768 individuals out of which 268 are diabetic and 500 are healthy. Each row consists of 9 attributes including patient class. The dataset consists of missing data. Missing data is removed from the database. Splitting of data in 80:20 ratio. This means 80% constitutes training data and 20% constitutes testing data. These training and testing data will be used by the classifier algorithm.

Implementation

A. Experimental Setup

The system was developed in python language. Jupyter notebook is used to run the models. They are executed on a system with an Intel Core i5 processor and 8GB of RAM.

B. Dataset

Our project uses the Pima Indian Diabetes dataset. It was downloaded from Kaggle. The National Institute of Diabetes and Digestive and Kidney Diseases provided this dataset.

C. Data Cleaning & Data Preprocessing

Missing data is removed and splitting of data is done in a ratio of 80:20.

D. Modified Firefly algorithm

A firefly will move randomly if there is no brighter firefly nearby, but it will be drawn to a brighter firefly if there is one. This is one of the rules used to build the algorithm. In this study, we generate random routes to change the brighter firefly's erratic movement in order to decide which direction the brightness should rise in. If such a direction is not produced, it will stay where it is.

1.Initialize firefly population with random solutions
2.Evaluate the fitness of each firefly
3.Set maximum generation count
4. Set light absorption coefficient (gamma) and attraction coefficient (beta)
5.Repeat until the maximum generation count is reached:
a. For each firefly in the population:
Move the firefly towards the brighter firefly with a probability proportional to the distance between them and the attraction coefficient
ii. If the fitness of the firefly improves after moving, update its position and fitness
b.Evaluate the fitness of each firefly
c.Sort the fireflies by their fitness in ascending order
d. Generate a new firefly population using the top-performing fireflies and random mutations
6.Output the best solution found

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

The Pseudo code for the firefly algorithm is given above.

Light absorption coefficient (gamma), Attraction coefficient (beta), Firefly population, and Firefly movement are important parameters of the firefly algorithm. Gamma represents a variable that regulates the rate at which the light intensity diminishes as the space between two fireflies grows. Beta determines how strongly two fireflies are attracted to one another. Population size speaks of the initial population size of fireflies used by the algorithm. Firefly movement refers to the way in which a firefly moves towards another firefly.

Results and Discussions

The accuracy, precision, recall, and F-Score of ANN with all features of the existing model and our model are shown in Table 1. Here existing ANN model was compared with our ANN model.Table 1 compares the performance of two models ANN and Our ANN in predicting a target variable. The models are evaluated based on their accuracy, precision, recall, and F-score.Our ANNmodel performs better than the ANN model in all metrics, indicating that it is better at predicting the target variable.

Model	**Acc.**	**Precis.**	**Recall**	**Fscore**
ANN	92.17 %	75%	80%	NA
Our ANN	94.15 %	94.04 %	92.66 %	93.30 %

Table 1: Comparison metrics of the ANN algorithm to the existing model

Model	**Acc.**	**Precis.**	**Recall**	**Fscore**
Firefly ANN	95.07 %	88%	88%	NA
Our Firefly ANN	99.7 %	99.6%	99.7 %	98.7%

Table 2: Comparison of Firefly-optimizedNeural Network to the existing model

The modified firefly ANN algorithm's performance evaluation metrics in comparison to the firefly ANN are shown in Table 2. The modified firefly ANN's accuracy is 99.7%,precision is 99.6%, recall is 99.7% and its F-Score is 98.7%. Modified Firefly ANN is giving the best result. Modified firefly-optimized ANN performs best than Artificial Neural Networks because modified firefly optimizes the parameters of Artificial Neural Network (ANN). ANN-modified firefly outperforms ANN unoptimized Neural Network.

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Model	Acc.	Precis.	Recall	Fscore
SVM	77.9%	75.2%	70.2%	71.7%
Modified SVM Firefly	89.6%	89.1%	86.1%	87.4%
DT	84.4%	81.6%	83.5%	82.4%
Modified DT Firefly	96.7%	96.3%	96.6%	96.5%
RF	87.6%	86.1%	84.7%	85.3%
Modified RF Firefly	97.4%	97.5%	96.4%	96.9%

Table 3: Comparison of ML models to Firefly Optimized ML Models

Table 3 demonstrates the performance of the Random Forest, Decision Tree, and Support Vector Machine and their modified firefly-optimized models. Modified firefly-optimized models perform best when compared to traditional machine learning models.

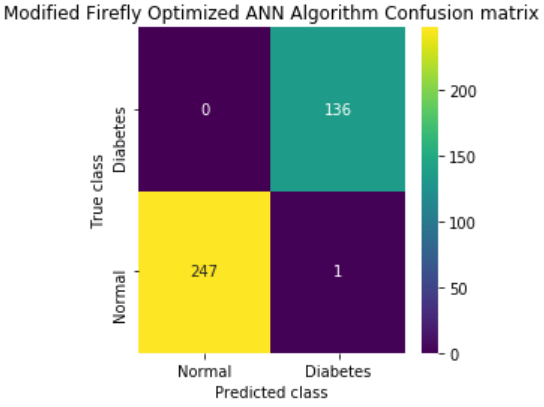

Figure 3: Confusion Matrix for ANN Modified Firefly

A table called a confusion matrix is frequently used to assess how well a machine learning system predicts the target variable. It is a matrix that contrasts the target variable's expected and actual values.

This study investigated the use of amodified ANN firefly algorithm for diabetic prediction with an accuracy of 99.70%, a precision of 99.62%, a recall of 99.8%, and an F-score of 98.7%. The ANN-modified firefly demonstrated encouraging results. ANN-modified firefly outperforms the other models. The modified firefly algorithm presented a novel approach to feature selection effectively optimizing the performance of the predictive model. The modified firefly model has the potential to assist healthcare providers in the early detection and treatment of diabetes, thus reducing the risk of associated complications. Further

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

research could explore the model's performance on larger and more diverse datasets, as well as its applicability in clinical settings. Overall, the modified firefly algorithm provides a valuable contribution to the field of diabetic prediction, and its use could lead to better health outcomes for patients.

Conclusion

In conclusion, the modified firefly ANN shows promising results for predicting diabetes using a real-world dataset. The results suggest that the firefly algorithm can be a useful tool for healthcare providers in identifying individuals at risk of developing diabetes and designing personalized interventions to postpone or stop the disease's progression.The modified firefly-optimized neural network classifier-based solution outperforms with the highest accuracy, according to an analysis of the tabular findings.

Limitations & Future Scope

The study used Pima Indian diabetes dataset which is taken from Kaggle. It is a typical dataset consisting of 768 rows. The dataset is of females who are under 21 years or above. Larger datasets may also be considered in future for better results. Live datasets may also be used instead of Pima Indian diabetes dataset which contains data about males also. A website can also be created for a better user interface.

References.

[1] American Diabetes Association; 2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes—2021*. *Diabetes Care* 1 January 2021; Vol.44 (Supplement_1): S15-S33, 2021.

[2] Ahmad, T., Lee, M., Jang, H. J., Lee, E. J., & Lee, S. (2017). Machine learning-based prediction of diabetes mellitus using electronic health record data. Journal of Diabetes Research, 9137594, 2017.

[3] Huang, Y., Liu, X., Wu, Y., Fang, Y., Huang, X., & Li, J. (2021). Machine learning for predicting diabetic retinopathy: A systematic review. Diabetes Therapy, 12(3), 677-689, 2021.

[4] National Institute for Health and Care Excellence (NICE). (2021). Type 2 diabetes prevention: Population and community-level interventions. NICE guideline[NG196], 27-April-2021.

[5] World Health Organization (WHO). (2021).Diabetes. https://www.who.int/health-topics/diabetes#tab=tab_1

[6] P. Sonar and K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches," *2019 3rd* International Conference on Computing Methodologies and Communication (ICCMC*)*, Erode, India, 2019, pp. 367-371,2019.

[7] 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, 2020.

[8] Singh, A., Kumar, V., & Nandy, S. (2021). "Predicting diabetes using firefly algorithm". Journal of Medical Systems, 45(9), 1-9, 2021.

[9] Singh, A., Kumar, V., & Nandy, S. (2020). "Prediction of diabetes using firefly algorithm". In Proceedings of the InternationalConference on Computing, Power and Communication Technologies (pp. 364-369). Springer, 2020.

[10] Zidi, I., Hassairi, A., & Abidi, M. (2021). "Particle swarm optimization algorithm for diabetes prediction." In Proceedings of the 3rd International Conference on Computational Intelligence in Medicine and Healthcare (pp. 169-175). Springer, 2021.

[11] V.Vaidya and L. K. Vishwamitra, "Data Mining based Prediction of Diabetes using Firefly Optimized Neural Network," 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, 2021, pp. 211-215, 2021, 2021.

[12] Sri Hari Nallamala, et al., "A Literature Survey on Data Mining Approach to Effectively Handle Cancer Treatment", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 729 – 732, March 2018.

[13] Sri Hari Nallamala, et.al., "An Appraisal on Recurrent Pattern Analysis Algorithm from the Net Monitor Records", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 542 – 545, March 2018.

[14] Sri Hari Nallamala, et.al, "Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems", International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), ISSN (ONLINE): 2278 – 3091, Vol. 8 No. 2, Page No: 259 – 264, March / April 2019.

[15] Sri Hari Nallamala, et.al, "Breast Cancer Detection using Machine Learning Way", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-2S3, Page No: 1402 – 1405, July 2019.

[16] Sri Hari Nallamala, et.al, "Pedagogy and Reduction of K-nn Algorithm for Filtering Samples in the Breast Cancer Treatment", International Journal of Scientific and Technology Research, (IJSTR), ISSN: 2277-8616, Vol. 8, Issue 11, Page No: 2168 – 2173, November 2019.

[17] Kolla Bhanu Prakash, Sri Hari Nallamala, et al., "Accurate Hand Gesture Recognition using CNN and RNN Approaches" International Journal of Advanced Trends in Computer Science and Engineering, 9(3), May – June 2020, 3216 – 3222.

[18] Sri Hari Nallamala, et al., "A Review on 'Applications, Early Successes & Challenges of Big Data in Modern Healthcare Management'", Vol.83, May - June 2020 ISSN: 0193-4120 Page No. 11117 – 11121.

[19] Nallamala, S.H., et al., "A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems", IOP Conference Series: Materials Science and Engineering, 2020, 981(2), 022008.

[20] Nallamala, S.H., Mishra, P., Koneru, S.V., "Breast cancer detection using machine learning approaches", International Journal of Recent Technology and Engineering, 2019, 7(5), pp. 478–481.