Kerbala Journal for Engineering Science
https://kjes.uokerbala.edu.iq/

Evaluation of the Use of ANFIS in Predicting Student's Performance

Zainab Rashid khudhair *, Mithaq N. Raheema **, Jabbar S. Hussein ***

*Department of Electrical and Electronic Engineering, Engineering, College of Engineering, Kerbala University, Karbala, Iraq.
E-mail: zainab.rashid@uokerbala.edu.iq
** Department of Prosthetic and Orthotic Engineering, College of Engineering, Kerbala University, Karbala, Iraq.
E-mail: methaq.n.rhiama@uokerbala.edu.iq
***Department of Prosthetic and Orthetic Engineering, College of Engineering, Kerbala University, Karbala, Iraq.
E-mail: jabbar.salman@uokerbala.edu.iq

Received: 19 June 2023; Revised: 09 July 2023; Accepted: 16 July 2023.

Abstract

This article suggests the implementation of an online distance education system as a way to enhance the academic performance of students. It emphasizes the importance of monitoring and predicting the Learning Academic Performance (LAP) to enable appropriate adjustments and ensure continuous academic progress. In this education mode, predicting LAP accurately is a challenging task. Adaptive Neuro-Fuzzy Inference System (ANFIS), executed by 'MATLAB R2020a', creates systems that can successfully achieve the aim. The suggested ANFIS models were used to forecast students' Cumulative Grade Point Averages (CGPAs). The interactions between input variables and their associated impacts on the output values were further investigated using the models. The results showed high accuracy levels, varying scores from 83.8% to 90%. Additionally, this study found a significant correlation between CGPA and the examination scores, indicating a direct proportional relationship. These findings emphasize the crucial role of examination scores in evaluating academic performance and the importance for students to prioritize and focus on their exam performance.

Keywords: Adaptive Neuro-Fuzzy Inference System (ANFIS), Performance Prediction, and Cumulative Grade Point Average.

1. Introduction

The use of artificial intelligence in education has increased dramatically over the last few years due to the new information technology provides about learners and the global community [1]. The development use of artificial intelligence in education focuses on creating strategies for examining special kinds of data generated by educational contexts. Standard databases may be only able to respond to questions such as "find the students who failed the exams". However, data mining could respond to more esoteric inquiries such as "identify the students who will probably pass the exams" [2]. Creating learner models that can forecast the learner's traits or success in their academic activities is one of the primary applications of artificial intelligence [3]. To help educators evaluate and improve the structure of their course material and track learners' academic achievement, academics have been looking into different data mining techniques [4]. It is beneficial to accurately anticipate students' academic success using artificial intelligence in a variety of educational settings [5], using the design expert (DX), Minitab, the artificial neural network (ANN), and others. Candidates accepted into a course via an online platform for distance learning may have precise estimates of their academic achievement, which will aid them in understanding how to study for and pass tests. Unsuitable exam outcomes might be the result of insufficient preparation. The number of accepted applicants has an impact on the institution's quality level since the educational institution's excellence is mostly represented in its research and teaching. By providing students with extra support such as individualized help and tutoring resources, educational managers can increase learners' academic performance [6]. In another words, online distance learning education opens up many possibilities for students, allowing them to participate in an excess of activities while following their academic goals. Yet, without proper supervision, these activities can easily distract students, causing their academic performance to suffer, semester after semester, ultimately affecting their cumulative grade point average (CGPA).

Despite several research deals, there is still a gap in our knowledge of how artificial intelligence, and more precisely, the ANFIS, could be used to predict academic success. Thus, it is essential to create a smart system that helps students and school administration track their academic performance and prepare for forthcoming tests. This intelligent system should provide students with the insight they need to achieve their desired CGPA and offer guidance on the steps they need to take to get there. Following that can allow students to stay focused, motivated, and on track, even during the pandemic.

2. Related Work

The effectiveness of ANFIS in dividing undergraduate students into two groups based on their Cumulative Grade Point Average (CGPA) at graduation was examined by Rusli et al. in 2008 [7]. The study showed that the Neuro-Fuzzy Predictive System outperformed artificial neural networks and logistic regression. In a similar study by Taylan et al. in 2009 [8], an ANFIS model was also created to forecast the academic performance of engineering students in terms of five CGPA categories (excellent, very well, good, mediocre, and unsatisfactory). A range of characteristics related to university students' performance throughout an academic year were used to build the specific model. In 2013, Do and Chen employed a neuro-fuzzy classifier that split the students into excellent, average, and bad groups to get an equivalent result [2]. The suggested approach surpassed typical classification techniques, including Decision Trees, Neural Networks, Support Vector Machines, and Naive Bayes classifier. In a related research by Hidayah et al. in 2013 [9], undergraduate students were divided into groups based on their academic achievement: very good, good, satisfactory, and poor, using several prediction models. Each model used the ANFIS methodology and was trained using data acquired from questionnaires on students' motivation, interests, and skills. Overall, the results were consistent with earlier studies. Arora and Saini introduced a fuzzy probabilistic neural network model in 2013 based on various input variables connected to family history, study habits, and regular attendance to solve the same issue [10]. A rule-mining approach for evaluating student performance based on the association rules was presented in 2013 by Ajith et al. [11]. Association rules were used to investigate the student dataset to collect essential data for assessing the student's performance.

The rapid development of educational technology helps educational institutions discover the reality of student performance prediction models [12, 13]. The student's performance was predicted using a variety of methods. To predict students' ultimate eventual success in remote higher education, Yildiz et al. in 2013 [14] used a genetic algorithm to develop a fuzzy model to optimize the membership functions. The gene-fuzzy model was trained using log data from the first eight weeks of the semester and data on student behaviour from the online Learning Management System (LMS). The experimental findings demonstrated the superiority of the suggested strategy over a knowledgeable and common fuzzy model. Abu-Naser et al., in 2015 [1], employed a model of an artificial neural network (ANN) to forecast how well engineering faculty students will do. They assessed the students' success using the course grades, the total amount of earned credits, and average cumulative grade points. The ANN model accurately predicted the students' performance. Hamsa et al., in 2016, used effectively a fuzzy genetic algorithm to identify postgraduate students who were "safe" and "at risk" [15]. In a recent study, Kostopoulos et al. employed a variety of active learning categorization models to provide early predictions about student performance (pass, fail) in distant learning [16]. To do this, the margin sampling query approach and pool-based sampling scenario were also employed, together with a large number of frequently used supervised classifiers as fundamental learners. The accuracy and Area Under Curve (AUC) findings were acceptable compared to the equivalent supervised approaches. Zhao et al., 2020 [17] adapted the fuzzy rule-based categorization technique to determine whether a student would learn anything in the first third of a semester based on log data from the Moodle LMS. Regarding accuracy, precision, recall, and F-measure, the new technique fared performed better than the previous one. Studies examining the development and use of active learning strategies in Education Data Mining (EDM) are scarce and far between.

3. Methodology

According to a thorough literature review and reliance on experts' opinions, a great deal of socioeconomic, environmental, biological, educational, and other linked parameters are expected to affect a university learner's performance, particularly regarding the learner's academic performance. The release of the data used in this study resulted from collecting several datasets from Kaggle.com based on the above variables. The collected data mainly focused on 395 students whose initial performances were used to determine whether they would end up with a Weak, Good, or Excellent cumulative grade point average. It is important to note that the data was based on factors that influenced student performance, including first- and second-period grades, weekly study time, family connection quality, father's education, and the number of absences from school. Additionally, as it was the desired outcome, the CGPA of each applicant was determined as the basis for these.

The data were neatly scrutinized and harmonized to a tolerable amount inside the confines of the modelling interface of ANFIS. These influencing factors were subdivided into input and output variables and then formatted appropriately for ANFIS modelling, optimization, and analysis.

ANFIS uses a learning approach to map the relationships between the input and output data in a Sugeno-type multilayer feed-forward network to fine-tune the parameters of the fuzzy inference system[10]. The ANFIS method is a neural network-trained fuzzy logic (FL) system in its most basic form. When finding the FL system's rules and membership functions, the adaptive method uses the ANN's impressive capacity for learning [18]. The data set used is called training data, and its primary purpose was to design a system for accomplishing the required nonlinear mapping based on the arrangement of the set of data. The target system's many input-output pairings constitute this data collection. The network parameters are adjusted during training, commonly called a data learning method, for better performance. [19]. The model's ability to generalize was tested by introducing a test data set that had not been utilized during training. The model architecture of ANFIS consists of five distinct layers (Fig. 1), as follows:

The crisp input data is for the first one,

First layer: Fuzzification neurons take a ‘bell’ activation function as displayed below:

$$y_i^{(1)} = \frac{1}{1+(\frac{x_i^{(1)}-a_i}{c_i})^{2bi}} \tag{1}$$

Where c_i, a_i and bi denote the centre, width and slope of the generalized Bell function, respectively, and x_i denotes the input.

In the second layer, the strength of the Sugeno-type fuzzy rule is calculated by the operator product as the third output of neuron (i) in the second layer obtained by:

$$y_i^{(2)} = \prod_{j=1}^{k} x_{ji}^{(2)} \tag{2}$$

In the third layer, the calculation of the normalized rule strength is shown:

$$y_i^{(3)} = \frac{x_{ii}^{(3)}}{\sum_{j=1}^{n} x_{ji}^{(3)}} = \frac{\mu_i}{\sum_{j=1}^{n} \mu_i} = \overline{\mu_i} \tag{3}$$

Where $x_{ji}^{(3)}$is the input from neuron j located in Layer 2 to neuron i in Layer 3, and n is the total number of rule neurons. The defuzzification operation in the fourth layer is calculated by:

$$y_i^{(4)} = x_i^{(4)}[k_{i0} + k_{i1}x_1 + k_{i2}x_2] = \overline{\mu_i}^4[k_{i0} + k_{i1}x_1 + k_{i2}x_2] \tag{4}$$

Where $x_i^{(4)}$ is the input and $y_i^{(4)}$ is the output of defuzzification neuron i in Layer 4, and k_{i0} , k_{i1} andk_{i2} are a set of rules i consequent parameters.

Layer 5, a single summation neuron and the fuzzification neurons in the two-layer that have a ‘bell’ activation function as displayed below:

$$y_i^{(5)} = \sum_{j=1}^{n} x_i^{(5)} = \sum_{j=1}^{n} \overline{\mu_i}\,[k_{i0} + k_{i1}x_1 + k_{i2}x_2] \quad (5)$$

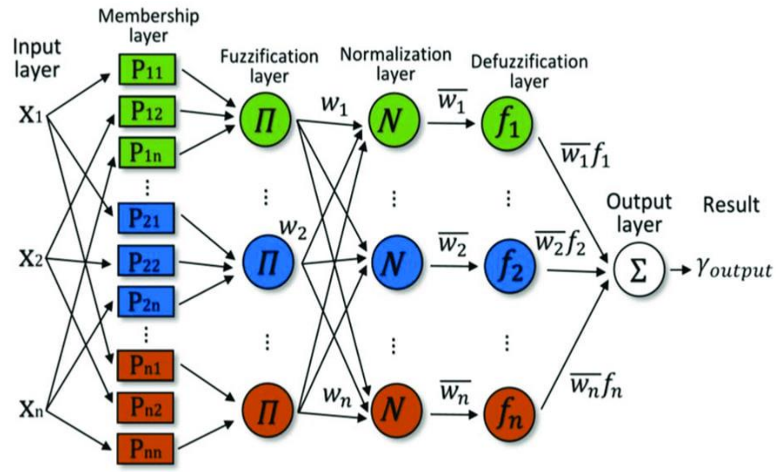

Figure1. ANFIS Model Architecture

4. ANFIS Modelling

In the ANFIS model, it is necessary first to identify the critical factors that affect student performance. The term "output variable" refers to the desired result, whereas "input variable" refers to the information utilized to direct the process. Given that the CGPA was the sole output necessary for this task, the ANFIS model seemed to function effectively. Below is a detailed explanation of the input/output mapping. Using the MATLAB (R2020a) user interface, the ANFIS model for predicting academic achievement in the classroom has been modified.

4.1 The Input Variables

The students' efforts resulted in scores and grades, and cumulative grade point averages were calculated. As a result, the stated input variables may be easily collected from the data gathered. Father's Educational Attainment (Fedu), Weekly Study Time (study time), Family Relationship Quality (famrel), University Absences (absences), and First and Second Period Grades (G1) and (G2) are all considered to be input variables, as shown in Table 1.

Table 1. Input and Output Information

No. S/N	Input Parameters
1	father's education (Fedu)
2	weekly study time(study-time)
3	quality of family relationship(famrel)
4	number of university absences(absences)
5	first period grade(G1)
6	second period grade(G2)
No. S/N	Output Parameters
1	CGPA

4.2 The Output Variables

Conversely, the output variable was students' cumulative grade point averages, corresponding to a candidate's levels of achievement under the current college grading system. Grade point average was the standard by which students were judged in the classroom, as presented in Table 2. Labelling the variables produced, Table 2 shows the domain categorization of the output variables as Weak, Mid, and Good. After all, the grading system at the university has already been using this method to categorize students.

Table 2. Output Variable Classification

No. S/N	Results Variable	CGPA
1	Weak	<50
2	Mid	50-70
3	Good	>70

5. The Data Set Grouping

Both the training and testing sets are used in the supervised training. The system can investigate the relationships between the input data and the subsequent outputs using the training set to forecast the relationship between the input and the predicted output. The data acquired from 395 students was classified into two categories, such that 315 students (70% of the total number of students) and 80 students (30% of the total number of students) were applied for the training set and testing set correspondingly, respectively.

6. ANFIS Model Prediction and Performance Evaluation

The input variables were subjected to fuzzification using one of three different Membership Functions (MFs): the Gaussian MF (Gaussmf), the triangular MF (Trimf), or the trapezium MF (Trapmf).

The developed ANFIS architectures were afterwards placed through a simulation study with a range of (MFs) for inputs, outputs, and epoch numbers that were looked at to find the best mixture so that the ideal system could be used. The models were ranked based on their relative accuracy.

7. Results and Discussion

In this section, the model performance of the ANFIS based on three different membership functions (MFs) will be presented as follows:

7.1 Results of ANFIS-Models Developments

The implementation of the model has been tested based on Gaussmf, Trimf and Trapmf MFs to train the dataset of the ANFIS models. The output MFs were tested based on the accuracy of the test and training dataset. ANFIS systems were accomplished using the training-data subset, as shown in Figs. 2 and 3. Figure 4 shows a three-dimensional surface that represents the output of the ANFIS model. The x-axis represents the father’s education with 15 different levels, the y-axis represents study time with 15 different levels, and the z-axis represents the predicted CGPA. Figure 4 shows how the predicted CGPA changes verse to two input parameters (the values of father education and study time variation as examples). This can be useful in understanding how these variables affect a student's performance and can be used to make decisions or predictions about future student performance based on these variables.

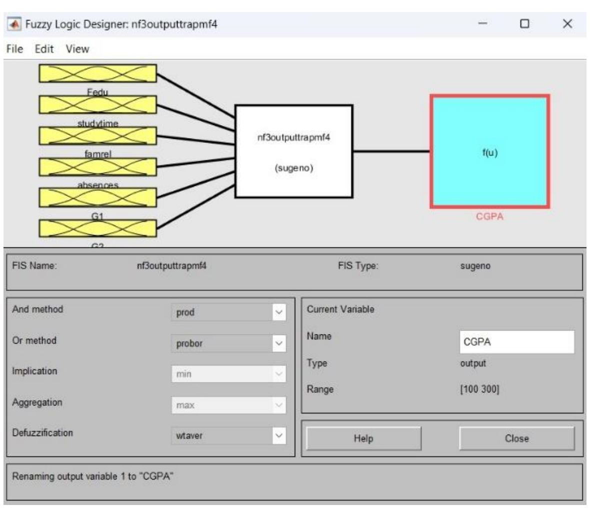

Figure 2. Designing of ANFIS model for semester’s GPA predictions.

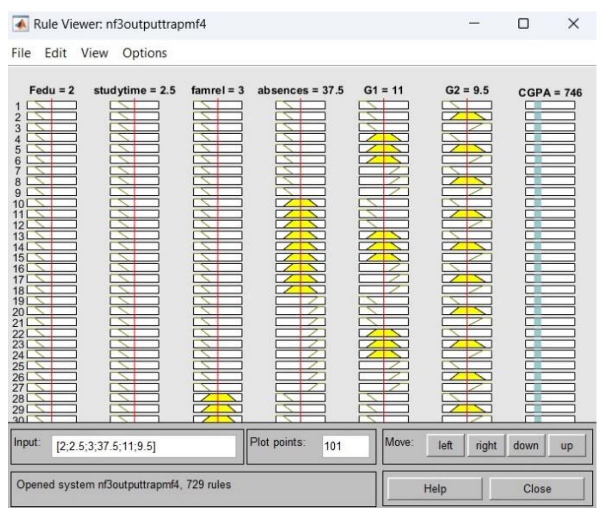

Figure 3. Set of rules in the ANFIS training.

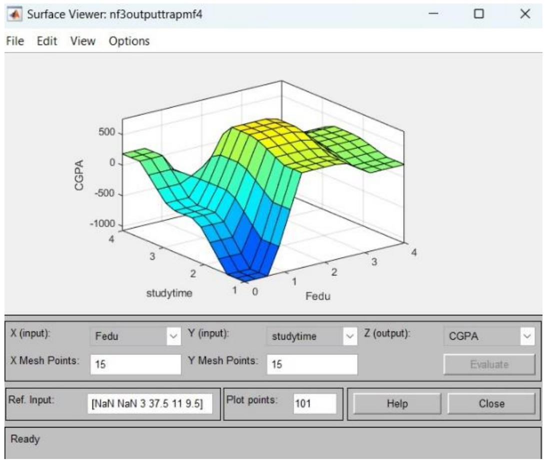

Figure 4. Three-dimensional surface of the ANFIS proposed model

7.2 ANFIS Models Performances and Evaluation

The ANFIS model's training and testing accuracy were calculated using various membership functions (MFs) with 300 fixed epochs. The three membership functions used were Gaussmf, Trimf and Trapmf. The results showed that the Gaussmf membership function achieved the highest accuracy in training and testing, with 92.7% and 90.0%, respectively. On the other hand, the Trimf membership function achieved a training accuracy of 89.8% and a testing accuracy of 83.8%. In comparison, the Trapmf membership function achieved a training accuracy of 87.3% and a testing accuracy of 83.8%. The results suggested that increasing the epoch number could potentially improve the accuracy of the ANFIS model. The results also indicated that the Gaussmf membership function performed accurately among the three membership functions. It achieved a training accuracy of 92.7% and a testing accuracy of 90.0%, suggesting that it could effectively capture the underlying patterns and relationships in the data.

On the other hand, the Trimf and Trapmf membership functions achieved lower accuracies in training and testing. The Trimf membership function attained a training accuracy of 89.8% and a testing accuracy of 83.8%, while the Trapmf membership function achieved a training accuracy of 87.3% and a testing accuracy of 83.8%. These lower accuracies indicate that these membership functions may not be effective in capturing the complexities of the data compared to the Gaussmf function.

It is worth mentioning that the choice of membership function depends on the specific problem and dataset. In this case, the Gaussmf function demonstrated superior performance, suggesting its suitability for this particular problem. However, further analysis and experimentation may be required to understand the reasons behind these differences in accuracy and to determine the most appropriate membership function for the given problem domain.

The results indicate that increasing the epoch number can enhance the accuracy of the ANFIS model. The model has more opportunities to learn and adjust its parameters based on the available data by conducting multiple epochs during training. This iterative process allows the model to refine its predictions and potentially achieve better accuracy as it captures more patterns and relationships within the dataset. The ANFIS model using the Gaussmf membership function achieved the highest accuracy for predicting student performance. This model can identify students needing additional support and interventions to improve their academic performance. Figure 5 shows the confusion matrices for the ANFIS model with Gaussmf membership with a high prediction accuracy of 90%.

Figure 5. Confusion Matrices for the ANFIS-Gaussmf.

Comparing the outcomes of two prior studies, the classification accuracy of the ANFIS model created in this research was evaluated. Fahd et al. [20] used a Random Forest classifier with booster ensemble tuning to achieve an accuracy of 85.7%. At the same time, Abu-Naser et al. [1] used ANN techniques to achieve an accuracy of 88%. Riyadi Yanto et al. [21] used a Fuzzy Soft Set classifier and reported an accuracy of 89%. The ANFIS model developed in this work achieved an accuracy of 90%, demonstrating performance improvement compared to the other two classifiers, as presented in Table 3. These results suggest that the ANFIS model developed in this work may effectively predict academic performance.

Table 3. Comparison of the accuracy results from various techniques.

Study	Classifier	Dataset	Accuracy
Fahd et al., [20]	Random Forest with booster ensemble tuning	112 students	85.7%
Abu-Naser et al.,[1]	ANN	150 students	88%
Riyadi Yanto et al., [21]	Fuzzy soft set	2068 students	89.3%
This work	ANFIS	395 students	90.00%

As a comparison of the accuracy results with other works that use the ANFIS model, the achieved accuracy in [22] was 83% compared with a significantly higher accuracy of 90% obtained in this work. These results suggest that the ANFIS model developed in this work can effectively predict students' academic performance and be a valuable addition to neuro-fuzzy research.

8. Conclusion

In this work, an ANFIS model was developed to predict student academic performance based on several variables, including first-and second-period grades, weekly study time, family connection quality, father's education, and the number of absences from school. The model was evaluated using Gaussmf, Trimf, and Trapmf membership functions. The ANFIS model using Gaussmf achieved the highest accuracy, with 92.7% and 90.0% for training and testing, respectively, at 300 epochs, which could be a promising tool for predicting student academic performance. The suggested predictor model's high accuracy demonstrates its efficacy in identifying students who need extra assistance and interventions to raise their academic performance. Compared with other classifiers in neuro-fuzzy research, the ANFIS model developed in this work outperformed the other classifiers with a 6% improvement in accuracy. Input parameters were identified as factors affecting the model's output, which can be used to make decisions or predict future student performance. In conclusion, the ANFIS model developed in this work has shown promising results in accurately predicting students' academic performance. These findings can be useful to teachers, policymakers, and other stakeholders to improve students' academic performance.

References

[1] S. S. Abu-Naser, I. S. Zaqout, M. Abu Ghosh, R. R. Atallah, and E. Alajrami, "Predicting student performance using artificial neural network: In the engineering and information technology faculty," 2015.

[2] Q. H. Do and J.-F. Chen, "A neuro-fuzzy approach in the classification of students' academic performance," vol. 2013, pp. 6-6, 2013.

[3] A. M. Shahiri and W. Husain, "A review on predicting student's performance using data mining techniques," vol. 72, pp. 414-422, 2015.

[4] I. E. Livieris, K. Drakopoulou, and P. Pintelas, "Predicting students' performance using artificial neural networks," pp. 321-328, 2012.

[5] J. Undavia, P. Dolia, and A. Patel, "Comparison of Decision Tree Classification Algorithm to Predict Students Post Graduate Degree in Weka Environment," vol. 1, no. 2, pp. 17-21, 2014.

[6] K. Colchester, H. Hagras, D. Alghazzawi, and G. Aldabbagh, "A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms," vol. 7, no. 1, pp. 47-64, 2017.

[7] N. M. Rusli, Z. Ibrahim, and R. M. Janor, "Predicting students’ academic achievement: Comparison between logistic regression, artificial neural network, and Neuro-fuzzy," in 2008 international symposium on information technology, 2008, vol. 1, pp. 1-6: IEEE.

[8] O. Taylan, B. J. C. Karagözoğlu, and I. Engineering, "An adaptive neuro-fuzzy model for predicting student’s academic performance," vol. 57, no. 3, pp. 732-741, 2009.

[9] I. Hidayah, A. E. Permanasari, and N. Ratwastuti, "Student classification for academic performance prediction using neuro-fuzzy in a conventional classroom," in 2013 International Conference on Information Technology and Electrical Engineering (ICITEE), 2013, pp. 221-225: IEEE.

[10] N. Arora and J. R. Saini, "A fuzzy probabilistic neural network for student’s academic performance prediction," vol. 2, no. 9, pp. 4425-4432, 2013.

[11] P. Ajith, B. Tejaswi, and M. Sai, "Rule mining framework for students performance evaluation," vol. 2, no. 6, pp. 201-6, 2013.

[12] C. Romero and S. Ventura, "Data mining in education," vol. 3, no. 1, pp. 12-27, 2013.

[13] S. Dawson, D. Gašević, G. Siemens, and S. Joksimovic, "Current state and future trends: A citation network analysis of the learning analytics field," in Proceedings of the Fourth International Conference on learning analytics and knowledge, 2014, pp. 231-240.

[14] O. Yildiz, A. Bal, and S. Gulsecen, "Improved fuzzy modelling to predict the academic performance of distance education students," vol. 14, no. 5, pp. 144-165, 2013.

[15] H. Hamsa, S. Indiradevi, and J. Kizhakkethottam, "Student academic performance prediction model using decision tree and fuzzy genetic algorithm," vol. 25, pp. 326-332, 2016.

[16] G. Kostopoulos, A.-D. Lipitakis, S. Kotsiantis, and G. Gravvanis, "Predicting student performance in distance higher education using active learning," in Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings, 2017, pp. 75-86: Springer.

[17] Q. Zhao, J.-L. Wang, T.-L. Pao, and L.-Y. Wang, "Modified fuzzy rule-based classification system for early warning of student learning," vol. 48, no. 3, pp. 385-406, 2020.

[18] A. Altaher and O. BaRukab, "Prediction of student's academic performance based on adaptive neuro-fuzzy inference," vol. 17, no. 1, p. 165, 2017.

[19] A. Ikuomola and O. Arowolo, "Evaluation Of Student Academic Performance Using Adaptive Neuro-Fuzzy Approach," vol. 11, no. 1, pp. 11-23, 2012.

[20] K. Fahd, S. J. Miah, and K. Ahmed, "Predicting student performance in a blended learning environment using learning management system interaction data," no. Ahead-of-print, 2021.

[21] I. T. R. Yanto, E. Sutoyo, A. Rahman, R. Hidayat, A. A. Ramli, and M. F. M. Fudzee, "Classification of student academic performance using fuzzy soft set," in 2020 International Conference on Smart Technology and Applications (ICoSTA), 2020, pp. 1-6: IEEE.
[22] M. Adil, F. Tahir, and S. Maqsood, "Predictive analysis for student retention using neuro-fuzzy algorithm," in 2018 10th Computer Science and Electronic Engineering (CEEC), 2018, pp. 41-45: IEEE.

تقييم استخدام ANFIS في التنبؤ بأداء الطلاب

الخلاصة: *تقترح هذه المقالة نظام التعليم عن بعد عبر الإنترنت كوسيلة لتحسين الأداء الأكاديمي للمتعلمين ويؤكد على أهمية مراقبة و توقع الأداء الأكاديمي التعليمي (LAP) والتنبؤ به، بحيث يمكن إجراء التعديلات اللازمة لضمان تقدمهم الأكاديمي المستمر. في هذا النمط من التعليم، يعد التنبؤ الدقيق بـ LAP مهمة صعبة. يقدم هذا العمل أنظامة يمكنها التنبؤ بشكل فعال بـ LAP باستخدام نظام الاستدلال العصبي الضبابي التكيفي (ANFIS) المطبق بواسطة MATLAB R2020a. تُستخدَم نماذج ANFIS المقترحة للقيام بالتنبؤ بمتوسط الدرجات التراكمي (CGPA) للمتعلمين. تم استخدام النماذج أيضاً لاستقصاء التفاعلات بين متغيرات الإدخال وتأثيراتها المقابلة على قيم المخرجات، CGPA. أظهرت النتائج مستويات دقة عالية مع درجات متفاوتة من 83.8٪ إلى 90٪. أستنتج هذا العمل أيضًا أن المعدل التراكمي المعدل (CGPA) كان يعتمد بشكل كبير على الدرجات في الاختبارات، مع ملاحظة علاقة تناسبية مباشرة. وهذا يسلط الضوء على أهمية درجات الاختبار في تحديد الأداء الأكاديمي ويؤكد على حاجة الطلبة إلى إيلاء اهتمام وثيق لأدائهم في الاختبارات.*

الكلمات الدالة: نظام الاستدلال العصبي الضبابي (ANFIS)،التنبؤ بالاداء،متوسط الدرجات التراكمي.