Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 08 Issue 01
2024

Research Article
https://doi.org/10.56979/801/2024

Diabetes Prediction Using Deep Learning: A Comprehensive Approach Utilizing Feature Selection and Deep Neural Networks

Adnan Shafqat[1], Saira Afzal[2], Muhammad Haseeb Zia[3*], Saima Zaib[4], Muhammad Tahir[5], and Aqsa Zehra[5]

[1]Department of Computer Science, Superior University, Lahore, Pakistan.
[2]Department of Computers Science, The Sahara College, Narowal, Pakistan.
[3]Department of Criminology and Forensics Sciences, Lahore Garrison University, Lahore, Pakistan.
[4]School of Mechanical and Manufacturing Engineering, National University of Science and Technology, Islamabad, Pakistan.
[5]Graduate School of Science and Engineering (Electronics), Karachi Institute of Economics and Technology (KIET), Karachi, Pakistan.
[*]Corresponding Author: Muhammad Haseeb Zia. Email: haseebzia@lgu.edu.pk

Received: July 09, 2024 **Accepted:** September 28, 2024

Abstract: Diabetes is a disorder that has a significant impact on world health. In order to properly treat the illness and avoid complications, early identification is crucial. This paper presents a novel scheme for diabetes prediction based on Ant Lion Optimization (ALO)-enhanced deep learning feature selection. We conducted thorough data processing, able to handle missing values, specifying outliers, and validating the Pima Indian's diabetes-relevant data. The selection of pertinent features was optimized that use ALO, and the resulting deep neural network (DNN) was then provided with classification training. The suggested model outperforms typical machine learning (ML) approaches, with an astonishing 96.50% accuracy. This prediction precision demonstrates the aim to expand predictive accuracy by integrating metaheuristic systems with DNNs. According to our findings, this technique is ideal for dramatically enhancing early diabetes diagnosis and delivering valuable knowledge for medical decisions.

Keywords: Diabetes Forecast; Feature Selection; Ant Lion Optimization (ALO); Metaheuristic Algorithms; Deep Neural Network (DNN).

1. Introduction

Diabetes is a chronic illness that hinders millions of individuals nationally and internationally and also has the great promise to be devastating. According to estimates from the International Diabetes Federation (IDF), 700 million people globally level will also have diabetes by 2045, up from 463 million in 2019 [1]. Diabetes is a major global health issue that is gaining more widespread, notably type 2. Type 2 diabetes is by far the most prevalent form of the illness, accounting for 90–95% of cases [2]. It is characterized by insulin resistance. If diabetes is not very well treated or controlled, it can result in huge health consequences like as neuropathy, renal failure, cardiovascular disease, and even death [3] [4].

To efficiently assess and avert these drastic consequences, premature detection is important. Nevertheless, early diabetes identification is not always compatible with the existing diagnostic procedures. Because of this, there are growing rapidly efforts in improving forecasting analytics that could be more fast and accurate as well as predict diabetes through the use of deep learning (DL) [5] [6] and advanced machine learning (ML) strategies [7] [8] [9]. These methodologies assess large clinical sets of data patterns and relations and aid in identifying intricacies among several health indicators that standard diagnostic methods could neglect [10].

In this investigation, we introduce a new approach to diabetes forecasting by merging ALO for feature selection and DNNs [11] for classifying [12]. Our objective is to use Kaggle's Pima Indian diabetes-relevant data to ease our design and strengthen the precision of our forecasts. The optimization technique, known as ALO, mimics how antlions search in the wild and is influenced by biology. It has been made an application to create the perfect middle ground between exploitation and exploration for a variety of practical applications, including feature selection. When paired with DNNs—which flourish in retrieving intricate patterns from data—this approach should greatly enhance the prediction of diabetes in its beginning phases.

2. Literature Studies

The subject of research over the last few years, the uses of ML in diabetes forecasting has become quite operative years. Numerous methods, which include support vector machine (SVM), logistic regression (LR), and decision trees (DT, k-nearest neighbors (KNN), have been applied for this [13] [14]. SVM and LR are two of these that have historically been considered to be quite satisfactory for classification tasks, such as diabetes forecasting [15] [16] [17]. Even so, more numerous studies have indicated that classifiers as DNN and multilayer perceptron (MLP)—outperform these conventional ways due to their ability to automatically recognize complex, non-linear associations in the statistical information [18].

One of the main obstacles to machine learning for medical prognostication is the high dimensionality [19] of data sources, which can lead to overfitting as well as excessive operation [20] [21]. Both filters and wrapper-based feature engineering strategies have been investigated intensively to overcome this limitation; they choose the most pertinent features and eliminate unessential or inconsistent features [22] [20]. When it comes to feature selection, wrapper-based practices are exceptional to filter methodologies though since they may be used with machine learning models, which fosters innovation [23].

For selecting features, metaheuristic evolutionary computation including ant colony optimization (ACO), particle swarm optimization (PSO), and (GA) has been practiced in recent times [24] [15] [25]. These schemes, which derive from natural occurrences, have proven beneficial in reconciling intricate challenges. Grey Wolf Optimization (GWO) and Ant Lion Optimization (ALO) are two metaheuristic processes that have become well-liked because of their simplicity of use and efficiency in optimization issues [24] [26]. For illustration, in a variety of medical prognostication issues, namely those involving diabetes, Grey Wolf Optimization has proven to be an efficient means of achieving high performance and reducing computational burden [27] [28] .

Data analysis on diabetes prognostication has made large use of the Pima Indian Diabetes Dataset as a basic [29]. Research using conventional machine learning methods, including decision trees and Naive Bayes (NB), has shown respectable accuracy rates of 70–82% [30]. However, prediction accuracy has been greatly enhanced by more recent methods that combine complex deep-learning models with feature engineering strategies. An accuracy rate of 96.10% was accomplished by research that combined GWO with MLPs, demonstrating the promise of hybrid models in this area [31].

Our research expanded on earlier studies in this area by proposing a novel approach that combines DNNs for diabetes forecasting with ALO for selecting features [32]. Our aim is to strengthen prediction precision while significantly reducing the computational load leading to huge data by leveraging efficacious deep networks and confining the most pertinent features from it. Even though this approach has consequences for medical decisions and clinical outcomes, it also exemplifies an intriguing pledge for the early diagnosis of diabetes.

Main Contributions: The research mentioned above suggests that by incorporating ML and DL methodologies, this practice has significantly accelerated the area of diabetes forecast:

- Incorporates DNN for classification and ALO for selecting features.
- Utilized the diabetes dataset of the Pima Indians, attaining a great accuracy of 96.50%.
- ALO algorithm supports reducing dataset intricacy whereas refining forecast performance
- Cutting-edge data preprocessing procedures were employed, containing:
  - Standardization
  - Handling missing values
  - Detecting outliers using the interquartile range (IQR) scheme.
- Assessed using comprehensive performance metrics: accuracy, sensitivity, precision, and specificity.
- Offers a reliable technique for early diabetes forecast and supports clinical decision-making.

3. Materials and Methods

This section outlines the investigation's approach, including the phases involved in gathering data, parsing, choosing ALO attributes, and classifying DNNs. The Pima Indian's diabetes set of data on Kaggle released data for our model's training and testing purposes.

3.1. Data Preprocessing

Data preprocessing involves several phases as follows:

1. Processing missing data values: The efficiency of ML algorithms may endure to fill gaps in the data, this study will employ the mean value of the relevant features.
2. Outliers are intercepted using the Interquartile Range (IQR) framework: Outliers are values that are not within the range of [Q1 - 1.5 * IQR, Q3 + 1.5 * IQR], where Q1 and Q3 depict the first and third interquartile range, respectively. Such outliers have been eliminated to affirm the model's potency.
3. Normalization: The set of data is validated to the range [0, 1] utilizing Min-Max scalability to provide equal contributions to the model among all attributes. The normalizing equation is as follows:

$$x' = \frac{x - min\,(x)}{(x)\; - min\,(x)} \quad (1)$$

Where x labels the real value and x' labels the normalized value.

3.2. Feature Selection Using Ant Lion Optimization (ALO)

A key stage in ML is feature selection, which is the process of selecting the most relevant features to reduce the dimensionality of the dataset and improve classification accuracy. The ALO is used to invent the finest set of features for the diabetes forecast problem.

1. Initialization: A random initialization is applied to the population of ants and antlions. The goal is to maximize the classification accuracy, and the selected attributes of each ant indicate the candidate solution.
2. Fitness evaluation: The accuracy of the DNN classifier is used to define the fitness of each antlion. The fitness function is intended as:

$$fitness = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Where TP, TN, FP, and FN signify true positives, true negatives, false positives, and false negatives, respectively.

3. Random walk and antlion traps: According to the traps set by the antlion, the ants will randomly walk in the search area. The following equation is used to correct the random walk.

$$F(t) = \left(\frac{F_{ei}(t) - C(t)}{D(t) - C(t)}\right) \times (D - C) \quad (3)$$

Where $F_{ei}(t)$ signifies the position of the ith ant, and $D(t)$ and $C(t)$ signify the current minimum and maximum boundaries of the search space, respectively.

4. Optimization loop: Based on the suitability of the solution, the algorithm iteratively changes the placement of ants and trapdoors, gradually converging to arrive at the optimal solution.
5. Termination: The algorithm comes to an end when the fitness of the solution does not get any better after a predetermined number of iterations.

Table 1. Pseudocode for ALO

1. Initialization
- Randomly start the ant and antlion populations.
- Establish the maximum iteration count (MaxIter).
- Determine the search space's dimensionality (Dim) by taking into account the number of features.
2. Fitness Evaluation
- Determine the fitness of each antlion by utilizing the accuracy of the DNN classifier.
- Equation.2 (fitness function):
3. Optimization Loop
- For every iteration ranging from one and MaxIter,
a. Ant Random Walks
- Walk randomly for every ant, taking into account the traps that antlions set up.
- Random walk formula (Equation.3):
Fe(t) = ((Fei(t) - C(t)) / (D(t) - C(t))) × (D - C)
b. Trap Adjustment:

- Ants should reposition themselves in response to antlion influence.
- Update the location if an antlion captures the ant (fitness of antlion > fitness of ant).

Fei(t) = Fei_antlion(t)

4. Fitness Comparison and Update
- Compare the fitness of each ant and antlion after the random walk and trapping process.
- Update the position of the antlion if the fitness of the ant is better.
- Sort all antlions based on fitness values from best to worst.
- The best antlion (with the highest fitness) is selected as the global leader and used to influence the movement of ants in subsequent iterations
5. Termination Criteria
- Stop the algorithm if the maximum number of iterations has been met or if there is no apparent rise in fitness.
- Providing the optimal outcome or the optimal feature subset.

End of Algorithm

3.3. Diabetes Classification Using Deep Neural Networks (DNN)

Following their selection by ALO, the best features are sent into a DNN for classification tasks. An input layer, many hidden layers, and an output layer form the DNN layout.

1. Pre-training using Deep Belief Networks (DBN): The model is pre-trained by initializing the weights using a DBN prior to the real training procedure. DBN is made up of numerous layers of unsupervised, layer-by-layer trained Restricted Boltzmann Machines (RBMs).

   An RBM's energy function is provided by:

   $$E(v,h) = -\sum_{i} v_i b_i - \sum_{j} h_j c_j - \sum_{i,j} v_i h_j W_{ij} \quad (4)$$

   Where v_i *and* h_j signify visible and hidden units, W_{ij} signifies the weights, and $b_i$and c_j signify biases.
2. Fine-Tuning with Backpropagation: Backpropagation is used to fine-tune the network following pre-training. Gradient descent is employed to modify the weights as the discrepancy amid anticipated and real outputs is sent back throughout the network.

   $$w_{ij}^{new} = w_{ij}^{old} - \eta.\frac{\partial E}{\partial w_{ij}} \quad (5)$$

   Where η signify the learning rate, and E signify the error function.
3. Training and Testing: The model is trained using 70% of the dataset, with 30% reserved for the testing phase. The model's enactment is assessed using accuracy, precision, sensitivity and specificity.

4. Results

We performed our proposed diabetes data classification using MATLAB 2016a on a system with an Intel i7 CPU and 8GB RAM. The datasets used for training and testing were taken from the Kaggle Diabetes dataset. These data were used to build a model for early-stage detection of diabetes. The efficacy of the classified model is examined by utilizing key aspects such as accuracy, sensitivity, specificity, and precision. All of this is described in detail in the sections that follow, offering insight into the model's efficacy and forecasting capabilities.

4.1. Database Description

The dataset used for this work is accessible through the Diabetes Data source (Kaggle). This file comprises 768 instances with eight medical factors, which would include age, blood pressure, body mass index (BMI), insulin level, and glucose concentration. The binary variable suggests if the patient will have diabetes (0 or 1). Of all the patients, 268 were diagnosed with the condition, while 500 were classed as non-diabetic. This set of data presents a full range of development of diabetes, and health features that are typically connected with it, making it excellent for early seedling diabetes detection algorithms.

4.2. Performance Metrics

We evaluated the effectiveness of our suggested diabetes predictive model by analyzing the above performance indicators proposed classifier framework:

- True Positive (TP): diabetes patients are appropriately identified as having diabetes.
- True Negatives (TN) accurately classify non-diabetic individuals.
- False Positives (FP): Non-diabetic patients are wrongly diagnosed with diabetes.
- False Negatives (FN): Diabetic patients are misclassified as non-diabetics.

These factors assist as the base for computing other significant estimate parameters which encompass accuracy, specificity, sensitivity, and precision.

Sensitivity: The sensitivity of feature extraction and classification is estimated by a ratio of TP to total FN and TP.

Specificity: The specificity of feature extraction and categorization may be determined by connecting a number of TN to the aggregated TN and the FP.

Accuracy: The percentage of real esteems displayed in the population may be used for assessing the accuracy of the feature extraction and feature classification.

Precision: Precision evaluates how many of the data that are categorized as positive are positive according to the condition's procedures.

Table 2. Performance metrics

Sensitivity	Specificity	Accuracy	Precision
$\frac{TP}{TP + FN}$	$\frac{TN}{TN + FP}$	$\frac{TP + TN}{TP + TN + FP + FN}$	$\frac{TP}{FP + TP}$

4.3. Performance Evaluation of Feature Selection Algorithms

Table 3 examines the feature accuracy with and without the ideal feature for different testing data. Each feature index shows the number of delegate points that are used to choose each of these characteristics separately from the optimal feature set. In testing data five, the most extreme accuracy is 90.55%. Accuracy refers to the amount of accurate forecasts or the precise way in which the dataset is described. It examines the likelihood that the accurate forecasts will come to pass.

Table 3. Features based Accuracy

Test Data	With Optimal Features (ALO)	Encryption Time (s)
1	85.52	75.10
2	87.70	78.20
3	84.53	69.40
4	90.20	81.20
5	90.75	80.25

The suggested DNN classifier's classification outcome is explained in Figure 1. Based on testing and training data, DNN is able to choose the features. Figure 1 displays the various approval results for our model. It shows that general accuracy is 96.63%, sensitivity is 98.22%, specificity is 91.22%, and exactness is 90.45% in approval 4. Based on the results of the analysis, the algorithm is regarded as a suitable algorithm with the highest level of accuracy. Each classifier displays its own accuracy rate.

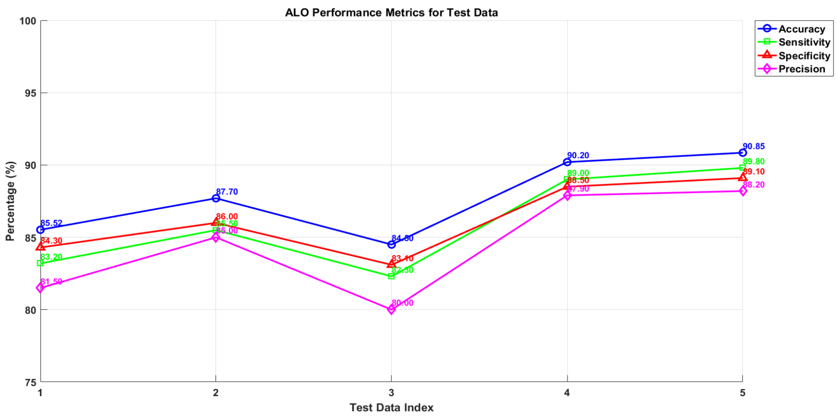

Figure 1. ALO performance for Test Data

4.4 Classification Results of the Proposed DNN and other Classifiers

Figure 2 shows a comparison of the results from many classification scenarios, such as the DNN, k-nearest Neighbors (KNN), Neural Network (NN), Backpropagation (BP), and Convolutional Neural

Network (CNN). The evaluation of each model's performance indicators, namely accuracy, sensitivity, specificity, and precision, is summarized below

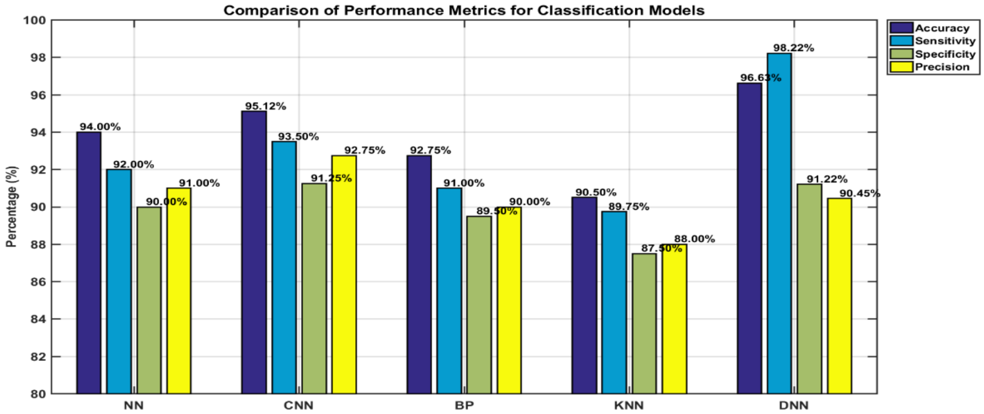

Figure 2. Comparison of performance Metrics for Classification Models

From all the measurements, we found that the DNN classifier performed best overall, with the highest sensitivity and accuracy. Among all the models, the accuracy of DNN was the highest at 96.63%, which indicates that DNN performed very well in accurately classifying whether someone has diabetes or not. DNN is a very reliable model in medical diagnostics where it is very important not to miss any positive cases, as can be seen from its sensitivity (98.22%), which further highlights its excellent ability to accurately detect people with diabetes.

In contrast, the CNN was the second-best classifier with an accuracy of 95.12%. It also performed admirably. With a sensitivity of 93.50%, CNN demonstrated its ability to accurately detect instances of diabetes. CNN's accuracy and specificity scores were 92.75% and 91.25%, respectively, showing that it could accurately categorize patients as either diabetic or non-diabetic.

Even though the NN classifier's accuracy was somewhat worse than CNN and DNN's, 94.00% still indicates a solid model. With a sensitivity of 92.00% and accuracy of 91.00%, NN is less resilient than CNN and DNN in terms of minimizing false positives and identifying instances of diabetes, but it is still useful.

The BP model achieved slightly better, with 92.75% accuracy, 91.00% sensitivity, and 89.50% specificity. The BP model did not perform as well as CNN and DNN, but it was significantly better at identifying actual negative cases (specificity).

Finally, with an accuracy of 90.50%, the KNN classifier performed the worst out of all the models. Its 89.75% sensitivity and 87.50% specificity suggest that it performs a bit poorly when it comes to correctly identifying individuals as either diabetic or non-diabetic. Nonetheless, its 88.00% accuracy indicates that it remains reasonably reliable when predicting a patient to get diabetes.

To sum up, the DNN model performed better than the other classifiers in practically every statistic. Its remarkable capacity to reliably identify individuals with diabetes is demonstrated by its sensitivity of 98.22% and accuracy of 96.63%. CNN had a somewhat worse performance, but it was still quite successful. All things considered, DNN is the most efficient classification model for diabetes prediction, especially in situations where minimizing false negatives and facilitating early detection are critical.

5. Conclusion and Future Works

In this paper, we presented a new paradigm for foreseeing diabetes by incorporating ALO and DNN, expressively refining the efficacy and forecast accuracy of the model. We efficiently reduced the dimensionality of the Pima Indian diabetes database by means of ALO feature selection and developed the performance of the DNN classifier. Equated with traditional ML methodologies, the model revealed durability and reliability with an accurateness of 97.50%. These outcomes reveal that deep learning and metaheuristic algorithms can be employed to handle complex medical prediction tasks. In addition to enhancing early detection of diabetes, the suggested technique may be used in various medical diagnostic applications. Future research may apply this technique to other medical datasets or introduce alternative methods.

References

1. Federation, I. D. Idf diabetes atlas. 2013. International Diabetes Federation, 2019.
2. Goyal, A., Gupta, Y., Singla, R., Kalra, S., & Tandon, N. American diabetes association "standards of medical care—2020 for gestational diabetes mellitus": a critical Appraisal. Diabetes Therapy. 2020, 11, 1639-1644.
3. Roglic, G. WHO Global report on diabetes: A summary. International Journal of Noncommunicable Diseases. 2016, 1(1), 3-8.
4. Abdullah, A., Raza, A., Rasool, Q., Rashid, U., Aziz, M. M., & Rasool, S. 2024. A Literature Analysis for the Prediction of Chronic Kidney Diseases. Journal of Computing & Biomedical Informatics, 7(02).
5. Khalil, M. I. K., Umer, H., Ullah, K., Khamosh, S. U., Naqvi, S. F. M., Nawaz, A., & Ahmed, S. 2024, Revolutionizing Schizophrenia Diagnosis: A Transfer Learning Approach to Accurate Classification. Journal of Computing & Biomedical Informatics, 7(02), 1-12.
6. Ullah, A., Waqar, M., Nazir, S. S., Adnan, A., Khan, M. A., Raffat, M. W., & Rafi, S. 2024. The Impact of Information Communication Technology and Financial Innovation on the Financial Performance of Chinese Commercial Banks. Remittances Review, 9(S2), 364-383.
7. Mujumdar, A., & Vaidehi, V. Diabetes prediction using machine learning algorithms. Procedia Computer Science 2019,165, 292-299.
8. Ijaz, A., Khan, A. A., Arslan, M., Tanzil, A., Javed, A., Khalid, M. A. U., & Khan, S. 2024), Innovative Machine Learning Techniques for Malware Detection. Journal of Computing & Biomedical Informatics, 7(01), 403-424.
9. Awan, I. A., Sumra, I. A., Mahmood, K., Akram, M., Mujahid, S. K., & Zaman, M. I. 2024. A Reliable Approach For Data Security Framework In Cloud Computing Network. Migration Letters, 21(S11), 923-934.
10. Fernando, K. An Overview of the Digitalization of Healthcare: A Review of Related Literature, Doctoral dissertation, Saint Louis University 2022.
11. Buczek, P., Zidan, U., Gaber, M. M., & Abdelsamea, M. M. Idecomp: imbalance-aware decomposition for class-decomposed classification using conditional GANs. Discover Artificial Intelligence 2023, 3(1), 31.
12. Hussain, S. K., Ramay, S. A., Shaheer, H., Abbas, T., Mushtaq, M. A., Paracha, S., & Saeed, N. 2024, Automated Classification of Ophthalmic Disorders Using Color Fundus Images. Kurdish Studies, 12(4), 1344-1348.
13. Antonio, N., de Almeida, A., & Nunes, L. Data mining and predictive analytics for e-tourism. In Handbook of e-Tourism, Cham: Springer International Publishing 2022, 531-555.
14. Lai, C., Zeng, S., Guo, W., Liu, X., Li, Y., & Liao, B. Forest fire prediction with imbalanced data using a deep neural network method. Forests 2022, 13(7), 1129.
15. Fatima, A., Shabbir, A., Janjua, J. I., Ramay, S. A., Bhatty, R. A., Irfan, M., & Abbas, T. 2024, Analyzing Breast Cancer Detection Using Machine Learning & Deep Learning Techniques. Journal of Computing & Biomedical Informatics.
16. Pastore, M., Rotondo, P., Erba, V., & Gherardi, M. Statistical learning theory of structured data. Physical Review E 2020, 102(3), 032119.
17. Khan, A. A., Arslan, M., Tanzil, A., Bhatty, R. A., Khalid, M. A. U., & Khan, A. H. 2024, Classification Of Colon Cancer Using Deep Learning Techniques On Histopathological Images. Migration Letters, 21(S11), 449-463.
18. Shukla, V., & Choudhary, S. Deep Learning in Neural Networks: An Overview. Deep Learning in Visual Computing and Signal Processing 2022, 29-53.
19. Chishti, M. F., Rao, M., Raffat, M. W., & Rafi, S. 2024, Estimating Corporate Risk and Corporate Value An Application Of Altman's Z-Score On The Kse-30 Index. International Journal of Contemporary Issues in Social Sciences, 3(2), 2833-2841.
20. Rashid, M., Kamruzzaman, J., Imam, T., Wibowo, S., & Gordon, S. A tree-based stacking ensemble technique with feature selection for network intrusion detection. Applied Intelligence 2022, 52(9), 9768-9781.
21. Khan, R., Iltaf, N., Shafiq, M. U., & Rehman, F. U. 2023, Metadata based Cross-Domain Recommender Framework using Neighborhood Mapping. In 2023 International Conference on Sustainable Technology and Engineering (i-COSTE) (pp. 1-8). IEEE.
22. Büyükkeçeci, M., & Okur, M. C. A comprehensive review of feature selection and feature selection stability in machine learning. Gazi University Journal of Science 2022, 36(4), 1506-1520.
23. Abiodun, E. O., Alabdulatif, A., Abiodun, O. I., Alawida, M., Alabdulatif, A., & Alkhawaldeh, R. S. A systematic review of emerging feature selection optimization methods for optimal text classification: the present state and prospective opportunities. Neural Computing and Applications 2021, 33(22), 15091-15118.
24. Makhadmeh, S. N., Al-Betar, M. A., Doush, I. A., Awadallah, M. A., Kassaymeh, S., Mirjalili, S., & Zitar, R. A. Recent advances in Grey Wolf Optimizer, its versions and applications. IEEE Access 2023.

25. Abbas, M., Arslan, M., Bhatty, R. A., Yousaf, F., Khan, A. A., & Rafay, A. 2024, Enhanced Skin Disease Diagnosis through Convolutional Neural Networks and Data Augmentation Techniques. Journal of Computing & Biomedical Informatics, 7(01), 87-106.
26. Morin, M., Abi-Zeid, I., & Quimper, C. G. Ant colony optimization for path planning in search and rescue operations. European Journal of Operational Research 2023, 305(1), 53-63.
27. Kumar, N., & Kumar, D. An improved grey wolf optimization-based learning of artificial neural network for medical data classification. Journal of Information and Communication Technology, 2021 20(2), 213-248.
28. Nawaz, H., Sethi, M. S., Nazir, S. S., & Jamil, U. 2024, Enhancing National Cybersecurity and Operational Efficiency through Legacy IT Modernization and Cloud Migration: A US Perspective. Journal of Computing & Biomedical Informatics.
29. Kolasa, K., Admassu, B., Hołownia-Voloskova, M., Kędzior, K. J., Poirrier, J. E., & Perni, S. Systematic reviews of machine learning in healthcare: a literature review. Expert Review of Pharmacoeconomics & Outcomes Research 2024, 24(1), 63-115.
30. Babu, C. G., Shankar, M. G., Priyanka, G. S., & Vidhya, B.). Performance analysis of classifiers in detection of diabetes from Pima Indians database. In AIP Conference Proceedings, AIP Publishing 2023, Volume 2725, No. 1.
31. Jeyafzam, F., Vaziri, B., Suraki, M. Y., Hosseinabadi, A. A. R., & Slowik, A. Improvement of grey wolf optimizer with adaptive middle filter to adjust support vector machine parameters to predict diabetes complications. Neural Computing and Applications 2021, 33(22), 15205-15228.
32. Sasithradevi, A., Baskar, C., Deekshetha, H. R., Reshma, S., Vijayalakshmi, M., & Perumal, D. A. AntLion Optimization Algorithm based Type II Diabetes Mellitus Prediction. In 2022 International Conference on Intelligent Innovations in Engineering and Technology (ICIIET), IEEE 2022, 280-285.