DOI: 10.15849/IJASCA.240730.11

Int. J. Advance Soft Compu. Appl, Vol. 16, No. 2, July 2024
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Kidney Disease Prediction using Elitist-Ant System Algorithm with Multilayer Perceptron Neural Network

[1]*Majd M. AL-Hawamdeh, [2]Ahlam Alsalahat, [2]Ghaith M. Jaradat, [1]Mahmoud A. Al Shugran, [1]Ahmad R. Ayasrah

[1]Faculty of Computer Science and Information Technology, Jerash University, Jerash, Jordan
e-mail: majd.alhawamdeh@jpu.edu.jo; m.shquran@jpu.edu.jo; a.ayasrah@jpu.edu.jo
[2]College of Computer Sciences and Informatics, Amman Arab University, Mubis-Amman, Jordan
e-mail: 202020254@aau.edu.jo; g.jaradat@aau.edu.jo

Abstract

A significant percentage of chronic kidney disease (CKD) patients has recently increased, prediction procedures have become effective, using machine learning models and optimizers to identify most relevant features and improve classification accuracy. The goal is to avoid overfitting, minimize computational time in the classification model, and provide a faster and cost-effective model. This study aims to improve the accuracy of the CKD prediction model by combining the Elitist-Ant System algorithm with a Multi-Layer Perceptron Neural Network (EAS-MLPNN). EAS selects significant features while maintaining an appropriate analytical result. The used dataset verifies the model's performance. Where observed results from the experiments showed that, the EAS-MLPNN model made better predictions than traditional classifiers including support vector machine, MLPNN, and random forest with 94.67% accuracy indicating a 2% enhancement of the classification accuracy.

Keywords: *feature selection; chronic kidney disease; prediction model; elitist-ant system, multilayer perceptron, neural network.*

1 Introduction

A chronic kidney disease (CKD) diagnostic model can help medical experts to make more efficient and accurate diagnosis. The model extracts useful information from mass amounts of medical datasets. Machine learning (ML) model guarantee early detection of CKD by analyzing large amounts of data with different variables, developed from computational learning theory and pattern recognition in artificial intelligence (AI).

CKD is a common condition that gradually damages the kidneys over time. It is difficult to diagnose early because it often has no symptoms in the initial stages. However, early detection is important as it reduces the risk of death, treatment costs, and the need for dialysis or kidney transplantation [1] [2]. CKD is identified in patients with high blood pressure, diabetes, or a family history of the disease.

Received 24 March 2024; Accepted 28 May 2024

EAS algorithm uses probability to solve computational problems and find the best path on a graph. It has advantages like rapid solution discovery and effectiveness in solving problems like the traveling salesman [3]. On the other hand, various ML-based decision support models have been developed to detect CKD, such as MLP, which is a feed-forward artificial neural network with hidden layers that uses backpropagation for training [2]. Challenges include missing data, irrelevant features, and prediction issues. The EAS-MLPNN addresses these problems by removing irrelevant features and maintaining prediction accuracy. This results in a faster and more efficient CKD diagnosis model.

Hence, this study aims to improve CKD model's prediction accuracy using elitist ant system (EAS) algorithm incorporated with multi-layer perceptron neural network (MLPNN) in the training phase of the model. Where EAS optimizer selects relevant features in the dataset for the MLPNN prediction model before the diagnostic process. The objective is to reduce dataset dimensions by ignoring irrelevant features, increasing prediction accuracy and reducing classification time. On the other hand, MLPNN offers processing, learning ability, fault tolerance, and high generalizability. Based on various metrics, the model's performance is evaluated on a well-known UCI dataset, which has 8820 instances and 34 features related to CKD diagnosis. The study benefits researchers and health professionals by assisting in the diagnosis of CKD and providing insight into optimization algorithms and ML methods.

2 Related Work

Many studies developed a variety of ML models used for diagnosing CKD and studying the effectiveness of predictive models are discussed in the literature. They emphasize the importance of using optimization algorithms to select significant features and improve the efficiency of prediction models. Additionally, they mentioned comparative studies of proposed methods and reviewed existing CKD prediction models.

2.1 Machine Learning with feature selection

ML is a subset of AI that involves training computers using annotated datasets to improve their performance over time. ML techniques are categorized into supervised and unsupervised learning to address pattern recognition challenges. In supervised learning, various classifiers such as k-nearest neighbor, decision trees, self-organizing maps, naïve bayes, random forest, neural network, and support vector machine [2] are employed to classify data, with training data used to create a function with input vectors and class labels. MLP and features selection were used in a study [2] for CKD prediction, with MLP being commonly used for classification and having high accuracy, and features selection used to select the most relevant features in the dataset.

2.2 Feature selection using optimization algorithms

Feature selection is a crucial aspect of data mining and AI, aiming to improve classification accuracy and reduce computational time [4]. Despite eliminating some features, they may still hold valuable information and significant statistical relationships with others. Various feature selection algorithms, such as filter, wrapper, and embedded methods, are widely employed in research [4]. Additionally, metaheuristic algorithms are powerful tools for addressing complex optimization problems, including feature selection, due to their efficient exploration of diverse solution spaces [4].

2.3 Elitist-Ant System

The EAS algorithm is based on a hybrid ant colony system algorithm and aims to balance search diversity and intensity [3]. It uses two mechanisms, diversification and condensation, to aid in search exploration and exploitation while maintaining solution quality. The algorithm was tested on the course schedule problem and produced better solutions than other ant systems by exploiting specific knowledge using an effective directed search. The results demonstrated the algorithm's ability to provide high-quality solutions [3].

2.4 Neural Network with Backpropagation

Neural networks are inspired by the human brain. They consist of processing elements, input, and output components. The basic model includes a summer weight and an activation function [5] [6]. The feed-forward method is used to propagate signals from input layer to output layer through hidden layers, combining the weights associated with all neurons. The learning phase aims to optimize input signal weights for better computational results. Neural networks are utilized in various fields such as medicine, pharmacology, weather forecasts, and economics predictions. They are widely used for predictive medical decisions and disease diagnosis, including CKD.

Backpropagation is a gradient descent-based strategy that calculates the loss gradient at the output and distribute it back through deep neural networks layers to adjust neuron weights. It is used for training a neural network based on its activity and is classified as a supervised learning algorithm [5] [6]. The backpropagation has several problems, such as follows:

- Overfitting in neural networks can occur due to prolonged learning, infrequent training examples, random features, or an excessive number of hidden neurons.
- The backpropagation strategy may not reach the global minimum due to numerous local minimums in the error function.

The algorithm cannot accurately determine the best solution and may require multiple attempts to find the global minimum. Backpropagation is a popular and useful algorithm in natural language processing, optical character recognition, and image processing [5]. Additional related works provide further information and insights for those interested [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18].

2.5 Commenting on the related studies

ML is valuable in health diagnostics, but analyzing the data uses resources and time. Not all features support disease diagnosis, so efficient optimization algorithms are needed to find the most relevant ones.

Based on table 1, we found that using optimization methods to select relevant features improved the accuracy and reduced the complexity and execution time of the CKD model. Most previous studies examined CKD prediction models using a dataset from the UCI machine-learning repository. The proposed EAS-MLPNN selects the most relevant features from a comprehensive dataset of 34 features and 400 records.

Table 1: Summary of related works

Methods	Results	Limitation
Decision tree, random forest, linear regression [19]	DT provided the best performance Accuracy = 99.7%	No feature selection No AROC, MSE, confusion matrix No comparison with or without FS
SVM with Fruit fly optimization for feature selection [20]	Accuracy = 98.5%	Local optima stagnation, long run time, and early convergence No AROC, MSE, and confusion matrix
XGBoost [21]	Test accuracy, sensitivity, & specificity = 100%, 1.000, & 1.000	No AROC, MSE, and confusion matrix
Deep NN with feature selection [22]	Accuracy = 98% High F1-measure & sensitivity	Local optima stagnation, long run time, and early convergence No AROC, MSE, & confusion matrix

3 The Proposed Model

In this study, we utilized supervised learning to train a neural network for accurate diagnosis of diseases including CKD [1]. Optimization algorithms were also used to improve accuracy, and execution time, while addressing overfitting. Therefore, EAS is used to select most relevant features related to CKD.

3.1 Dataset

This study used a dataset from UCI repository with 8,819 individuals, where a CKD test conducted for each individual. The dataset was randomly split into training and testing sets, containing 6,000 and 2,819 cases respectively. There are 34 features per record, including the target representing the probability of infection.

3.2 Implementation

As shown in Fig. 1, there are many steps that we have taken to implement the EAS-MLPNN.

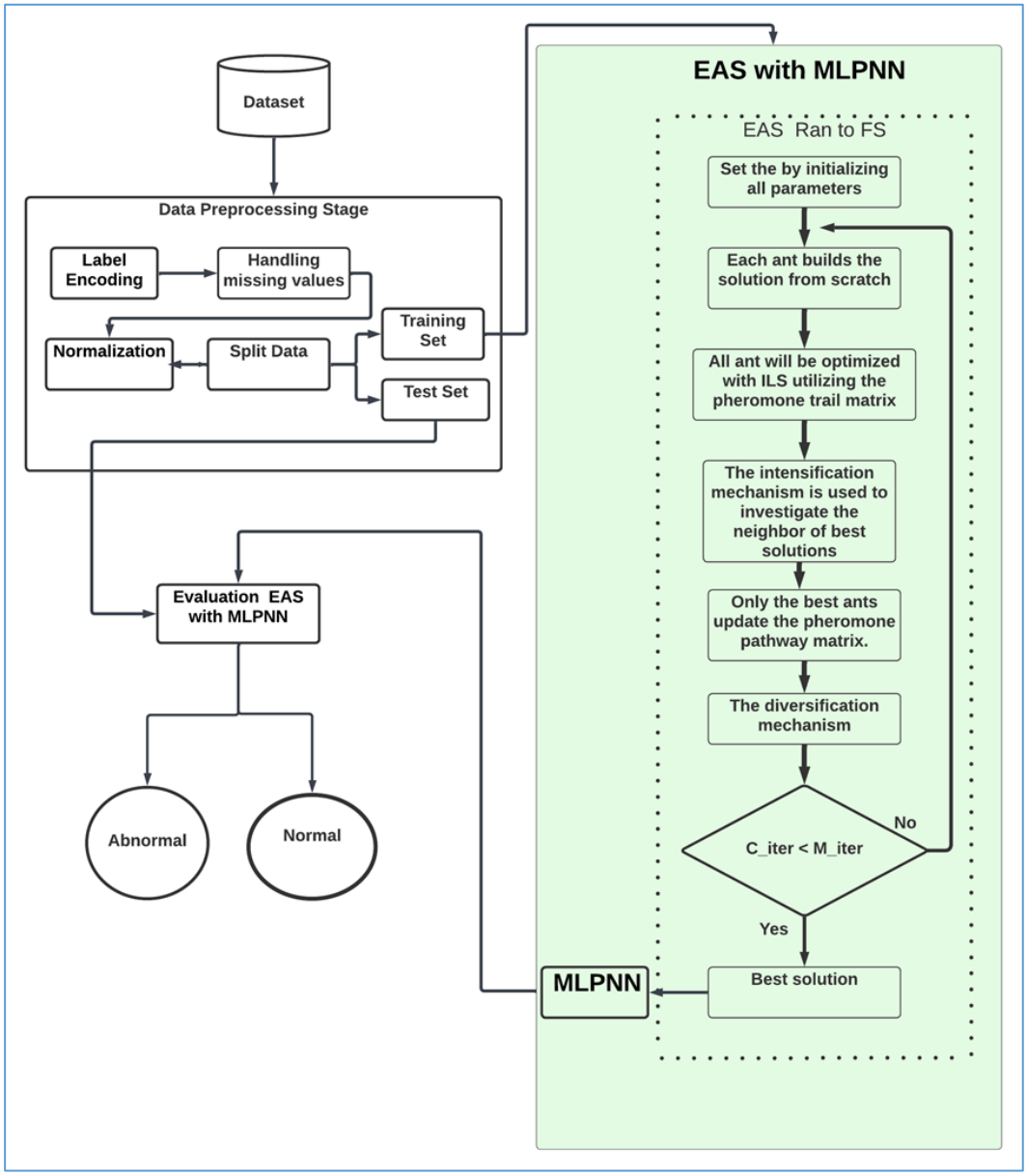

Figure 1: EAS-MLPNN model for CKD prediction

3.2.1 Dataset Preparation

There are many steps for dataset preparation including:

- Step 1: label encoding is used to convert nominal variables into numerical forms [23] including categorical features (nominal or ordinal), such as converting ethnic categories into numeric attributes using ordinal notation.
- Step 2: mean value of features is used to fill in the gaps of features with missing values [24].
- Step 3: before feature selection, normalization is used to reduce attribute values to a limited range [25] to simplify attribute comparison and improve algorithm

learning. The dataset's non-uniformly distributed features required scaling to ensure values were within the range of 0 to 1 after encoding [26].

3.2.2 Elitist-Ant System for Features Selection

The EAS is utilized for feature selection, which selects a subset of features with high performance. EAS is based on the behavior of real ants. It creates a relationship between the search for real ants and possible solutions to the problem, using pheromone pathways for both exploration and exploitation. The pheromone pathway is updated based on the current pheromone level and path evaporation coefficient. This process helps in selecting components and increasing the mixture of pheromone trail values and objective partial function evaluations [3] [27]. This study used the EAS algorithm to select features affecting the accuracy of the CDK prediction model. It also utilized an external memory for elite solutions (as in [3]), allowing the EAS to handle robust exploration.

3.2.3 MLPNN's Architecture

Creating a mathematical model for NN is challenging, where all neurons are synchronized, and signal processing is the same for all neurons. Each neuron has a transfer function that defines its output based on input power, and this function is time-independent. The signal changes in a linear shape when it crosses synapses, with value multiplied by a synaptic weight. This weight changes over time, allowing the brain to respond differently to the same inputs. The following is a review of the components of MLPNN [24]:

1. input layer is made up of neurons that receive input and pass it to the next layer. The proposed EAS-MLPNN model contains 34 input neurons matching the features.
2. hidden layers are made up of hidden neurons, transmitting signals between input and output neurons.
3. output layer consists of neurons converting signals into values, classifying if a patient has CKD.
4. momentum impacts how past weights influence current weights, helping to avoid local minimums and improve performance. Selecting the right momentum value is crucial and should be determined empirically through experiments equals to 0.5.
5. learning rate in training is a configurable hyperparameter that determines the stride size of weight updates. The best value for learning rate is found through experimentation, with 0.15 being the best value in the proposed model.

MLPNN accuracy in predicting infection depends on training, experimenting, and parameter comparison, as presented in the following section.

4 Results and Discussion

4.1 Experimental settings

There are 8 evaluation metrics employed to assess the proposed model based on the description of [28]: (i) Accuracy to find the percentage of correct predictions of test data. (ii) Area under the receiver-operating characteristic (ROC) is a discriminatory power

estimation of species distribution models. It is measured by how accurate a quantitative diagnostic test is. (iii) Execution time taken by prediction models to diagnose CKD. (iv) Geometric mean is a product of a series of numbers by the inverse of the total length of the series. It is useful when the numbers in the series are dependent or if the numbers tend to cause large fluctuations [29]. It is most proper for series that display serial correlation. (v) F1-Score combines accuracy and recall by taking a harmonic mean for comparing two different classifiers. (vi) Precision to determine how close prediction results are to each other. (vii) Recall (sensitivity) to find true positives proportion that the model correctly predicts. (viii) Specificity to find true negatives proportion that the model correctly predicts. The proposed model was implemented using Matlab on a Core i7 machine with 8GB RAM running 64-bit Microsoft Windows.

4.2 Results Analysis

Two methods for diagnosing and predicting CKD are studied, analyzed, and compared. The first method is a CKD prediction using traditional standalone MLPNN, SVM, and RFC without FS, while the second is the EAS-MLPNN.

4.2.1 CKD predictions without FS

A set of random configurations for MLPNN is performed as shown in table 2. We concluded that the best configuration parameters are (3, 4, 0.25, 500, 0.03) for the number of hidden layers, neurons number in each layer, learning rate, epochs number, and momentum_alpha, respectively, that achieved 92.6% accuracy and 0.176 AROC.

Table 2: Performance comparison of MLPNN configurations

Config.	Neurons	Layers	Epochs	Learning rate	Momentum α	Accuracy	Recall	Precision	Specificity	F1-Score	AROC	Geomean	Time (s)
1	10	4	2000	0.25	0.03	89.0	0.35	0.35	0.214	0.92	0.26	0.85	331
2	10	1	100	0.4	0.01	92.3	0.23	0.28	0.965	0.25	0.23	0.89	11
3	4	1	500	0.25	0.03	89.6	0.23	0.18	0.936	0.20	0.23	0.86	54
4	10	1	1000	0.25	0.03	91.6	0.23	0.25	0.958	0.24	0.23	0.88	118
5	4	1	1000	0.5	0.03	89.6	0.23	0.18	0.936	0.20	0.23	0.86	100
6	10	3	500	0.25	0.03	90.3	0.23	0.20	0.943	0.21	0.23	0.87	84
7	4	4	500	0.25	0.03	90.6	0.23	0.23	0.211	0.94	0.22	0.87	82
8	4	4	1000	0.25	0.03	90.3	0.23	0.23	0.200	0.94	0.21	0.87	188
9	10	4	100	0.25	0.03	92.0	0.17	0.17	0.231	0.96	0.20	0.89	13
10	10	4	100	0.1	0.03	89.3	0.23	0.23	0.174	0.93	0.20	0.85	21
11	10	1	50	0.25	0.03	92.3	0.17	0.25	0.968	0.20	0.17	0.89	6
12	10	1	2000	0.25	0.03	89.6	0.17	0.15	0.940	0.16	0.17	0.86	238
13	10	3	50	0.4	0.03	91.0	0.17	0.18	0.954	0.18	0.17	0.87	9
14	**4**	**3**	**500**	**0.25**	**0.03**	**92.6**	**0.17**	**0.27**	**0.972**	**0.21**	**0.17**	**0.90**	**69**
15	10	3	2000	0.25	0.03	87.3	0.17	0.11	0.915	0.13	0.17	0.83	334
16	10	2	2000	0.25	0.03	90.0	0.11	0.11	0.947	0.11	0.11	0.86	295
17	10	3	1000	0.25	0.03	91.6	0.11	0.16	0.965	0.13	0.11	0.88	168
18	10	4	1000	0.1	0.03	83.6	0.11	0.11	0.056	0.88	0.07	0.78	198
19	4	4	2000	0.25	0.03	90.3	0.05	0.05	0.071	0.95	0.06	0.87	331
20	10	2	50	0.4	0.03	90.3	0.05	0.07	0.954	0.06	0.05	0.87	7

4.2.2 CKD prediction using EAS-MLPNN

Three versions of EAS (EAS, EAS2, and EAS2VTanh) with various configurations are performed (as shown in table 3) to determine the best EAS methods, solution number, max iteration, and dimensions. From the table, we concluded that the best version is EAS2, which achieved 94.67% accuracy, and 0.294 AROC, with 1000 epochs (iterations), 4 neurons (solutions), and 32nd feature (dimensions), and it selected 8 relevant features. Furthermore, most of experiments selected the 32nd feature, which represents anemia disease because it is a common complication of CKD.

Table 3: Performance comparison of EAS versions

EAS models	Neurons	Layers	Epochs	Features	Accuracy	Recall	Precision	Specificity	F1-Score	AROC	Geomean	Time (s)
EAS2	**4**	**1**	**1000**	**8**	**94.6**	**0.294**	**0.556**	**0.986**	**0.385**	**0.294**	**0.927**	**71.39**
EAS2VTanh	4	1	500	6	94.3	0.235	0.500	0.986	0.320	0.235	0.923	34.09
EAS2VTanh	10	1	500	6	94.3	0.235	0.500	0.986	0.320	0.235	0.923	34.09
EAS2VTanh	4	2	1000	14	94.0	0.235	0.444	0.982	0.308	0.235	0.919	81.18
EAS	4	1	2000	8	92.0	0.176	0.231	0.965	0.200	0.176	0.892	253.35
EAS2	4	3	2000	9	90.3	0.176	0.167	0.947	0.171	0.176	0.870	258.37
EAS	4	1	500	10	90.3	0.118	0.125	0.951	0.121	0.118	0.870	65.54
EAS	4	1	1000	7	93.0	0.118	0.250	0.979	0.160	0.118	0.905	127.55
EAS	4	1	1000	13	92.6	0.118	0.222	0.975	0.154	0.118	0.901	79.71
EAS	**10**	**1**	**50**	**6**	**94.3**	**0.059**	**0.500**	**0.996**	**0.105**	**0.059**	**0.923**	**8.75**
EAS	4	1	1000	6	93.0	0.059	0.167	0.982	0.087	0.059	0.905	67.95
EAS2	10	3	500	3	91.6	0.059	0.100	0.968	0.074	0.059	0.888	80.05
EAS2	4	1	1000	6	94.0	0.059	0.333	0.993	0.100	0.059	0.919	67.89
EAS2VTanh	**4**	**1**	**500**	**7**	**94.6**	**0.059**	**1.000**	**1.000**	**0.111**	**0.059**	**0.927**	**30.56**
EAS2VTanh	4	4	1000	9	93.6	0.059	0.250	0.989	0.095	0.059	0.914	72.75

To confirm the effectiveness of the proposed model, we implemented k-fold cross-validation technique, where k=10 determined experimentally, see table 4. We can gain a considerably more accurate visualization of a model's performance by averaging all of its evaluations. However, there is a drawback; the number of validation sets multiplies the training time. We can conclude the best version is EAS2, which achieved 92% accuracy, and 0.364 AROC, with 500 epochs (iterations), 10 neurons (solutions), and 32nd features (dimensions).

Table 4: Performance comparison of EAS versions using 10-fold cross validation

EAS models	Neurons	Layers	Epochs	Features	Accuracy	Recall	Precision	Specificity	F1-Score	AROC	Geomean	Time (s)
EAS	10	2	100	9	88.0	0.11	0.20	0.95	0.14	0.11	0.84	14.6
EAS2	4	1	1000	9	88.0	0.18	0.40	0.96	0.25	0.18	0.82	145.6
EAS	4	1	1000	5	91.0	0.27	0.75	0.98	0.40	0.27	0.87	30.4
EAS2VTanh	4	3	1000	7	90.0	0.11	0.33	0.97	0.16	0.11	0.86	63.4
EAS2	**10**	**1**	**500**	**11**	**92.0**	**0.36**	**0.80**	**0.98**	**0.50**	**0.36**	**0.86**	**44.7**
EAS2	4	4	1000	13	90.0	0.20	0.50	0.97	0.28	0.20	0.85	77.8
EAS2VTanh	10	1	500	11	88.0	0.11	0.20	0.95	0.14	0.11	0.84	44.1
EAS2VTanh	4	3	1000	8	88.0	0.20	0.33	0.95	0.25	0.20	0.82	135.8
EAS	10	1	1000	13	89.0	0.10	0.33	0.97	0.15	0.10	0.85	145.5

EAS2VTanh	4	2	500	9	88.1	0.18	0.40	0.25	0.25	0.18	0.82	67.1

4.2.3 Comparison of EAS-MLPNN with other classifiers

Table 5 illustrates the experimental results of our proposed model compared to traditional MLP, SVM, and RFC with feature selection. Notice the proposed model outperforms other classifiers in all evaluation metrics. In detail, SVM achieved 93.3% accuracy and 0.217 AROC; MLP achieved 92.6% accuracy and 0.176 AROC; and RFC achieved 27% accuracy and 0 AROC. On the other hand, the proposed model achieved 94.67% accuracy and 0.294 AROC.

Table 5: Comparison of classifier performances

classifiers	Accuracy	Recall	Precision	Specificity	F1-Score	AROC	Geomean	Time (s)
SVM	93.3	0.217	0.714	0.993	0.333	0.217	0.888	0.87
RFC	27.0	0.000	0.000	0.286	Nan	0.000	0.205	0.20
MLPNN	92.6	0.176	0.273	0.972	0.214	0.176	0.901	69.54
EAS-MLPNN	94.6	0.294	0.556	0.986	0.385	0.294	71.397	0.92

In addition, using FS to predict CKD, notice the proposed model outperforms the MLPNN. In detail, the proposed model achieved 94.67% accuracy when epochs = 1000, while MLP achieved 92.6% accuracy when epochs = 500. Therefore, we conclude that the accuracy increases over selecting the most relevant features in the proposed model. Based on AROC, we can notice the proposed model outperforms the MLPNN, where it achieved 0.294 when epochs = 1000, while MLPNN achieved 0.176 when epochs = 500. It is concluded that EAS has significantly improved the classification accuracy of MLPNN. Therefore, Table 6 shows the remaining metrics enhancement percentages. The improvement rate on the accuracy was not noticeable, but there was an improvement percentage on the AROC and F1-Score. The reason behind this is the accuracy alone does not reflect the model effectiveness, so other performance measurement was used in this study such as Geomean.

Table 6: Enhancement ratio of the proposed model

Model	Accuracy	AROC	F1-Score	Geomean
MLPNN	92.6	0.176	0.214	0.901
EAS-MLPNN	94.67	0.294	0.385	0.028
Enhancement	0.021	0.670	0.799	78.241

The outcomes of this study proved that the proposed model outperformed the MLPNN, SVM, and RFC in all evaluation metrics. The EAS has proven its ability to select best features while reducing their number. As a result, the proposed model facilitates understanding the dataset because it reduces the number of features.

5 Conclusion

The study used the EAS-MLPNN to predict CKD, emphasized the importance of data preprocessing, and feature selection in analyzing CKD data. EAS is a feature selection algorithm that selects relevant features to enhance accuracy. The study aimed to select the most relevant features needed for MLPNN and compared the proposed model with other

classifiers. The results show that the proposed model outperforms others in terms of accuracy, AROC, recall, precision, specificity, F1-score, geomean, execution time, and total number of selected features. Additionally, there is an enhancement in accuracy and AROC compared to MLPNN without feature selection. Future work suggests combining MLP with other optimization algorithms to identify effective features for further improvement of kidney disease prediction models.

References

[1] Aydin, Z. E., & Ozturk, Z. K. (2021). XGBoost Feature Selection on Chronic Kidney Disease Diagnosis. *International Conference on Data Science and Applications (ICONDATA'20)*, June 4-6, 2021, Turkey.

[2] Majid, M., Gulzar, Y., Ayoub, S., Khan, F., Reegu, F. A., Mir, M. S., & Soomro, A. B. (2023). Using ensemble learning and advanced data mining techniques to improve the diagnosis of chronic kidney disease. International Journal of Advanced Computer Science and Applications, 14(10), 470-480.

[3] Jaradat, G. M., & Ayob, M. (2010). An elitist-ant system for solving the post-enrolment course timetabling problem. *In Database theory and application, bio-science and bio-technology* (pp. 167-176). Springer, Berlin, Heidelberg.

[4] Parthiban, R., Usharani, S., Saravanan, D., Jayakumar, D., Palani, D. U., StalinDavid, D. D., & Raghuraman, D. (2021). Prognosis of chronic kidney disease (CKD) using hybrid filter wrapper embedded feature selection method. European Journal of Molecular & Clinical Medicine, 7(9), 2511-30.

[5] Singam, A. K., & Shahid, M. (2023). A just resembling translations of feed forward and back propagation neural networks. Authorea Preprints.

[6] Alzoubi, S., Aldiabat, K., Al-diabat, M., & Abualigah, L. (2024). An extensive analysis of several methods for classifying unbalanced datasets. Journal of Autonomous Intelligence, 7(3).

[7] Salekin, A., & Stankovic, J. (2016). Detection of chronic kidney disease and selecting important predictive attributes. *In 2016 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 262-270). IEEE.

[8] Alasker, H., Alharkan, S., Alharkan, W., Zaki, A., & Riza, L. S. (2017). Detection of kidney disease using various intelligent classifiers. *In 2017 3rd international conference on science in information technology (ICSITech)* (pp. 681-684). IEEE.

[9] Polat, H., Danaei Mehr, H., & Cetin, A. (2017). Diagnosis of chronic kidney disease based on support vector machine by feature selection methods. Journal of medical systems, 41(4), 1-11.

[10] Pasadana, I. A., Hartama, D., Zarlis, M., Sianipar, A. S., Munandar, A., Baeha, S., & Alam, A. R. M. (2019). Chronic kidney disease prediction by using different decision tree techniques. *In Journal of Physics: Conference Series* (Vol. 1255, No. 1, p. 012024). IOP Publishing.

[11] Almansour, N. A., Syed, H. F., Khayat, N. R., Altheeb, R. K., Juri, R. E., Alhiyafi, J., & Olatunji, S. O. (2019). Neural network and support vector machine for the prediction

of chronic kidney disease: A comparative study. Computers in biology and medicine, 109, 101-111.

[12] Alaiad, A., Najadat, H., Mohsen, B., & Balhaf, K. (2020). Classification and association rule mining technique for predicting chronic kidney disease. Journal of Information & Knowledge Management, 19(01), 2040015.

[13] Elhoseny, M., Shankar, K., & Uthayakumar, J. (2020). Intelligent Diagnostic Prediction and Classification System for Chronic Kidney Disease (vol 9, 9583, 2019). SCIENTIFIC REPORTS, 10(1).

[14] Hamedan, F., Orooji, A., Sanadgol, H., & Sheikhtaheri, A. (2020). Clinical decision support system to predict chronic kidney disease: A fuzzy expert system approach. International journal of medical informatics, 138, 104134.

[15] Jain, D., & Singh, V. (2021). A two-phase hybrid approach using feature selection and adaptive SVM for chronic disease classification. International Journal of Computers and Applications, 43(6), 524-536.

[16] Krishnamurthy, S., Ks, K., Dovgan, E., Luštrek, M., Gradišek Piletič, B., Srinivasan, K., & Syed-Abdul, S. (2021). Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan. In Healthcare, 9(5): 546. Multidisciplinary Digital Publishing Institute.

[17] Mazaherinezhad, A., & Rawassizadeh, R. (2021). A diagnostic prediction model for chronic kidney disease in internet of things platform. Multimedia Tools and Applications, 80(11), 16933-16950.

[18] Senan, E. M., Al-Adhaileh, M. H., Alsaade, F. W., Aldhyani, T. H., Alqarni, A. A., Alsharif, N., & Alzahrani, M. Y. (2021). Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques. Journal of Healthcare Engineering, 2021.

[19] Gupta, R., Koli, N., Mahor, N., & Tejashri, N. (2020). Performance Analysis of Machine Learning Classifier for Predicting Chronic Kidney Disease. *In 2020 International Conference for Emerging Technology (INCET)* (pp. 1-4). IEEE.

[20] Jerlin Rubini, L., & Perumal, E. (2020). Efficient classification of chronic kidney disease by using multi-kernel support vector machine and fruit fly optimization algorithm. International Journal of Imaging Systems and Technology, 30(3), 660-673.

[21] Ogunleye, A., & Wang, Q. G. (2019). XGBoost model for chronic kidney disease diagnosis. IEEE/ACM transactions on computational biology and bioinformatics, 17(6), 2131-2140.

[22] Shankar, K., Manickam, P., Devika, G., & Ilayaraja, M. (2018, December). Optimal feature selection for chronic kidney disease classification using deep learning classifier. *In 2018 IEEE international conference on computational intelligence and computing research (ICCIC)* (pp. 1-5). IEEE.

[23] Revuelta, J., Maydeu-Olivares, A., & Ximénez, C. (2020). Factor analysis for nominal (first choice) data. Structural Equation Modeling: A Multidisciplinary Journal, 27(5), 781-797.

[24] Mahdi, R. O., & Hacham, W. S. (2024). Applying Scikit-Learn of Machine Learning To Predict Consumed Energy in Al-Khwarizmi College of Engineering, Baghdad, Iraq. Al-Khwarizmi Engineering Journal, 20(1), 76-88.

[25] Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. Applied Soft Computing, 97, 105524.

[26] Ali, P. J. M., Faraj, R. H., Koya, E., Ali, P. J. M., & Faraj, R. H. (2014). Data normalization and standardization: a technical report. Mach Learn Tech Rep, 1(1), 1-6.

[27] Kashef, S., & ElShaer, R. (2021). A Review of Implementing Ant System Algorithms on Scheduling Problems. Egyptian Journal for Engineering Sciences and Technology, 36(2), 43-52.

[28] David, M. W. (2020). Powers. *Evaluation: from precision, recall and f-measure to roc, informedness, markedness and* correlation. 2 (1) (2011) 37–63.

[29] Al-Suwaihel , O. (2023). Teacher's Role in Attaining the Pillars of Digital Citizenship in Teaching Students at Government Shools in The State of Kuwait. Journal of Education /Al Mejlh Altrbwyh, 37(146), 47-79. doi:10.34120/0085-037-146-003