Bulletin of Electrical Engineering and Informatics
Vol. 11, No. 6, December 2022, pp. 3590~3597
ISSN: 2302-9285, DOI: 10.11591/eei.v11i6.4269

A proposed approach for diabetes diagnosis using neuro-fuzzy technique

Maher Talal Alasaady[1,2], Teh Noranis Mohd Aris[1], Nurfadhlina Mohd Sharef[1], Hazlina Hamdan[1]
[1]Department of Computer Science, Faculty of Computer Sciences and Information Technology, Universiti Putra Malaysia, Selangor, Malaysia
[2]Department Computer Systems, Northern Technical University, Mosul, Iraq

Article Info

Article history:

Received Jun 16, 2022
Revised Jul 27, 2022
Accepted Aug 05, 2022

Keywords:

ANFIS
Clinical decision support
Diabetes diagnosis
Fuzzy inference system
Neuro-fuzzy inference system
Pima Indians

ABSTRACT

Diabetes is a chronic disease characterized by a decrease in pancreatic insulin production. The immune system will be harmed due to this condition, which will raise blood sugar levels. However, early detection of diabetes enables patients to begin treatment on time, therefore reducing or eliminating the risk of severe consequences. One of the most significant challenges in the healthcare unit is disease diagnosis. Traditional techniques of disease diagnosis are manual and prone to inaccuracy. This paper proposed an approach for diagnosing diabetes using the adaptive neuro-fuzzy inference system (ANFIS) based on Pima Indians diabetes dataset (PIDD). The three stages of the proposed approach are pre-processing classification and evaluation. Normalization, imputation, and anomaly detection are part of the pre-processing stage. The pre-processing was done by normalizing the data, replacing the missing values, and using the local outlier factor (LOF) technique. In the classification stage, ANFIS classifiers were trained using the hybrid learning algorithm of the neural network. Finally, the evaluation procedures use the last stage's sensitivity, specificity, and accuracy metrics. The obtained classification accuracy was 92.77%, and it seemed rather promising compared to the other classification applications for this topic found in the literature.

Corresponding Author:

Teh Noranis Mohd Aris
Faculty of Computer Sciences and Information Technology, Universiti Putra Malaysia
43400 UPM Darul Ehsan, Serdang, Selangor, Malaysia
Email: nuranis@upm.edu.my

1. INTRODUCTION

Diabetes mellitus (DM) is a chronic disease defined by an insufficient amount of insulin produced by the pancreas. Insulin is a hormone that maintains a healthy blood sugar balance. If insulin does not function properly, it will cause a rise in the blood glucose level [1]. Diabetes has several consequences, including an increased risk of blindness, high blood pressure, renal damage, and heart disease [2]. The World Health Organization (WHO) studied that over 37 crore people worldwide presently suffer from this chronic condition, which is anticipated to double by 2030 [3]. Recent advancements in the healthcare industry have improved diabetes early detection; however, over half of diabetic individuals are ignorant of their condition. It might take ten years or more to diagnose them. As a result, an expert system capable of processing imprecise and unclear data, as is the situation in the medical area, is required for early diabetes diagnosis.

One of the most significant challenges in the healthcare unit is disease diagnosis. Traditional disease diagnosis techniques are manual and prone to inaccuracy [4]. Compared to human competence alone, prediction approaches based on artificial intelligence (AI) allow auto-diagnosis and decrease the detection of mistakes [5]. Diabetes prediction algorithms often encounter noisy, missing, irrelevant, and inconsistent data [6], [7]. The model's effectiveness is determined by the accuracy of the diabetic data provided; as a result, the researcher must submit credible data to the classifier for the disease to be predicted accurately [8]. In the medical area, fuzzy logic (FL) algorithms are well-suited to handle ambiguity and uncertainty in large datasets that are ideal for decision-making in diabetes diagnosis [9]. FL's adjustment constraint during the learning process was solved by introducing adaptive neuro-fuzzy inference system (ANFIS), which combines the benefits of fuzzy control interpolation with adaptability through neural network backpropagation [10].

Numerous studies have predicted diabetes using a variety of diabetic datasets utilizing machine learning (ML), FL, and hybrid techniques. We included only studies based on the PIDD. One of the works in which researchers used fuzzy logic for diabetes diagnosis is Aamir *et al.* [11]. They used FL to create a model with two fuzzy classifiers: fuzzy logic and the cosine amplitude approach. Manikandan [12] suggested a model using the FL and grey wolf optimization (GWO) method. Thungrut and Wattanapongsakorn [13] describe two fuzzy and genetic algorithms were used to increase classification accuracy.

Regarding the works in which researchers used ANFIS technology, an ANFIS classification was used by Priyadarshini *et al.* [14]. They used 240 rules generated in MATLAB. Alby *et al.* [15] developed a method using ANFIS with GA. Kalaiselvi *et al.* [16] proposed ANFIS-based k-nearest neighborhood (KNN).

Regarding how researchers used ML techniques in diagnosing diabetes, Khanam and Foo [5] examined multiple ML approaches, such as support vector machines (SVM), random forest (RF), KNN, and artificial neural network (ANN). Pradhan *et al.* [17] proposed an ANN algorithm to identify diabetes. Table 1 illustrates the comparison of related work with the accuracy that achieved.

This study aims to detect type 2 diabetes in its earliest stages to ensure that patients get quick treatment and avoid serious consequences. Additionally, it is aimed to achieve a high degree of classification accuracy. The study proposes a neuro-fuzzy approach to predict diabetes with well-known PIDD data. The rest of this paper is as follows: section 2 outlines the research method. Section 3 highlights the findings acquired using our proposed approach. Finally, section 4 concludes this work.

Table 1. Comparison of classifiers from the literature

References	Year	Method	Accuracy (%)
[12]	2019	Fuzzy rules & GWO	81
[13]	2018	Fuzzy rules & GA	87
[14]	2020	ANFIS	86
[15]	2018	ANFIS + GA	96
[18]	2017	ANFIS	85
[16]	2014	ANFIS + KNN	80
[5]	2021	ANN	88
[17]	2020	ANN	86
[19]	2019	DL	86
[20]	2018	NB + DT + SVM	76

2. METHOD

The approach consisted of three stages: Pre-processing, classification, and evaluation. First, normalization, imputation, and anomaly detection are steps of pre-processing. Next, the ANFIS was used with backpropagation (BP) and least square estimation (LSE) estimation in the classification stage to get the best fuzzy rules possible. Finally, the evaluation stage uses accuracy, sensitivity, and specificity criteria. The proposed approach architecture and operations carried out by each architecture component during diagnosis are shown in Figure 1. The following subsections outline the suggested strategy in further depth.

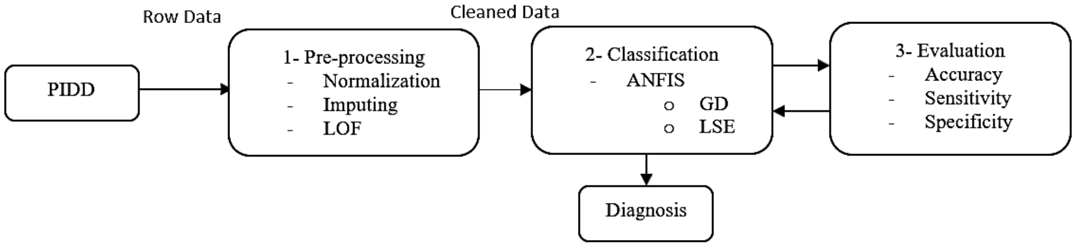


Figure 1. The proposed approach architecture

2.1. Dataset

The dataset most commonly used to compare diabetes diagnosis algorithms was obtained from [21]. The data set included a total of 768 instances. The dataset is divided into two groups, denoted by the codes 1 and 0 for diabetes and healthy. There are 268 instances in class 1 (34.9%) and 500 occurrences in class 0 (65.1%). There are eight continuous features: (1) Number of pregnancies. (2) Plasma glucose concentration measured over two hours during an oral glucose tolerance test. (3) Blood pressure in the diastole (mm Hg). (4) Thickness of the triceps skinfold (mm). (5) 2-hour serum insulin concentration (mu U/ml). (6) Body mass index (BMI). (7) Diabetes pedigree function (8) Age (years).

2.2. Pre-processing stage

Data pre-processing enables the generation of a robust classification model with high accuracy [22]. At this stage, some initial operations are performed on the PIDD to improve classification precision.

2.2.1. Normalization

Normalization is a widely used data preparation method that enables the values of numeric columns in a dataset to be converted to a standard scale [23]. Because the range of different attributes in PIDD varies, one of the pre-processing steps is to normalize attributes such that their normalized values lie within the range of (0,1). This processing allows for more accurate comparisons. We employed min-max scalar (MMS) [24] as the normalization model in our proposed approach. The MMS is shown in (1):

$$x_{Scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

2.2.2. Imputation

Imputation of missing data is a process that substitutes probable values for the missing value [25]. The different imputation approaches are designed to generate reliable estimates of population parameters [26]. The quantity of missing data determines the optimum approach for missing data [27]. Although there is no rule for what percentage of data is unacceptable, comparing findings before and after imputation is usually preferable when more than 25% of data is missing [28]. Mean imputation is the most often used technique for replacing missing data.

2.2.3. Anomaly detection

Anomalies are patterns within a dataset that deviate from well-defined and expected behavior [29]. Anomaly detection is the process of identifying these patterns [30]. These anomalies often have a negative impact on classification accuracy [31]. The local outlier factor (LOF) is a practical unsupervised machine learning approach that finds outliers in local regions rather than the overall data distribution [32]. After calculating outlier ratings for each data point, the data points may be sorted to identify outliers in the dataset. In (2) expresses the relative density of a data point X with k neighbors:

$$Relative\ density\ of\ X = \frac{Density\ of\ X}{Average\ density\ of\ all\ data\ points\ in\ the\ neighborhood} \quad (2)$$

X density is proportional to the average distance between the k-nearest data points [33].

2.3. Classification stage

Due to the fuzziness of the dataset's properties and the benefits of the ANFIS approach, ANFIS was employed for classification allowing for further modification or optimization. ANFIS was proposed by Jung [34]. It combines ANN and fuzzy inference system (FIS), where ANN algorithms determine the fuzzy system's parameters and expressively model uncertainty. The FL algorithm maps each parameter in the dataset to linguistic labels using a membership function (MF). This operation is utilized to trace incoming data to output data during computational analyzes of the ANN component [35], [36]. ANFIS determines parameters using a hybrid learning strategy that merges BP and LSE [37]. The ANFIS architecture could be described as a five-layer neural network. The input layer for our proposed approach comprises eight attributes, each with three MFs, and the fuzzy set was (low, medium, and high), resulting in layer one having 24 nodes. ANFIS structure is shown in Figure 2 [34], whereas Figure 3 illustrates its structure for diabetes classification.

a. The first layer (IF part)

All nodes in this layer are represented by adaptive nodes, representing the function of belonging to the system's entry. Each node has the following activation functions:

$$O_{1,i} = \mu_{A_i}(X_1), for\ i = 1,2 \quad (3)$$

$$O_{1,i} = \mu_{B_{i-2}}(X_2), for\ i = 3,4 \quad (4)$$

X_1, X_2 are the input nodes i, and A, B are the linguistic labels for this node, and the MFs are $\mu(X_1)$ and $\mu(X_2)$.
b. The second layer (Rules)
In this layer, the output of nodes displays the firing strength of the rule and is symbolized by (π). The result is indicated in (5):

$$O_{2,i} = W_i = \mu_{A_i}(X_1).\ \mu_{B_i}(X_2).\ \mu_{C_i}(X_3).\ \mu_{D_i}(X_4).\ \mu_{E_i}(X_5).\ \mu_{F_i}(X_6).\ \mu_{G_i}(X_7).\ \mu_{H_i}(X_8) \quad (5)$$

c. The third layer (Norm)
The nodes of the third layer are also fixed and symbolized by (N), and the nodes represent the normalized firing strength of each rule as (6):

$$O_{3,i} = \bar{W} = \frac{W_i}{\sum_{i=1}^{8} W_i}, for\ i = 1,2 \ldots 8 \quad (6)$$

d. The fourth layer (Then part)
Nodes in this layer are adaptive, and a node function displays the contribution of the rules toward the total output. The nodes multiply the normalized weight of each fuzzy rule by the latter part of that rule.

$$O_{4,i} = \bar{W}_i . f_i = \bar{W}_i \,(p_i X_1 + q_i X_2 + \cdots + r_i) \quad (7)$$

p_i, q_i, and r_i are specified parameters known as consequent parameters.

e. The fifth layer (Output)
This layer consists of a single circular (fixed) node, symbolized by (∑). This node handles all output values from the fourth layer nodes and delivers them into the network as (8):

$$O_{5,i} = \sum_{i=1}^{n} \bar{W}_i . f_i = \frac{\sum_{i=1}^{n} W_i . f_i}{\sum_{i=1}^{n} W_i} \quad (8)$$

2.4 Evaluation stage

The proposed approach results were evaluated using sensitivity, specificity, and accuracy. These statistical measurements reveal a test's essential reliability in the medical diagnostic test [38]. Sensitivity shows the ability to separate positive patients from all patients in a test. Specificity is an ability of the separates to be real sturdy from within the all sturdy. Accuracy identified the diagnostic test by excluding a specific condition. To calculate these metrics, we first compute certain concepts such as false negative (FN), true positive (TP), false positive (FP), and true negative (TN) using the definitions in Table 2.

Table 2. Sensitivity, specificity, and accuracy concepts [39]

Outcome	Positive	Negative	Row total
Positive	TP	FP	TP + FP (total number of subjects with a positive test)
Negative	FN	TN	FN + TN (total number of subjects with negative tests)
Column total	TP + FN (total number of subjects with given condition)	FP + TN (total number of subjects without given condition)	N = TP + TN + FP + FN (Total number of subjects in study)

3. RESULTS AND DISCUSSION

The dataset was first normalized, implying that each value falls between (0 and 1). The MMS technique is used for normalization in equation (1). This adjustment helps mitigate the adverse effects of specific attributes prevailing, particularly undesirable ones due to their more extensive value ranges. Next, data imputation is applied, as many features have zero value (For example, the lowest Blood Pressure value is 0 (which is not conceivable)). As a result, erroneous information is provided. We can replace such values with medians since we cannot discard such observations (which results in a 4.6 percent data loss). We have chosen to impute as we have a small dataset (768 observations). We should note that some features must not be imputed (i.e., Pregnant) because the zero value here does not represent a missed value. Finally, anomaly detection is performed using the LOF technique, and (68) outlier values are removed from PIDD after applying this technique.

The ANFIS classifiers were implemented at the classification stage using the MATLAB script. The dataset was randomly divided into two parts: The first is training part contains (470) records which are 67% of the dataset, and the testing part contains (230) records which are 33% of the dataset since the total number of the dataset is (700) after removing the outliers. The training data is used to train the model, while the testing data is used to verify the trained model's accuracy and effectiveness. The parameters of FIS are tuned using a hybrid optimization learning method. This strategy trains the model using LSE and BP techniques.

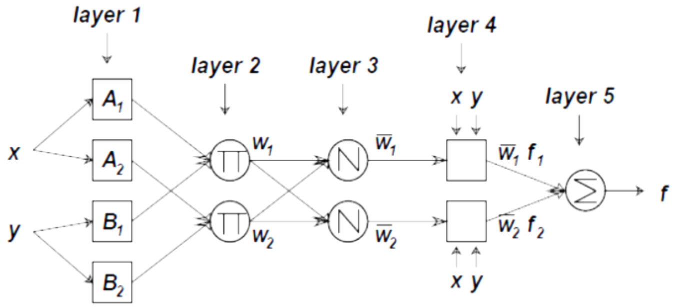


Figure 2. ANFIS general structure

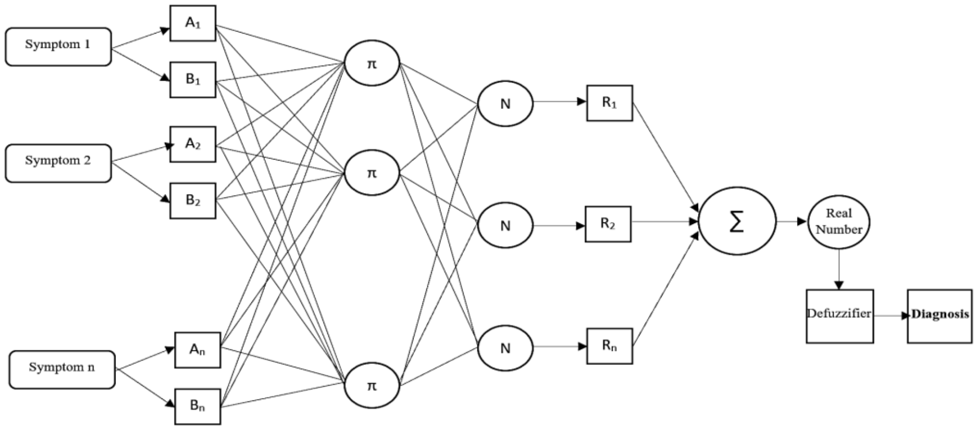


Figure 3. ANFIS architecture for diabetes diagnosis

Root Mean Square Error (RMSE) values were calculated using equation (9), where *n* denotes the number of data points, *p* represents the predicted value, and *o* represents the actual value, to verify the model's performance during both training and testing.

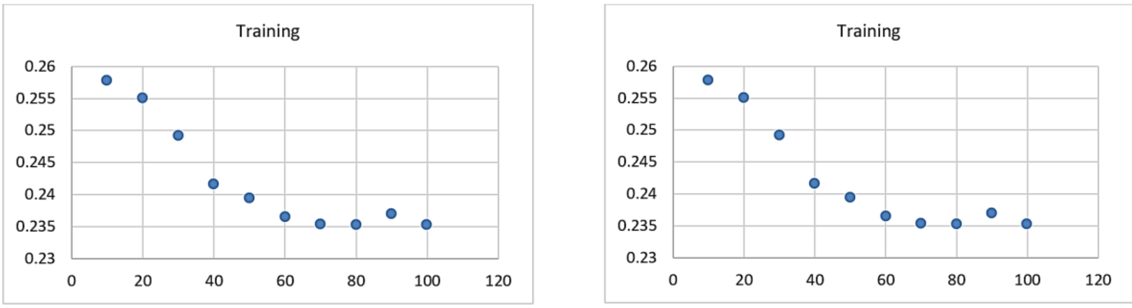
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2} \tag{9}$$

In order to identify the best number of epochs for training the FIS, we utilized the model complexity graph. The early epochs demonstrate a reduction in training and testing error, implying that training should be continued. When the error in the testing increases or decreases, the training should be interrupted, this is the ideal trade-off between epochs and model complexity. We began with (100) epochs with an error tolerance of (0.01) and gradually decreased to the optimal (70) epochs. Figure 4(a) depict the complexity graph for training, while Figure 4(b) depict the complexity graph for testing. Both training and testing processes were repeated for different epochs, as shown in Table 3.

The evaluation stage determines the classifier's test performance by computing its sensitivity, specificity, and accuracy. The results of the statistical of the ANFIS model were sensitivity (97.65%), specificity (84.30%), and accuracy (92.77%). To demonstrate the efficacy of the proposed approach strategy, we generated classification models on the same dataset using another five popular learning classifiers and compared them with the proposed model; the validation results are illustrated in Table 4 and Figure 5. The proposed approach outperforms all widely used learning classifiers as shown in Table 1 to compare these classifiers with our method.

Table 3. Performance of ANFIS during epochs

During training		During testing	
Epoch	RMSE	Epoch	RMSE
20	0.255062	20	0.259686
40	0.241584	40	0.251125
60	0.236524	60	0.240353
80	0.235229	80	0.232319
100	0.235232	100	0.229453



(a) (b)

Figure 4. Complexity graph for training and testing data, (a) training graph and (b) testing graph

Table 4. Comparison of accuracies with other standard classifiers

Classifier	Sensitivity (%)	Specificity (%)	Accuracy (%)
Neural network	82	74	75
Naïve Bayes	81	75	74
KNN	73	70	71
Decision tree	64	70	70
Random forest	78	72	73
Proposed approach	97	84	92

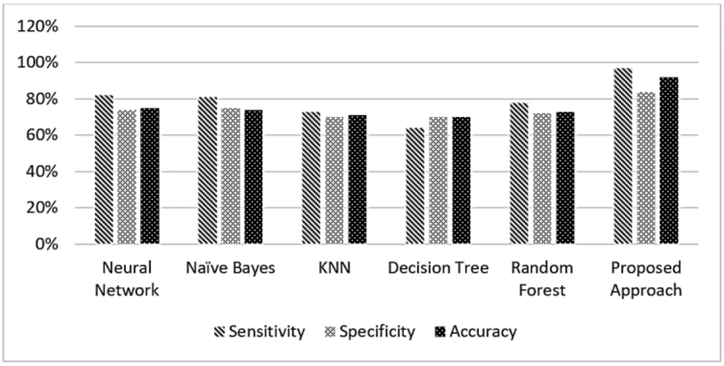


Figure 5. Graphical comparison of accuracies with other common classifiers

4. CONCLUSION

This research proposes a multistage classification approach to acquire accurate results using ANFIS for classifying patients with diabetes based on PIDD. The pre-processing was done by normalizing the data, replacing the missing values, and using the LOF technique to eliminate data anomalies. ANFIS classifiers were implemented in the MATLAB script. Finally, the evaluation stage uses accuracy, sensitivity, and specificity criteria. The results of the statistical of the ANFIS model were sensitivity (97.65%), specificity (84.30%), and accuracy (92.77%). There is still exciting work ahead; this includes: Adding an optimization process, using deep learning algorithms, and using classification for other diseases.

REFERENCES

[1] D. Mellitus, "Diagnosis and classification of diabetes mellitus," *Diabetes Care*, vol. 28, no. S37, pp. S5–S10, 2005, doi: 10.2337/DIACARE.28.SUPPL_1.S37.

[2] Centers for Disease Control and Prevention, "National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011," *Atlanta, GA US Dep. Heal. Hum. Serv. centers Dis. Control Prev.*, vol. 201, no. 1, pp. 2568–2569, 2011.

[3] WHO, "World Diabetes Statistics, Geneva, Switzerland," 2014. http://www.who.int/diabetes/en/index.html

[4] R. Krishnamoorthi *et al.*, "A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques," *J. Healthc. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/1684017.

[5] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, 2021, doi: 10.1016/j.icte.2021.02.004.

[6] G. Swapna, S. Kp, and R. Vinayakumar, "Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals," *Procedia Comput. Sci.*, vol. 132, pp. 1253–1262, 2018, doi: 10.1016/j.procs.2018.05.041.

[7] A. Mohebbi, T. B. Aradottir, A. R. Johansen, H. Bengtsson, M. Fraccaro, and M. Mørup, "A deep learning approach to adherence detection for type 2 diabetics," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 2896–2899.

[8] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Predicting healthcare trajectories from medical records: A deep learning approach," *J. Biomed. Inform.*, vol. 69, pp. 218–229, 2017, doi: 10.1016/j.jbi.2017.04.001.

[9] A. M. Ahmed, A. H. Ahmed, R. W. Daoud, and O. K. Ahmed, "Optimization of Simple Solar Still Performance Using Fuzzy Logic Control," in *2020 6th International Engineering Conference "Sustainable Technology and Development" (IEC)*, 2020, pp. 205–210. doi: 10.1109/IEC49899.2020.9122819.

[10] A. N. H. Raid R. A. -Nima, F. S. Abdullah, "Design a Technology Based on the Fusion of Genetic Algorithm, Neural network and Fuzzy logic," *arXiv*, vol. 2102.08035, 2021, doi: 10.48550/arXiv.2102.08035.

[11] K. M. Aamir, L. Sarfraz, M. Ramzan, M. Bilal, J. Shafi, and M. Attique, "A fuzzy rule-based system for classification of diabetes," *Sensors*, vol. 21, no. 23, 2021, doi: 10.3390/s21238095.

[12] K. Manikandan, "Diagnosis of diabetes diseases using optimized fuzzy rule set by grey wolf optimization," *Pattern Recognit. Lett.*, vol. 125, pp. 432–438, 2019, doi: 10.1016/j.patrec.2019.06.005.

[13] W. Thungrut and N. Wattanapongsakorn, "Diabetes classification with fuzzy genetic algorithm," in *International Conference on Computing and Information Technology*, 2018, pp. 107–114.

[14] L. Priyadarshini and L. Shrinivasan, "Design of an ANFIS based Decision Support System for Diabetes Diagnosis," *Proc. 2020 IEEE Int. Conf. Commun. Signal Process. ICCSP 2020*, pp. 1486–1489, 2020, doi: 10.1109/ICCSP48568.2020.9182163.

[15] S. Alby and B. L. Shivakumar, "A prediction model for type 2 diabetes using adaptive neuro-fuzzy interface system," *Biomed. Res.*, vol. 2018, no. Special Issue ComputationalLifeSciencesandSmarterTechnologicalAdvancement, pp. S69–S74, 2018, doi: 10.4066/biomedicalresearch.29-17-254.

[16] C. Kalaiselvi and G. M. Nasira, "A new approach for diagnosis of diabetes and prediction of cancer using ANFIS," in *2014 World Congress on Computing and Communication Technologies*, 2014, pp. 188–190.

[17] N. Pradhan, G. Rani, V. S. Dhaka, and R. C. Poonia, "Diabetes prediction using artificial neural network," in *Deep Learning Techniques for Biomedical and Health Informatics*, Elsevier, 2020, pp. 327–339. doi: 10.1016/B978-0-12-819061-6.00014-8.

[18] O. Geman, I. Chiuchisan, and R. Toderean, "Application of Adaptive Neuro-Fuzzy Inference System for diabetes classification and prediction," in *2017 E-health and bioengineering conference (EHB)*, 2017, pp. 639–642.

[19] K. Kannadasan, D. R. Edla, and V. Kuppili, "Type 2 diabetes data classification using stacked autoencoders in deep neural networks," *Clin. Epidemiol. Glob. Heal.*, vol. 7, no. 4, pp. 530–535, 2019, doi: 10.1016/j.cegh.2018.12.004.

[20] D. Sisodia and D. S. Sisodia, "Prediction of diabetes using classification algorithms," *Procedia Comput. Sci.*, vol. 132, pp. 1578–1585, 2018.

[21] R. S. Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., & Johannes, "Pima Indians Diabetes Dataset," 1988. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

[22] C. C. Olisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective," *Comput. Methods Programs Biomed.*, vol. 220, p. 106773, 2022, doi: 10.1016/j.cmpb.2022.106773.

[23] M. T. Alasaady, M. G. Saeed, and K. H. Faraj, "Evaluation and Comparison Framework for Data Modeling Languages," in *2019 2nd International Conference on Electrical, Communication, Computer, Power and Control Engineering (ICECCPCE)*, 2019, pp. 68–73.

[24] T. Phaladisailoed and T. Numnonda, "Machine learning models comparison for bitcoin price prediction," in *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, 2018, pp. 506–511.

[25] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976, doi: 10.2307/2335739.

[26] B. A. Othman and S. Mustafa, "Predict the Risk Level in Iraqi Governorates According to the Spread of COVID-19 Using Data Mining Long Short-Term Memory," *NTU J. Pure Sci.*, vol. 1, no. 2, pp. 22–28, 2022, [Online]. Available: https://journals.ntu.edu.iq/index.php/NTU-JPS/article/view/203

[27] A. Q. Saeed, S. N. H. S. Abdullah, J. C. Hamzah, and A. T. A. Ghani, "Accuracy of Using Generative Adversarial Networks for Glaucoma Detection: Systematic Review and Bibliometric Analysis.," *J. Med. Internet Res.*, vol. 23, no. 9, p. e27414, Sep. 2021, doi: 10.2196/27414.

[28] A. Jadhav, D. Pramod, and K. Ramanathan, "Comparison of Performance of Data Imputation Methods for Numeric Dataset," *Appl. Artif. Intell.*, vol. 33, no. 10, pp. 913–933, Aug. 2019, doi: 10.1080/08839514.2019.1637138.

[29] R. R. Saeed, O. M. Yaseen, M. M. Rashid, and M. R. Ahmed, "Applications of Machine Learning in Battling Against Novel COVID-19," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022, pp. 1–6. doi: 10.1109/HORA55278.2022.9799969.

[30] M. L. Shahreza, D. Moazzami, B. Moshiri, and M. R. Delavar, "Anomaly detection using a self-organizing map and particle swarm optimization," *Sci. Iran.*, vol. 18, no. 6, pp. 1460–1468, 2011, doi: 10.1016/j.scient.2011.08.025.

[31] S. Fathi, M. Ahmadi, B. Birashk, and A. Dehnad, "Development and use of a clinical decision support system for the diagnosis of social anxiety disorder," *Comput. Methods Programs Biomed.*, vol. 190, no. 4, p. 105354, 2020, doi: 10.1016/j.cmpb.2020.105354.

[32] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.

[33] V. Kotu and B. Deshpande, "Chapter 13 - Anomaly Detection," in *Data Science (Second Edition)*, Second Edi., V. Kotu and B.

Deshpande, Eds. Morgan Kaufmann, 2019, pp. 447–465. doi: 10.1016/B978-0-12-814761-0.00013-7.
[34] J. R. Jang, "ANFIS : Adap tive-Ne twork-Based Fuzzy Inference System," vol. 23, no. 3, 1993.
[35] L. Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm," *Int. J. Biol. Med. Sci.*, vol. 3, no. 3, pp. 157–160, 2008, doi: 10.5281/zenodo.1082439.
[36] S. U. Ghumbre and A. A. Ghatol, "An intelligent system for hepatitis b disease diagnosis," *Int. J. Comput. Appl.*, vol. 32, no. 4, pp. 455–460, 2010, doi: 10.2316/Journal.202.2010.4.202-2874.
[37] A. Mellit, S. A. Kalogirou, S. Shaari, H. Salhi, and A. Hadj Arab, "Methodology for predicting sequences of mean monthly clearness index and daily solar radiation data in remote areas: Application for sizing a stand-alone PV system," *Renew. Energy*, vol. 33, no. 7, pp. 1570–1590, 2008, doi: 10.1016/j.renene.2007.08.006.
[38] A. Baratloo, M. Hosseini, A. Negida, and G. El Ashal, "Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity.," *Emerg. (Tehran, Iran)*, vol. 3, no. 2, pp. 48–49, 2015.
[39] P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 24, no. 1, pp. 27–40, 2012, doi: 10.1016/j.jksuci.2011.09.002.

BIOGRAPHIES OF AUTHORS

Maher Talal Alasaady received his Bachelor Degree in Software Engineering from the University of Mosul in 2007. He received his Master of Science in Software Engineering from the University of Mosul in 2012 and is currently a Ph.D. candidate in Artificial Intelligence from Universiti Putra Malaysia (UPM). Currently, he is a Lecturer at Technical Institute/Mosul, Northern Technical University (NTU), Mosul, Iraq. Former Media Department Manager, Former Head of Computer Systems Department, Technical Institute/Mosul. His research interests include artificial intelligence, machine learning, deep learning, software agent, and software engineering. He can be contacted at email: maher.alasaady@ntu.edu.iq.

Teh Noranis Mohd Aris received her Bachelor Degree in Information Technology (Hons.) from University Utara Malaysia (UUM), in 1994. She received her Master of Science in Artificial Intelligence from University Putra Malaysia (UPM) in 2002, and a Ph.D. Degree in Computer Science from University Kebangsaan Malaysia (UKM) in 2008. She is an Associate Professor at the Faculty of Computer Science and Information Technology, UPM. Her research interests include Artificial Intelligence, Software Agent, and Programming Science. Students are invited to join her research work in various intelligent computing areas such as image processing, forecasting, performance prediction, computer programming education, bio-informatics, etc. She can be contacted at email: nuranis@upm.edu.my.

Nurfadhlina Mohd Sharef is an Associate Professor at the Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM), and is currently serving as the Deputy Director (Innovation in Teaching and Learning) at the Centre for Academic Development (CADe) in UPM. She obtained her Ph.D. from the University of Bristol, the UK, in Intelligent Systems and MSc in Software Engineering from Universiti Putra Malaysia, Malaysia. Her bachelor study was on B. IT (Science and System Management). Dr. Fadhlina's main research interests are in artificial intelligence and data science, especially in solving sentiment analysis, question answering chatbot and recommendation system problems. She can be contacted at email: nurfadhlina@upm.edu.my.

Hazlina binti Hamdan PhD. (Artificial Intelligence), University of Nottingham, United Kingdom. Master of Computer Science (Artificial Intelligence), University of Malaya Kuala Lumpur, Malaysia. Bachelor of Science (Hons.) (Computer Science), Universiti Kebangsaan Malaysia-Universiti Teknologi MARA, Malaysia. Her research interest is in computational intelligent techniques and applications, primarily in the context of medical. My work to date is on utilizing a neural network and fuzzy inference systems to model the prediction of cancer patients. The main interest is to transfer the medical intelligent systems into clinical use by applying and evaluating the intelligent techniques to see the mechanisms of implementation. Machine learning in clustering, classification, etc., are other areas of my research interest. She can be contacted at email: hazlina@upm.edu.my.