



Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)



Neuro-Fuzzy System for Heart Failure Prediction

Wijdan A. Khaleel^a, Adala M. Chiad^b

a b University of Al Basra, Al Basra, Iraq, College of Computer Science and Information

a Email: wejdanedani@gmail.com,

b Email: adala.gyad@uobasrah.edu.iq.

ARTICLE INFO

Article history:

Received: 13 /09/2022

Revised form: 22/10/2022

Accepted : 26 /10/2022

Available online: 01/12/2022

Keywords:

Heart Failure, Feature Extraction, Neuro-Fuzzy System (NFS), Artificial Neural Network (ANN).

ABSTRACT

Heart failure is one of the dangerous heart diseases that infect humans and may cause death. This disease causes damage to the heart muscle, and it becomes unable to pump blood in the body as well as it should. Therefore, the condition of heart failure patients must be predicted as soon as possible in order to help the patients to live longer lives by offering appropriate therapy. Based on that, the aims of this paper are to use medical records to predict the state of a patient with heart failure if he/she will die or not, and extract the important features that have a direct effect on the patient's state. This paper used a dataset of 299 heart failure patients and applied the Neuro-Fuzzy systems (NFS) to this dataset. This prediction is made by testing each two feature together in the dataset and feeding it to the NFS system to determine its effect on the patients. In this paper, the accuracy and confusion matrix is used to evaluate the system's performance. The experimental results show that the system yielded 100% accuracy when the two-feature, serum creatinine and ejection fraction, are tested together, so it can be used alone to predict whether patients with heart failure will survive.

MSC..

<https://doi.org/10.29304/jqcm.2022.14.4.1084>

1. Introduction

Heart failure is one of the severe diseases that kills millions of people around the world [1]. According to the "Centers for Disease Control and Prevention", more than 6 million people in the United States suffer from heart failure. In addition, heart failure is not limited to adults but includes children. "Heart failure occurs when the heart muscle cannot be able to pump blood around the body properly" [2]. There are many causes that lead to heart

*Corresponding author

Email addresses:

Communicated by 'sub editor'

failure, such as ("Heart attack, High blood pressure (hypertension), Faulty heart valves, Cardiomyopathy (a disease of the heart muscle), Inherited heart problems (congenital heart problems), Abnormal heart rhythms (arrhythmias) and Viral infection in the heart muscle") [3].

Medical records are considered a very informative resource since they reveal the hidden relationships between patients' data. The diagnosis of heart failure disease (HF) is dependent on examinations, the interpretation of which is time-consuming and often limited by the level of clinical experience of doctors. With the development of computer science and the artificial intelligence (AI) technique and as well as the big data presence, technological innovations have occurred in the field of heart medicine, where AI has been used to assist precise classification of the state of the patients with heart failure by integrating a variety of medical data [4]. In this study, we try to identify the risk factors or the important features in the dataset. This process is done by taking each two features in the dataset and testing them together to determine their effect by monitoring the accuracy after testing every two features in the dataset together. The features with the highest accuracy are considered the most important features in the dataset. After that, we used these important features only and fed them to the Neuro-Fuzzy System (NFS) again to predict the patient's state. To our knowledge, this study is the first study in this field to predict the heart failure patient's state by using this data set with the NFS system.

Literature Review

Several studies have been presented in the field of heart failure prediction. Therefore, in this part, we will present some of these studies that focused on predicting heart failure using nuro-fuzzy system techniques and artificial neural networks.

The authors in [5] The authors in made predictions about the survival of heart failure patients based on the heart failure clinical health record. They used the ANFIS system and the Genetic Algorithm to improve the performance of the ANFIS classification by the Genetic Algorithm through the attribute selection process. The experimental result shows the highest ANFIS accuracy is produces 94.444%. While the ANFIS algorithm, after features selection using the Genetic Algorithm, has the highest accuracy of 96.667%. In [6] suggested a machine learning methodology that uses an artificial neural network (ANN) to predict heart failure. They also suggested a unique wrapper-based feature selection method using a grey wolf optimization (GWO) to cut down on the number of necessary input attributes. Computing findings demonstrate that far fewer features are required, and higher prediction accuracy of about 87% may be attained. The authors of [7] presents a model for the classification of medical diseases. The crisp inputs are fuzzified by using linguistic variables of the membership function to deal with imprecise input data. The model becomes complex because of the fuzzification process. These expanded fuzzified values are passed to the model to extract the relevant features that have a significant effect on the result. The diseases will be classified by passing the reduced features again to the ANN-based model. In [8] the authors in his study proposes an Adaptive Neuro-Fuzzy Inference Systems (ANFIS) to diagnose heart disease. The genetic algorithm is applied to optimize the parameters related to membership functions in ANFIS. The experiment was conducted on the public UCI heart disease datasets. The experimental result shows that 91.25% accuracy was obtained on the testing set. Table 1 illustrates the related works on the prediction heart failure using various methods.

Table 1: A summary of the above previous researches.

Research	Year	Classifiers for heart failure prediction	Best Accuracy
[5]	2021	They used the ANFIS system and the Genetic Algorithm to improve the performance of the ANFIS classification by the Genetic Algorithm through the attribute selection process.	96.667%.
[6]	2021	They used an artificial neural network (ANN) to predict heart failure. They also proposed a novel wrapper-based feature selection by the GWO to reduce the number of features.	87%
[7]	2020	They presents a model for the classification of medical diseases. The crisp inputs are fuzzified by using linguistic variables of the membership function to deal with imprecise input data. These expanded fuzzified values are passed to the model to extract the relevant features that have a significant effect on the result. The diseases will be classified by passing the reduced features again to the ANN-based model	
[8]	2020	The authors in his study proposes an Adaptive Neuro-Fuzzy Inference Systems (ANFIS) to diagnose heart disease. The genetic algorithm is applied to optimize the parameters related to membership functions in ANFIS.	91.25%

2. Neuro-Fuzzy System (NFS)

Neuro-Fuzzy system is a term that refers to a models that combine between two techniques which are: neural networks and fuzzy systems [9]. NFS is one of the hybrid systems that result from explicitly combining the fuzzy systems' transparency and generalization capabilities of the dynamic behavior of the neural networks [10]. This hybrid system is one of the most commonly used systems in difficult real-life problems like system identification problem. Therefore, the researcher derived to find new methods to simplify the design and to get high-speed convergence of dynamic systems. One of these methods is called the Neuro-fuzzy system. This method can choose the network weights by the fuzzy (IF_THEN) rules. The universal approximation capability of fuzzy systems, explain their successful application on many complex fields, in which dynamically changing knowledge base is needed. Identification and modeling of nonlinear dynamic systems, and adaptive predication of time series [11]. , one of the first works was proposed by Keller and Hunt, where a combination of neural network learning methods with the concepts of fuzzy systems was proposed [12]. They proposed an approach to improve the stability in the learning algorithm of classification problems using fuzzy techniques. They introduced a fuzzy membership of data items to the searched classes to improve the convergence of the learning algorithm. The architecture of a NFS system is shown in Fig. 1.

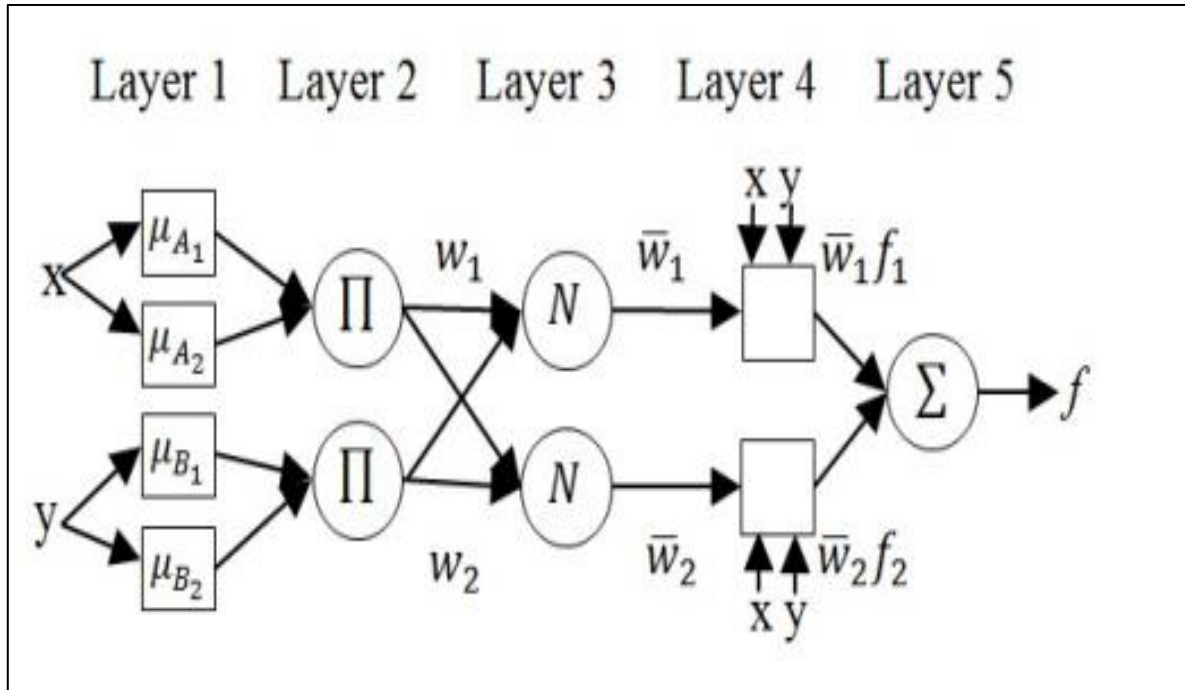


Fig. 1 - Architecture of NFS system[13].

NFS system consist of five layers. It describe, as follows:

1. **Input Layer:** This layer receives crisp inputs, which is the features of the dataset.
2. **Fuzzification Layer:** The crisp values are Fuzzification in this layer; the process done by transforming each feature of the input is into its corresponding membership values. Here, the Gaussian membership function with three linguistic variables is used: (low, medium, and high) to compute the membership values of each feature of the input pattern.
3. **Rule Layer:** The firing strength of each rule is computed in every node of this layer.
4. **Normalization Layer:** The normalized firing strength is computed in this layer.
5. **Consequent Layer:** The contribution of each rule in the NFM final result is reflected in each consequent node in the Consequent Layer [13].

3. Evaluating Metrics

Many evaluating metrics, in this paper will be used, it will explained as follows:

3.1 Confusion Matrix

A confusion matrix is a technique that shows the performance of a classification model. A confusion matrix is drawn as a table representing the results, including the correct and incorrect classified samples. It is used for binary classification [14]. A (2*2) matrix is employed to describe the confusion matrix [14]. Table 2 illustrates the confusion matrix.

Table 2 - The confusion matrix.

Confusion Matrix		Actual Class	
		Positive (p)	Negative (N)
Predicted Class	Positive (p)	True Positive (TP)	False Positive (FP)
	Negative (N)	False Negative (FN)	True Negative (TN)

- **True Positive (TP):** It represents the positive samples that are correctly identified by the model.
- **False Negative (FN):** It represents the positive samples that are wrongly classified by the model.
- **False Positive (FP):** It represents the negative sample mistakenly classified by the model.
- **True Negative (TN):** It represents the negative samples correctly classified by the model [15].

3.2 Accuracy

Accuracy is considered a widely metric. It is the represent ratio of correctly categorized samples to the total number of samples for a given test data set , and it is denoted mathematically [16],[17] as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

4. The Proposed System

The proposed system will be discussed in this part. Our system contains four steps: the first is the data collection, the second step is data pre-processing, which involves data splitting and feature scaling, the third step is the feature extraction and the last step is the prediction step. The following paragraphs will explain each step in detail.

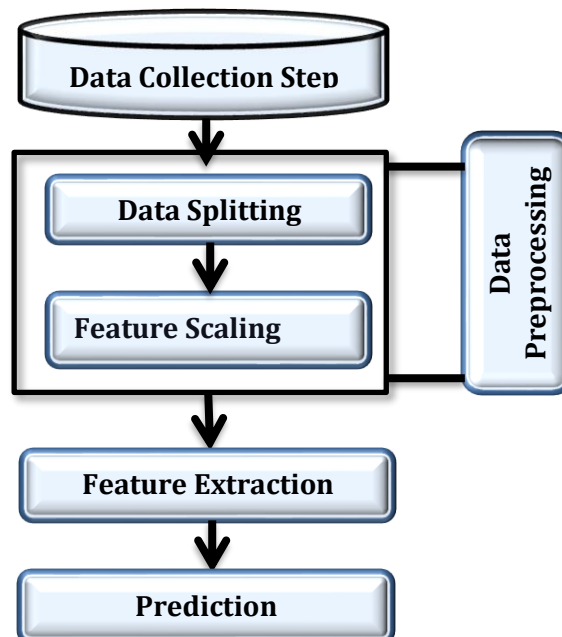


Fig. 2 - The proposed System for Heart Failure Prediction.

4.1 Data Collection Step

The dataset is explained in this step. The "Heart Failure dataset 2015" is used in this study. Available at the University of California, Irvine (UCI). This dataset consists of 13 columns and 299 rows. It represents the medical records of patients with heart failure. Table 2.

It is essential to mention that the dataset is imbalanced as there are 203 positive samples (death event = 0) and 96 negatives (death event = 1).

Table 2: Describes the elements in the dataset.

Feature	Description	Measurement	Range
Age	"Age of the patient"	Years	[40, ...,95]
Anaemia	"Decrease of red blood cells or hemoglobin"	Boolean	0,1
High blood pressure	"If the patient has hypertension"	Boolean	0,1
creatinine phosphokin ase (CPK)	"The CPK enzyme Level in the blood"	mcg/L	[23, ...,7861]
Diabetes	"If the patient has diabetes"	Boolean	0,1
Ejection fraction	"Percentage of blood leaving the heart at each contraction"	%	[14, ...,80]
Sex	"Woman or man"	Binary	0, 1
Platelets	"Platelets in the blood"	kiloplatelets/MI	[25.01, ...,850.00]
Serum creatinine	"Level of creatinine in the blood"	mg/Dl	[0.50, ...9.40]
Serum sodium	"Level of sodium in the blood"	mEq/L	[114, ...,148]
Smoking	"If the patient smokes"	Boolean	0,1
Time	"Follow-up period"	Days	[4, ...,285]
[target] death event	"If the patient died during the follow-up period"	Boolean	0,1

4.2 Data pre-processing

This step includes two stages as follows:

4.2.1 Data Splitting

In the splitting part, the dataset is divided into two separate sets: the training set and the testing. The dataset was split in a numeric manner into (training data=235) and (testing data=64).

4.2.2 Feature scaling

The dataset contains variables with widely varying scales, which necessitates the use of this technique. As a result, the dataset needs to be convert to a scaled form that can be understood by Neuro-Fuzzy system. So, the feature vectors will be scaled. In this work, the MinMaxScaler was used to scale each input variable separately to the range 0-1. Mathematically, Equation (2) is used to represent it.

$$Z = \frac{(x - \min)}{(\max - \min)} \quad (2)$$

Where:

1. **Z** = the scaled value of the feature.
2. **x** = the original value.
3. **min** = the minimum value.
4. **max** = the maximum value.

4.3 Feature extraction

In this step, the essential features in the dataset are selected and used to predict the patient's state. This process begins with testing every two features in the dataset together to determine how much they impact on the system's accuracy. This is done by monitoring the accuracy after each test. Lastly, the features that achieve the highest accuracy prediction will be considered the most important features in the dataset, as it directly affects on the patient's state, and these features can be used alone to predict whether the patients with heart failure will survive or not.

4.4 Prediction

In this step, the features that we obtained from the feature extraction step are used to prediction the patient's state This process done by feed these features alone as a input to the NFS again.

5. Experimental Result

This part presents the results collected from the ANFIS model after every two features are tested together. The performance of the NFS is evaluated with many metrics, including testing accuracy, and the confusion matrix, which will be represented for only the most important features in the dataset. Table 3 illustrates the NFS accuracy results based tested every two features from the dataset

Table 3-The NFS model results with all the experiments.

	age	anaemia	Cratinine phosphokinase (CPK)	Diabetes	Ejection fraction	High blood pressure	Platelete	Serum Creatinine	Serum sodium	smoking
age										
anaemia	56%									
Cratinine phosphokina se (CPK)	56%	60%								
Diabetes	50%	38%	61%							
Ejection fraction	81%	84%	78%	81%						
High blood pressure	56%	62%	55%	70%	79%					
Platelete	44%	55%	59%	55%	81%	51%				

Serum Creatinine	74%	79%	77%	73%	1.00	80%	79%			
Serum sodium	70%	75%	72%	70%	94%	78%	71%	91%		
sex	44%	50%	61%	50	77%	61%	59%	83%	66%	
smoking	49%	51%	50%	37%	60	55%	50%	66%	61	41%

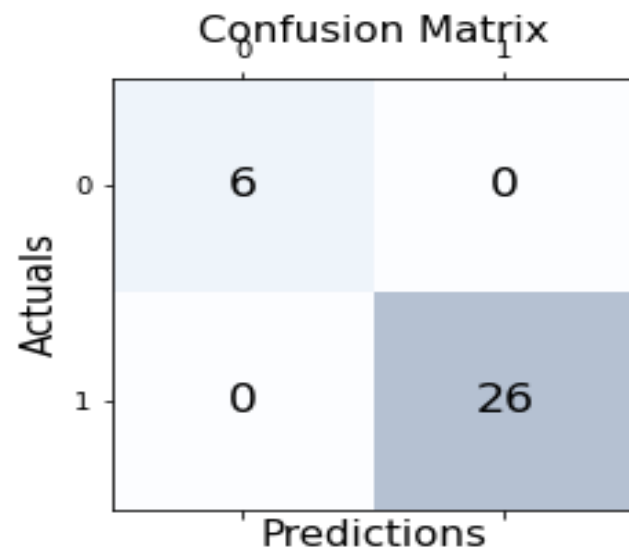


Fig. 3 - The confusion matrix for testing ejection fraction and serum creatinine.

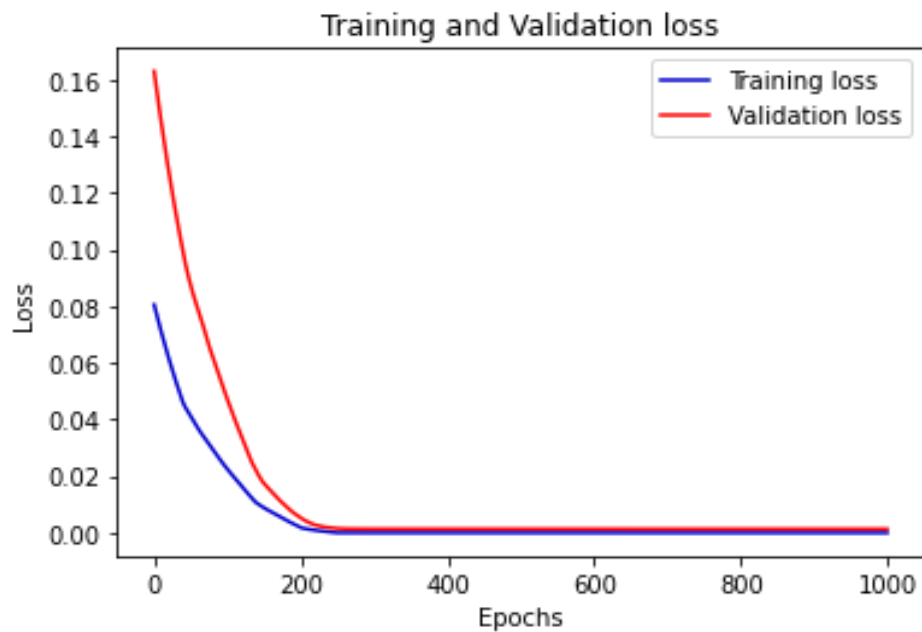


Fig. 4 - Illustrate the training loss and validation loss across epochs for testing ejection fraction and serum creatinine.

6. Discussion

This part will discuss the experimental result of the proposed system. After all the experiments, we found many features give high prediction accuracy when we test them together. Still, what we found is that when the features (ejection fraction, serum creatinine) are tested together, a 100% prediction accuracy is obtained. As a result, these features are considered the most important features in the dataset and can be used alone to predict the patient's state. Therefore, doctors could still forecast patient survival by merely looking at the "ejection fraction and serum creatinine" rates in the patient's electronic health record

7. Conclusion

Heart failure is a very dangerous disease that may lead to death. Early detection of heart failure will help to save the patient's life. Therefore, this paper aims to design a model that manually selects the most important features from the dataset and predicts the patient's state. This prediction is made by using Neuro-Fuzzy System (NFS). The dataset used in this study represents the "medical records" of patients with heart failure, consisting of 12 features. All features were used independently from the time feature. Every two features in the dataset are entered manually into the NFS model and monitored to know which features have the major impact on the accuracy after they are tested together. The experimental results show that the "serum creatinine and ejection fraction" when tested together (1.00%) accuracy is obtained. Therefore, these features are considered the most important features in the dataset and can be used alone to predict the patient's state if he/she will die or not.

REFERENCES

- [1] World Health Organization. (2019). *Cardiovascular diseases*. Available: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1
- [2] L. National Heart, and Blood institute,. (2022, March 24). *What Is Heart Failure?* Available: <https://www.nhlbi.nih.gov/health/heart-failure>
- [3] Mayo clinic. (2021, Des 10). *Heart failure*. Available: <https://www.mayoclinic.org/diseases-conditions/heart-failure/symptoms-causes/syc-20373142>
- [4] S. Shu, J. Ren, and J. J. C. J. Song, "Clinical application of machine learning-based artificial intelligence in the diagnosis, prediction, and classification of cardiovascular diseases," vol. 85, no. 9, pp. 1416-1425, 2021.
- [5] D. A. Korzhakin and E. J. S. J. o. I. Sugiharti, "Implementation of Genetic Algorithm and Adaptive Neuro Fuzzy Inference System in Predicting Survival of Patients with Heart Failure," vol. 8, no. 2, pp. 251-257, 2021.
- [6] M. T. Le, M. T. Vo, N. T. Pham, S. V. J. I. J. o. E. E. Dao, and C. Science, "Predicting heart failure using a wrapper-based feature selection," vol. 21, no. 3, pp. 1530-1539, 2021.
- [7] H. Das, B. Naik, and H. J. I. i. M. U. Behera, "Medical disease analysis using neuro-fuzzy with feature extraction model for classification," vol. 18, p. 100288, 2020.
- [8] J. Feng, Q. Wang, and N. Li, "An intelligent system for heart disease prediction using adaptive neuro-fuzzy inference systems and genetic algorithm," in *Journal of Physics: Conference Series*, 2021, vol. 2010, no. 1, p. 012172: IOP Publishing.
- [9] R. Czabanski, M. Jezewski, and J. Leski, "Introduction to fuzzy systems," in *Theory and Applications of Ordered Fuzzy Numbers*: Springer, Cham, 2017, pp. 23-43.
- [10] Z. J. Viharos and K. B. J. M. Kis, "Survey on neuro-fuzzy systems and their applications in technical diagnostics and measurement," vol. 67, pp. 126-136, 2015.
- [11] X.-H. J. J. o. C. i. C. E. Jin, "Neurofuzzy decision support system for efficient risk allocation in public-private partnership infrastructure projects," vol. 24, no. 6, pp. 525-538, 2010.

- [12] D. D. Nauck and A. Nürnberger, "Neuro-fuzzy systems: A short historical review," in *Computational intelligence in intelligent data analysis*: Springer, 2013, pp. 91-109.
- [13] A.-F. Attia, H. A. Ismail, H. M. J. A. Basurah, and S. Science, "A Neuro-Fuzzy modeling for prediction of solar cycles 24 and 25," vol. 344, no. 1, pp. 5-11, 2013.
- [14] S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, "Machine learning and deep learning techniques for cybersecurity: a review," in *The International Conference on Artificial Intelligence and Computer Vision*, 2020, pp. 50-57: Springer.
- [15] G. S. Ohannesian and E. J. J. I. Harfash, "Epileptic Seizures Detection from EEG Recordings Based on a Hybrid system of Gaussian Mixture Model and Random Forest Classifier," vol. 46, no. 6, 2022.
- [16] N. M. A.-M. M. Al and R. S. J. I. Khudeyer, "ResNet-34/DR: A Residual Convolutional Neural Network for the Diagnosis of Diabetic Retinopathy," vol. 45, no. 7, 2021.
- [17] S. F. Raheem and M. J. I. Alabbas, "Dynamic Artificial Bee Colony Algorithm with Hybrid Initialization Method," vol. 45, no. 6, 2021.