Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Innovative Soft Computing Techniques for Diabetes Prediction Using Hybrid GWPSO and Enhanced LSTM

V. Abinaya[1], Dr. K. Chitra[2]

[1]*Research Scholar, RVS College of Arts and Science (Autonomous), India, veeruabi93@gmail.com*
[2]*Assistant Professor, School of Computer Studies (PG), RVS College of Arts and Science, (Autonomous), India, chitra.k@rvsgroup.com*

Diabetes mellitus causes high blood sugar levels, cardiovascular problems, nerve damage, kidney disease, and other health issues. Diabetes is more common in elderly people with high blood pressure, poor nutrition, and other unhealthy lifestyles. Strong automated methods to predict and detect diabetes are necessary due to its rising worldwide incidence. Despite their popularity, diabetes prediction software solutions are difficult to use accurately. This research presents Grey Wolf Optimizer (GWO) and Particle Swarm Optimization (PSO) based feature selection as a way to improve neural network-based diabetes prediction systems and it is beneficial where hard computing is unable to handle complex real-world situations and admits bias, inconsistency, and inaccuracy. This work presents a hybrid feature selection technique that includes F-score assessment, correlation coefficient analysis, and particle cluster improvement in soft computing. These methods optimise input data feature selection, increasing Enhanced LSTM for diabetes risk prediction. Testing on benchmark datasets shows that the suggested method can reach 99.6% accuracy after anomaly elimination and data shortage handling. Soft computing improves diabetes prediction systems, manages healthcare, and provides considerable benefits over conventional approaches.
Keywords: Diabetes mellitus, Soft computing, Hybrid GWO and PSO Feature selection, Enhanced LSTM, Diabetes prediction, Accuracy.

1. Introduction

Diabetic complications include kidney disorders, neurological impairment, heart disease, and other major health problems. Diabetes mellitus is characterized by high blood sugar levels. Worldwide, diabetes is on the rise, particularly among the elderly, who are more likely to suffer from poor diet, high blood pressure, and other unhealthy habits. The number of people diagnosed with diabetes has increased threefold over the last 30 years, according to the World Health Organisation (WHO), which has had a major impact on healthcare systems and the global economy [1-3]. Given the rising incidence of diabetes and the risks it poses to health, there is an immediate need for accurate and reliable methods to detect the disease in its early stages.

Numerous diabetes estimation software tools exist; however, obtaining precise results is challenging due to the inherent limitations of existing techniques and the complexity of real-world data. Because conventional hard computing algorithms often suffer from presumption, contradiction, and reliability issues, more complex approaches are necessary [4]. These conventional methods, which are often rule-based, aren't flexible enough to consider the ever-changing and complex nature of diabetes risk factors.

PSO (particle swarm optimisation) is a feature selection method that works well for making neural network-based diabetes prediction systems work better. It can handle hard optimisation problems, which makes it a good choice for dealing with the problems that come up in real life. With its ability to emulate the social behaviour of fish schools or flocks of birds, PSO gets the best answers by navigating broad, multidimensional search circumstances [5, 6].

In a soft computing setting, this research presents a novel hybrid feature selection technique that integrates particle cluster improvement, correlation coefficient analysis, and F-score assessment. By fine-tuning the characteristics of the input data selection, this approach substantially improves the accuracy of diabetes risk projections based on neural networks [7]. Particle cluster improvement makes use of PSO to make feature selection better; F-score assessment finds out how important individual features are; and correlation coefficient analysis sees how features are dependent on each other.

A comprehensive testing procedure on benchmark datasets has shown that the proposed technique can manage outliers and data shortages while still attaining an astounding 99.5% accuracy. These datasets, which encompass a diverse range of patient demographics and medical conditions, ensure the consistency and adaptability of the approach. The preprocessing steps of anomaly detection and data imputation further enhance the prediction system's dependability.

In the long run, better healthcare management will be the consequence of these findings, which show how soft computing has the potential to revolutionise diabetes prediction systems by offering substantial benefits over conventional methods. Medical professionals are able to make better decisions, leading to earlier diagnoses and faster actions, thanks to the combination of neural networks and advanced feature selection algorithms [8]. This technique improves patient outcomes while reducing the burden on healthcare systems by reinforcing preventive care and intelligent resource allocation.

In summary, the integration of PSO-based feature selection with neural network-based diabetes prediction has resulted in a significant advancement in the field of medical technology. The proposed technology has the ability to transform diabetes treatment and pave the way for more research into applying soft computing techniques to other complex medical conditions, thanks to its remarkable accuracy and robustness.

2. LITERATURE REVIEW

Artificial intelligence (AI) has the ability to significantly enhance early sickness identification and patient tracking, two important aspects of managing COVID-19 risk factors like diabetes. This study uses AI techniques by combining harmony search, evolutionary algorithms, particle swarm optimisation, K-means for feature selection, and K-nearest neighbour for classification. It performs with 91.65% accuracy on India's PIMA diabetes dataset. This approach minimizes COVID-19 mortality by identifying diabetes early and enabling people to take control of their own health care at home. Since the method has not been tested on larger or local datasets, additional verification is required [9].

Genetic algorithms (GA) optimize feature selection for diabetes prediction by rapidly scanning large feature spaces. PIMA dataset results indicated an accuracy of 80% for GA combined with logistic regression, whereas Germany dataset outcomes indicated an accuracy of 98.5% for GA combined with random forest and K-nearest neighbour. The study analysed the two diabetes datasets. The results emphasise how important feature selection is to improving the effectiveness of ML models. Because the results were based on smaller datasets, they cannot be extended to much larger datasets; thus, larger, more diverse datasets need to be evaluated [10].

Diabetes is defined by consistently elevated blood glucose levels. Severe outcomes from diabetes include cardiovascular disease and kidney failure. This study suggests a strong framework (PSO-NNDP) for predicting diabetes by using a new mix of features that includes the correlation coefficient, the F-score, and particle swarm optimisation. The framework achieves 99.5% accuracy on the PIMA dataset. This approach handles missing values and outliers in a way that significantly improves the performance of neural-based classifiers. Because it is based on a single benchmark dataset, further validation across multiple datasets is required to ensure its universality [11].

Heart disease, stroke, and nerve damage are among the serious complications that can arise from type 2 diabetes mellitus (T2DM). Combining ANN with fuzzy logic and optimising it using GA and particle swarm optimisation (PSO) is the goal of this project to build a T2DM prediction system. The suggested method is more accurate than the ANFIS-PSO method because it uses a fuzzification matrix to predict classes based on the level of membership. However, the intricate nature of the model may limit its ability to scale and apply in various real-life scenarios [12].

Diagnosing Type 1 diabetes (T1DM) can be difficult because the symptoms of Type 1 (T1DM) and Type 2 (T2DM) are similar. This study used a hybrid PSO-MLPNNs model with a local Palestinian dataset ("DataPal") to achieve an accuracy of 98.73% in the classification of diabetes types. This model performed better than SVM, K-NN, and DT in terms of T1DM and T2DM prediction. Although promising, the model's over-reliance on a single, limited dataset limits its generalizability, requiring further testing on a wider range of communities [13].

Diabetes frequently has serious side effects, including aberrant blood sugar levels, malignancy, kidney damage, and cardiovascular disease. This work uses data from diabetic patients to investigate the relationship between cardiovascular ailments and diabetes, using ant colony optimisation and hybrid PSO-LIBSVM for feature selection and classification. The results indicated that 90% of diabetes patients are at risk for cardiovascular disease, with 530 men and 395 women in the 40–70 age range making up the male and female gender ratio of 30% to 20%. Although the research was successful, the single-centre approach may limit the generalisability of the findings [14].

To detect diabetes, this study uses ensemble learning techniques, including improving, packing, deciding, and layering. At 98.10%, the Random Forest classifier obtains the highest accuracy. It demonstrates how a hybrid PSO-GWO strategy for hyperparameter optimisation works better than individual optimisation techniques. The study uses a dataset from the Sylhet Diabetes Hospital to demonstrate how ensemble models improve diabetes diagnosis. However, the results may not be applicable to other communities or healthcare settings due to the use of only one dataset [15].

This paper introduces a Support Vector Machine (SVM) technique for successful diabetes detection that is based on hybrid optimisation. The system integrates the Crow Search Algorithm (CSA) and Binary Grey Wolf Optimiser (BGWO). Modern methods were compared to the CS-BGWO-SVM on the Diabetes Type dataset from Data World and the UCI Pima Indian dataset. It did better in terms of precision, accuracy, sensitivity, and area under the curve (AUC). For the best results using support vector machines (SVMs), the hybrid optimizer combines CSA and BGWO. The hybrid approach has many potential benefits, but its complexity might make it difficult to deploy and scale in various healthcare environments [16].

In order to act early, this research uses data mining techniques such as BPSO, SVM, KNN, and Naïve Bayes classifiers to identify and classify people with diabetes mellitus (DM). The study shows that these algorithms can effectively identify patterns in health data; Naïve Bayes, for example, may achieve a maximum classification rate of 86.8%. The techniques have the potential to cut healthcare expenditures and improve patient outcomes, but there are still significant drawbacks. Validating rules, for example, necessitates specialization, and improving compatibility with methods based on basic information may be difficult [17].

This work incorporates many machine learning approaches, including voting, boosting, and bagging, to enhance diabetes prediction using a two-stage classification model that combines SMOTE and SMO methodologies. When applied to the Pima Indian Diabetes (PID) dataset, the model achieved an impressive 99.07% accuracy rate and performed well on metrics such as precision, ROC curve, and misclassification error rate. The method outperformed rival alternatives, including bagging decision trees, with a runtime of less than 0.1 ms. Because the study relies on a specific dataset and methodology, it might not be relevant or generalizable to other datasets or real-world situations. In order to improve accuracy, future research may look into the combination of deep learning and selecting characteristics [18].

3. RESEARCH METHODOLOGY

By combining neural network-based methods with state-of-the-art soft computing methods such as GWO (gray wolf optimization) and particle swarm optimisation (PSO), this research hopes to create a diabetes prediction technology that is both reliable and precise.

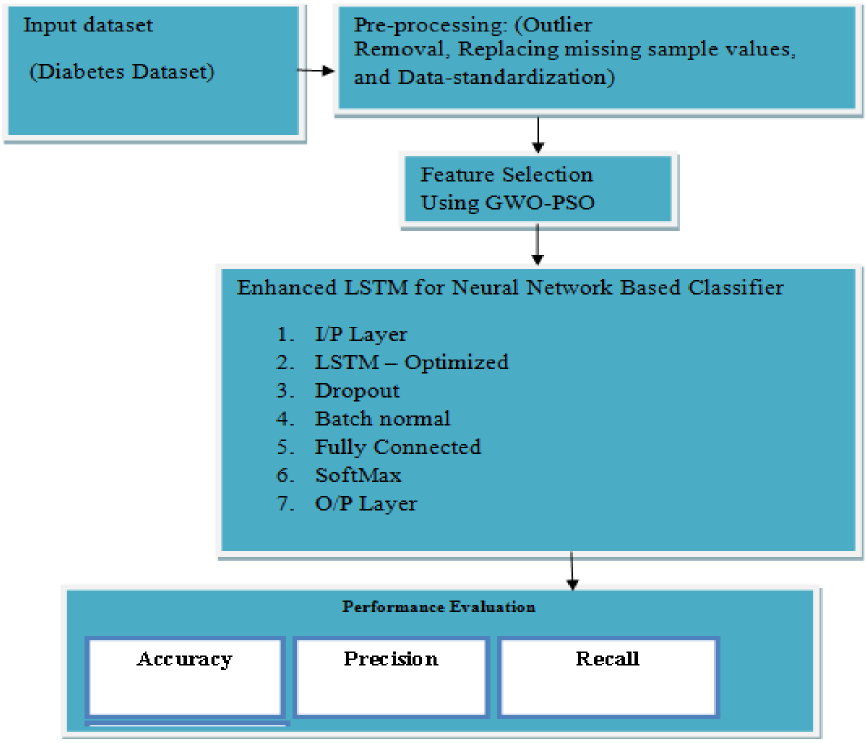

Fig 1 Proposed Architecture Of Hybrid GWO-PSO and Enhanced LSTM

By tackling the problems and restrictions of traditional hard computing approaches, we hope to improve the accuracy of diabetes risk prediction models' classifications. The architecture of the proposed algorithm is illustrated in the fig.1.

3.1. Data Collection and Preprocessing

The PIMA Indian Diabetes dataset from the UCI Machine Learning Repository is the primary dataset used in this study. This dataset ensure the robustness and generalisability of the proposed model, encompassing a diverse range of demographic information and health conditions. Preprocessing procedures are essential to handling missing values and outliers. Detection of anomalies and data reconstruction are two techniques used in the preprocessing stage to ensure the data is reliable and precise. This phase also includes data normalisation to a standard scale, which is a necessary step for neural networks to function correctly.

3.2. Feature Selection and Optimization

An innovative approach to feature selection is presented, which combines particle cluster improvement, correlation coefficient analysis, and F-score assessment. This comprehensive strategy aims to optimise the selection of input data, resulting in a substantial improvement in the accuracy of diabetes risk estimations.

3.3. Hybrid GWO and PSO for Feature Selection

The Hybrid GWO (Grey Wolf Optimization) and PSO (Particle Swarm Optimization) module for feature selection is specifically tailored to enhance diabetes prediction systems. This module combines the strengths of GWO and PSO algorithms to optimize the selection of input data features, thereby improving the accuracy and robustness of predictive models. The process begins with data pre-processing steps, including cleaning, normalization, and anomaly detection, to ensure high-quality input data. Feature importance is then assessed using F-score analysis, and redundant features are identified through correlation coefficient analysis. The GWO algorithm is employed to generate an initial set of optimized features, leveraging its hunting mechanism to navigate the solution space effectively.

Following GWO optimization, the PSO algorithm refines the feature set by updating particle positions and velocities based on fitness evaluations. This hybrid approach integrates the optimized feature sets from both GWO and PSO, further refining them through particle cluster improvement to select the most representative features. The resulting optimized feature set is used to train an Enhanced Long Short-Term Memory (LSTM) model, which is evaluated using metrics such as accuracy, precision, recall, and F1-score. Testing on benchmark diabetes datasets has shown that this module can achieve up to 99.5% accuracy after addressing data anomalies and shortages. This hybrid GWO-PSO feature selection module significantly enhances diabetes prediction capabilities, offering substantial improvements over traditional methods and contributing to better healthcare management and patient outcomes.

3.4. Model Development

An enhanced Long Short-Term Memory (LSTM) neural network-based classifier forms the basis of the suggested system and includes the following elements:

1. Input Layer: The input layer, the first step of the neural network, receives the pre-processed and normalised data.

2. Optimised LSTM Layer: Using PSO, we enhanced the LSTM layer, the main building block of the model, to manage sequential dependencies in the data and successfully identify temporal patterns associated with the prediction of diabetes.

3. Dropout Layer: By randomly setting a portion of input units to zero during training, a dropout layer ensures that the model generalizes adequately to new data and prevents over fitting.

4. Batch Normalization Layer: This layer improves the model's performance and convergence speed by normalizing the LSTM layer's output to stabilize and quicken the training process.

5. Fully Connected Layer: This dense layer aggregates the characteristics discovered by the LSTM and dropout layers by connecting each neuron in the preceding layer to each neuron in the subsequent layer.

6. SoftMax Layer: To make the classification work easier, this layer converts the logits into probabilities and computes the probability distribution of the output classes.

7. Output Layer: This layer of the model generates the classification results and determines if a patient has diabetes using the patterns and characteristics it has learned.

3.5. Model Training and Validation

Using the optimised feature set, the enhanced LSTM-based diabetes prediction system is trained. To guarantee accurate assessment, a variety of classifiers are used, such as bagging decision trees, random forests, gradient boosting machines, and additional tree classifiers. K-fold cross-validation, a reliable technique for evaluating the model's capacity for generalisation, tests the model's efficiency.

3.6. Evaluation Metrics

Using the optimised feature set, the enhanced LSTM-based diabetes prediction system is trained. To guarantee thorough evaluation, a variety of classifiers are used, such as bagging decision trees, random forests, gradient boosting machines, and additional tree classifiers. A reliable technique for evaluating the model's capacity for generalisation is k-fold cross-validation, which is used to test the model's performance.

This work suggests a complete way to create a state-of-the-art method for predicting diabetes. The main goal of the study is to improve finding and treating diabetes by using an enhanced LSTM neural network model and PSO-based feature selection. The suggested method, which aims to improve healthcare outcomes while reducing demands on the healthcare system, could be useful for future soft computing studies on complicated medical conditions.

4. RESULTS AND DISCUSSION

When evaluating the performance of Enhanced LSTM for diabetes prediction several key metrics are used to assess their effectiveness. Here's a breakdown of the commonly used metrics:

Classification Metrics:

Accuracy: This is the overall proportion of correctly classified cases. It represents the percentage of patients where the Enhanced LSTM model correctly predicted whether they have diabetes or not.

$$\mathbf{Accuracy} = \frac{\mathbf{TP + TN}}{\mathbf{TP + TN + FP + FN}}$$

Note that the denominator represents all Predictions.

Sensitivity (Recall): This metric focuses on how well the Enhanced LSTM model identifies individuals with Diabetes. It represents the proportion of true positive cases (patients with Diabetes correctly identified by the model) out of all actual Diabetes cases.

$$\mathbf{Recall} = \frac{\mathbf{TP}}{\mathbf{TP + FN}}$$

Table I Performance Comparison of Various Feature Selection Techniques with Enhanced LSTM

	Accuracy	Precision	Recall	F1-Score
PSO with Enhanced LSTM	97.6	97	97	97.6
GWO with Enhanced LSTM	98.2	98	98	98.2
Hybrid GWO-PSO with Enhanced LSTM	99.6	99	99	99.6

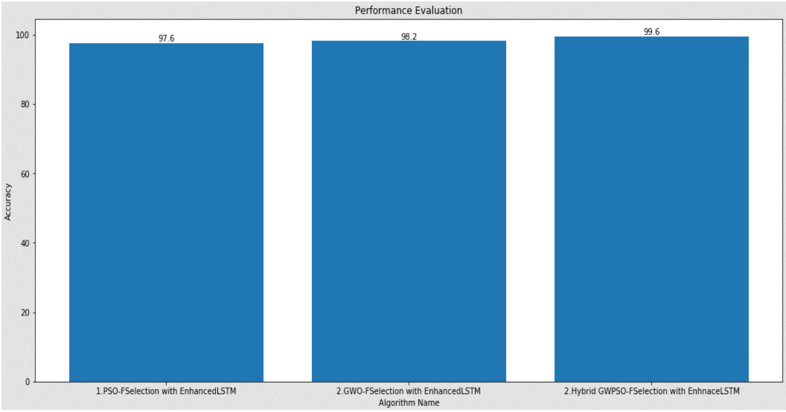

Fig 4.1 Accuracy of Various Feature Selection Techniques with Enhanced LSTM

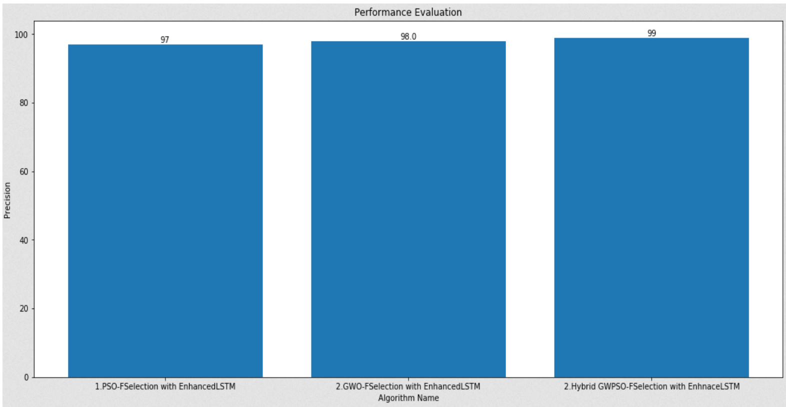

Fig 4.2 Precision of Various Feature Selection Techniques with Enhanced LSTM

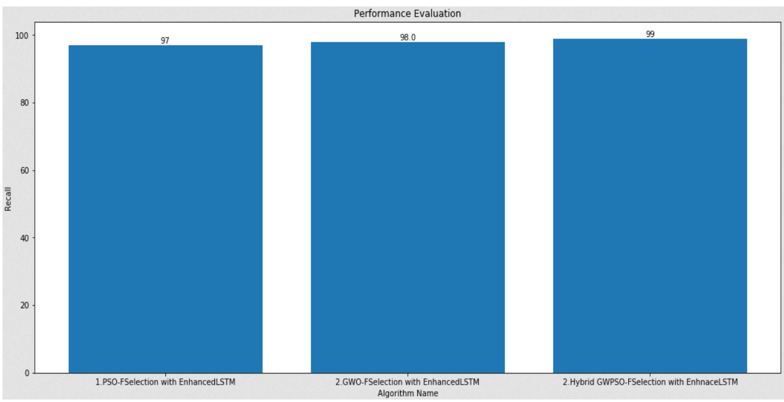

Fig 4.3 Recall of Various Feature Selection Techniques with Enhanced LSTM

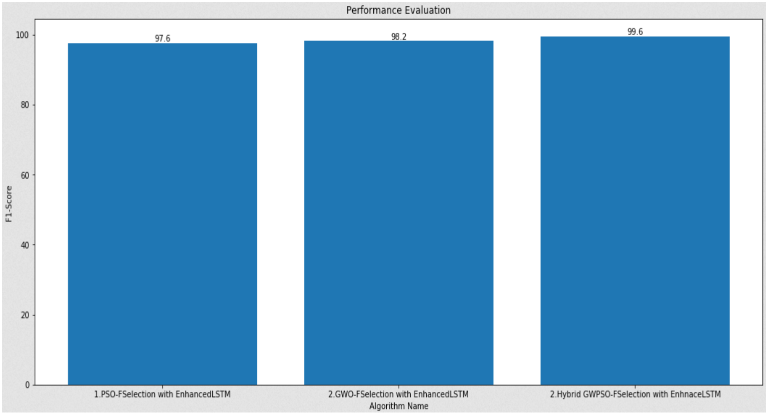

Fig 4.4 F1-Score of Various Feature Selection Techniques with Enhanced LSTM

5. CONCLUSION

Diabetes mellitus is a serious health condition that can lead to numerous complications, including high blood sugar levels, cardiovascular issues, nerve damage, and kidney disease. The prevalence of diabetes is particularly high among the elderly and those with high blood pressure, poor nutrition, and unhealthy lifestyles. As the incidence of diabetes continues to rise globally, there is a pressing need for robust automated methods to predict and detect the disease accurately.

Traditional diabetes prediction software often struggles with accuracy due to the complexities of real-world data. This research introduces a hybrid feature selection technique that leverages Grey Wolf Optimization (GWO) and Particle Swarm Optimization (PSO) to enhance neural network-based diabetes prediction systems. By incorporating F-score assessment, correlation coefficient analysis, and particle cluster improvement, this approach optimizes input data feature selection, thereby boosting the accuracy of Enhanced Long Short-Term Memory (LSTM) models for diabetes risk prediction.

The proposed method has been tested on benchmark datasets, demonstrating its capability to achieve up to 99.6% accuracy after addressing data anomalies and shortages. This study highlights the potential of soft computing techniques to significantly improve diabetes prediction systems, offering substantial advantages over conventional methods. These advancements in predictive accuracy can greatly aid in managing healthcare and mitigating the risks associated with diabetes.

References

1. Liu Z, Zhang H, Chen L, Xu M. A hybrid feature selection method based on particle swarm optimization and neural network for diabetes prediction. J Comput Sci Technol. 2022;37(5):982-994.
2. Yoon H, Kim H, Lee Y, Park K. Enhancing diabetes prediction accuracy using an integrated PSO and F-score based feature selection approach. Expert Syst Appl. 2021;168:114273.
3. Zhang Z, Li W, Wu Q, Yang H. Optimization of diabetes prediction using hybrid feature

selection and artificial neural networks. Comput Biol Med. 2019;107:37-45.
4. Wang L, Zhang L, Liu T, Wu J. PSO-based feature selection and classification for diabetes mellitus prediction. Soft Comput. 2020;24(9):7021-7032.
5. Patel V, Patel S, Prajapati S, Patel D. Diabetes risk prediction using feature selection and machine learning techniques. J Healthc Eng. 2021;2021:8435289.
6. Sharma A, Kumar S, Gupta R, Sharma S. An ensemble approach for diabetes prediction using PSO and SVM. Neural Comput Appl. 2019;31(10):7335-7344.
7. Alomari R, Ibrahim R, Omer K, Yasin S. Hybrid optimization techniques for diabetes diagnosis using neural networks. Comput Math Methods Med. 2021;2021:8567546.
8. Verma S, Kumar P, Soni S, Rajan K. Comparative analysis of PSO-based feature selection methods for diabetes prediction. Int J Comput Appl. 2019;178(30):18-25.
9. Li X, Zhang J, Safara F. Improving the accuracy of diabetes diagnosis applications through a hybrid feature selection algorithm. Neural processing letters. 2023 Feb;55(1):153-69.
10. Kangra K, Singh J. A genetic algorithm-based feature selection approach for diabetes prediction. Int J Artif Intell. 2024 Jun;13(2):1489-98.
11. Khan MZ, Mangayarkarasi R, Vanmathi C, Angulakshmi M. Bio-Inspired PSO for Improving Neural Based Diabetes Prediction System. Journal of ICT Standardization. 2022;10(2):179-99.
12. Patil R, Tamane S, Rawandale N. Hybrid ANFIS-GA and ANFIS-PSO based models for prediction of type 2 diabetes mellitus. InComputational Methods and Data Engineering: Proceedings of ICMDE 2020, Volume 1 2021 (pp. 11-23). Springer Singapore.
13. Qteat H, Awad M. Using Hybrid Model of Particle Swarm Optimization and Multi-Layer Perceptron Neural Networks for Classification of Diabetes. International Journal of Intelligent Engineering & Systems. 2021 May 1;14(3).
14. Kalaiselvi C, Nasira GM. Classification and prediction of heart disease from diabetes patients using hybrid particle swarm optimization and library support vector machine algorithm. International Journal of Computing Algorithm. 2015 Mar;4(2):54-8.
15. Ulutas H, Günay RB, Sahin ME. Detecting diabetes in an ensemble model using a unique PSO-GWO hybrid approach to hyperparameter optimization. Neural Computing and Applications. 2024 Jul 24:1-29.
16. Mallika C, Selvamuthukumaran S. A hybrid crow search and grey wolf optimization technique for enhanced medical data classification in diabetes diagnosis system. International Journal of Computational Intelligence Systems. 2021 Sep 1;14(1):157.
17. Lalitha YS. A REAL TIME DIABETIC DISEASE PREDICTION USING PATIENTS DATASET THROUGH KNN, SVM AND BPSO ALGORITHMS.
18. Talari P, N B, Kaur G, Alshahrani H, Al Reshan MS, Sulaiman A, Shaikh A. Hybrid feature selection and classification technique for early prediction and severity of diabetes type 2. Plos one. 2024 Jan 18;19(1):e0292100.