



A Novel Metaheuristic Approach to Optimization of Neuro-Fuzzy System for Students' Performance Prediction

Kashif Hussain^{1*}, Noreen Talpur², Muhammad Umar Aftab³, Zakria³

¹Institute of Fundamental and Frontier Sciences,
 University of Electronic Science and Technology of China, Chengdu, 610054, CHINA

²Faculty of Science and Information Technology,
 Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750, MALAYSIA

³School of Information and Software Engineering,
 University of Electronic Science and Technology of China, Chengdu, 610054, CHINA

*Corresponding Author

DOI: <https://doi.org/10.30880/jscdm.2020.01.01.001>

Received 20th December 2019; Accepted 10th February 2020; Available online 1st March 2020

Abstract: Data mining is being increasingly leveraged in educational settings for achieving various different outcomes including students' learning patterns, course and teaching outcome assessment, and students' expected achievement prediction. Utilizing data collected from daily curricular and non-curricular activities, machine learning techniques have benefited administrators in making efficient decisions. Based on students' behavioral information, this research proposes student performance prediction model using fuzzy-based neural network (FNN) trained by a novel metaheuristic approach. Because original gradient-based learning method associated with FNN limits its performance, this research employs Henry Gas Solubility Optimization (HGSO) algorithm for tuning FNN parameters. The empirical analysis suggests superiority of results produced by the proposed approach as compared with the FNN trained by the competitive methods.

Keywords: Fuzzy systems, student performance prediction, Henry gas solubility optimization

1. Introduction

Today, technology is aggressively adopted by educational institutes to maximize output of existing resources, to benefit both students and the teachers. The use of technology-based learning makes it possible to collect enormous amount of data pertaining to student and teacher performance. This data is then used for extracting meaningful information that deems useful for administration in making efficient decisions. For this purpose, literature shows significant advantages of data mining and machine learning techniques with regards to educational data mining [1,2,3]. Utilizing behavioral and performance data of students and teachers, researchers have applied these methods to discover knowledge in hidden patterns, to assess the effectiveness of courses, teaching methodologies, curriculum, and overall education system [4]. More specifically, the educational data mining is now being used to predict students' performances in order to identify the vulnerable ones at risk of failing the upcoming examinations so that more focus could be drawn on them to solve any difficulties [5,6].

*Corresponding author: k.hussain@uestc.edu.cn

Students' performance prediction (SPP), in the context of machine learning and data mining, is considered NP-hard problem as it involves several multivariate data types spread on highly dynamic hyperspace. This makes SPP non-convex and non-linear, difficult for simple statistical approaches [7]. That said, there can be found extensive research with applications of data mining techniques. Realizing the importance of valuable data generated from technology-enhanced learning systems, Buenano-Fernandez *et al.* [7] utilized decision trees for predicting performance of computer engineering students based on their historical academic data. In [8], semi-supervised machine learning approach was used to predict performance of secondary school students. The authors contend that robust prediction results can be achieved by using semi-supervised machine learning when the data is largely unlabeled. Based on comparative analysis performed in [9], Gradient Boosting (GB) achieved better prediction results than Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF) when implemented on classifying students with poor performance. Cazarez and Martin [10] employed Artificial Neural Network (ANN) for predicting performance of students attaining online education. The researchers found that ANN predicting accuracy was higher than statistical methods. Similarly, a previous study in [11] also solved SPP problem by ANN while using students' behavioral and performance related data from various different countries.

Fuzzy-based systems have also been leveraged in this research line; for instance, Son and Fujita [12] proposed the use of multiple Adaptive Neuro-Fuzzy Inference Systems (ANFIS) with representative sets. Based on empirical analysis of three different test cases, the research affirmed superiority of the prediction accuracy as compared to standard ANFIS, OneR, and Random Tree methods. Likewise, [13,14,15] also employed ANFIS systems on the same problem and claimed robust results in comparison with ANN and Multiple Linear Regression (MLR) techniques. However, generally, it is suggested that ANFIS suffers from high computational cost and network complexity due to curse of dimensionality problem. The performance of ANFIS suffers when the input-size is considerably large [16]. On the other hand, gradient-based learning approach in standard ANFIS is prone to stuck in local optima when problem is highly non-linear [17]. For addressing learning drawbacks, several studies have used metaheuristic algorithms, as these methods have robust search strategies that do not need gradient information for finding globally optimum solutions [18]. Keeping in view the aforementioned issues in standard ANFIS, this current study modifies its architecture for enhanced prediction accuracy and less computational complexity. The study employs recently introduced metaheuristic algorithm, known as Henry Gas Solubility Optimization (HGSO) [19], for updating adjustable parameters in ANFIS network. Moreover, the large rule-base in architecture is optimized by selecting only the useful rules that have effective contribution in enhancing model accuracy. Later, the proposed modified ANFIS is then implemented on solving SPP problem. The empirical results suggest robustness of the proposed approach as compared to several other methods used in this research.

This paper is organized as follows. The upcoming section explains methodology adopted in this study for solving SPP problem. In this section, ANFIS and HGSO are elaborated in detail. Furthermore, the proposed approach of training ANFIS parameters and optimizing its rule-base is also presented in this section. Section 3 presents experimental environment along with detail of the dataset related to SPP problem. This section also reports results and competitive analysis. Finally, Section 4 duly concludes this study.

2. Methodology

This section explains standard ANFIS algorithm with gradient-based learning approach. Later in this section, the metaheuristic approach HGSO is elaborated in detail. The application of ANFIS trained by HGSO on SAP problem is also elaborated here.

2.1 Adaptive Neuro-Fuzzy Inference System (ANFIS)

Adaptive neuro-fuzzy inference system (ANFIS) was introduced by Jang in 1993 [20]. This model has produced efficient results as compared to several other machine learning techniques [21]. The architecture of ANFIS comprises of five layers which transform inputs to the approximated outcome, by performing fuzzification, generating rules, and defuzzification for computing output. Since, it is an adaptive technique, it learns from training data by adjusting its weights in membership functions layer and defuzzification layer. For the model training, original ANFIS employs two-pass learning algorithm utilizing gradient-based learning approach. In the forward pass, ANFIS uses least squared estimation method for tuning consequent parameters, on the other hand, gradient-descend algorithm is used for minimizing error measures by adjusting membership function parameters. The layer-by-layer description of ANFIS architecture is given as following, and illustrated via Fig. 2.

The first layer, after inputs, is a fuzzification layer which transforms crisp input values into related membership degree. Here, it is noteworthy to mention that the choice of membership function varies from triangular to Gaussian to trapezoidal. Several other types of membership functions can also be used based on application suitability. The number of membership functions m per input is also problem dependent, and chosen by the user. This layer contains $m \times n$ membership functions. The membership degree for i th input using j th membership function can be calculated as Eq. (1):

$$O_{ij}^1 = \mu_{ij}(x_i), i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, m\} \quad (1)$$

where O_{ij}^1 is membership degree for i th input x_i in layer 1.

Once membership degrees for each input is calculated, these values are input to the product layer which performs product operation for generating rules and corresponding weights. In grid partitioning method, the number of rules generated by ANFIS can be determined as m^n . The weight of each rule r can be computed as Eq. (2):

$$O_r^2 = w_r = \prod_{i=1}^n \mu_{ir}(x_i), r \in \{1, 2, \dots, k\} \quad (2)$$

where O_r^2 refers to weight for r th rule in layer 2. The weights in this layer are then normalized in third layer to determine rules' firing strength against all k rules in ANFIS model, using Eq. (3):

$$O_r^3 = \bar{w}_r = \frac{w_r}{\sum_{i=1}^k w_i}, r \in \{1, 2, \dots, k\} \quad (3)$$

where \bar{w}_r is normalized weight for r th rule. \bar{w}_r is now applied on fourth layer in defuzzification operation for each rule to compute the aggregated consequent part for each rule, using Eq. (4):

$$O_r^4 = \bar{w}_r \times f_r, f_r = \sum_{i=1}^n x_i q_{ir}, r \in \{1, 2, \dots, k\} \quad (4)$$

where O_r^4 is consequent part of r th rule generated in layer 4, f_r is polynomial function maintaining sum of all inputs x and corresponding consequent parameters q . The output of defuzzification for all the k rules is then aggregated in fifth layer to generate single output value approximating target class or value in actual training instance, as $O_r^5 = \sum_{i=1}^k O_r^4$.

Since gradient-based learning approach has serious drawbacks when applied on highly non-convex problem, using metaheuristic algorithms produces better results instead [18]. Therefore, this research employs recently introduced efficient optimization method HGSO on training all the adjustable parameters of the proposed ANFIS model. Moreover, this research also modifies ANFIS architecture by adding additional layer for finding optimized rule-set that maintains efficient rules that produce better approximated values. The proposed ANFIS architecture is explained later in this section.

2.2 Henry Gas Solubility Optimization (HGSO)

Fatma *et al.* [19] introduced a novel metaheuristic algorithm in 2019, inspired from a well-known Henry's law of physics related to gas solubility, hence called Henry Gas Solubility Optimization (HGSO) algorithm. Based on extensive experimental analysis, the research proved that HGSO not only outperformed other recent and established metaheuristic algorithms, but also maintained outstanding ability of finding global optimum locations. According to authors, the proposed HGSO maintains significant balance between exploration and exploitation abilities. A greater detail related to Henry's law and HGSO inspiration can be found in original work in [19]; however, a precise description is given in this section.

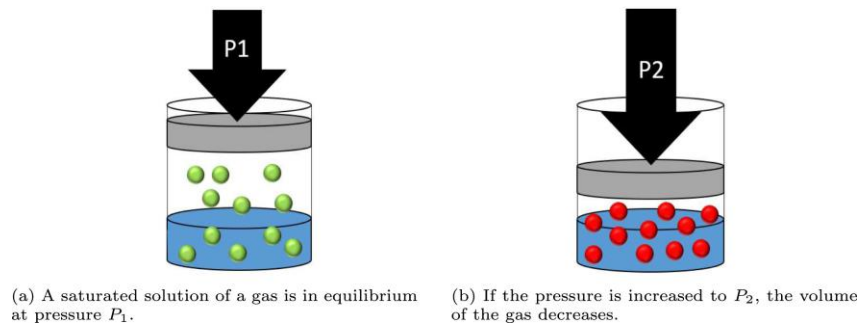


Fig. 1 - Gas solubility under certain pressures [19]

Similar to several other population-based metaheuristic algorithms, HGSO also initiates search by generating random candidate solutions which are gases that are meant to be dissolved in a solvent with a specific pressure (Fig. 1). The population of N gases are initialized with random positions using Eq. (5):

$$x_i = lb_i + rand \times (ub_i - lb_i), i \in \{1, 2, \dots, N\} \quad (5)$$

where x_i , lb_i , and ub_i are respectively i th candidate solution, lower bound and upper bound of search space. $rand$ is randomly generated number within 0 and 1. Apart from position of each gas particle, there are other properties of it which have to be initialized using Eq. (6):

$$\begin{aligned} H_j^t &= l_1 \times rand, \\ P_{ij}^t &= l_2 \times rand, \\ C_j^t &= l_3 \times rand, \\ l_1 &= 0.05, l_{21} = 100, l_3 = 0.01, i \in \{1, 2, \dots, N\}, j \in \{1, 2, \dots, K\} \end{aligned} \quad (6)$$

where H_j^t is Henry's coefficient of j th type, P_{ij}^t is particle pressure on i th gas in j th cluster, and C_j^t is a constant for cluster j . In Eq. (6), l_1 to l_3 are different constant for controlling randomization in HGSO, whereas t is iteration counter.

After initialization step, the gas particles are clustered into K groups and evaluated using fitness function. Each cluster j maintains its best gas particle. During evaluation, the gas particle with best fitness value in whole population is also remembered in x_{best} . After evaluation, new generation ($t+1$) of candidate solution is generated. For that, it is important to update Henry's coefficient using Eq. (7):

$$H_j^{t+1} = H_j^t \times \exp\left(-C_j^t \times \left(\frac{1}{T^t} - \frac{1}{T^\theta}\right)\right), T^t = \exp\left(\frac{t}{t_{\max}}\right) \quad (7)$$

where T^t and T^θ are respectively temperature at t iteration and constant with value 298.15, whereas t_{\max} is maximum iterations number. The Henry's constant computed in Eq. (7) is then used to compute gas solubility S_{ij} for i th gas particle in j th cluster, as in Eq. (8):

$$S_{ij} = \delta \times H_j^{t+1} \times P_{ij}^t \quad (8)$$

Since, all the required properties related to gas solubility have been calculated, it is now turn to update position of every i th gas particle in j th cluster, as in Eq. (9):

$$\begin{aligned} x_{ij}^{t+1} &= x_{ij}^t + F \times rand \times \gamma \times (x_{j,best}^t - x_{ij}^t) + F \times rand \times \alpha \times (S_{ij}^t \times x_{best} - x_{ij}^t), \\ \beta \times \exp\left(-\frac{f(x_{best}) + \eta}{f(x_{ij}^t) + \eta}\right), \eta &= 0.05 \end{aligned} \quad (9)$$

where a and β are constant values, F is flag that determines search direction, $f(x_{best})$ and $f(x_{ij}^t)$ are fitness values of best solution in population and i th candidate solution respectively. To avoid trapping in local optimal regions, the HGSO algorithm destroys worst candidate solutions and re-initializes them using Eq. (5). The number of worst solutions N_w is computed using Eq. (10):

$$N_w = N \times [rand \times (c_2 - c_1) + c_1], c_1 = 0.1, c_2 = 0.2 \quad (10)$$

2.3 Proposed ANFIS using HGSO

For solving SPP problem, the proposed ANFIS model takes 16 inputs with two membership functions each, of Gaussian type. The inputs are features or attributes from A_1 to A_{16} listed in Table 1 and depicted in Fig. 2. The predicted output value is “Pass” or “Fail”, which is expected result that may help administration of the educational systems take proper corrective actions deemed necessary to avoid possible failure of any student. Because, the proposed ANFIS model uses grid partitioning method, hence total number of rules in the system are 65536.

Instead of applying original gradient-based learning approach, the proposed method employs metaheuristic algorithm HGSO on tuning both the premise and consequent parameters. Because, ANFIS suffers from curse of dimensionality issue, the significantly large inputs generated highly complex network due to numerous rules as the case in this study. It is important to apply rule-filtering method for selecting only those rules that contribute more in predicted output. The rules which do not contribute effectively in output value are considered weak rules in this study, and are filtered using threshold error measure. This rule-filtering approach optimizes the knowledge-base or rule-base by selecting strong rules only. As depicted in Fig. 2, the five-layer architecture is modified by adding additional layer, namely “Optimize Rules”, before output layer. Therefore, the proposed ANFIS architecture becomes six-layer network instead of five-layer system. The proposed ANFIS architecture not only produces efficient approximation but also optimizes the network complexity by reducing the rule-base when input-size is large enough.

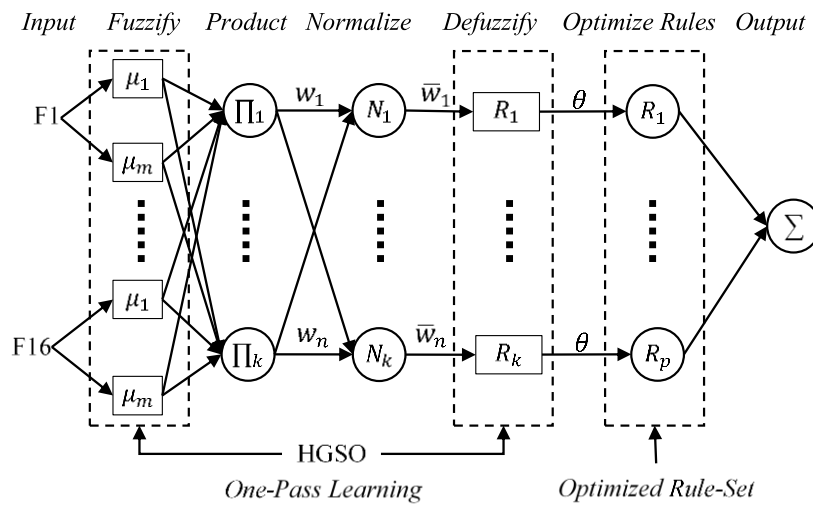


Fig. 2 - Proposed ANFIS-HGSO for solving SPP problem

3. Experimental Results

3.1 Experimental Environment

The proposed ANFIS-HGSO was implemented on SPP problem using Kalboard 360 dataset collected from Learning Management System (LMS) [22], which contains multivariate data collected from students with 9 different countries of origin including both male and female students. The dataset detail is given in Table 1. The dataset contains 480 instances with 16 attributes having no missing data. From attribute A_1 to A_4 are related to demographic features, A_5 to A_{10} belong to academic information, A_{11} to A_{12} provide information about parents participation in learning process, and remaining A_{13} to A_{16} are about student behaviors towards learning.

For effective learning purpose, the dataset was split into training and testing sets with 70:30 ratio; meaning that 336 instances were used for training and the remaining 144 for testing. The model was implemented on MATLAB 2017b on computing environment with Intel® Core™ i5 (3.40 GHz) CPU with RAM 8GB, and operating system Microsoft Windows 10. The parameter settings of the used methods are listed in Table 2. Apart from ANFIS, other machine learning algorithms were also tested on the same problem using WEKA [cite{b23}], for results comparison maximum iterations 50.

3.2 Evaluation Measures

In the study, performance of the proposed and competitive approaches is evaluated using different measures. The integration of ANFIS with metaheuristic algorithms is evaluated in terms of accuracy, optimized rule-base, and number of iterations the method took to converge. Moreover, with the help of confusion matrix (Table 3), four different evaluation criteria are also used: accuracy, precision, recall, and F-measure. Accuracy refers to instances correctly predicted against total instances, Eq. (11) defines it mathematically. Precision tells about reliability of the predicted

value by a classifier, it is measured as Eq. (12). Recall, measured as in Eq. (13), affirms the quality of predicted values. Additionally, F-Measure is weighted average of precision and recall, calculated as in Eq. (14).

Table 1 - Student data features and types

	Features	Type	Sample Data
A ₁	Gender	Nominal	Male or Female
A ₂	Nationality	Nominal	i.e. Kuwait, USA, Jordan, etc.
A ₃	Place of birth	Nominal	i.e. Kuwait, USA, Jordan, etc.
A ₄	Parent responsible for student	Nominal	Mother or Father
A ₅	Educational Stage	Nominal	Low level, middle school, high school
A ₆	Grade Level	Nominal	G-01 to G-12
A ₇	Section ID	Nominal	A, B, C
A ₈	Semester	Nominal	First or Second
A ₉	Topic	Nominal	i.e. English, Spanish, Science, etc.
A ₁₀	Student Absence Days	Nominal	above-7, under-7
A ₁₁	Parent Answering Survey	Nominal	Yes or No
A ₁₂	Parent School Satisfaction	Nominal	Yes or No
A ₁₃	Discussion groups	Numeric	[0, 100]
A ₁₄	Visited Resources	Numeric	[0, 100]
A ₁₅	Raised Hand on Class	Numeric	[0, 100]
A ₁₆	Viewing Announcements	Numeric	[0, 100]

Table 2 - Parameter settings of the selected algorithms

Methods	Parameter Settings
BA	N=50, Frequency=[0,5], A ₀ =0.95, r ₀ =0.9
PSO	N=50, C ₁ =2, C ₂ =2, w=[0.9,0.4]
ABC	N=50, Limit=N×D
CS	N=50, P _a =0.25
HGSO	N=50, Cluster size=5, M ₁ and M ₂ = 0.1 and 0.2, β=a = K=1
ANFIS	Membership functions=Gaussian type; 2 for each input

Table 3 - Confusion matrix for evaluation measures

		Actual Value	
		Positive	Negative
Predicted Value	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$Precision = \frac{TP}{TP + FN} \quad (12)$$

$$Recall = \frac{TP}{TP + FP} \quad (13)$$

$$F\text{-measure} = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (14)$$

3.3 Results

The results of ANFIS trained by HGSO and other metaheuristic algorithms, while solving SPP problem, are depicted via Fig. 3. It can be observed that ANFIS trained by HGSO produced highest accuracy as compared to ANFIS trained by BA [29], PSO [30], ABC [31], and CS [32] algorithms. Moreover, the proposed approach generated 288

rules in optimum rule-set out of total 65536 rules, which is smaller than the rule-set produced by PSO, ABC, and CS. However, these rules are greater than 256 rules optimized by BA. Nevertheless, the rule-set of ANFIS trained by HGSO achieved the best accuracy among other algorithms. Fig. 3 also reveals that the ANFIS trained by HGSO converged in 38 iterations with 96.04% accuracy, whereas the ANFIS trained by BA, PSO, and CS consumed maximum iterations. ANFIS trained by ABC, however, converged in 46 iterations by achieving second best accuracy of 92.29%.

The statistics generated from confusion matrix are presented in Table 4. As shown in Table 4, ANFIS trained by HGSO outperforms other approaches. The best accuracy is recorded 96.04% and 74.38% as the worst in case of HGSO and BA, respectively. ANFIS training approach using ABC on SPP problem obtained 92.29% accurate predictions. On the other hand, ANFIS trained by PSO and CS remained mediocre performers with 87.29% and 90.63% accuracies, respectively. As far as the precision of prediction results generated by the competitive methods, Table 4 affirms superiority of ANFIS optimized by HGSO. Out of 480 instances in the dataset, the number of instances correctly predicted by ANFIS trained by BA, PSO, ABC, CS, and HGSO are respectively 357, 419, 443, 435, and 461. The highest recall value is also reported for ANFIS trained by HGSO, which indicates that the proposed method generated rule-set that successfully predicted 461 instances to the unpredicted and predicted instances. A combined effect of precision and recall is shown by F-measure value which is 95.49 for ANFIS trained by HGSO, and it is least 36.50 with training approach utilizing BA algorithm.

Compared to various other data mining techniques applied in this research using Weka, as well as, from recent literature, Table 5 presents the comparable accuracies. According to the comparative analysis, it can be suggested the most accurate prediction results were achieved by the proposed approach in this study. The other machine learning methods including Naive Bayes, k -NN, standard ANFIS with gradient-based learning, and other ANFIS training approaches. The experiments performed in this study confirm that least performer was k -NN algorithm with 71.93% accurate prediction results. Naive Bayes achieved 78.77% accuracy as compared to 88.64% in case of standard ANFIS algorithm. From recent literature, the study performed in [24] achieved second best accurate results with 94.10%. Overall, from comparisons, it can be easily affirmed that the proposed ANFIS training method using HGSO not only outperformed recently introduced methods from literature, but also various other machine learning algorithms implemented on SPP in this study.

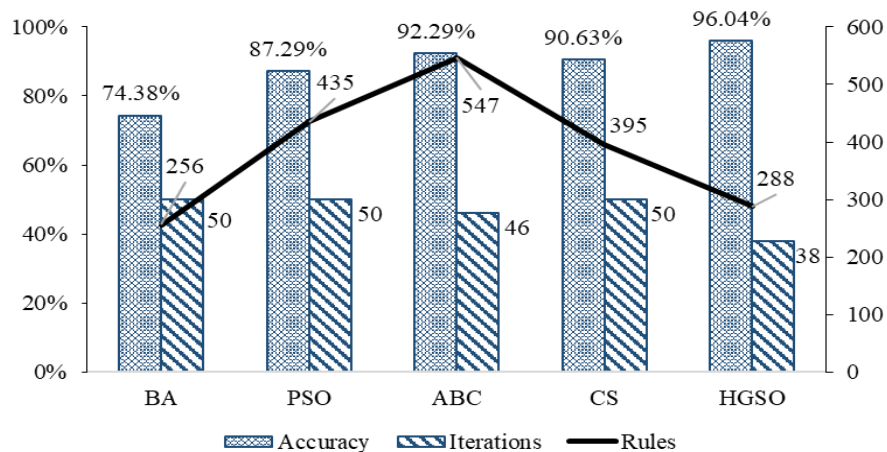


Fig. 3 - ANFIS trained by HGSO and other metaheuristic algorithms for solving SPP problem

Table 5 - ANFIS trained by HGSO and other metaheuristic algorithms for solving SPP problem

	GA	PSO	ABC	CS	HGSO
Accuracy	74.38	87.29	92.29	90.63	96.04
Precision	61.14	84.50	91.44	88.94	95.71
Recall	66.05	84.92	91.86	88.50	95.26
F-measure	63.50	84.71	91.65	88.72	95.49

Table 5 - Comparative results of the proposed and other methods on SPP problem

Method	Ref.	Accuracy
Random Forest	[28]	85.00%
ANN	[27]	81.70%
Hybrid Stacking ensemble technique	[26]	84.30%
DNN	[25]	86.01%
ADDE	[25]	94.10%
Tech	[24]	84.30%
Naive Bayes	This Study	78.77%
k-NN	This Study	71.93%
ANFIS	This Study	88.64%
ANFIS-GA	This Study	74.38%
ANFIS-PSO	This Study	87.29%
ANFIS-ABC	This Study	92.29%
ANFIS-CS	This Study	90.63%
ANFIS-HGSO	Proposal model	96.04%

4. Conclusion

Student performance prediction plays important role in taking appropriate actions, in advance, in order to achieve optimum results from students. It helps administrators of educational systems enhance quality of educational facilities. This study solved student performance prediction problem using ANFIS trained by recently introduced efficient metaheuristic technique HGSO. Because of large input size in the problem under consideration, this study proposed optimizing rule-base by selecting potential rules that produced efficient approximation results. For comparative analysis, the study also used other metaheuristic algorithms like BA, PSO, ABC and CS and found that the proposed approach outperformed the competitive methods. Besides, several other machine learning algorithms like Naive Bayes, k-NN, and standard ANFIS were also employed on the same problem. Additionally, comparison with other methods applied in recent literature, like Random Forest, ANN, Hybrid Stacking, DNN, and ADDE, also suggested that the proposed approach achieved 96.04% accuracy which was the highest among all other approaches compared in this study. The empirical results also suggested that the proposed ANFIS rule-base optimization technique is suitable for applications with large input-size.

Acknowledgement

The authors would like to acknowledge the University of Electronic Science and Technology of China, Chengdu and Universiti Teknologi PETRONAS, Bandar Seri Iskandar.

References

- [1] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," *Procedia Computer Science*, vol. 72, pp. 414-422, 2015.
- [2] H. Wan, K. Liu, Q. Yu et al., "Pedagogical Intervention Practices: Improving Learning Engagement Based on Early Prediction," *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 278-289, 2019.
- [3] R. Patil, S. Salunke, M. Kalbhor, et al., "Prediction System for Student Performance Using Data Mining Classification," In *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Aug. 2018, pp. 1-4.
- [4] I.E. Livieris, T. Kotsilieris, V. Tampakas et al., "Improving the evaluation process of students' performance utilizing a decision support software," *Neural Computing & Applications*, vol. 31, no. 6, pp. 1683-1694, 2019.
- [5] P. Kamal and S. Ahuja, "An ensemble-based model for prediction of academic performance of students in undergrad professional course," *Journal of Engineering, Design and Technology*, vol. 17, no. 4, pp. 769-781, 2019.
- [6] I. E. Livieris, T. Mikropoulos, P. Pintelas, "A decision support system for predicting students' performance," *Themes in Science & Technology Education*, vol. 9, no. 1, pp. 43-57, 2016.
- [7] D. Buenano-Fernandez, D. Gil, S. Lujan-Mora, "Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study," *Sustainability*, vol. 11, no. 10, pp. 2071-1050, 2019.

- [8] I. E. Livieris, K. Drakopoulou, V. T. Tampakas, *et al.*, "Predicting Secondary School Students' Performance Utilizing a Semi-supervised Learning Approach," *Journal of Educational Computing Research*, vol. 57, no. 2, pp. 448-470, 2019.
- [9] A. Polyzou and G. Karypis, "Feature Extraction for Next-Term Prediction of Poor Student Performance," *IEEE Transactions on Learning Technologies*, vol. 12, no. 2, pp. 237-248, 2019.
- [10] R. L. U. Cazarez and C. L. Martin, "Neural Networks for Predicting Student Performance in Online Education," *IEEE Latin America Transactions*, vol. 16, no. 7, pp. 2053-2060, 2018.
- [11] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Preprocessing and analyzing educational data set using X-API for improving student's performance," In *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, Nov. 2015, pp. 1-5.
- [12] L. H. Son and H. Fujita, "Neural-fuzzy with representative sets for prediction of student performance," *Applied Intelligence*, vol. 49, no. 1, pp. 172-187, 2019.
- [13] S. Maitra, S. Madam, and P. Mahajan, "An Adaptive Neural Fuzzy Inference System for prediction of student performance in Higher Education," In *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Oct. 2018, pp. 1158-1163.
- [14] J. -F. Chen and Q. H. Do, "Prediction of student academic performance using an ANFIS approach," *International Journal of Information and Management Sciences*, vol. 25, no. 4, pp. 371-389, 2014.
- [15] J. -F. Chen and Q. H. Do, "A cooperative Cuckoo Search – hierarchical adaptive neuro-fuzzy inference system approach for predicting student academic performance," *Journal of Intelligent & Fuzzy Systems*, vol. 27, no. 5, pp. 2551-2561, 2014.
- [16] N. Talpur, M. N. M. Salleh, K. Hussain, H. Ali, "Modified ANFIS with Less Model Complexity for Classification Problems," In *Advances in Intelligent Systems and Computing*, S. Omar, H. W. Suhaili, S. Phon-Amnuaisuk (eds), Springer Cham, vol. 888, pp. 36-47, Oct. 2018.
- [17] M. Behmanesh, M. Mohammadi, V. S. Naeini, "Chaotic time series prediction using improved ANFIS with imperialist competitive learning algorithm," *International Journal of Soft Computing and Engineering*, vol. 4, no. 4, pp. 25-33, 2014.
- [18] D. Karaboga and E. Kaya, "Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey," *Artificial Intelligence Review*, pp. 1-31, 2018.
- [19] F. H. Hashim, E. H. Houssein, M. S. Mabrouk, *et al.*, "Henry gas solubility optimization: A novel physics-based algorithm," *Future Generation Computer Systems*, vol. 101, pp. 646-667, 2019.
- [20] J. -R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665-685, 1993.
- [21] M. N. M. Salleh, N. Talpur, and K. Hussain, "A Modified Neuro-Fuzzy System Using Metaheuristic Approaches for Data Classification," *Artificial Intelligence - Emerging Trends and Applications*, Marco Antonio Aceves-Fernandez, IntechOpen, 2018.
- [22] A. Elaf, H. Thair, and A. Ibrahim, "Mining Educational Data to Predict Student's academic Performance using Ensemble Methods," *International Journal of Database Theory and Application*, vol. 9, pp. 119-136, 2016.
- [23] A. Rohit and S. Suman, "Comparative Analysis of Classification Algorithms on Different Datasets using WEKA," *International Journal of Computer Applications*, vol. 54, no. 13, pp. 21-25, 2012.
- [24] S. -S. M. Ajibade, N. B. Ahmad, S. M. Shamsuddin, "A Data Mining Approach to Predict Academic Performance of Students Using Ensemble Techniques," In *Advances in Intelligent Systems and Computing*, A. Abraham, A. Cherukuri, P. Melin, N. Gandhi (eds), Springer Cham, vol. 940, pp. 749-760, Apr. 2018.
- [25] S. -S. M. Ajibade, N. B. Ahmad, S. M. Shamsuddin, "A Novel Hybrid Approach of Adaboostm2 Algorithm and Differential Evolution For Prediction Of Student Performance," *International Journal of Scientific & Technology Research*, vol. 8, no. 7, pp. 65-67, 2019.
- [26] Bendangnugsung and P. Prabu, "Students' Performance Prediction Using Deep Neural Network," *International Journal of Applied Engineering Research*, vol. 13, no. 2, pp. 1171-1176, 2018.
- [27] O. W. Adejo and T. Connolly, "Predicting student academic performance using multi-model heterogeneous ensemble approach," *Journal of Applied Research in Higher Education*, vol. 10, no. 1, pp. 61-75, 2018.
- [28] M. M. Alam, K. Mohiuddin, A. K. Das, *et al.*, "A Reduced feature based neural network approach to classify the category of students," In *Proceedings of the 2nd International Conference on Innovation in Artificial Intelligence*, pp. 28-32, 2018, ACM.
- [29] X. -S. Yang, "A New Metaheuristic Bat-Inspired Algorithm," In *Studies in Computational Intelligence*, Gonzalez J.R., Pelta D.A., Cruz C., Terrazas G., Krasnogor N. (eds), Springer Berlin, vol. 284, pp. 65-74, 2010.
- [30] R. C. Eberhart, J. Kennedy, "A new optimizer using particle swarm theory," In *Proceedings of the sixth international symposium on micro machine and human science*, vol. 1, pp. 39-43, 1995, New York.
- [31] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Report, technical report-tr06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.
- [32] X. -S. Yang, S. Deb, "Cuckoo search: recent advances and applications," *Neural Computing & Applications*, vol. 24, no. 1, pp. 169-174, 2011.