

An Intelligent Hybrid Approach for Predicting the Academic Performance of Students using Genetic Algorithms and Neuro Fuzzy System

Altyeb Altaher¹ and Omar M. Barukab²

^{1,2}Department of Information Technology, Faculty of Computing and Information Technology-Rabigh, King Abdulaziz University, Jeddah, Saudi Arabia

Summary

Higher education institutions aim to offer high quality education to its students, by monitoring and analyzing the performance of the students to provide appropriate means to the students with low performance to meet their needs in early stages during their academic career. The availability of large volume of data in educational databases has made the prediction of the students' academic performance a challenging and quite important task. To overcome this problem, utilizing soft computing is one of the most effective and efficient methods for solving different kinds of problems, especially predicting the student's academic performance. Among many soft computing approaches, the neuron fuzzy inference models which are considered the most appropriate approach. This paper presents an intelligent hybrid genetic neuro-fuzzy inference system (HGANFIS) for student's academic performance prediction. HGANFIS intelligently integrates the learning and reasoning capabilities of the Adaptive neuro fuzzy inference system (ANFIS) with the powerful optimization of the genetic Algorithm (GA). The proposed approach HGANFIS utilizes the previous exam results of the students as input variables to ANFIS, and incorporates the GA into ANFIS to acquire the optimal ANFIS parameters in order to obtain accurate prediction of student's performance. The experimental results were appealing and showed that the proposed HGANIS obtained highest student's performance prediction accuracy when compared with other approaches.

Key words:

Student's performance prediction, Genetic algorithm, adaptive fuzzy inference, data mining.

1. Introduction

Educators at different academic institutions are keeping an eye on the KPI's that help in monitoring the status of their program's student's outcomes. Student performance prediction has significant influence in offering education that is tagged with high quality. Therefore, educational institutions are attempting to augment the prediction of student's performance into their educational culture for more advanced support to the students, by planning extra plans especially for the students with unsatisfactory performance in order to help them. Machine learning techniques are used extensively in many administrations as

a capable approach for mining the volumous amount of data which may lead to finding significant information that can be utilized in supporting the decisions making processes. Educational institutions utilize machine learning methods to improve the performance of the students and the quality of education. Recently, many educational institutes adopted the Educational data mining techniques, to extract important data to supervise and enhance the learning process, by mining the immense amounts of data in the educational arena [1-3].

The rapid growing number of available educational databases, encompass valuable hidden information that are not utilized to enhance the academic performance of the students. Educational data mining studies huge amounts of educational records stored in databases, to get the smeared valuable information for more effective utilization of this information. The extracted information is beneficiary for several educational processes like the prediction of students' performance [4]. Moreover it will assist in enabling the educators to better know their students and where they stand. The regression, clustering classification and visualization are different methods used for mining the information unseen in different educational databases.

The widely used technique in educational data mining is the classification. It is a technique that allocates data items to dissimilar Classes defined previously, the aim of classification is to forecast the right class for each data item. There are many methods for data classification such as support vector machine, artificial neural network and Bayesian classifiers [5]. These methods used extensively in many systems related to education [6,7]. The Neuro- fuzzy inference system is a hybrid computing approach which associates the powerful reasoning of fuzzy logic and the learning ability of neural network, hence the neuro-fuzzy inference system avoid the shortcomings of both fuzzy systems and neural networks when used independently [8]. NFIS has been utilized effectively in several applications such as classification and control [9, 10].

In this paper, a genetic neuro-fuzzy inference system is proposed for predicting the academic performance of the students to assist them to advance their academic

attainments. The proposed approach (HGANFIS) integrates the reasoning ability of the Adaptive neuro-fuzzy inference system (ANFIS) with powerful optimization of the Genetic Algorithm (GA). The main contributions of this paper are as follows:

- Integrating the Genetic algorithm technique with the Adaptive Neuro-Fuzzy Inference System for accurate prediction of student's performance.
- Determining the optimum parameters of the Adaptive Neuro-Fuzzy Inference System for more accurate prediction of students performance.
- Investigating the potential of the neuro-fuzzy inference systems in predicting the student performance.

This paper is structured as follows, the literature review is presented in Section 2. Section 3 reviews (ANFIS) and (GA) techniques, section 4 explain the proposed approach (GANFIS). Results and discussion are presented in section 5. Section 6 concludes our paper.

2. Literature Review

The fast improvements in the recent technologies for education help the educational institutes to utilize several models for predicting the academic performance of the students [4]. A number of students' performance prediction approaches have been proposed, In [11] Barber and Sharkey employed the logistic regression scheme to predict the students' performance using data obtained from the students' information management systems. In [12], Wolff et al. employed the decision-tree method to distinguish the students in danger academic circumstance. Jai Ruby & David in [13] utilized several data mining methods to predict the students' performance, they conclude that the neural network approach performs better than the other data mining approaches and attained the best prediction of student academic performance accuracy of 74.8%.

In [14], Abu Naser et al., employed the Artificial Neural Network (ANN) for students' performance prediction, they used the courses scores, amount of obtained credit hours and the cumulative grade point average as inputs to estimate the performance of the students, the ANN based model predicted the performance of the students with an accuracy of 80%. In [15], Ajith et al., suggested a rule mining based approach for the students' performance evaluation. Association Rules used to mine the student database for finding important information for the evaluation of student's performance.

Altayeb and Omer in [16] proposed an approach for student's performance prediction based on the Adaptive neural fuzzy Inference System (ANFIS), they used three ANFIS models by utilizing various membership functions.

Their results show that ANFIS-GbellMF model achieved the best performance when compared with the other ANFIS models.

3. The Adaptive Neuro Fuzzy Inference and Genetic Algorithms

3.1 Adaptive Neuro-fuzzy Inference System

Currently, the fuzzy inference systems are witnessing advanced developments and used in many applications. The fuzzy inference system (ANFIS) associates the advantages of fuzzy logic and neural network [9], and has the capability of fast convergence due to its hybrid learning. These advantages equipped ANFIS with better tracking and adaptive capabilities more than any other controller [17].

ANFIS uses training dataset to build a fuzzy inference system with membership functions parameters that could be tuned using the back propagation algorithm with the recursive least squares algorithm or back propagation algorithm only [18].

Suppose we considered a fuzzy inference system with two inputs x and y and only one output z . In this fuzzy inference system, two fuzzy if-then rules based on Takagi and Sugeno's type [9] stated as follows:

Rule 1: IF x is L_1 and y is T_1 , then

$$O_1 = v_1x + s_1y + q_1 \quad (1)$$

Rule 2: IF x is L_2 and y is T_2 , then

$$O_2 = v_2x + s_2y + q_2 \quad (2)$$

The inputs to the ANFIS system represented by x and y , the fuzzy sets denoted by L_i and T_i , the fuzzy rule output is O_i , whereas v_i , s_i and q_i denote the parameters identified during the training process. ANFIS consists of 5 layers and each layer performs certain function as follows:

Layer 1: input data into the ANFIS will be expressed as linguistic expressions using membership function (μ), there are different forms of membership functions. In this research, Gaussian membership function is employed due to its capability in generating a model with smoothed performance. The Gaussian membership function is expressed as in (3):

$$OP_i^1 = \beta(x) = \exp\left(-\frac{1}{2} \frac{(x-c)^2}{\sigma^2}\right) \quad (3)$$

Where O_i 's represent the layer's outputs. C and σ denote the center of the Gaussian membership function the variance.

Layer 2: This layer defines the accuracy level of the antecedent parts by the following equation.

$$O_i^2 = w_i = \mu_{L_i}(x) * \mu_{T_i}(y), i = 1, 2 \quad (4)$$

Layer 3: all the firing strengths calculated in the previous layer (w_i) are normalized in this layer, by finding the ratio of the i th rule's firing strength to the all rules firing strengths as follows.

$$O_i^3 = \frac{w_i}{\sum_i w_i} \quad (5)$$

Layer 4: the consequent portions of the fuzzy rules are included in this layer, the impact of each rule in the ultimate output is specified as follows:

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i (d_i X_1 + k_i X_2 + v_i) \quad (6)$$

Where d_i , k_i , and v_i denote the consequent parameters. The antecedent parameters and the consequent parameters are adjusted by ANFIS during the learning phase to reduce the differences between the target result and the actual output. Layer 5: in is layer all the fuzzy rules produced for single output are grouped and defuzzified as outputs of numerical values, as follows:

$$O_i^5 = Y = \sum_i \overline{w}_i f_i = \overline{w}_1 f_1 + \overline{w}_2 f_2 = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (7)$$

3.2 Genetic Algorithms

The Genetic Algorithm was introduced by Holland in 1975. It is a common technique for solving the diversity of constrained and unconstrained optimization issues [19]. GA can also be employed to solve a variety of optimization difficulties that are not appropriate for standard optimization algorithms. Several researchers also proposed the GA as a global search method for improving the performance of the artificial neural networks ANN [20-22].

The first population will be adjusted to reach an improved answer. At each step, the GA chooses individuals (chromosomes) from the recent population (parents) arbitrarily and utilizes them to create the children for the next generation. After some generations, based on the spirit of the GA, it attempts to jump to the good solution. At each step, the GA employs three main kinds of rules to

generate the next generation from the present population. The rules types are presented as follows:

Selection of rules: select the parent rules which compose the population for the next generation.

Crossover rules: associate the chromosomes in order to create the next generation.

Mutation rules: modify and change chromosome value [23].

4. The Proposed Genetic Adaptive Neuro-Fuzzy Inference Approach

The proposed HGANFIS approach utilizes the powerful optimization capability of the genetic algorithm for intelligent tuning of ANFIS parameters. The architecture of GANFIS is shown in Fig. 1.

ANFIS contains two types of parameters that could be tuned to improve its performance, the parameters of the antecedent and the consequent part. The parameters of the antecedent (IF-part) and consequent (THEN-part) are used as weights in the fuzzy rules and significantly affect its accuracy. The antecedent parameters depends on the type of membership function used to construct the ANFIS structure, the proposed approach GANFIS uses the Gaussian membership function in its structure.

The ANFIS Parameters were initialized randomly in first step and then are being optimized using GA algorithm. GA algorithm transforms the parameters of ANIS models into chromosomes and adjusts them via reproduction, mutation and crossover operations. In this study, the Root Mean Square Error (RMSE) was used as objective function in genetic algorithm to perform the optimization by minimizing the error between measured values and model estimates as shown in the following equation.

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (X_k - Y_k)^2} \quad (8)$$

Where X_k is observed value and Y_k is model predicted value for sample k .

The steps taken for implementing the proposed GANFIS are shown in the fig. 2.

5. Results and Discussion

The proposed HGANFIS approach that integrates the GA for ANFIS optimization is applied for student's performance prediction. The inputs to GANFIS are the student's marks and the output is the predicted student's academic performance in terms of Grade Point Average (GPA), which is standard method for measuring student's academic achievement in the universities. The dataset contains 100 records for computer science students. In the

dataset there are 5 required subjects, marks obtained by students in these required subjects are in the range between 0 and 100. The dataset preprocessing is the first step in the proposed approach, as it is significant to minimize the error rate throughout the learning phase, and to get clear inputs for further processing by data mining techniques [19]. The normalization process is performed by calculating the standard deviation and mean of the data items. Equation 9 shows the normalized item X_i based on the standard deviation σ and the mean μ .

$$X_i = \frac{x_i - \mu}{\sigma} \tag{9}$$

The dataset was separated into two parts randomly for training (80%) and testing (20%) respectively.

The performance of the proposed HGANFIS approach is determined after its training is completed. The student performance prediction of the both the training and the testing sets is evaluated based on the Root Mean Square Error (RMSE) measure of goodness-of-fit. RMSE is a generally used measure of the difference between the real values observed from the modeled environment and the values predicted by the model as defined in (6).

In the proposed HGANFIS, the initial ANFIS model is built based on the Fuzzy C-means method using the settings shown in table 1. After the initial ANFIS model is developed, the optimization of its parameters is carried out by using genetic algorithm. Based on the settings in Table 1, the genetic algorithm searches the entire solution space to find the optimum parameters for ANFIS model. In order to determine the optimum parameters of GA for tuning the ANFIS, many experiments were conducted by varying the crossover and mutation percentages. Table 2 shows that the crossover percentage of 0.4 with the mutation percentage of 0.7 achieved the lowest testing RMSE of 0.104, while the crossover percentage of 0.3 with the mutation percentage of 0.8 achieved the highest value of the testing RMSE, which is 0.116.

Table 1: Settings of the proposed HGANFIS

Technique	Type	Value
ANFIS	Membership function	Gaussian
	Number of Epochs	100
	Fuzzy inference system training method	Hybrid
Genetic Algorithm	Population size	25
	Maximum number of iterations	2000
	Crossover Percentage	0.4
	Mutation Percentage	0.7
	Number of Offsprings	20

Table 2: The performance of the proposed HGANFIS using different crossover and mutation percentages.

Crossover Percentage	Mutation Percentage	Testing RMSE	Training RMSE
0.1	0.7	0.115	0.106
0.2	0.7	0.112	0.115
0.3	0.7	0.110	0.129
0.4	0.7	0.104	0.101
0.5	0.7	0.107	0.108
0.1	0.8	0.108	0.109
0.2	0.8	0.114	0.097
0.3	0.8	0.116	0.096
0.4	0.8	0.114	0.131
0.5	0.8	0.111	0.098

Fig. Fig 3 shows the actual output and target output by the proposed HGANFIS for student performance prediction using the testing data. It can be observed from the fig. 3 that the predicted student performance and the actual student performance coincide well, indicating that the optimized ANFIS has strong generalization capability. Fig. 4 shows the root mean error fluctuations of the proposed GANFIS using the testing data.

Fig. 5 shows the Root Mean Error fluctuations of the proposed HGANFIS using the training data. The training RMSE of the proposed HGANFIS is 0.101. Fig. 6 shows the actual output and target output by the proposed GANFIS for student performance prediction using the training data. It can be observed from the fig. 5 that the predicted student performance and the actual student performance coincide well, indicating that the optimized ANFIS has strong generalization capability.

In order to evaluate the performance of the proposed HGANFIS, it has to be compared with other approaches. Table 3 presents a comparison of student’s performance prediction accuracy between the proposed HGANFIS, our previous research and Neural Networks (NN) in terms of RMSE.

It is clear from the table 3 that proposed HGANFIS outperform the other approaches and achieved the lowest testing RMSE of 0.101. The results of HGANFIS are better than those of ANFIS and NN, due to the fact that the proposed HGANFIS utilizes the genetic algorithm effectively for optimizing the ANFIS parameters. The neural network technique obtained the worst testing RMSE of 0.43.

Table 3: Comparison of prediction accuracy between the proposed HGANFIS , ANFIS and NN in terms of RMSE

Models	Training Data RMSE	Testing Data RMSE
Neural Network	0.467	0.436
Our previous approach [16]	0.146	0.193
The proposed HGANFIS	0.101	0.104

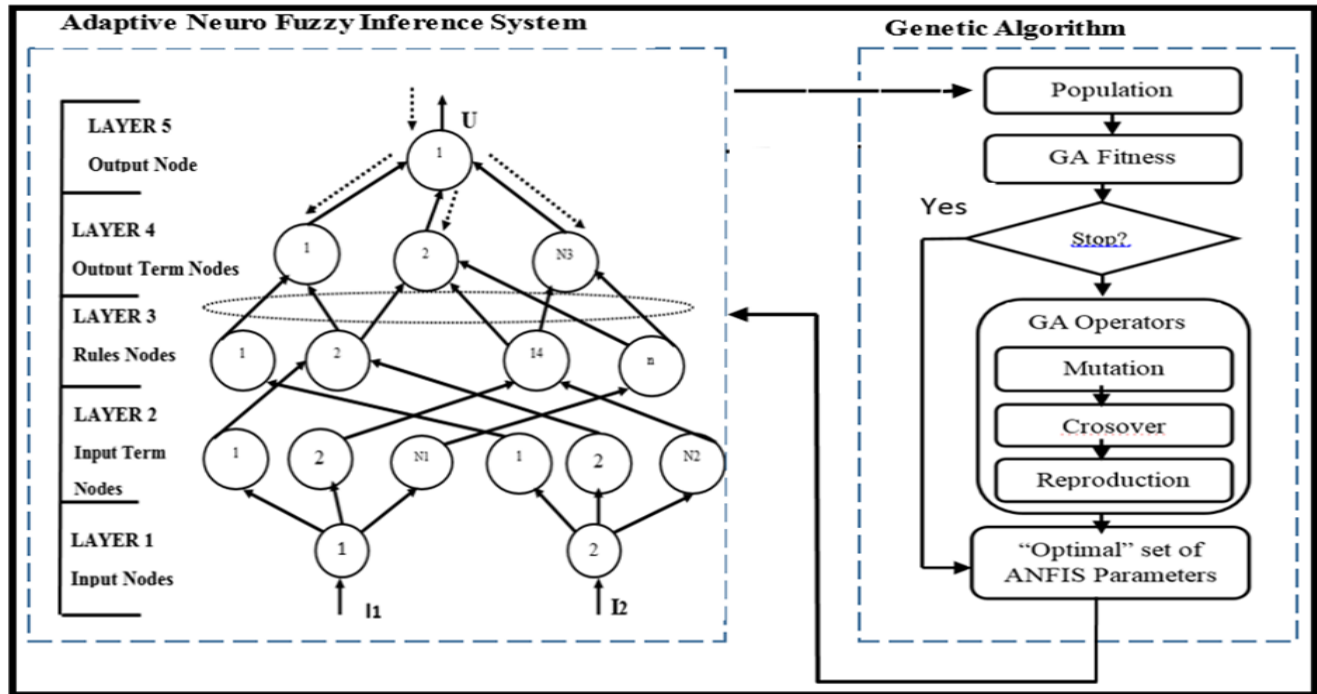


Fig. 1 The proposed HGANFIS approach for student performance prediction.

Step 1: Initialization:

Initialize the ANFIS parameters the Fuzzy C-means method and encode them as chromosomes;

Initialize the Genetic algorithm parameters as follows

Set the Maximum number of iterations = 2000

Set Crossover Percentage = 0.4

Set Mutation Percentage = 0.7

Step 2: Use the training dataset and the initial parameters to train the ANFIS model, compute the objective function of ANFIS model using the equation (8)

Step 3: Evaluate the fitness: Compute the current fitness value (i) of the i th chromosome and use the roulette wheel selection algorithm to find the best chromosome.

Step 4: Compute the difference between the best fitness(k) and the best fitness($K+1$)

If the stopping criteria is not satisfied, go to next step

Else

Go to **step 7**.

Step 5: Perform the crossover and mutation operations to generate the offspring population.

Step 6: Update the population, go to step 2.

Fig. 2 The steps of implementing the proposed HGANFIS

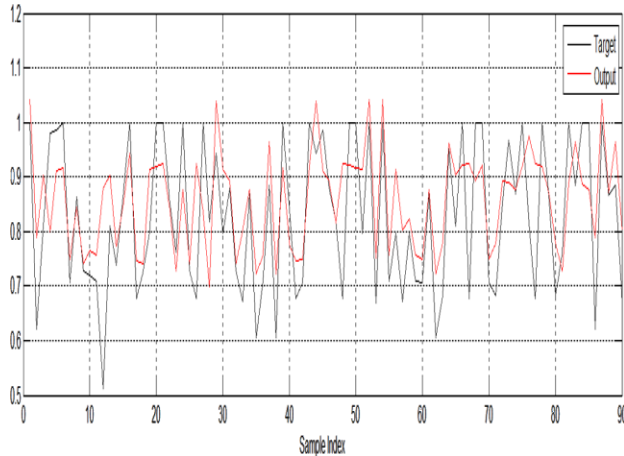


Fig. 3 The actual output and target output by the proposed HGANFIS for student performance prediction in testing phase.

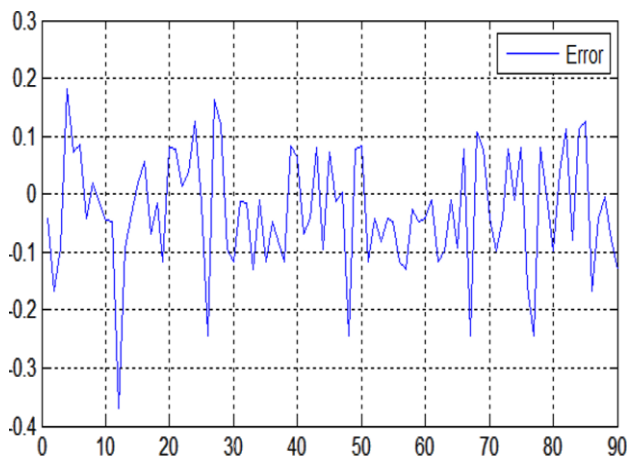


Fig. 4 The testing error of the proposed HGANFIS

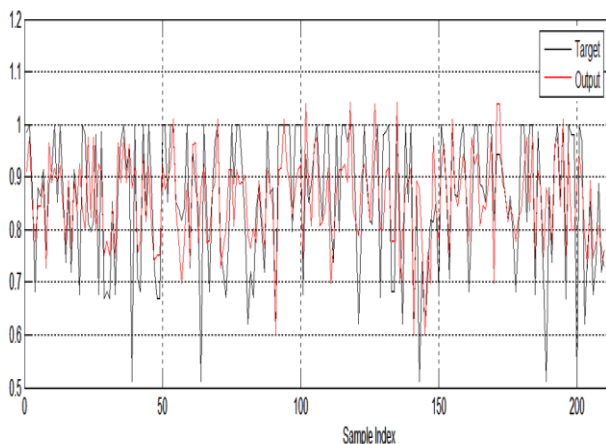


Fig. 6 The actual output and target output by the proposed HGANFIS for student performance prediction in the training phase.

6. Conclusion

Student performance prediction is one of the most effective solutions to improving the student's academic performance. If the academic institution can early properly predicts the student's performance, it can arrange and take the appropriate action accordingly. This, in turn, contributes significantly to the student's success in advancing his/her career. In this paper, an improvement in the prediction power of the adaptive fuzzy inference system is presented by integrating the genetic algorithms with the adaptive neuro fuzzy inference system for the accurate prediction of student's academic performance. The genetic algorithm can optimize ANFIS parameters that can improve the prediction results for best performance.

The comparison results of the student performance predicted by HGANFIS and ANFIS models show that the proposed hybrid approach HGANFIS is very promising, because it gives advantages over the other models in terms of accuracy. Future work will extend the scope by considering advanced pre-process and analysis approaches for further improvement of the prediction accuracy.

Acknowledgement

This Project was funded by the Deanship of Scientific Research (DSR), at King Abdulaziz University, Jeddah, Saudi Arabia, under grant no. G-350-830-1439. The authors, therefore, acknowledge with thanks DSR for technical and financial support.

References

- [1] Baker, Ryan Shaun, and Paul Salvador Inventado. "Educational data mining and learning analytics." Learning analytics. Springer New York, 2014, pp. 61-75.
- [2] C. Romero, S. Ventura, Educational data mining: A review of the state of the art, *Trans. Sys. Man Cyber Part C* 40 (6) ,2010, pp. 601–618.
- [3] Chamizo-Gonzalez, Julian, et al. "Educational data mining for improving learning outcomes in teaching accounting within higher education." *The International Journal of Information and Learning Technology* 32.5, 2015, pp. 272-285.
- [4] Yukselturk, Erman, Serhat Ozekes, and Yalın Kılıç Türel. "Predicting dropout student: an application of data mining methods in an online education program." *European Journal of Open, Distance and E-learning* 17.1, 2014, pp. 118-133.
- [5] Sharma, Anju, et al. "A comparative study of support vector machine, artificial neural network and bayesian classifier for mutagenicity prediction." *Interdisciplinary Sciences: Computational Life Sciences* 3.3, 2011, pp.232-239.
- [6] B. K. Bhardwaj and S. Pal, "Data mining: a prediction for performance improvement using classification," *International Journal of Computer Science and Information Security*, vol. 9, no.4, 2011, pp. 1–5.

- [7] S. Huang and N. Fang, "Predicting student academic performance in an engineering dynamics course: a comparison of four types of predictive mathematical models," *Computers & Education*, vol. 61, 2013, pp. 133–145.
- [8] R. Singh, A. Kainthola, and T. N. Singh, "Estimation of elastic constant of rocks using an ANFIS approach," *Applied Soft Computing Journal*, vol. 12, no. 1, 2012, pp. 40–45.
- [9] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 3, 1993, pp. 665–685.
- [10] K. S. Pal and S. Mitra, *Neuro-Fuzzy Pattern Recognition: Methods in Soft Computing*, JohnWiley & Sons, New York, NY, USA, 1999.
- [11] Barber, R., & Sharkey, M. Course correction: Using analytics to predict course success. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, 2011, pp. 259–262.
- [12] Wolff A, Zdrahal Z, Nikolov A, Pantucek M. Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In *Proceedings of the third international conference on learning analytics and knowledge 2013 Apr 8*, pp. 145-149.
- [13] Ruby, Jai, and K. David. "Predicting the Performance of Students in Higher Education Using Data Mining Classification Algorithms—A Case Study." *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 2014.
- [14] Naser, Samy Abu, et al. "Predicting student performance using artificial neural network: in the Faculty of Engineering and Information Technology." *International Journal of Hybrid Information Technology* 8.2, 2015, pp.221-228.
- [15] Ajith P., Tejaswi B, Sai MSS. Rule Mining Framework for Students Performance Evaluation. *International Journal of Soft Computing and Engineering*, 2(6), 2013, pp. 201–206.
- [16] Altaher, Altyeb, and Omar BaRukab. "Prediction of Student's Academic Performance Based on Adaptive Neuro-Fuzzy Inference." *International Journal of Computer Science and Network Security (IJCSNS)* 17.1 ,2017, pp. 165-169.
- [17] Jain, M., Singh, M., Chandra, A., & Williamson, S. S. Sensorless control of permanent magnet synchronous motor using ANFIS based MRAS. In *Electric Machines & Drives Conference (IEMDC), 2011 IEEE International*, 2011, pp. 599-606.
- [18] Sharifian, M. B. B., Shafiei, M., Sadeghi, M. S., & Golestaneh, F. Direct torque control of brushless DC motor drives based on ANFIS controller with fuzzy supervisory learning. In *Electrical Machines and Systems (ICEMS), 2011 International Conference on*, 2011, pp. 1-6
- [19] Goldberg, David E., and John H. Holland. "Genetic algorithms and machine learning." *Machine learning* 3.2 (1988): 95-99.
- [20] McInerney M, Dhawan AP. Use of genetic algorithms with backpropagation in training of feedforward neural networks. In *Neural Networks, 1993., IEEE International Conference on 1993* (pp. 203-208). IEEE.
- [21] Alemdag, S., Z. Gurocak, A. Cevik, A. F. Cabalar, and C. Gokceoglu. "Modeling deformation modulus of a stratified sedimentary rock mass using neural network, fuzzy inference and genetic programming." *Engineering Geology* 203,2016,pp. 70-82.
- [22] Asadi, Ehsan, et al. "Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application." *Energy and Buildings* 81,2014,pp. 444-456.
- [23] Castillo, P. A., Merelo, J. J., Prieto, A., Rivas, V., & Romero, G.G-Prop: Global optimization of multilayer perceptrons using GAs. *Neurocomputing*, 35(1), 2000,pp. 149-163.