

Simulating and Predicting Students' Academic Performance Using a New Approach based on STEAM Education

Nibras Othman Abdulwahid

(Miracle Laboratory, Enet'com, Technopole of Sfax, Sfax University, Road Tunis km 10, 3018 Sfax, Tunisia)

 <https://orcid.org/0000-0002-8109-3151>, nebrasse.toubane@enetcom.u-sfax.tn)

Sana Fakhfakh *

(Department of information systems, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University Al-Kharj, 11942, Saudi Arabia.

Miracle Laboratory, Sfax University, Tunisia

 <https://orcid.org/0000-0003-1845-4353>, s.fakhfakh@psau.edu.sa)

Ikram Amous

(Miracle Laboratory, Enet'com, Technopole of Sfax, Sfax University, Road Tunis km 10, 3018 Sfax, Tunisia)

 <https://orcid.org/0000-0002-5893-9833>, ikram.amous@enetcom.usf.tn)

Abstract: In many countries, particularly in Iraq, the students' academic performance (SsAP) system is based on the final grade scores in high school. This final high school grade may not reflect the students' intelligence level or the interests that link the student to a relevant university. Also, skills are not used to predict score-related school or college. In this research, a seven-subject, one-grade, one-output (SOO) model was proposed to simulate the classic SsAP system to show that the predicting system is completely based on the previous year's score and not on the students' interests and skills. Moreover, a seven-subject, twelve-year, seven-output (STS) model, which used seven parallel deep neural networks with a scaled conjugate learning algorithm, was employed to determine the students' science, technology, engineering, art, and mathematics (STEAM) skills and interests across 12 grades and predict their corresponding most appropriate school. This article contributed to constructing two models: SOO model which simulates the classical Iraqi education system, and the STS model which predicts the acceptance of students according to the STEAM system, which is what makes it different from previous research. The results revealed that the SOO model properly simulated the classic SsAP system. Furthermore, the new approach based on STEAM education successfully predicted students' academic performance in line with their skills and interests over a twelve-year period. The overall accuracy rate of the two proposed models (SOO and STS) is about 99% with 10-5 histogram errors between the target and the actual output. However, the optimized epochs of the SOO model are 1000 epochs while the STS model got 10–600 epochs.

Keywords: computational model, artificial neural network, deep neural network, auto encoder, predicting student's academic performance, Semantic Web, STEAM Education

Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1

DOI: 10.3897/jucs.86340

* Corresponding author

1 Introduction

Basically, the primary objective of education by science, technology, engineering, art and mathematics (STEAM) is to predict student performance, grades, averages, and academic acceptance from the early stages. STEAM also suggests job selections for early-stage students. STEAM teaching is a technology that benefits the teaching practices of educators and educators by integrating the arts that seek to engage and motivate students with STEAM education [Mandalapu and Gong, 2019].

STEAM teaching is a modern technology in teaching that deepens field understanding, as in developed countries that use this technology in teaching, Canada tops the list, followed by Russia, Japan, North Korea, the United States, Ireland, and the United States. Kingdom, Australia, Finland, and Luxembourg. These countries tend to set up factories and laboratories for the employment of graduate students, each according to his specialization and the skill acquired from education through STEAM because this method of teaching gives the student specialization in his academic field and thus has a career field in the same field of specialization [Olalekan et al, 2020].

Using a variety of data mining approaches, several researchers have investigated student academic performance in supervised and unsupervised learning. To gain sufficient prediction ability, neural networks frequently require a larger collection of observations. Due to the rise in the number of poor-performing graduates, it is vital to build a system that will help address this concern and lower the number of students who will have to retake classes due to poor performance or must drop out in the middle of their schooling. As a result, it is critical to examine each one, as well as their benefits and drawbacks, to determine which is more efficient and when one should be favored over the other. This research intended to create a system that uses an artificial neural network to predict student performance based on demographic traits, thereby assisting universities in selecting candidates (students) with a high probability of admission success based on admitted students' previous academic records, resulting in the institution having quality graduates [Adeniyi et al, 2021].

Iraq has a long and illustrious history of prestigious universities and high educational standards [Issa et al, 2010]. Iraq's levels of education are as follows: pre-school (ages 4–5), primary (ages 6–11), secondary (ages 12–14), preparatory/vocational (ages 15–17), and higher education (ages 18–21). Preschool, primary, and secondary education, as well as teacher training institutes and the Open College of Education, are all under the Ministry of Education. Higher education, including universities and technical institutes, is overseen by the Ministry of Higher Education and Scientific Research [Alborz et al, 2013].

In studies of college graduation performance, a great deal of attention is paid to the discovery of important predictor variables/factors, as well as the development of mathematical models that use these variables to predict successful college completion. Several studies in the literature examine indicators that might predict whether a student will graduate from college. This matter is divided into two categories: pre-admission and post-admission. Academic and non-academic criteria can be used to classify readmission factors. High school rank, high school grade point average, and standardized test scores are all common academic pre-admission considerations [Lesinski et al, 2016].

Academic failure is an important issue at a time when higher education is becoming increasingly important to economic success. Moreover, for higher education institutions whose goal is to contribute to improving higher education quality, the success of human capital creation is a continuous matter of analysis. Predicting student performance is one strategy to improve such quality [Albarka, 2019].

Student achievement is influenced by a number of interconnected elements. The application of new technological advancements in educational displacement offers limitless possibilities. One of these developments is the use of analytics and data mining to forecast student academic achievement and performance. Machine learning (ML) technologies such as artificial neural networks (ANNs) may be continually enhanced based on the current literature [Mohamed et al, 2022].

However, proposed computational neural networks [Khalid et al, 2020] have shown the effect of forwarded internal representations from one neuron to another in guiding behavior, thereby hastening the learning procedure. Forwarding internal representations may reduce the number of epochs required to achieve optimized performance [Khalid et al, 2020]. Such representations can be used to represent student skills and interests while simulating the students' academic performance system.

The motivation for simulating the traditional SsAP system is to determine the dependency of such an accepting system solely on the final high school grade. This research aims to prove that the SsAP approach works regardless of the students' interests and skills. Furthermore, this study proposes a new approach that is based on the students' interests and skills while considering their scores across all grades.

This research intends to construct a supervised artificial neural network that simulates the classic SsAP system using a seven-subject, one-grade, one-output (SOO) model. It also proposes a new approach by combining STEAM education criteria with deep artificial neural networks to build a seven-criteria, twelve-grade, seven-output (STS) model.

2 Related Works and Research Gap

The artificial neural network (ANN) has a wide range of applications [Khalid et al, 2020]. Recently, many researchers presented their proposals for applying the ANN techniques to study the effect of STEAM education on graduate students' academic careers [Mandalapu and Gong, 2019].

However, some studies investigated SsAP prediction using data mining, deep neural networks, and backpropagation learning algorithm [Nabil et al, 2021]; [Selvia et al, 2021]. Others simulated the SsAP system for tertiary institution students [Olalekan et al, 2020]. Also, data mining techniques were used to simulate the SsAP system to implement the decision-making algorithms [Mengash, 2020].

Predicting high school students' university admission results using deep neural networks has become important for their careers and post-graduation lives [Santana et al, 2020]. Other researchers have used different algorithms and tools, such as the augmented reality technique [Ang and Hann, 2019].

On the other hand, some scholars employed supervised learning algorithms to demonstrate the effect of machine learning algorithms in simulating STEAM education-based SsAP systems [Banadaki, 2020]. Other researchers also applied some proposed criteria based on artificial neural networks in STEAM education [Nguyen et al, 2020].

Recently, some researchers successfully simulated the SsAP system by proposing predicting models that classify the undergraduate students' academic admission using back-propagation [Rodríguez-Hernández et al, 2021]; [Prasetyawan et al, 2018]. Also, other researchers used the perceptron concept to implement an admission system with supervised neural networks [kurniadi et al, 2021]; [Putra et al, 2018].

The success of an educational institution is linked to the success of its students in the time allotted to them without wasting school years. Therefore, a student's expected

academic performance, from its early stages, is one of the predictors of university and college student success in higher education institutions.

Based on previous research, it was found that there are two ways to predict student performance: the traditional method and a modern educational approach based on STEAM education. The traditional approach starts with the conventional method of education and ends with the final cumulative average. The Grade Point Average (GPA) predicts students' academic performance, which allows them to know their averages from the early stages, thereby preventing many students from deferring their studies [Sabukunze Didier et al, 2021]. Moreover, a student's graduation and obtainment of the required cumulative average on time are two of the most important admission factors in the field of computer science and information technology [Sabukunze Didier et al, 2021].

The second forecasting approach relies on modern educational methods based on the STEAM educational system. It yields predictions of students' academic performance based on their abilities and preferences. As STEAM education determines the students' tendencies in line with their skills and interests, this method reveals their academic future and career path. Moreover, STEAM education is high-quality learning that emphasizes teamwork and the ability to explore and solve problems. It also helps with social integration, which is one of the key elements for eliminating poverty and societal disparities, as well as achieving sustainable progress, facilitating access to good careers and thus a better life [Sabukunze Didier et al, 2021].

This part will present previous research that used the latest methods in predicting students' performance, grades, and academic acceptance. In addition, these studies suggested a profession that employs STEAM technology, as STEAM education forecasts students' academic performance and recommends a career path. However, these past works have some limitations, which are listed as follows:

1. Most of the articles [Bujang et al, 2021]; [Mathew et al, 2021]; [Qazdar et al, 2019]; [Helal et al, 2018]; [Iqbal et al, 2017] applied the proposed system solely in a single location, whether it is a university, department, college, institute, or school for a particular specialty. They also employed a certain category and a specific age.

2. Most of the previous studies [Ahammad et al, 2021]; [Spyropoulou et al, 2020]; [Tatar and Dilek, 2020]; [Zhang et al, 2020]; [Mandalapu and Gong, 2019] relied on data from internal sources such as the Internet, soft skills, faculty members, individual systems, as well as most surveys and registrations. However, all these sources are thought to have a limited amount of data.

3. Some of the studies [Mengash, 2020]; [Iqbal et al, 2017] focused on predicting student performance using the final GPA. They did not focus on students' interests, preferences, and skills despite the importance of these aspects in discerning academic and career choices.

According to this paper's analysis, most of the past studies did not address students' academic problems, such as dropping out of school for social, economic, and cultural reasons. Most articles also did not provide solutions for at-risk students. The provision of a non-profit learning environment based on a new STEAM educational system, which could prevent most students from leaving school, was not found. Moreover, it is necessary to introduce students to the concept of programming and its timely significance, thereby ensuring programming literacy for the progress of future generations. This enables them to make pre- and post-admission decisions that guarantee their scientific and professional futures, which happens to be one of this paper's objectives. In addition, this study's proposed system covered all higher education institutions, as well as teachers, professors, educators, administrative staff, and students across all K-12 academic levels. This study also relied on big data to predict student acceptance based on their skills, preferences,

and GPA.

2.1 Related works based on student academic performance prediction

Predicting students' academic performance is valuable to any educational institution seeking to improve performance [Yang et al, 2020]. Therefore, predictive analytics is one of the most widely used applications in higher education institutions to ensure high-quality performance [Jin et al, 2020].

[Bujang et al, 2021] compared the accuracy of six well-known machine learning techniques: J48 decision tree (DT), support vector machine (SVM), naïve Bayes (NB), k-nearest neighbors (k-NN), logistic regression (LR), and random forest (RF). Also, they proposed a multiclass prediction model to reduce overfitting and misclassification. Their results were caused by imbalanced multiclassification based on the synthetic minority oversampling technique (SMOTE) with two feature selection methods. The dataset consisted of 1,282 real student course grades from a 2016–2019 course. This study's dataset was also considered limited. The methodologies used were J48, SVM, NB, k-NN, LR, and RF.

[Qazdar et al, 2019] presented a framework for predicting student performance based on a machine learning algorithm at H. E. K. high school in Morocco from 2016 to 2018. The dataset was based on student data collected from the school management system "MASSAR" (SMS-MASSAR). The dataset used in this study covered the school years 2015–2016, 2016–2017, and 2017–2018 and concerned 478 Physics students. [Qazdar et al, 2019] used an interprofessional standard data mining process (CRISP-DM). The dataset was limited and used only one school to test the model.

[Polyzou et al, 2019] investigated the problem of predicting student performance at the end of the semester before they started a course. They built a model with various feature subsets. The original dataset (students' grades) was obtained by the University of Minnesota and spanned 13 years. The methodologies used were decision tree, gradient boost, random forest, and support vector machine. The developed models performed poorly for failing students, which was identified as a limitation of these approaches [Polyzou et al, 2019].

[Ahammad et al, 2021] conducted a comparative study of different machine learning techniques for predicting student results. The dataset was collected at Bangladesh's Feni Model High School. It includes the student's marks for different class subjects of Grade 9–10 students during the academic years 2013–2014 and 2016–2017. After eliminating incomplete data, the dataset comprised 400 students. The methodologies used were naïve Bayes, k-nearest neighbors, support vector machine, XG-boost, and multilayer perceptron. The dataset was found to be limited. Also, with a large dataset, different neural network structures such as the convolutional neural networks (CNN), recurrent neural networks (RNN), etc., should be used [Ahammad et al, 2021].

[Tatar and Dilek, 2020] proposed a more relevant predictor of student graduation academic performance. They investigated whether it was the individual course grades or the grade average. The dataset included records for 357 students admitted to the CCSIT at IAU from Fall 2011 to Fall 2013 (included), resulting in three student batches. The methodologies used were LR, RF, and NB. The data set was found to be limited [Tatar and Dilek, 2020].

Another study analyzed the different modern techniques widely applied for predicting students' performance, together with the objectives they must reach in a specific field. The dataset was collected using software tools for technology-enhanced learning. The

methodologies used were SVM, RF, NB, DT, and collaborative filtering algorithms [Rastrollo et al, 2020].

[Mengash, 2020] suggested utilizing data mining classification methods to forecast applicants' academic success at a university to help institutions make admission decisions. The researcher also used a dataset of 2,039 students enrolled in the Computer Science and Information College of a Saudi state university from 2016 to 2019 to build and evaluate four prediction models employing data mining methods. This study differed from others in predicting student performance because the researcher used the ANN, DT, SVM, and NB methodologies, even though the paper used the educational data mining technique (EDM) to extract information from an enormous educational database. This study is limited to a single department and university's implementation system [Mengash, 2020]].

2.2 Related works based on career choice

Predicting a student's profession is one of the modern aims that depend on excellent education [Mandalapu and Gong, 2019]]. STEAM education is one of the best modern learning methods for predicting a student's future career [Mandalapu and Gong, 2019]. A good education enhances students' self-confidence and skill acquisition so they could obtain careers. STEAM education provides good instruction that aims to equalize women with men at work and leads to gender quality, leading to entrepreneurship and elimination of unemployment in the society. A country's economic health depends on solid infrastructures based on science, technology, engineering, and mathematics [Banadaki, 2020].

STEAM occupations represent more than 50% of employment in major industries in America. Therefore, all STEAM students must have access to high-quality education to ensure the economic growth of countries. Big data and decision-making analysis in the United States predicted that sports and computing professions would grow the most in the ten years leading up to 2022. Moreover, the prediction of future computer and information technology job opportunities could reach 77%. African Americans account for 11% of the workforce, with 55% working STEAM jobs [Banadaki, 2020].

[Mandalapu and Gong, 2019] analyzed the effect of various attributes collected by ASSISTments' online learning platforms on the performance of machine learning algorithms in predicting students' occupational fields at an early stage. The dataset consists of 1,709 students from various career paths: 591 students in this, 466 non-STEM students, and 125 STEM-field students. They used enhanced gradient tree, deep learning, AutoMLP, random critical, and logistic regression methodologies. The results proved that the tree-enhanced scaling and deep learning methods are among the best types of prediction techniques when compared to the other methods used. However, this study's limitation is related to the use of clickstream data, which depends on multiple factors such as time spent on the system. Also, the dataset analyzed was from a single platform – ASSISTments [Mandalapu and Gong, 2019].

[Santana et al, 2020] used deep learning to solve the problem of a project (Brazilian semi-arid green technology) that encourages high school students to pursue STEAM jobs. Adopted deep learning activities were used to determine the amount of forage needed to feed goats in areas with a semi-arid climate. The dataset consisted of 67 secondary school students (14 to 17 years old) from Brazilian public schools. Students who participated in this activity became more involved and aware of scientific and technical issues.

[Spyropoulou et al, 2020] reported on a case study that aimed to increase student (boys and girls, aged 14–16 years old) motivation in STEAM education and increase their

chance of obtaining a STEAM-based occupation. This was done by using the internet of things (IoT), which is one of the STEAM learning techniques. In seven different locations within an integrated educational framework, more than 150 secondary students and nine teachers engaged in practical activities and teamwork. The economic and social data indicated that perception of IoT was generally successful, with enhanced awareness and skills in STEAM education.

2.3 Related works based on the STEAM approach in solving educational problems

Because problem-solving is one of the STEAM goals, STEAM students are best able to discover and solve problems [Spyropoulou et al, 2020]. In addition, the technique indicates the development of scientific knowledge using modern technologies [Spyropoulou et al, 2020]. [Ang and Hann, 2019] proposed AURE, a mobile platform that combines augmented reality and machine learning. It aims to improve education through STEAM. The Google Cloud Platform has been used to collect STEAM data due to its ability to retrieve information in the form of 3D images. The purpose of using this feature in education is to ignite students' passion and curiosity about specific topics in science, technology, engineering, and mathematics. Google Cloud Tensor Processing Units (TPU) were used to train the dataset with Cloud Vision API, while Kit for Firebase was used to host models such as Tenser-Follow-Life for better accuracy. Finally, the information in AR was displayed in the mobile app using Sceneform SDK from ARcore. In the future, they can expand this application to include all scientific materials, enabling students to learn through an interactive platform [Bertrand and Namukasa, 2020].

[Banadaki, 2020] developed Supervised Research Experiences (SURE) to engage STEAM students in machine learning research. The development of several interdisciplinary graduation projects, including mechanical engineering, biology, physics, and cybersecurity, enabled STEAM students to solve problems (automating microscopic image analysis, quantitative optical mode determination, and IoT penetration detection). This study revealed the significance of machine learning for STEAM students at the undergraduate level and the importance of improving the introduction of computing curricula and big data into these disciplines. In addition, this study provided the basic principles for effectively educating STEAM students to efficiently solve big data problems in their disciplines and meet new challenges in a computer-based world.

[Nguyen et al, 2020] suggested criteria for an intelligent problem-solving (IPS) model. The IPS model can automatically solve problems or teach a person how to solve them. It also enables learners to state the hypotheses and objectives of the problems; they can either ask the program to solve the issues automatically or give instructions so it can help them solve the concerns. In addition, they built a Rela-Ops model, which represents a combination of knowledge from relationships and operators. The Rela-Ops model was developed using an objectivist and existential approach [Nguyen et al, 2020]. Every file inside Rela-Ops was able to solve the problems.

[Bertrand and Namukasa, 2020] presented a case study on student quality in Ontario, Canada, using interviews, observations, and data analysis. This study aimed to better understand STEM educational programs for students delivered by nonprofit organizations and publicly-funded schools. There was a total of 103 participants (19 adults, principals, trainers, and teachers, as well as 84 students). Training under each STEAM specialization was based on character building, discipline, problem-solving, teamwork, communication, creativity, and innovation. The results focused on students' learning to develop perseverance, adaptability, and transferable skills. In addition, this study concentrated on the future, aiming to develop and implement STEAM programs that enhance teaching,

skills education, and interaction in the workplace [Bertrand and Namukasa, 2020]. For further details, see Table 1 for a list of other related studies that used ANN and ML for STEAM education.

Reference	Datasets size	Machine learning algorithms	Best algorithm	Source data	Proposed	Results	Accuracy	Limitations
[Tatar and Dilek, 2020]	550	Support Vector Classifier (SVM) and Naive Bayes (NB)	SVC (SVM)	Nizwa University (Real data)	Created a machine-based learning model to predict a student's educational performance	The model showed a very accurate accuracy rate that can be adopted	87%	number of samples (size of data)
[Altubrawee et al, 2019]	500	ANN, Decision Tree, Random Forest, Bagging Voting, Boosting	F1-Score of Random Forest with Genetic algorithm	Online data (Kaggle)	Proposed method based on the Genetic Algorithm (GA) with classification method is to predict student academic performance	The proposed method makes an impressive result in the predictions of student academic performance	81.18%	number of samples (size of data)
[Uddin et al, 2019]	500	Decision tree and K-Nearest Neighbor	Decision tree	Online data (Kaggle)	Compared the performance of two classifiers, namely C4.5 and k-Nearest Neighbor (KNN) and applies the SMOTE preprocessing method in the classification of student academic performance	The C4.5 Decision Tree method resulted in better prediction	71.09%	1- A new variable might enhance an underperforming algorithm in the research, and vice versa. 2- Researchers compared supervised machine learning techniques for illness prediction. 3- In comparing numerous supervised machine learning algorithms, they didn't consider hyperparameters from different papers.
[Selvia et al, 2021]	500	Nave bayes, decision table, MLP, and J48 Ensemble methods like Bagging, Random sub space, and AdaBoost	AdaBoost with MLP technique	Online data (Kaggle)	1- The aim to build the prediction model by different ML techniques. 2- Implementations of different ensemble meta-based model is presented with different classification algorithms of ML techniques	The ensemble meta-based technique (AdaBoostM1) gained a superior accuracy performance with Multilayer Perceptron ML technique	80.33%	number of samples (size of data)
[Mengash, 2020]	1073	Deep Dense Neural Network (DDNN), Decision Tree, KNN, MLP, SGD, Random Forest, and NB	DDNN	Open University	Evaluated the efficiency of deep dense neural networks with a view to early predicting failure-prone students in distance higher education	The Deep Learning methods may contribute to building more accurate predictive models	78-81%	1- data collected from a single university in Saudi Arabia. 2- More studies need to look at pre-admission factors that affect how well a student will do in the future, such as the student's personality, background, family, and communication skills.
[Santana et al, 2020]	6807	Random forest, Logistic regression, Support vector classifier, Voting, decision Tree, Bagging MLP, and AdaBoost	F1-score of Random Forest	Technical institute (Real Data)	Compares the two models: one built using academic parameters only and another using demographic parameters.	The results allow us to conclude that only the combination of academic and non-academic parameters can give us the most appropriate prediction model	93.8%	Despite full engagement, a student's academic vocation is revealed by their preferences in one stage.

[Spyropoulou et al, 2020]	388	Random forest, Logistic regression, and K-Nearest Neighbor	Random forest	Virtual learning Environment	Used ML methods to detect students who do not submit assignments on time	Shows that Random Forest is a best option for predicting students who do not submit an assignment on time	93%	1-Because of the sample's cognitive background and geographical location, it's representative of Greek secondary education. This may restrict generalizing our results to other educational systems, places, and cultures. 2- Some students have difficulty with electronics and programming and require extra time from professors.
[Nguyen et al, 2020]	337	Backpropagation neural Network (BPNN), MLP	ANN with BP and MLP	From previous student subject scores (Real data)	Aims to predict the final score of courses of students who are undergoing education in higher education.	The results obtained RMSE value of 0.049929 and MSE of 0.001675	93.43%	The model's formality is limited. COB's mathematical base isn't obvious
[Bertrand and Namukasa, 2020]	200000	ANN, K-fold Cross-Validation, fast large margin, decision tree, random forest, gradient boosted tree, general linear model, logistic regression	Artificial Neural Network	Record of Colombia university student	The first objective to test a systematic procedure for implementing artificial neural networks to predict academic performance in HE. The second to analyze the importance of several well-known predictors of academic performance in HE	It is possible to systematically implement artificial neural networks to classify students' academic performance	high (accuracy of 82%) or low (accuracy of 71%)	There is a need for more research sites and data being collected over a longer period of time.

Table 1: Related works overview.

3 The proposed method

In this research, two different models (classical and STEAM) were used to predict student academic performance in the Iraqi system, as shown in Figure 1. Both models seek to predict students' academic performance but in different aspects. The classical system was simulated via a seven-subject, one-grade, one-output (SOO) model based on the final year scores. On the other hand, the STEAM model was generated using a seven-criteria, twelve-year, seven-group (STS) model based on the influence of STEAM education on students' interests and skills.

The overall framework in Figure 1 illustrates the dataset analysis procedure for both SOO and STS models, as well as their respective workflows. Initially, the dataset was a three-layer matrix consisting of seven subjects, 12 grades, and 50,000 students. The SOO model is linked to the final high school grades, which are reshaped to obtain the 7-subject and 50,000-student dimensions. The supervised and multilayer artificial neural network (ANN) was then employed to predict the relevant institute according to the input scores. On the other hand, the STS model rearranged the input dataset due to the STEAM criteria. Different criteria represented different subjects. Thus, each criterion has a different row number but fixed columns. The criteria matrices were then normalized in the 1–100 range before being mapped with the related subjects. Afterward, seven parallel fully connected neural networks (FCNN) predicted the STEAM outputs based on the overall grade input criteria.

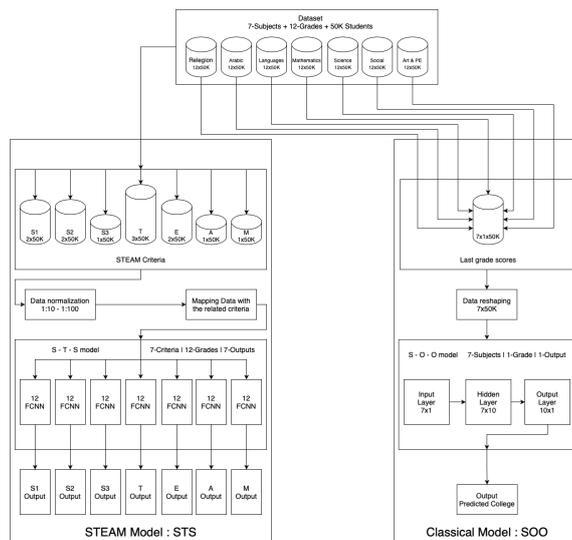


Figure 1: SOO and STS models' frameworks.

3.1 Input dataset

This research was based on the proposed scores of 50,000 students from random schools in Baghdad, Iraq. The dataset is a three-dimensional matrix that consists of 50,000 students' seven grades for 12 grades. Figure 2 depicts a histogram representation of the 50,000 students with scores ranging from 50–100 for all stages. The dataset has been divided as 70% for training, 15% for testing, and 15% for validation.

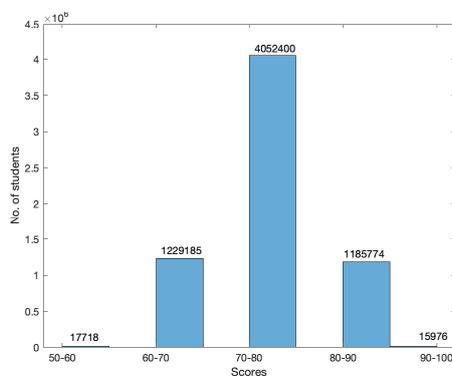


Figure 2: Input dataset histogram.

The procedure of setting up the dataset is shown in Table 2.

Step Procedure	
1	Set a three dimensional array with 7 rows, 12 columns, and 50,000 layers.
2	Set the rows as the subjects of every grade.
3	Set the subjects to 7 only, as follows: (Religion, Arabic, English and Foreign languages, Mathematics, Science, Social, and Art and Physical education).
3	Subjects normalization: If there are one or more foreign languages beside the English language subjects, then get the average score for all of them. After grade 4, the social subject contains the average scores of: Geography, History, and Nationality subjects. Before grade 4, the scores are in range 1–10, thus they have been normalized to be in range 1–100. After grade 7, the science subject contains the average scores of: Biology, Chemistry, and Physics subjects.
4	Set the columns as the grades numbers: Grade 1 – Grade 6 stand for primary school, Grade 7 – Grade 9 stand for middle school, Grade 10 – Grade 12 stand for high school.
5	Set the third dimension layer to the number of students.
6	All the scores for all the grades should be equal or grater than 50 (success threshold).

Table 2: Dataset setup.

3.2 The proposed SOO model

The SOO model is a multilayer supervised feed-forward neural network that used the Levenberg-Marquardt learning algorithm. The input layer contained seven neurons that are linked to Grade 12 scores (final year of high school). Thus, the input layer has seven rows for 50,000 samples. The hidden layer has ten neurons while the output layer only has one.

The SOO model (framework shown in Figure 3 maps the last year of high school with the previous predicted student academic performance to forecast each student's academic path.

However, all the neurons in the SOO model have the same input–output structure, as illustrated in Figure 4. Consequently, the output of any neuron can be estimated as follows:



Figure 3: SOO model framework.

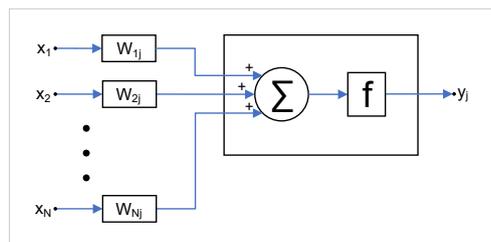


Figure 4: Input-Output structure.

$$y_j = f\left(\sum_{i=1}^N x_i W_{ij}\right), \tag{1}$$

where y_j is the output of neuron j ; x_i is the i^{th} input of neuron j ; W_{ij} is the weight of the links to neuron j from i ; and $f(\cdot)$ is the activation function that is usually *logsig* function. Noting that, the bias of all the neurons has been omitted in our model. Because we are considering the effects of the input scores on the STEAM criteria away from any other additional effects.

However, the weights of the fully connected layers have been updated using Levenberg-Marquardt learning algorithm. Backpropagation is used to calculate the Jacobian jX of performance *perf* with respect to the weight variables X . Each variable is adjusted according to Levenberg-Marquardt, as follows:

$$jj = jX * jX, \tag{2}$$

$$je = jX * jE, \tag{3}$$

$$\Delta W = -(jj + I * \mu) \setminus je \tag{4}$$

$$W = W + \Delta W \tag{5}$$

$$E = y - d, \tag{6}$$

where W is the weight of the network, y and d are the actual output and desired output, respectively; and E is all errors and I is the identity matrix. μ is the learning rate [Marquardt, D., 1963, Hagan, M.T., and M. Menhaj, 1994, Hagan, M.T., H.B. Demuth, and M.H. Beale, 1996].

3.3 The proposed STS model

The STS model is a fully connected deep neural network model, consisting of seven parallel deep neural networks, each with twelve fully connected networks. The seven lines represent the seven criteria, and the twelve fully connected networks signify the 12 grades, shown in Figure 5.



Figure 5: STS model framework.

The seven criteria have been assigned based on the subjects related to STEAM education concepts, as shown in table 3. The subjects were assigned to STEAM groups based on their interests and skills in line with their academic destination.

Criteria	Subject
S1	Religion and Arabic language.
S2	Science, English and foreign languages.
S3	History and Geography.
T	Biology, Physics, Chemistry, English and foreign languages.
E	Mathematics, Physics, Chemistry, English and foreign languages.
A	Physical education and Art.
M	Mathematics.

Table 3: STEAM criteria

3.3.1 Optimization procedure

A scaled conjugate gradient is a supervised learning algorithm for neurofeedback networks that avoid time-consuming linear search in the conjugate direction in the rest of the algorithms. The basic idea is to combine two approaches (one of which uses the Levenberg-Marquardt algorithm with the conjugate gradient approach). Moreover, the algorithm trains the network if its functions, weight, net income, and transmission are all useful derivative functions. Additionally, backpropagation is used for weight and stimulation variables [Aich et al, 2019]. Most function-minimization optimization approaches employ the same strategy. Minimization is a local iterative method that minimizes a function approximation in the proximity of the current point in weight space. A first- or second-order Taylor expansion of the function is frequently used to approximate the function. The strategy's concept is exhibited in the following pseudo method, which minimizes the error function $E(w)$:

$$E(w) = \left(\dots, \sum_{p=1}^P \frac{dE_p}{dw_{ij}}, \dots \right) \quad (7)$$

where P is the number of patterns presented to the network during training and E_p is the error associated with pattern p [Moller, 1993], as shown in table 4.

Step	Procedure
1	Choose initial weight vector w_1 and set $k = 1$.
2	Determine a search direction p_k and a step size α_k so that $E(w_k + \alpha_k p_k) < E(w_k)$.
3	Update vector: $w_{k+1} = w_k + \alpha_k p_k$.
4	If $E(w_k) \neq 0$ then set $k = k + 1$ and go to 2, else return w_{k+1} as the desired minimum.

Table 4: Optimization procedure

3.3.2 The STS model algorithm

The STS model, following the steps mentioned in Table 5 to set up the input dataset and Table 6 to learn all the networks using the SCG algorithm, is shown below:

Step	Procedure
1	Mapping the input matrix regarding the criteria mentioned in Table 3, as follows: S1 input is a 2x50000 matrix, S2 input is a 2x50000 matrix, S3 input is a 1x50000 matrix, T input is a 2x50000 matrix, E input is a 3x50000 matrix, A input is a 1x50000 matrix, M input is a 1x50000 matrix,
2	The Target is a 1x50000 matrix, Map the target matrix for every input, as follows: Target = 1 when the related score has an average value greater than 90, Target = 0 when the related score has an average value less than 90,
3	The output values should be either 0 or 1, as follows: If S1 = 1, then the related student eligible to S1 academic group, If S2 = 1, then the related student eligible to S2 academic group, If S3 = 1, then the related student eligible to S3 academic group, If T = 1, then the related student eligible to T academic group, If E = 1, then the related student eligible to E academic group, If A = 1, then the related student eligible to A academic group, If M = 1, then the related student eligible to M academic group, else, the related student is not eligible to the related criteria.

Table 5: STS model input initialization and setup

Step Procedure	
1	Choose weight vector w_1 and scalars $\sigma > 0, \lambda_1 > 0$ and $\lambda_1 > 0$. Set $p_1 = r_1 = -E(w_1), k = 1$ and $success = true$.
2	If $success = true$ then calculate second order information: $\sigma_k = \frac{\sigma}{ p_k },$ $s_k = \frac{E(w_k + \sigma_k p_k) - E(w_k)}{\sigma_k},$ $\delta_k = p_k^T s_k.$
3	Scale s_k : $s_k = s_k + (\lambda_k - \bar{\lambda}_k) p_k,$ $\delta_k = \delta_k + (\lambda_k - \bar{\lambda}_k) p_k ^2,$
4	If $\delta_k \leq 0$ then make the Hessian matrix positive definite: $s_k = s_k + (\lambda_k - 2 \frac{\delta_k}{ p_k ^2}) p_k,$ $\bar{\lambda}_k = 2(\lambda_k - \frac{\delta_k}{ p_k ^2}),$ $\delta_k = -\delta_k + \lambda_k p_k ^2, \lambda_k = \bar{\lambda}_k.$
5	Calculate step size: $\mu_k = p_k^T r_k, \alpha_k = \frac{\mu_k}{\delta_k}.$
6	Calculate the comparison parameter: $k = \frac{2\delta_k(E(w_k) - E(w_k + \alpha_k p_k))}{\mu_k^2}.$
7	If $k \geq 0$ then a successful reduction in error can be made: $w_{k+1} = w_k + \alpha_k p_k,$ $r_{k+1} = -E(w_{k+1}),$ $\lambda_k = 0, success = true,$ If $k \bmod N = 0$ then restart algorithm: $p_{k+1} = r_{k+1}$ else create new conjugate direction: $\beta_k = \frac{ r_{k+1} ^2 - r_{k+1}^T r_k}{\mu_k},$ $p_{k+1} = r_{k+1} + \beta_k p_k.$ If $k \geq 0.75$ then reduce the scale parameter: $\lambda_k = \frac{1}{2} \lambda_k$, else a reduction in error is not possible: $\bar{\lambda}_k = \lambda_k, success = false$.
8	If $k < 0.25$ then increase the scale parameter: $\lambda_k = 4\lambda_k$.
9	If the steepest descent direction $r_k \neq 0$ then set $k = k + 1$ and go to 2, else terminate and return w_{k+1} as the desired minimum.

Table 6: Scaled conjugate algorithm

4 Results

4.1 The SOO model

Figure 6 depicts that though the SOO model had an above-average performance, it had an extremely slow learning rate (1000 epochs). Nevertheless, its high accuracy is reflected by the histogram of the errors shown in Figure 7. Figure 8, in particular, shows the identical training, testing, and validation regressions. Moreover, the SOO model displayed a good training state with a plausible gradient value, as shown in Figure 9.

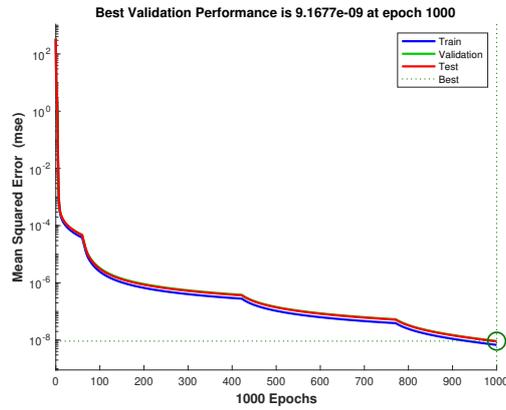


Figure 6: Best epochs with best performance of the SOO model.

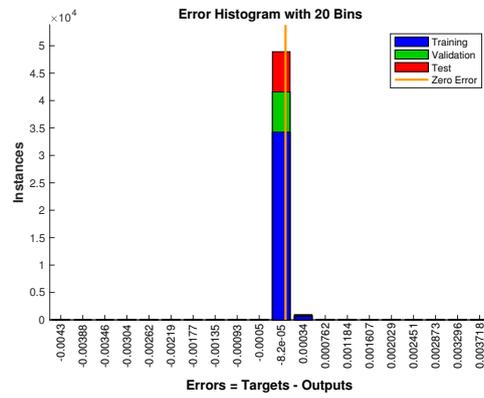


Figure 7: Errors of the SOO model.

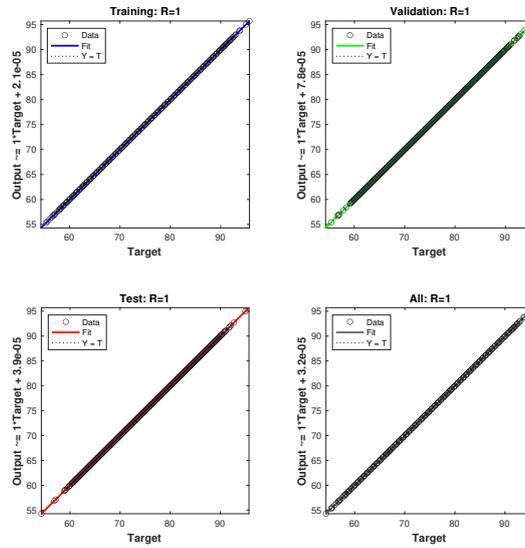


Figure 8: Regressions of the SOO model.

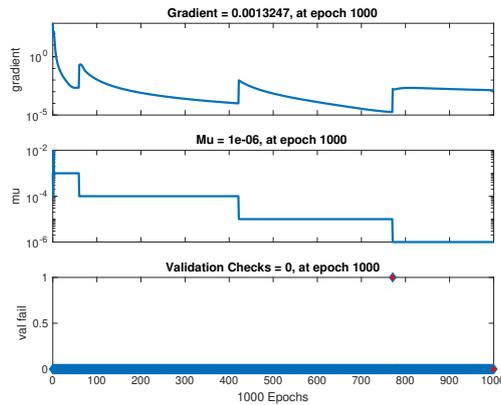


Figure 9: Training states of the SOO model.

4.2 The STS model

Results of the STS model exhibited reasonable behavior in terms of performance, error, fitting curves, training state gradient, regression, and the best number of epochs to achieve the best performance. Figure 10 displays the total number of epochs required to reach the best performance. The chart depicts seven lines for seven criteria: Science 1, Science

2, Science 3, Technology, Engineering, Art, and Mathematics, across the twelve grade levels. All the criteria's training was completed with a high number of epochs in the first year, but the number rapidly decreased to less than ten or 20 in some cases. This indicates the effect of forwarding internal representations, which improved the training speed of the preceding networks.

Relatively, Figures 11a–11g show the performance of the first-year layers across all seven criteria. All the performance behaviors (training, test, and validation data) have rapidly decreased to reach their best value with a plausible number of epochs. Figure 11h displays the overall percentage of the performance values, with more than 80% of the values being close to zero.

Accordingly, the performance values are related to the mean square errors, as shown in Figures 12a–12g. The histogram values for the three sets of data (training, test, and validation) are all around zero. Figure 12h illustrates the overall percentage of error values, with more than 70% of the error values approaching zero, while the rest were almost close to it.

Figures 13a–13g indicate how the training states are gradually decreasing to reach the best performance. Figure 13h shows that the overall percentage of the gradient values is about zero, with the percentage decreasing when the gradient values increase.

Moreover, Figures 14a–14g depict the correspondence between the training, test, and validation data and the target. Fifty percent of the output data is identical to the target data, with an approximate value of R around 0.995 for all the criteria networks, while the remaining data has regression values between 0.975 to 0.99, as shown in Figure 14h.

Also, Figures 15a–15g confirm the alignment of the training, test, and validation outputs with the targets, along with the reference fitting curve and error values. These charts reflect the STS model's accuracy, which is about 99%, as shown in Figure 15h.

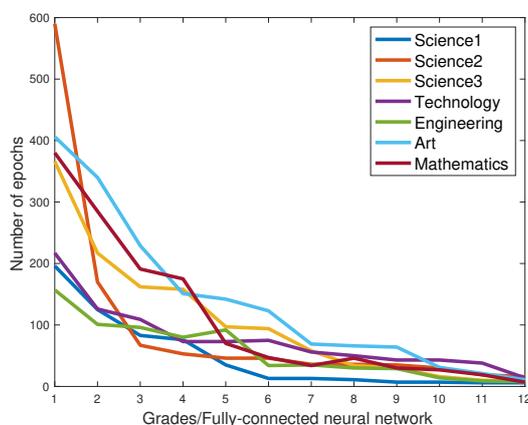


Figure 10: Best epochs with best performance.

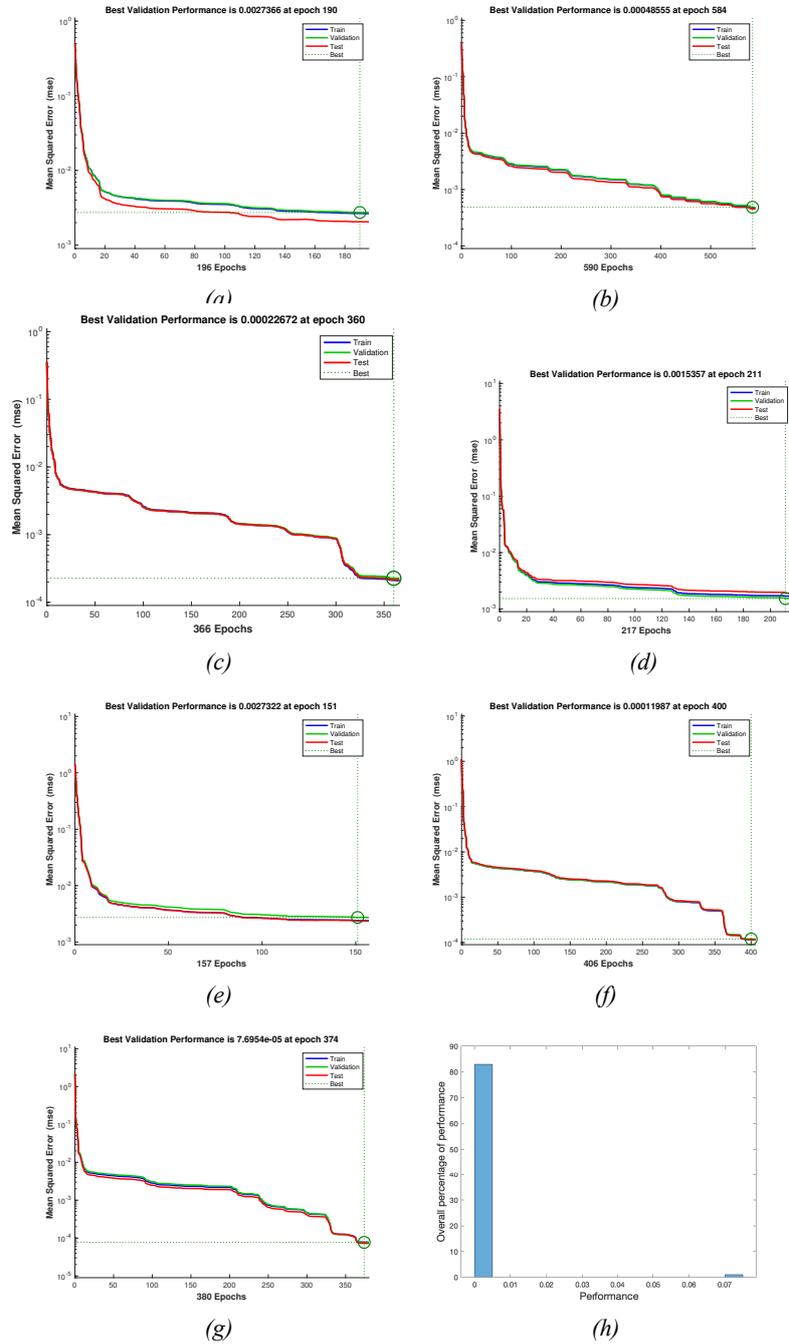
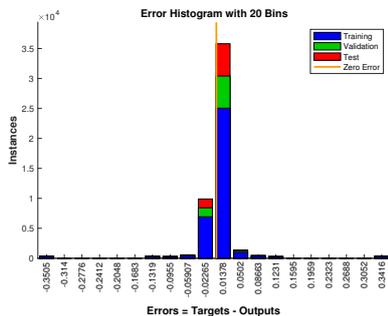
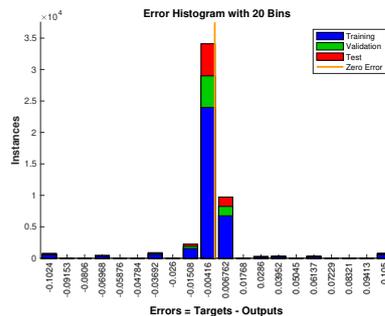


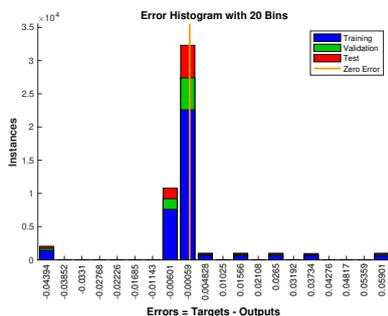
Figure 11: Performance of: (a) science1, (b) science2, (c) science3, (d) technology, (e) engineering, (f) art, (g) mathematics networks, and (h) overall percentage.



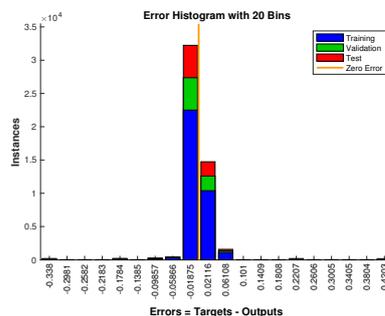
(a)



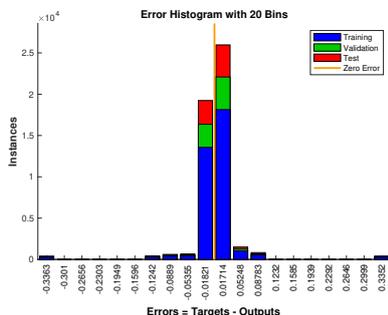
(b)



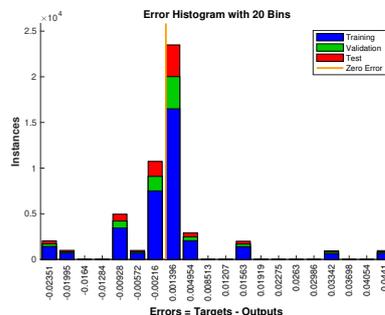
(c)



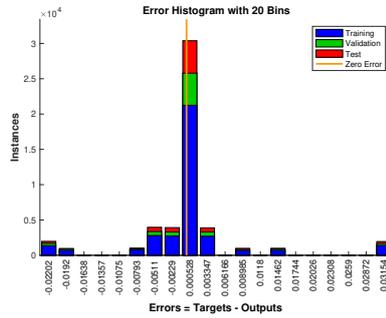
(d)



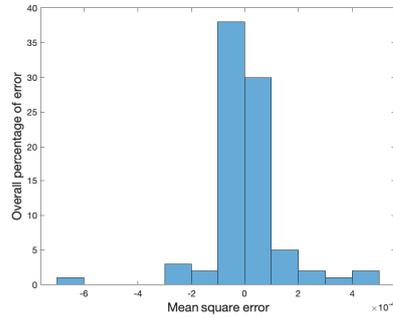
(e)



(f)

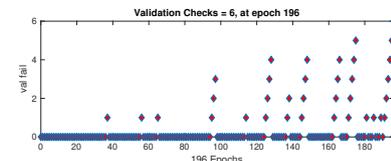
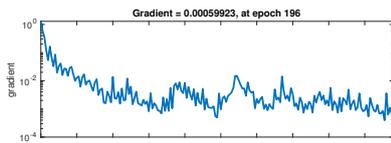


(g)

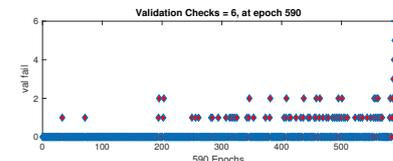
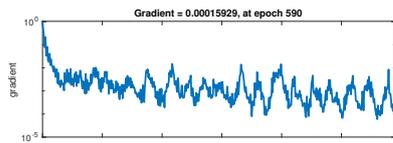


(h)

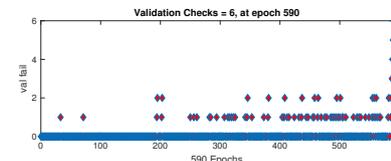
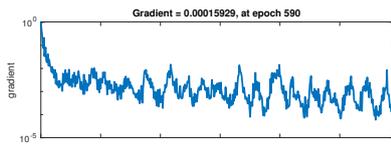
Figure 12: Mean square error of: (a) science1, (b) science2, (c) science3, (d) technology, (e) engineering, (f) art, and (g) mathematics networks, and (h) the overall percentage.



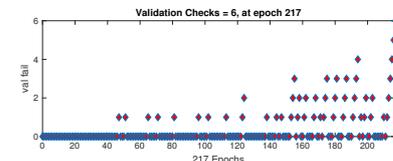
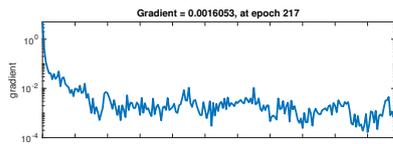
(a)



(b)



(c)



(d)

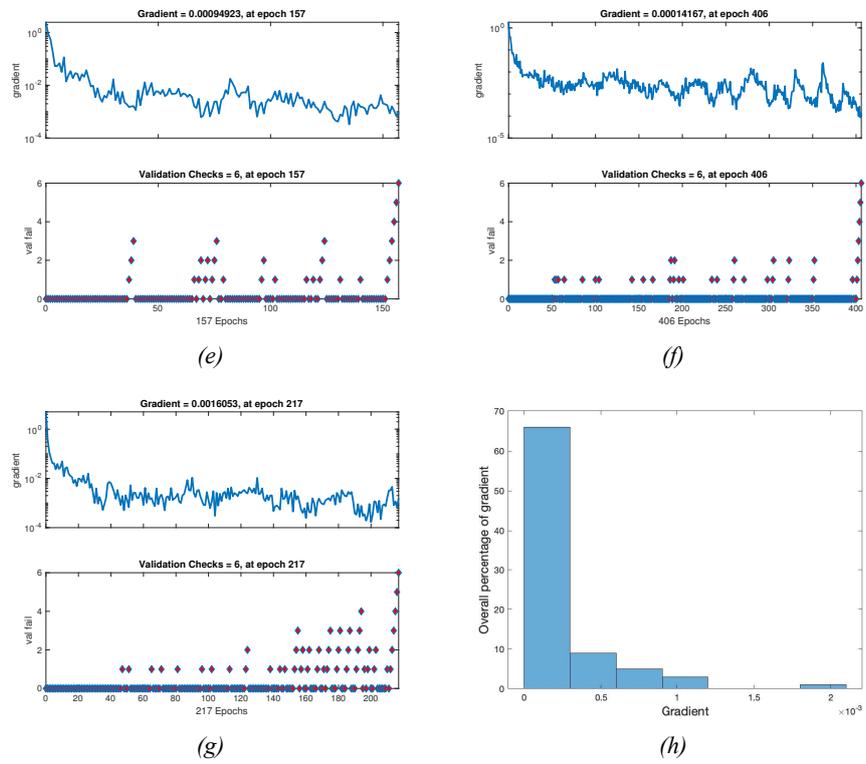
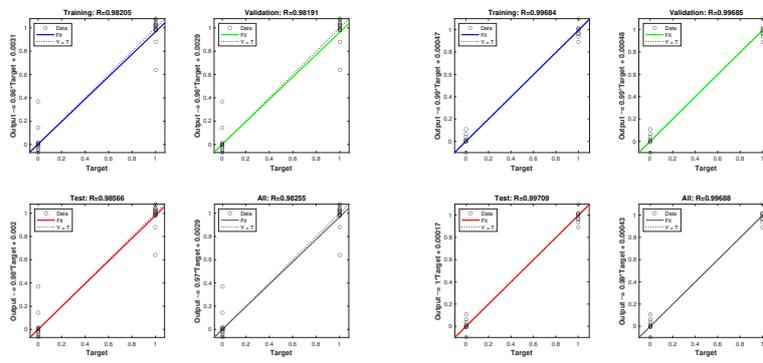
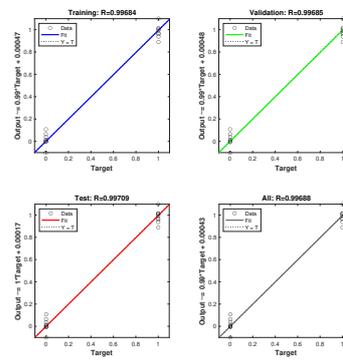


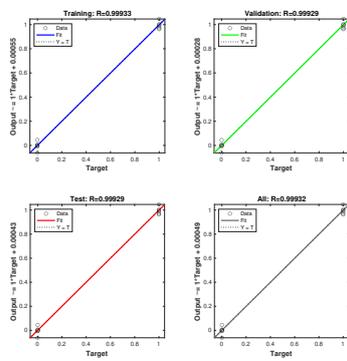
Figure 13: Training states of: (a) science1, (b) science2, (c) science3, (d) technology, (e) engineering, (f) art, and (g) mathematics, networks, and (h) the overall percentage.



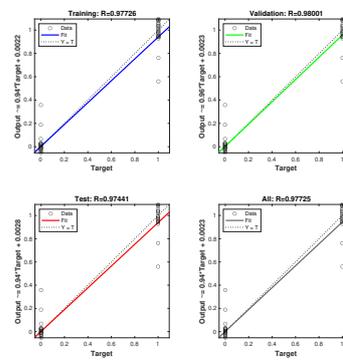
(a)



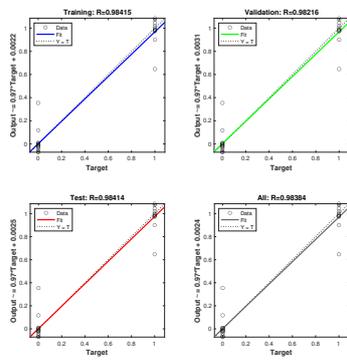
(b)



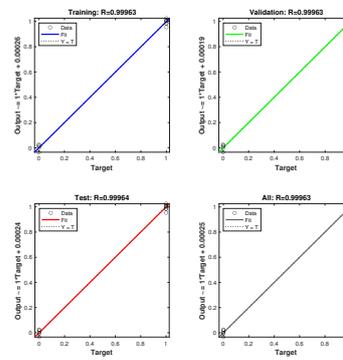
(c)



(d)



(e)



(f)

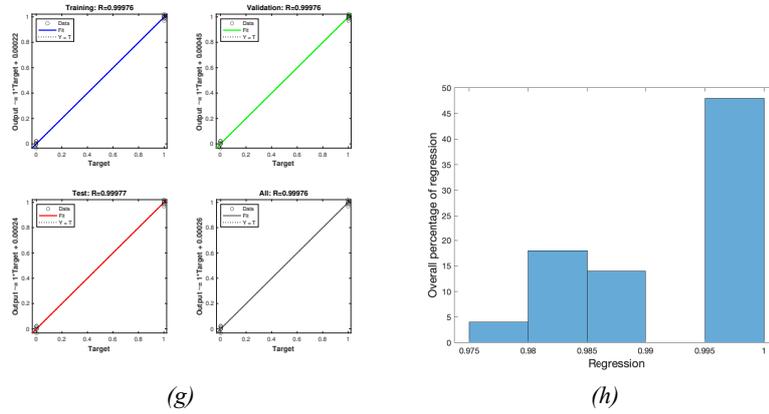
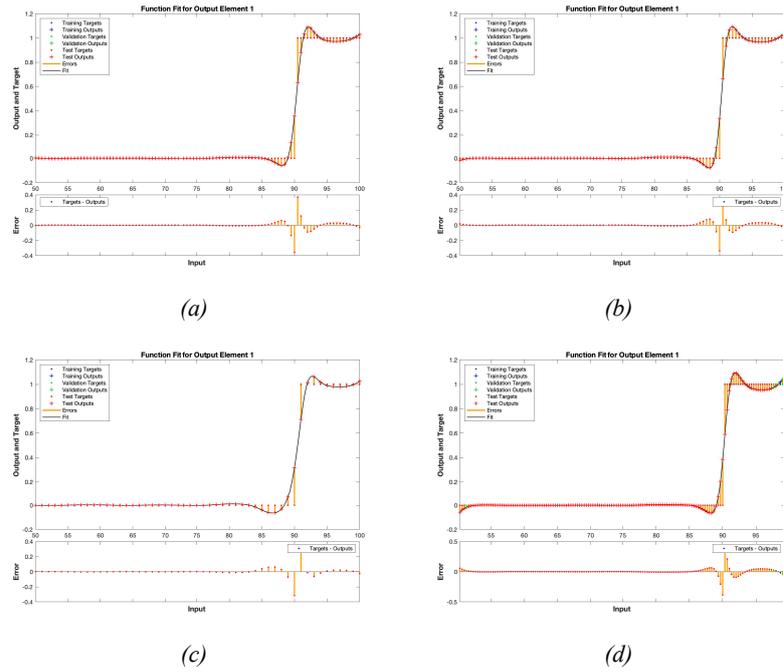


Figure 14: Regression of: (a) science1, (b) science2, (c) science3, (d) technology, (e) engineering, (f) art, and (g) mathematics, networks, and (h) the overall percentage.



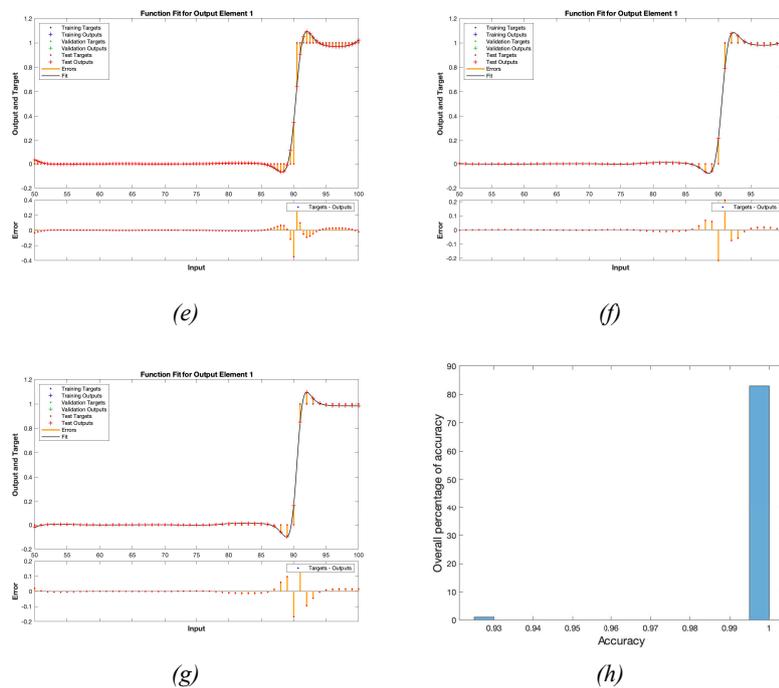


Figure 15: Regression of: (a) science1, (b) science2, (c) science3, (d) technology, (e) engineering, (f) art, and (g) mathematics, networks, and (h) the overall percentage.

The results of both models demonstrated a highly satisfactory performance and accuracy, denoting the ability of the proposed models to simulate both systems: classical and STEAM. Notably, in the SOO model, the maximum number of epochs required to achieve the best performance is 1,000. In the STS model, the maximum number of epochs for every fully connected neural network is inversely proportional to the grade number. The best number of epochs in the STS model for the first year is no more than 600 epochs for all criteria, while for the last year, the number is less than 20 epochs. The decreasing number of epochs in the STS model for the subsequent years is due to the cumulative forwarding of internal representations from the first grade to the following grades.

The other results of both models, such as the regressions, errors, performance, training states, and fitting curves, demonstrate their best response to obtain the best output. Particularly, the STS model's regression has a value of approximately 0.99 between the output and the target. This value represents the predicted academic results, which are entirely dependent on the input criteria's interests and skills.

Notably, the SOO and STS models can be compared to other related models, as shown in Table 7. Although this study's proposed models had substantial dataset entries, they provided more accurate responses (about 99%), with optimized performance and a very low error.

We can reference tables just like images. Here is an example of a reference to Table 7.

5 Discussion

To simulate the SsAP system and improve its performance using reliable and plausible parameters, two models have been proposed in this research:

1. SOO model: A supervised multilayer artificial neural network consisting of three layers: (a) an input layer with seven inputs for seven subjects, (b) a hidden layer, and (c) an output layer to predict the proper academic prediction. The input database only contained the final high school grade scores of 50,000 students. Such a proposal would be similar to the traditional SsAP system, which relies solely on the final year scores, regardless of the students' interests and skills.

2. STS model: Seven deep fully connected neural networks work in parallel to simulate a new approach based on STEAM criteria. The proposed model has seven criteria, twelve grades, and seven output stages. The seven criteria are derived from the students' interests and skills as a result of their STEAM education. The suggested model considers all the twelve grades to provide an accurate academic prediction based on deep academic background. The seven outputs that are divided into seven groups depending on the seven STEAM criteria are for the use of schools and institutions. The results of both models demonstrated a highly satisfactory performance and accuracy, denoting the ability of the proposed models to simulate both systems: classical and STEAM. Notably, in the SOO model, the maximum number of epochs required to achieve best performance is 1,000. In the STS model, the maximum number of epochs for every fully connected neural network is inversely proportional to the grade number. The best number of epochs in the STS model for the first year is no more than 600 epochs for all criteria, while for the last year, the number is less than 20 epochs. The decreasing number of epochs in the STS model for the subsequent years is due to the cumulative forwarding of internal representations from the first grade to the following grades.

The other results of both models, such as the regressions, errors, performance, training states, and fitting curves, demonstrate their best response to obtain the best output. Particularly, the STS model's regression has a value of approximately 0.99 between the output and the target. This value represents the predicted academic results, which are entirely dependent on the input criteria's interests and skills.

Notably, the SOO and STS models can be compared to other related models, as shown in Table 7. Although this study's proposed models had substantial dataset entries, they provided more accurate responses (about 99%), with optimized performance and a very low error.

The behavior of the STS model implies its ability to consider every single subject of every grade level to predict a proper academic path for students, The datasets, however, pose a challenge to such a proposal. Every student should have complete records from all their schools across their entire academic history. [Lesinski et al, 2016]

Parameter	[Putra et al, 2018]	[Prasetyawan et al, 2018]	[kurniadi et al, 2021]	[Rodriguez-Hernández et al, 2021]	SOO	STS
Dataset samples	1,318	21,731	337	162,030	50,000	50,000
Input layer neuron	9	8	4	122	7	7
Hidden layer neuron	1 hidden layer	1 hidden layer	1 hidden layer	1 hidden layer	1 hidden layer	12x7 FCNN
Output layer neuron	11	5	3	50	1	7
Activation function	Sigmoid	Sigmoid	Sigmoid	Sigmoid/Softmax	Sigmoid	Sigmoid
Error	N/A	0.001	0.040929 and 0.001675	N/A	0.00001	0.00001
Optimized Epochs	N/A	500	1000	200	1000	10-600
Learning rate	0.7	N/A	0.2	0.0001	0.000001	0.000001
Accuracy rate	89.56	85	93.43	82	99	99

Table 7: Comparison of related works

6 Conclusion

This study is a prospective investigation for those who expect academic and career performance in higher education, particularly with the continuous increase in the use of the latest machine learning technologies and big data. This research focused on building two models, the first simulating the classical Iraqi education system, where the student's acceptance depends on his cumulative average. The second predicts the acceptance of the student according to the global STEAM system, where the approval of the student depends on his skills and preferences in addition to his cumulative average. Wherever, the focus has been on the use of big data in our work, as opposed to previous work that used small samples of data. The proposed system also predicts the student's academic and career prospects. Furthermore, this research targets all age groups of students, from primary school to university and beyond. All formations and cadres of education and higher education institutions are represented by teachers, administrators, professors, and teachers. Besides, the focus was on using the best artificial intelligence techniques, machine learning, deep learning, and big data in contrast to what was found in previous works, which focused on one or more technologies.

Thus, this research is a starting point for the modernization of the education system by the Ministry of Education, the student admission system by the Ministry of Higher Education, and the job distribution system by the Ministry of Planning in Iraq. However, the measured conjugate algorithm is a supervised learning algorithm, which is a significant limitation. It cannot be used to rank students on background scores without specifying the target. In future work, we are looking for a specific and in-depth classification of individual colleges/institutes, not as general criteria or majors.

References

- [Albarka, 2019] Mubarak Albarka Umar, Student Academic Performance Prediction using Artificial Neural Networks: A Case Study, *International Journal of Computer Applications*, volume 178, pages 24-29,(2019).
- [Alborz et al, 2013] Alison Alborz, Roger Slee,Susie Miles, Establishing the foundations for an inclusive education system in Iraq: reflection on findings from a nationwide survey, *International Journal of Inclusive Education*, volume 17, pages 965-987,(2013).
- [Ang and Hann. 2019] Ivan Jie Xiong Ang, King Hann Lim, Enhancing STEM education using augmented reality and machine learning, *International Conference on Smart Computing & Communications (ICSCC)*,pages 1-5, (2019).
- [Adeniyi et al, 2021] Adeniyi Jide Kehinde, Abidemi Emmanuel Adeniyi, Roseline Oluwaseun Ogundokun, Himanshu Gupta, Sanjay Misra, Prediction of Students performance with Artificial Neural Network using Demographic Traits, *CoRR*, volume 855, pp 613–624, (2021).
- [Altabrawee et al, 2019] Hussein Altabrawee, Osama Ali, Samir Qaisar, Predicting students' performance using machine learning techniques, *JOURNAL OF UNIVERSITY OF BABYLON for pure and applied sciences*, volume 27,pages 194-205, (2019).
- [Ahammad et al, 2021] Khalil Ahammad, Partha Chakraborty, Evana Akter, Umme Honey Fomey, Saifur Rahman, A comparative study of different machine learning techniques to predict the result of an individual student using previous performances, *International Journal of Computer Science and Information Security (IJCSIS)*, volume 19, (2021).
- [Aich et al, 2019] Ankit Aich, Amit Dutta, Aruna Chakraborty, A scaled conjugate gradient backpropagation algorithm for keyword extraction, *Information Systems Design and Intelligent Applications*, pages 674-684, (2019).

- [Banadaki, 2020] Yaser M. Banadaki, Enhancing the Role of Machine Learning in STEM Disciplines through Supervised Undergraduate Research Experiences, Literacy Information and Computer Education Journal, volume 11 ,pages 3417-3423,(2020).
- [Bujang et al, 2021] Siti Dianah Abdul Bujang, Ali Selamat, Roliana Ibrahim, Ondrej Krejcar, Enrique Herrera-Viedma, Hamido Fujita, Nor Azura Md. Ghani, Multiclass prediction model for student grade prediction using machine learning, IEEE Access, IEEE, volume 9,pages 95608-95621, (2021).
- [Bertrand and Namukasa, 2020] STEAM education: student learning and transferable skills Marja G. Bertrand, Immaculate K. Namukasa, STEAM education: student learning and transferable skills, Journal of Research in Innovative Teaching and Learning, Emerald Publishing Limited, Volume 13, (2020).
- [Helal et al, 2018] Sumyea Helal, Jiuyong Li, Lin Liu, Esmail Ebrahimie, Shane Dawson, Duncan J. Murray, Qi Long, Predicting academic performance by considering student heterogeneity, Knowledge-Based Systems, volume 161, pages 134-146, (2018).
- [Hagan, M.T., and M. Menhaj, 1994] M T Hagan, M B Menhaj, Training feed-forward networks with the Marquardt algorithm, IEEE Transactions on Neural Networks, volume 5, No. 6, pages 989–993, (1994).
- [Hagan, M.T., H.B. Demuth, and M.H. Beale, 1996] Martin T. Hagan, Howard B. Demuth, Mark Hudson Beale, Orlando De Jesús, Neural Network Design, Boston, MA: PWS Publishing, (1996).
- [Issa et al, 2010] Jinan Issa, Hazri Jamil, Overview of the education system in contemporary Iraq, European Journal of Social Sciences, volume 14,pages 360-368, (2010).
- [Iqbal et al, 2017] Zafar Iqbal, Junaid Qadir, Adnan Noor Mian, Faisal Kamiran, Machine learning based student grade prediction: A case study, <https://doi.org/10.48550/arXiv.1708.08744>,(2017).
- [Jin et al, 2020] Lianchao Jin, Fuxiao Tan, Shengming Jiang, Generative adversarial network technologies and applications in computer vision, Computational Intelligence and Neuroscience, Hindawi, (2020).
- [Khalid et al, 2020] Mustafa Khalid, Jun Wu, Taghreed M. Ali, Ahmed A. Moustafa, Qiuguo Zhu, Rong Xiong, Green model to adapt classical conditioning learning in the hippocampus, Neuroscience, volume 426, pages 201-219, (2020).
- [Khalid et al, 2020] Mustafa Khalid, Jun Wu, Taghreed M. Ali, Ahmed A. Moustafa, Qiuguo Zhu, Rong Xiong, Green model to adapt classical conditioning learning in the hippocampus, Neuroscience, volume 426, pages 201-219, (2020).
- [kurniadi et al, 2021] Prediction of courses score using Artificial Neural Network with Backpropagation algorithm, IOP Conference Series: Materials Science and Engineering, volume 1098, (2021).
- [Lesinski et al, 2016] Gene Lesinski, Steven Corns, Cihan Dagli :Application of an Artificial Neural Network to Predict Graduation Success at the United States Military Academy, Procedia Computer Science, volume 95,pages 375-382, (2016).
- [Marquardt, D., 1963] Donald W. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, SIAM Journal on Applied Mathematics, volume 11, No. 2 pages 431–441, (1963).
- [Moller, 1993] Martin Møller , A Scaled Conjugate Gradient Algorithm For Fast Supervised Learning. Neural Networ, volume 6,pages 525–533, (1993).
- [Mohamed et al, 2022] Yahia Mohamed, Gamal Alkaws, Abdulsalam Salihu Mustafa, Ammar Ahmed Alkahtani, Yazan A. Alsariera, Abdulrazzaq Qasem Ali, Toward Predicting Student's Academic Performance Using Artificial Neural Networks (ANNs), Applied Sciences, volume 12, (2022).

- [Mandalapu and Gong, 2019] Varun Mandalapu, Jiaqi Gong, Studying Factors Influencing the Prediction of Student STEM and Non-STEM Career Choice, Proceedings of the 12th International Conference on Educational Data Mining, (2019).
- [Mengash, 2020] Hanan Abdullah Mengash, Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems, IEEE Access, volume 8, pages 55462-55470, (2020).
- [Mathew et al, 2021] Amitha Mathew, Amudha Arul, S. Sivakumari, Deep learning techniques: an overview, Mathew, Amitha and Amudha, P and Sivakumari, pages, 599-608, (2021).
- [Nabil et al, 2021] Aya Nabil, Mohammed Seyam, Ahmed Abou-Elfetouh, Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks, IEEE Access, volume 9, pages 140731-140746, (2021).
- [Nguyen et al, 2020] Hien D. Nguyen, Nhon V. Do, Nha P. Tran, Xuan Hau Pham, Vuong T. Pham, Some criteria of the Knowledge Representation method for an Intelligent Problem Solver in STEM education, Applied Computational Intelligence and Soft Computing, (2020).
- [Olalekan et al, 2020] Ajinaja Micheal Olalekan, Ojonukpe Sylvester Egwuche, Sylvester Oluyemi Olatunji, Performance Evaluation of Machine Learning Techniques For Prediction Of Graduating Students In Tertiary Institution, International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS), pages 1-7, (2020).
- [Polyzou et al, 2019] Agoritsa Polyzou, George Karypis, Feature extraction for next-term prediction of poor student performance, IEEE Transactions on Learning Technologies, IEEE, volume 12, pages 237-248, (2019).
- [Prasetyawan et al, 2018] Purwono Prasetyawan, Imam Ahmad, Rohmat Indra Borman, Ardiansyah, Yogi Aziz Pahlevi, Dwi Ely Kurniawan, Classification of the Period Undergraduate Study Using Back-propagation Neural Network, International Conference on Applied Engineering (ICAE), pages 1-5, (2018).
- [Putra et al, 2018] R E Putra, A I Nurhidayat, A Y Wicaksono, Implementation of neural network to determine the new college students, IOP Conference Series: Materials Science and Engineering, volume 288, (2018).
- [Qazdar et al, 2019] Aimad Qazdar, Brahim Er-Raha, Chihab Cherkaoui, D. Mammass, A machine learning algorithm framework for predicting students performance: a case study of baccalaureate students in Morocco, Education and Information Technologies, Springer, volume 24, pages 3577-3589, (2019).
- [Rodríguez-Hernández et al, 2021] Carlos Felipe Rodríguez-Hernández, Mariel Musso, Eva Kyndt, Eduardo Cascallar, Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation, Computers and Education: Artificial Intelligence, Volume 2, (2021).
- [Rastrollo et al, 2020] Analyzing and predicting students' performance by means of machine learning: A review. Applied sciences, 10(3), p.1042, (2020).
- [Selvia et al, 2021] Selvia Lorena Brginting, M. Aditya Fathur Rahman, Data Mining, Neural Network Algorithm To Predict Student'S Grade Point Average: Backpropagation Algorithm, Journal of Engineering Science and Technology, volume 16, pages 2028-2037, (2021).
- [Sabukunze Didier et al, 2021] Igor Didier Sabukunze, Yohanes Alvinika, Billy Josef Waroruntu, Paulus Mudjihartono, Prediction of Students' GPA Using Levenberg-Marquardt Backpropagation Algorithm, International Conference for Convergence in Technology (I2CT), pages 1-5, (2021).
- [Santana et al, 2020] Otacilio Antunes Santana, Barbara Alves de Sousa, Sandra Razana Silva do Monte, Mayara Lopes de Freitas Lima, Caina Ferraz e Silva, Deep learning practice for high school student engagement in STEM careers, IEEE Global Engineering Education Conference (EDUCON), pages 164-169, (2020).

[Spyropoulou et al, 2020] Natalia Spyropoulou, Dimitrios Glaroudis, Athanasios Iossifides, Ioannis D. Zaharakis, *ostering Secondary Students' STEM Career Awareness through IoT Hands-On Educational Activities: Experiences and Lessons Learned*, IEEE Communications Magazine, volume 58, pages 86-92, (2020).

[Tatar and Dilek, 2020] Ahmet Emin Tatar , Dilek Düşteğör , *Prediction of Academic Performance at Undergraduate Graduation: Course Grades or Grade Point Average?*, Applied Sciences, volume 10, pages 4967, (2020).

[Uddin et al, 2019] Shahadat Uddin, Arif Khan, Md Ekramul Hossain, Mohammad Ali Moni ,*Comparing different supervised machine learning algorithms for disease prediction*, BMC medical informatics and decision making, volume 19, pages 10-16, (2019).

[Yang et al, 2020] Mei Yang, Shah Nazir, Qingshan Xu, Shaukat Ali, *Deep learning algorithms and multicriteria decision-making used in big data: a systematic literature review*, Complexity, Hindawi, (2020).

[Zhang et al, 2020] Yupei Zhang, Yue Yun, Huan Dai, Jiaqi Cui, Xuequn Shang, *Graphs regularized robust matrix factorization and its application on student grade prediction*, Applied Sciences, volume 10, pages 1755, (2020).