Hindawi

Research Article

Personalized Hybrid Education Framework Based on Neuroevolution Methodologies

Wenjing Yin [ID]

Zhengzhou Preschool Education College, Zhengzhou, 450000, China

Correspondence should be addressed to Wenjing Yin; wenjing_yin66@126.com

The future pedagogical systems need anthropocentric inclusive educational programs in which the goal should be adjustable according to the knowledge requirements, intelligence, and learning objective of each student. Prioritizing these needs, innovative AI methods are required to assist and ensure the making of conscious educational decisions, in terms of clear identification and categorization with high accuracy of various forms of skills and knowledge of each student. This paper proposes a neuroevolution emerging technique that combines the searchability of evolutionary computation and the learning capability of a hybrid artificial neural networks method. Specifically, the proposed growing semiorganizing neural gas (GsONG) is a practical AI methodology utilizing advanced clustering techniques to enhance the learning experience by categorizing the true abilities, skills, and needs of learners, in an inclusive differentiated learning framework. It is a neural network architecture that includes competing and cooperating neurons with an unstructured mode whereby a cooperation-competition process delimits the topological neighborhood of neurons in a grid to identify patterns for which their classes are not known. To optimize the above process, a heuristic method was used that investigates the space of an objective function by regulating the optimal topologies of neurons that form pathway segments in a semi-contemplative manner. Based on the extensive experiments and results obtained from the GsONG clustering approach, the proposed algorithm can compensate with high accuracy for difficulties in multicriteria grouping and differentiation of uncertainty structures such as in small or tiny data sets.

1. Introduction

Inclusive education as a part of differentiated teaching is an organized strategy that is a product of interdisciplinary collaboration and that as a pedagogical approach places at the center the uniqueness of each student, their unique skills, and their needs [1, 2]. It is concerned with the fact that teaching and learning should start from the level of students, instead of being based on a specific and predetermined action plan, which does not include students' readiness, interest, and learning profile [3].

It is essentially the organizational and pedagogical adaptation of inclusive education to meet the various needs of students in mixed grades (lean, mediocre, more capable, and charismatic students) while taking into account the differences of students in terms of readiness, learning style, and interests. In a learning environment characterized by student diversity, each student's personal abilities and strengths are recognized, and he is given opportunities to utilize and demonstrate his various skills through a variety of assessment techniques. It is important to say that differentiated teaching in a learning environment where there is a diversity of students is properly organized and planned and in no way is a product of a random or unbalanced process of allocating learning resources [4].

In this sense, inclusive differentiated teaching is more qualitative than quantitative, with the emphasis on the fact that some students are not given more work than others but that the work is tailored to their specific needs. In this context, interconnected and adequately planned educational activities are offered based on the uniqueness of each student, their skills, and their level. In this way, it is possible to have multiple approaches to the content, the course, and the result of the educational process.

As appealing as teaching differentiation is, it is true that in order for this differentiation to occur and work in a classroom, the curriculum must be adapted. Achieving it is directly related to the functional ways in which students' critical thinking development can take place, as well as through the opportunities offered to students to demonstrate what they have learned [5]. In conclusion, effective differentiated teaching utilizes various methods, means, and materials manage to meet the special needs of all students by increasing their learning opportunities, offering a comfortable and positive environment, where all students achieve high academic achievements [6].

The aim of this paper is to propose a technological system that utilizes computer intelligence algorithms to facilitate decision-making related to curriculum adaptation and categorization of students, based on individual assessments of their unique characteristics. This methodology provides the technological background for automated differentiation of teaching, taking into account characteristics such as the treatment of each student as a person with a history of learning in and out of school, modeling of student metacognitive development, linking to prior knowledge, and building knowledge with gradual progress, according to the perceiving level of each student [7]. Specifically, a differentiated learning framework is proposed, which, with the extensive use of machine learning algorithms, as well as optimization methods, creates a clear framework for multicriteria assessments, to classify students into small groups. The students belonging to each group will have uniform features of learning ability, difficulties in the cognitive object, and their psychosocial and perceptive summary.

The objective of the learning activity determines whether to form groups with students of comparable or mixed ability. If the goal of a group learning activity is to assist struggling pupils, research reveals that heterogeneous groups are the most effective. On the other side, homogeneous grouping might be preferable if the goal is to promote medium-ability groups to learn at high levels. In an educational setting, homogeneous grouping is defined as grouping students of similar instructional levels together so that they can work on materials that are best suited to their individual strengths and areas for improvement. Assessment and instructor observation are frequently used to identify these ability levels [8]. Homogeneous groups enable teachers to create lesson plans that are suited to their students' skills and save time by addressing individual requirements. In any event, because student ability levels differ by subject, categorizing students based on their skills ensures that they are ranked appropriately even within homogeneous groups [7].

On the other hand, heterogeneous groups are preferable if the goal is to raise difficult pupils and help them develop the independence and collaborative skills that come with reduced reliance on the teacher. However, this technique will reduce the value of the exercise for talented and at-grade pupils, even if they will have the opportunity to practice communication and leadership abilities. While some bright children enjoy the benefits and responsibilities that varied groups offer, others dislike them. Knowing the hierarchy of competency and talents, as well as what their peers are

capable of, such students are aware that they may be responsible for a bigger portion of the work while dealing with apathetic or disruptive group members. Above-grade-level pupils are unlikely to be challenged by an exercise and will instead become annoyed by everyone's inability to accomplish it as well or as quickly as they could. While they will have the opportunity to develop other talents, they will miss out on the opportunity to learn as much as possible from the task at hand. In the worst-case situation, if a student is very spirited or requires regular stimulation, they may become disruptive and obstruct the learning of the rest of the group or perhaps the entire class. There is a danger that pupils in grade level and below will take a back seat in heterogeneous groups and allow their more gifted classmates to do all the work. This is especially true if the more capable students have assertive personalities, a strong desire to complete the assignment as quickly as possible, and a dismissive attitude toward their peers' talents and efforts.

The proposed system can be used to implement individualized teaching programs in homogeneous learning groups, based on the diversity that characterizes each student and person in general.

2. Related Literature

The related literature of the proposals on the educational methods to provide differentiated and personalized training mainly focuses on theoretical proposals. The potential of being able to provide personalized training in the educational process is great because it gives to the learner the best possible learning experience. Personalized e-learning encompasses a variety of educational technology and pedagogical methods that take into account individual student variations [4] and can customize the generic virtual training environment to meet their own needs. The competence of professors and educational material quality are key factors in e-learning programs [9], but instructors' competency is revealed through their approach, teaching style, and assistance throughout the entire online education experience. All these elements contribute to the positive influence on students' emotions and subsequently trigger flow. Ardura and Artola [10] intend to contribute to the flow and fill the hole that has been identified in terms of knowledge possessed in personalized e-learning settings in their extensive investigation. They also look at how education-related and subjective factors interact to create flow in a personalized e-learning environment, as well as evidence supporting the moderating influence of individual differences in gender and academic achievement. Even though this is detailed research, the limitations are that the timeframe for the calculations is limited making it impossible to say if the reported associations will alter over time as new e-learning methods for customization emerge. Furthermore, the sample of the study is limited to just one online college leaving the field open for contribution and enhancement of this model and the generalization of the results through further study.

In his paper, Nganji [11] provides a concept for a learning environment that allows learners to create their own online learning spaces and engage with pooled

materials. This model is called Flexible and Accessible User Constructed Learning Environment (FAUCLE) and essentially is a learner-centered model that connects the elements that contribute to the e-learning process and creates relationships among them. These are the learner, the educator, the accessibility expert, the e-learning programmer, the learning environment, and the learning content apps. The prerequisite for this model to take place is for the institution that provides the training to make applications and other resources available to these students in a form that is both accessible and useable, allowing them to create their learning environments with flexibility. The fact that this model is theoretical means that it needs to be implemented and tested to yield empirical data that can be compared to other studies. Milicevich and Ivanovic [7] in their work aim to address the research by examining aspects of tailored e-learning, as well as intelligent and interactive technologies. They show online education systems that are currently state-of-the-art that are powered by artificial intelligence at the end. Their study is also theoretical and serves as a suggestion for institutions and organizations who want to adopt these new technologies and approaches in e-learning. Maghsudi et al. [12] in their study give a quick rundown of current research, look into the challenges of AI/ML-based individualized education, and propose possible solutions. They conclude that "personalized education" is one of artificial intelligence's most valuable educational merits because it significantly improves education quality in several dimensions by adapting to the unique characteristics and expectations of each learner, such as personality, talent, objectives, and background. In addition, in unusual circumstances such as the COVID-19 outbreak [13] or natural disasters, online teaching is invaluable. Indeed, traditional education requires substantially more resources than online education in terms of classroom space, scheduling, and human resources, making it vulnerable to failure in the event of even minor changes in circumstances. As a result, new alternatives are unavoidable [14, 15]. Personalized education, despite the potential for a dramatic shift from traditional to modern education paradigms, is fraught with difficulties.

From the above literature, the conclusion is that the research community focuses on finding ways to combine technology with the traditional educational processes to provide a differentiated experience. What this paper does is go a step further and propose a practical methodology utilizing machine learning to enhance the learning experience.

3. The Gsong

To create an intelligent framework for categorizing the true abilities, skills, and needs of learners, a differentiated learning framework is proposed, which makes extensive use of nonsupervised machine learning technologies to achieve its goals [16]. Unsupervised learning is the ability to identify patterns for which their classes are not known, and the system generates predictions based on some distribution or some quantitative measures to evaluate and characterize the similarity of data in corresponding groups called clusters.

The general clustering technique [17] is based on the idea that a data set $D = \{x_1, x_2, \ldots, x_n\}F$, where $x_i = (x_{i1}, x_{i2}, \ldots, x_{ir})$ is a characteristic of the set $X \in R^r$ and r is the number of dimensions in data. After defining the set of blocks K, assign each point $x^{(i)}$ of the data set to a block C_k so that the $S_{core}(C, D)$ is maximized or minimized as appropriate. How to calculate a function of this type is given by the following formula [17]:

$$S_{core}(C, D) = \sum_{k=1}^{K} d(x, c_k), \tag{1}$$

where $c_k = 1/n_k \sum_{x \in c_k} x$ and $d(x, y) = \|x - y\|^2$.

Elements that belong to a cluster show greater or relatively greater similarity. The training of a computational intelligence model with the method of clustering is called to calculate and finally classify into clusters, data $x_1, \ldots, x_n$ without giving the values $f(x_1), \ldots, f(x_n)$. In this work, the proposed model applies clustering based on competitive learning and is specifically implemented using artificial neural networks [14, 18].

Specifically, the methodology provides for the classification of students into homogeneous groups based on their particular characteristics [19]. For this reason, a neural network was created that includes a competitive layer of competing neurons. Each competing neuron i is characterized by a weight vector $w_i = (w_{i\,d}, \ldots, w_{i\,d})^T$, $i = 1, \ldots, M$, and computes a similarity measure among the input information $x_i = (x_{i\,d}, \ldots, x_{i\,d})^T$ $x \in R$ and in w_i weights vector. Each time a student's characteristics appear as an entry in the network, competition is created between competing-level neurons to identify the winning neuron whose weight vector shows the greatest similarity to that input. The winning neuron m sets the output of $o_m = 1$, while the other neurons give $o_i = 0$, where $i = 1, \ldots, M$ and $i \neq m$. As a measure of similarity to find the winner neuron, a function inversely proportional to the Euclidean distance $\|x - w_i\|$ of the input vector x^n from the vector of w_i weights was used. Therefore the proposed neural network implements a representation of the input x, dimension d, in the coordinates of the grid $r_m = (z_{m1}, z_{m2})^T$.

More specifically, the proposed neural system forms a self-organizing map of its structures, starting from the process of initializing the weights $w_i = (w_{il}, \ldots, w_{i\,d})^T$. To achieve this, small weight values generated by a random number generator are given. The weight table of the hidden layer H is calculated as follows [20]:

$$H = g(\omega x + b). \tag{2}$$

The output weights β are calculated based on the following function:

$$\beta = \left(\frac{I}{C} + H^T H\right)^{-1} H^T X, \tag{3}$$

where $H = [h_1, \ldots, h_N]$ is the hidden level outputs and $X = [x_1, \ldots, x_N]$ is the input data. β can also be calculated from the general relation as follows:

$$\beta = H^+ T. \tag{4}$$

After this initialization, there are three basic procedures:

Competition. For each x^n, the grid neurons calculate the similarity function. The winner is the neuron with the highest similarity value. The Euclidean distance between $x = (x_1, \ldots, x_d)^T$ $x \in R$ and $w_i = (w_{il}, \ldots, w_{i\,d})^T$ of the opposing neurons is used as a function of similarity.

Cooperation. The winner-neuron delimits the topological neighborhood of neurons in the grid, which will adjust their weights to the input vector. $h_{j,i}$ denotes the topological neighborhood centered on the winner-neuron i, which includes a set of neurons, one of which is denoted as j. Also denoted by $d_{j,i}$ is the distance between the winner neuron I and a neuron j. Thus, the topological neighborhood satisfies the above constraints [17, 21].

$$h_{j,i}(x) = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right), \tag{5}$$

where parameter σ is the topological neighborhood's effective width, which defines how many neurons in the winner's neighborhood participate in the training process. This parameter decreases in each season n at an exponential rate according to the relation [21, 22]:

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_1}\right), n = 0, 1, 2, \ldots, \tag{6}$$

where parameter σ_0 is the initial value of the active amplitude and τ_1 is the polarity constant of the network.

Synaptic Adaption. In this last stage of the training process, the weights of the competing-level neurons are updated. The amount of this change is given by the following relation [23, 24]:

$$\Delta w_j = \eta h_{j,i}(x)\left(x - w_j\right), \tag{7}$$

where i is the winning neuron and j is the neuron in the neighborhood of i. Finally, given the vector of weights $w_j(n)$ for a given time n, the new vector for the time $n + 1$ can be calculated from the following relation:

$$w_j(n + 1) = w_j(n) + \eta(n) h_{j,i}(x)(n)\left(x(n) - w_j(n)\right). \tag{8}$$

From the above relation, it follows that the learning rate $\eta(n)$ depends on time. More specifically, it starts from an initial value of η_0 and decreases exponentially with increasing time n [25]:

$$\eta(n) = \eta_0 \exp\left(-\frac{n}{\tau_2}\right), n = 0, 1, 2, \ldots, \tag{9}$$

where τ_2 is the polarization constant of the network.

In addition, the above process is divided into two phases:

Ordering phase is the initial phase, and it is during this phase that the competing-level weights are topologically arranged. During this phase, the learning rate $\eta(n)$ starts
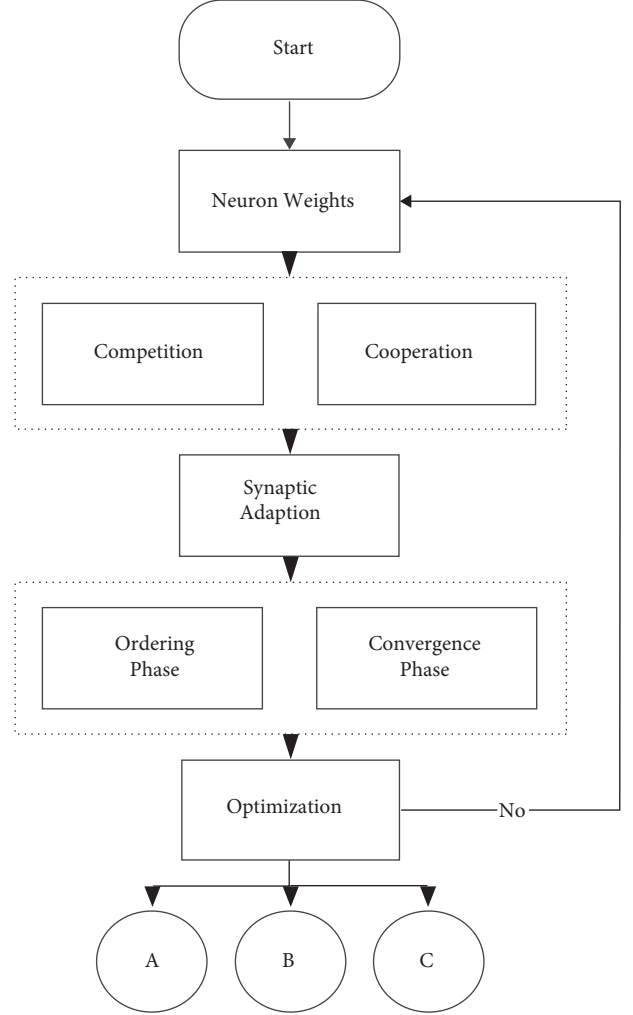

FIGURE 1: The logical data flow of GsONG.

TABLE 1: Statistical analysis of data set.

	V1	V2	V3	V4	V5
Mean	21.12	32.08	8.63	2.31	4.17
1st Q	23.09	34.72	9.40	2.35	4.38
Median	24.34	35.19	9.21	2.48	4.63
Mean	24.44	35.87	9.04	2.42	4.60
Std	3.45	4.89	0.97	0.49	0.59
3rd Q	25.51	37.82	9.67	2.73	4.81
Max	26	39	10	3	5

from a value around 0.1 and gradually decreases, up to the value of 0.01. These values are achieved as follows:

$$\eta(n) = \eta_0 \exp\left(-\frac{n}{\tau_1}\right), n = 0, 1, 2, \ldots,$$
$$\text{with } \eta_0 = 0.1 \text{ and } \tau_2 = 1000. \tag{10}$$

The following network polarization values were used to calculate the learning rate and active amplitude values in each iteration:
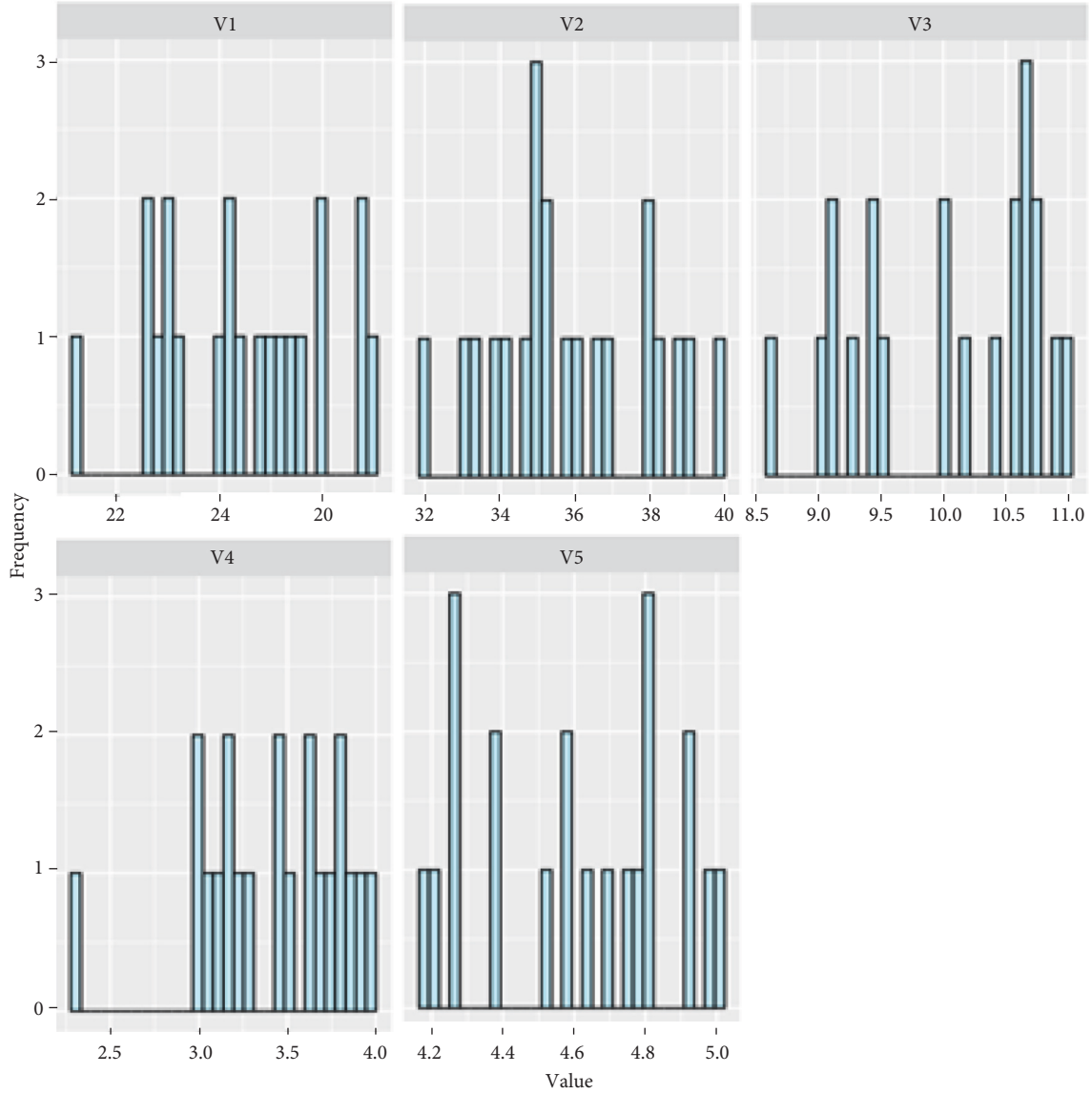
FIGURE 2: Histogram for each feature of data set.

$$\tau_2 = \frac{n_0}{lnln(100 \cdot \eta_0)} \text{ and } \tau_1 = \frac{n_0}{lnln(\sigma_0)}, \quad (11)$$

where n_0 is the number of repetitions of the phase of the device, η_0 is the initial learning rate, and σ_0 is the initial value of the active amplitude that in turn results from the following relation:

$$\sigma_0 = \sqrt{w^2 + h^2}, \quad (12)$$

where w and h are the length and height of the two-dimensional grid, respectively.

Also, the topological neighborhood function $h_{j,i(n)}$ initially includes almost all competing-level neurons centered on the winning neuron and is gradually limited to a few neurons or even just the winning neuron. Considering a two-dimensional frame, the value of the "radius" of the grid was taken as the initial value σ_0 of the active width and as the value of the parameter τ_1 of the above relation [26]:

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_1}\right), \, n = 0, 1, 2, \ldots \text{ with } \tau_1 = \frac{1000}{\log \, \log \sigma_0}. \quad (13)$$

Convergence phase is the phase in which the weights acquire their final values better coordinated in the training examples. In this phase, the number of repetitions was determined by the dimension of the network inputs. The learning rate $\eta(n)$ remained constant at values close to 0.01, and finally, the neighborhood $h_{j,i(n)}$ was limited to the nearest neighbors of the winner-neuron, ending up containing only the winner neuron.

In the second phase (convergence), the values of the learning rate and the active amplitude remained constant and equal to 0.01 and 0.0001, respectively.

Regarding the learning rate, the variables that were selected are λ_i and λ_f that control the rate at which the neural network learns, while ε_i and ε_f define the initial and final rate
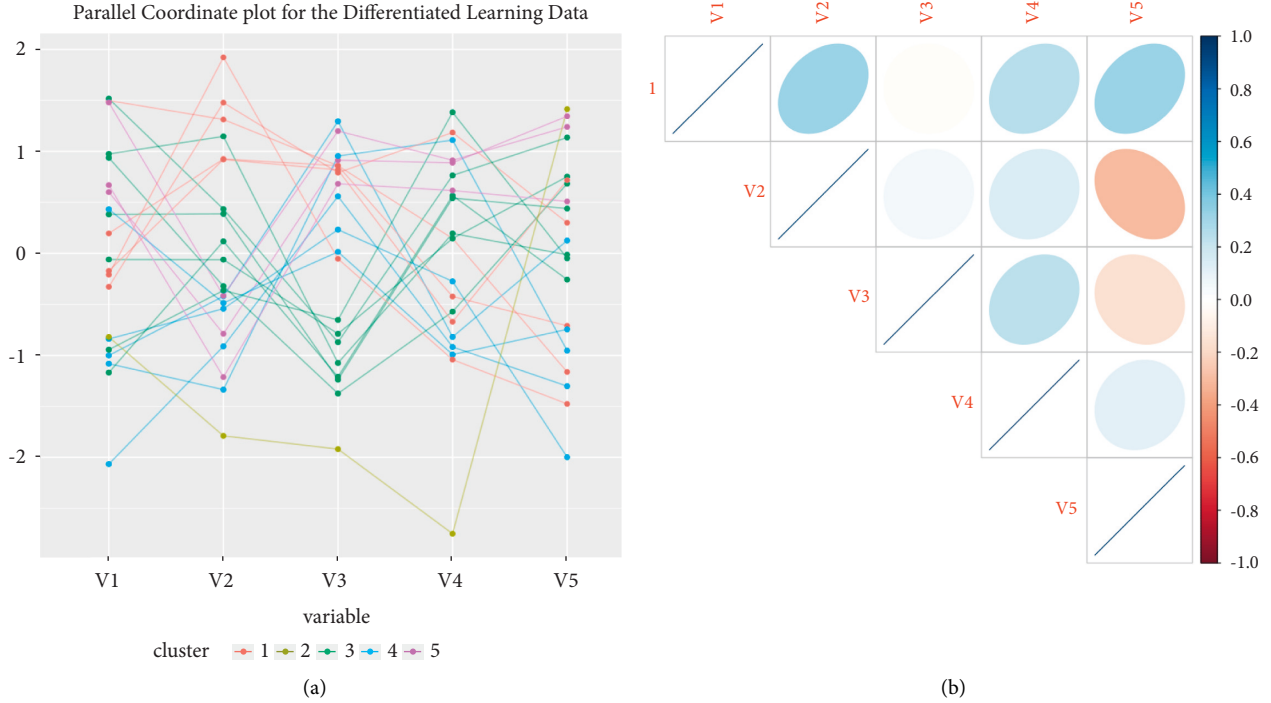
(a)

(b)

Figure 3: (a) Parallel coordinates plot between features and (b) correlation matrix between features.

according to which the neural network is trained. The variable t_{max} is the maximum number of execution times. With t the current season, t_{max} the total number of seasons, $\overrightarrow{x}$ the input signal generated at the beginning of each season, n each network node, n_w the vector that carries each neuron, and k is the degree of each node once, it has been classified into steps. All nodes were sorted in ascending order based on the Euclidean distance of their vectors from the input signal as follows [21, 26]:

$$\left\| \overrightarrow{x} - \overrightarrow{n_w} \right\|^2. \tag{14}$$

The weights of the nodes were adjusted in the order we have arranged them so that [27]

$$\overrightarrow{n_w} \leftarrow \overrightarrow{n_w} + \left[\overrightarrow{n_w} \times e(t) \times h(k) \times (\overrightarrow{x} - \overrightarrow{n_w}) \right], \tag{15}$$

where

$$h(t) = \exp\left(\frac{-k}{\sigma^2(t)}\right),$$

$$\sigma^2(t) = \lambda_i \times \left(\frac{\lambda_f}{\lambda_i}\right)^{t/t_{max}}, \tag{16}$$

$$\varepsilon(t) = \varepsilon_i \times \left(\frac{\varepsilon_f}{\varepsilon_i}\right)^{t/t_{max}}.$$

To optimize the above process, a heuristic method was used that investigates the space of an objective function by regulating the optimal topologies of neurons that form pathway segments in a semi-contemplative manner. Each neuron is attracted to the position of the best location found
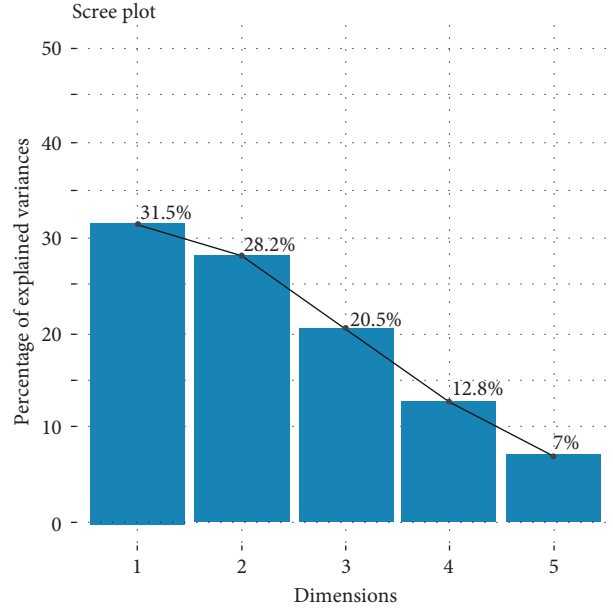

Figure 4: Principle component analysis.

by the heuristic function and the best location it has encountered, while, at the same time, it tends to move randomly.

Specifically, when an entity i discovers a locality that is superior to the previous ones it located, then it upgrades it to the best current for i. There is a current best for all n entities at any time t, during iterations. The goal is to find the best overall until the position of the neuron can no longer be improved [28].
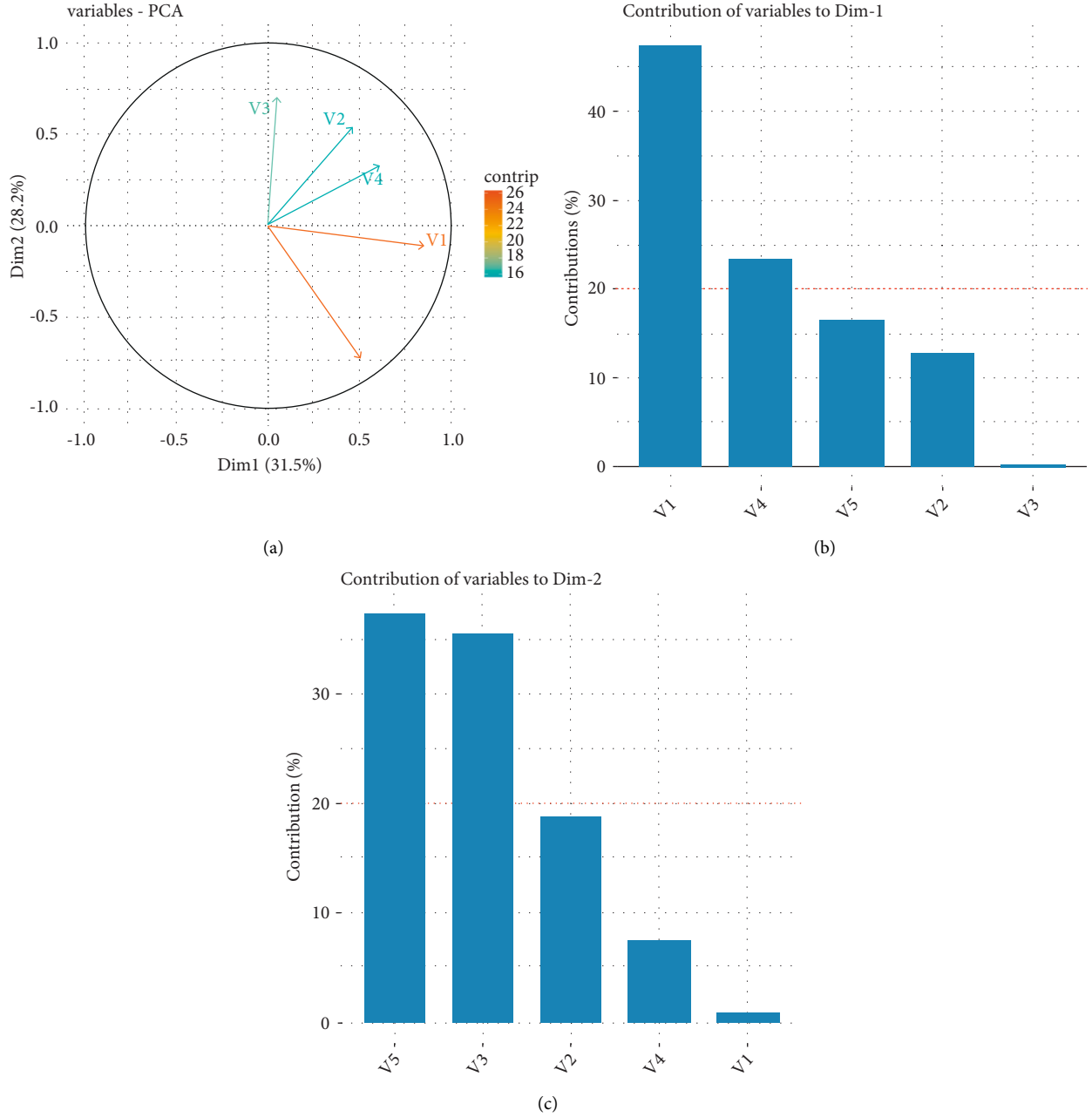
(a)

(b)

(c)

FIGURE 5: (a) Plot of contribution of PCA, (b) plot of contribution Dim-1, and (c) plot of contribution Dim-2.

Let p and u be the position and velocity for entity i, respectively. The new velocity vector is identified by the following formula [29]:

$$u_{n,m}^{\text{new}} = u_{n,m}^{\text{old}} + \Gamma_1 \times r_1 \times \left(p_{n,m}^{\text{local}_{\text{best}}} - p_{n,m}^{\text{old}} \right) + \Gamma_2 \times r_2 \times \left(p_{n,m}^{\text{global}_{\text{best}}} - p_{n,m}^{\text{old}} \right), \quad (17)$$

where $u_{n,m}$ represents the convergence speed, r_1, r_2 represents independent random numbers, Γ_1, Γ_2 represents learning parameters, $p_{n,m}^{\text{local\_best}}$ represents the best local solution, and $p_{n,m}^{\text{global\_best}}$ represents the best total solution.

The heuristic optimization algorithm renews the convergence speed component and then adds speed to the position component. This renewal depends on both the optimal solution/position discovered and the one used by all active neurons. If, at some point, the best solution discovered is better than that of the population, it replaces it. The initial locations of all neurons consist of being evenly distributed so that they are a sample for most areas of the search space. It is also possible for the original vector of an entity to be taken as zero. The new location is described by the following equation [22, 29] (Figure 1):

$$p_{n,m}^{\text{new}} = p_{n,m}^{\text{old}} + u_{n,m}^{\text{new}}, \quad (18)$$

where u is delimited to a range $[0, u_{\text{max}}]$.
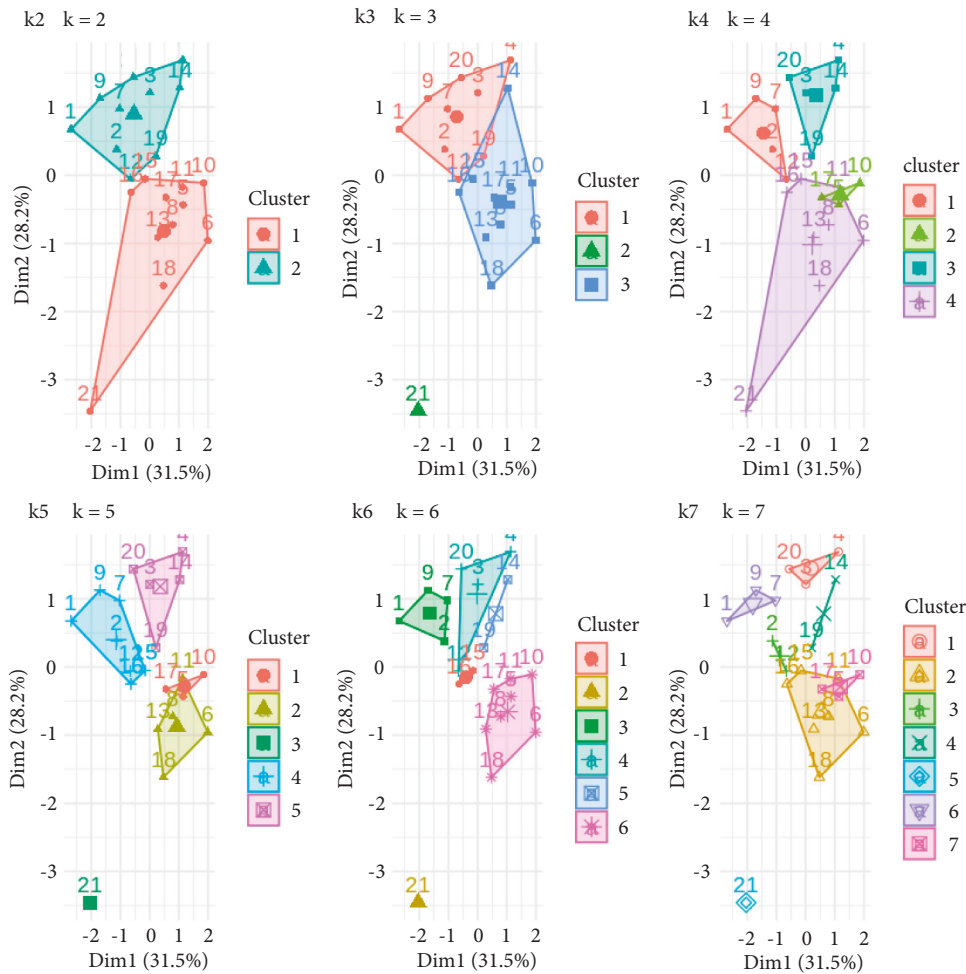
FIGURE 6: Plot of $k = [2–7]$ centers.

4. Experiments

The aforementioned algorithm was applied to student assessment data to handle the problem of categorizing students into heterogeneous groups with comparable features at the group level, where we assume that we have students who should be classified into groups at most.

The data used relate to quantitative individual performance and psychosocial data of students of a heterogeneous class of students, to implement the multicriteria test procedures.

Specifically, the data refer to the holistic assessment of an elementary school student class with a total capacity of 21 people, where Raven's IQ test was used to assess general mental ability (V1), and the math performance test for primary school students was used to determine mathematical skills (V2). The student's grade point average was used to evaluate performance (V3). The learning disability scale was used to determine social or emotional skills or difficulties (V4). Finally, the psychosocial adjustment tool was used to assess social or emotional skills or difficulties

(V5). It is important to emphasize that the problem is trying to be identified on a completely realistic basis, based on the real process that a teacher would follow in a department applying differentiated learning but within the children who are already part of a department. Specifically, the 21 children who participate in this classification process based on the proposed system are a real school class.

Specifically, the 21 children who participate in this classification process based on the proposed system are a real school class. The separation attempted is a realistic approach where the groups of 3, 4, 5, or 6 children that may arise are a fully satisfactory sample of children with homogeneous elements to whom personalized learning techniques can be applied. It should be emphasized that the resulting groups do not form new classes, but groups that receive training materials, instructions, exercises, and so on depending on the level in which they were classified but all within the same classroom.

In order to prove the correct use of this data set, a thorough preprocessing of the data was performed for the purpose of validation checks that prove the reliability of the data set under consideration, before the use of the proposed algorithm. This process is necessary as the initial data often suffer from various kinds of problems, such as conflicting
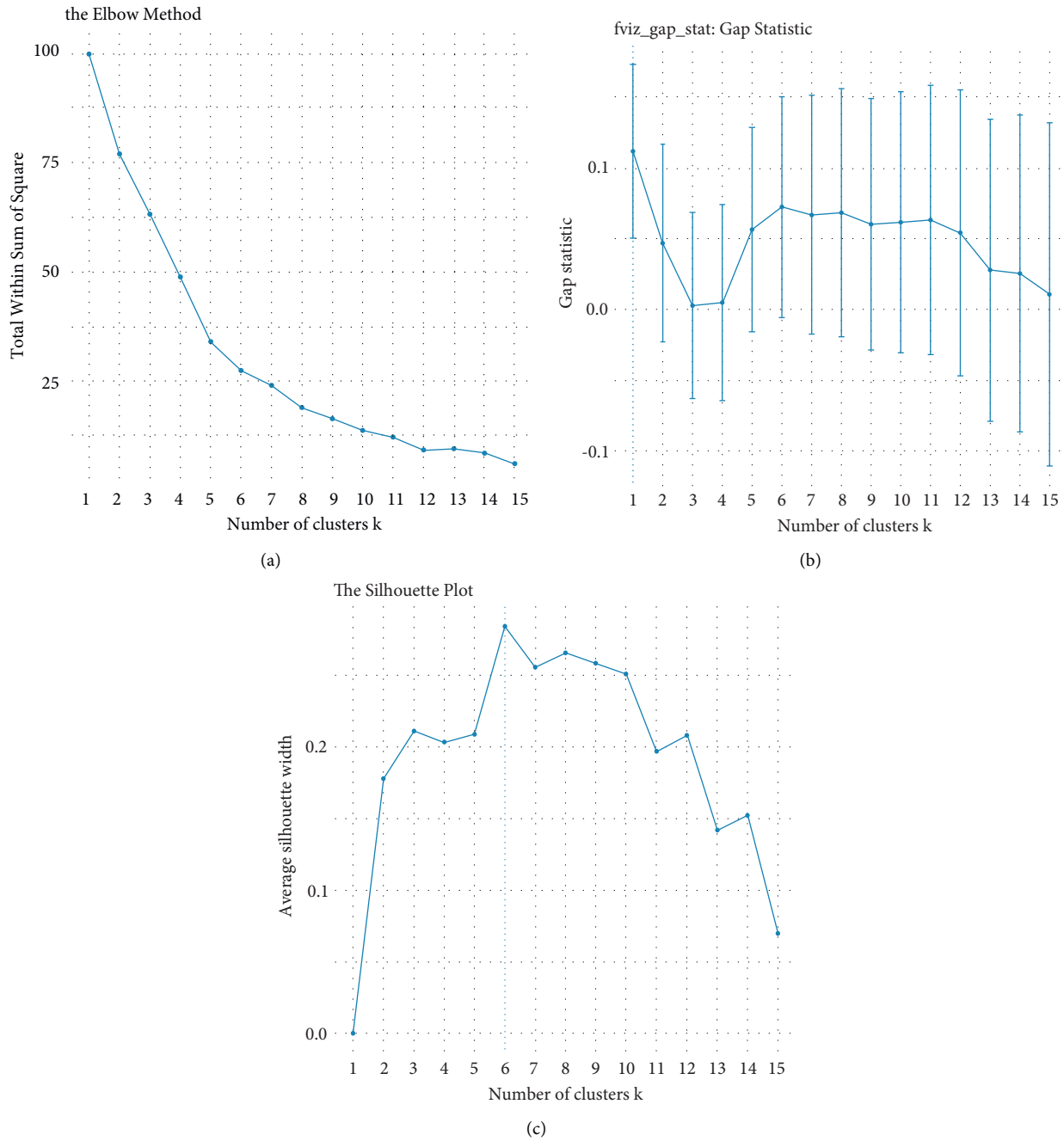
(a)

(b)

(c)

FIGURE 7: (a) "Elbow" method, (b) "Gap" statistic, and (c) "Silhouette" method.

information, coding inconsistencies, noise, and extremes, but also in addressing specific requirements that require data transformation, such as the discretization, the normalization, the reduction of their dimensions, or the selection of the most appropriate characteristics.

Initially, an indicative statistical analysis of the data set was performed. The main object of the above statistical analysis is the analysis and interpretation of the data used with the ultimate goal of drawing safe conclusions for making correct decisions. Specifically, Table 1 shows the probability for each sample to belong to a specific subset and if the sample space is made up of discrete random variables

for which a cumulative probability function can be used to determine the distribution. The statistical analysis of all 21 students is presented in Table 1.

For the clear and distinct localization of the fluctuation of their values, the graphs of the statistical frequencies of the price ranges of each feature used in the data set are presented in histograms. The height of each region is equal to the ratio of the frequency to the range of values represented by the rectangle. All five features are presented in Figure 2.

Because the data set used is multivariate numeric data, the parallel coordinates plot is listed in Figure 3(a), which is an imaging technique that facilitates the comparison of
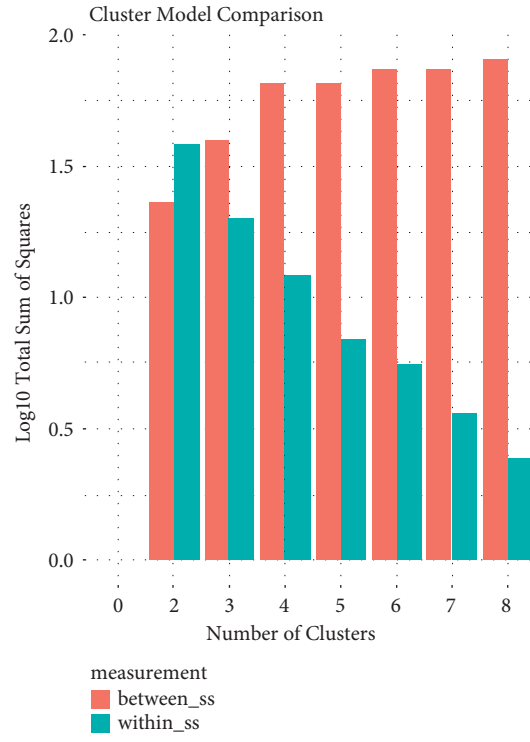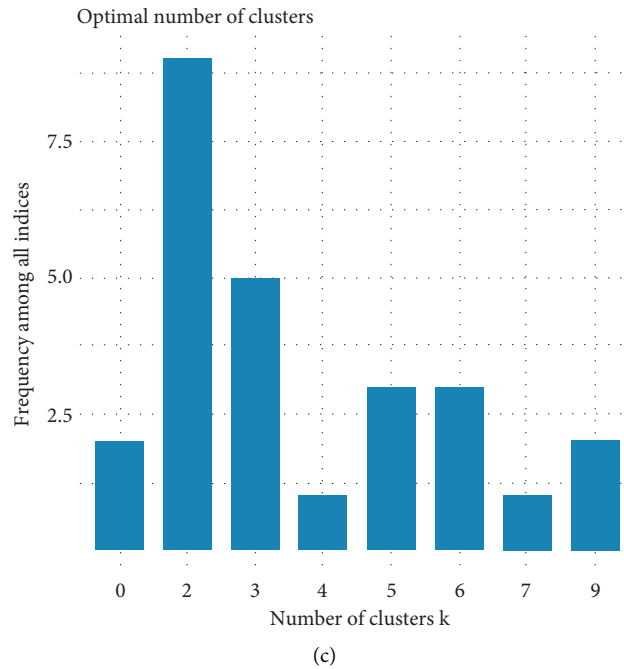
(a)
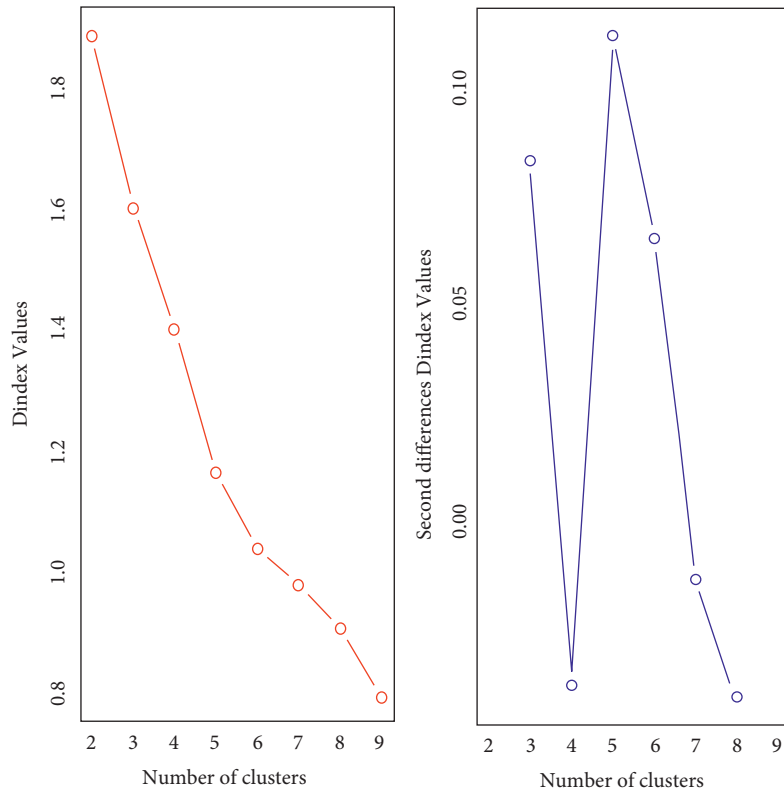
(b)

FIGURE 8: Continued.

(c)

FIGURE 8: (a) Sum of squares method, (b) Clustree method, and (c) NbClust method.

multiple quantitative variables simultaneously in order to identify patterns, similarities, complexes, and positive and negatives or neutral data relationships.

To investigate the two-way relationships between the features of the data set, a correlation analysis was performed, and Figure 3(b) shows the resulting correlation matrix.

Correlations are useful because they can indicate a predictive relationship that can be exploited in practice, although statistical dependence is not sufficient to prove the presence of a causal relationship (i.e., the correlation does not imply causality). A principle component analysis (PCA) test was then performed to detect data covariance and to apply if parameter reduction is required. As can be seen from the scree plot in Figure 4, the principal components retain less than 60% of the statistical data from the original data, so no parameter reduction is required.

From Figures 5(a)–5(c), it is understood that V1 and V2 tend to increase together in the first dimension, while in the second dimension, V1, V2, and V3 increase together. These two groups of features have a homogeneous and corresponding correlation.

The process of pretreatment of the set performed proves and ensures the quality of the data to be used by the proposed algorithm. Then, in order to identify the appropriate groups of students that will be the uniform clusters of differentiated learning, sequential analyzes were performed with various clustering methods (such as k-means, k-Medoids, and k-Centroids) [17, 22]. Specifically, successive configurations were performed with the available data, from 2 to 7 groups (clusters centers), in order to identify the best. For example, in the example of Figure 6 where he presents the configurations using the k-means algorithm, as this algorithm is sensitive to the initial positions of the
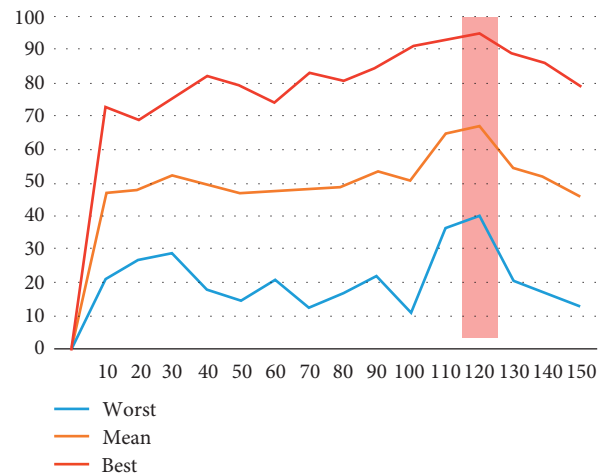

FIGURE 9: Optimization parameters.

centers of the clusters, 10 initial configurations were created, and then, all the results were calculated on average.

The above visual assessment gives clear explanations of where the demarcations between clusters occur; however, no information is given on the optimal number of clusters. To determine the optimal number of clusters, the method "Elbow" was used sequentially, in which the sum of the squares for each number of blocks is calculated and formed and the optimal number results in the abrupt change of inclination (Elbow), as in Figure 7(a).

"Gap" statistics method is also used that compares the total variance within clusters for different center values with their expected values under zero data reference distribution. The estimation of the optimal clusters is the value that

maximizes the statistical element of the gap, that is, that gives the largest statistical gap, which means that the clustering structure is far from the random uniform distribution of points. The "Gap" statistic is shown in Figure 7(b).

The "Silhouette" method was also used that calculates the average silhouette of the observations for different clusters values. The optimal number of blocks is the one that maximizes the average silhouette in a range of possible values. The "Silhouette" method is shown in Figure 6(c).

Another validation method used is to select the optimal number of clusters by minimizing the sum of squares within a cluster (how tight each cluster is) and by maximizing the sum of squares between the clusters (how sparsely the clusters are distributed). This method is shown in Figure 8(a).

Also, the Clustree statistical method produces a single score that takes into account only one set of clusters at a time considering how the samples change groups as the number of clusters increases. This is useful for showing which groups are different and which are unstable. The methodology is shown in Figure 8(b).

Another very interesting measurement comes from the NbClust method for determining the relative number of clusters, which proposes the best scheme from the different results obtained from the evaluation of 30 indicators.

The specific measurement is shown in Figure 8(c).

A Davies–Bouldin index was used to evaluate the candidate solutions [17, 22]. We consider that $R_{i,j}$ is an evaluation measure of each cluster that is calculated by the following equation:

$$R_{i,j} = \frac{s_i + s_j}{M_{i,j}}, \qquad (19)$$

where s_i, s_j are the dispersions of the i and j blocks, respectively, which are calculated from the following equation:

$$s_i = \left(\frac{1}{T_i} \sum_{j=1}^{T_i} \left| X_j - A_i \right|^q \right)^{1/q}, \qquad (20)$$

where T_i is the number of vectors in the i block, X_j is the vector of each student's attributes, and A_i is the center of the i block. Each student is classified in the cluster whose center is closer to his own. $M_{i,j}$ is the Minkowski metric for the distance of i and j blocks, which are calculated from the following equation [21, 30, 31]:

$$M_{i,j} = \left\{ \sum_{k=1}^{N} \left| a_{k,i} - a_{k,j} \right|^p \right\}^{1/p}, \qquad (21)$$

where $a_{k,i}$ is the k element of A_i and $a_{k,j}$ is the k element of A_j. The value of the objective function is finally defined as follows:

$$\underline{R} = \frac{1}{N} \sum_{i=1}^{N} R_i, \qquad (22)$$

where R_i is the maximum value of $R_{i,j}$ for $i \neq j$.

If a group of students was too small or too large, an error was introduced into the objective function, doubling its value, making all groups four to six students.

Table 2: Optimization process results.

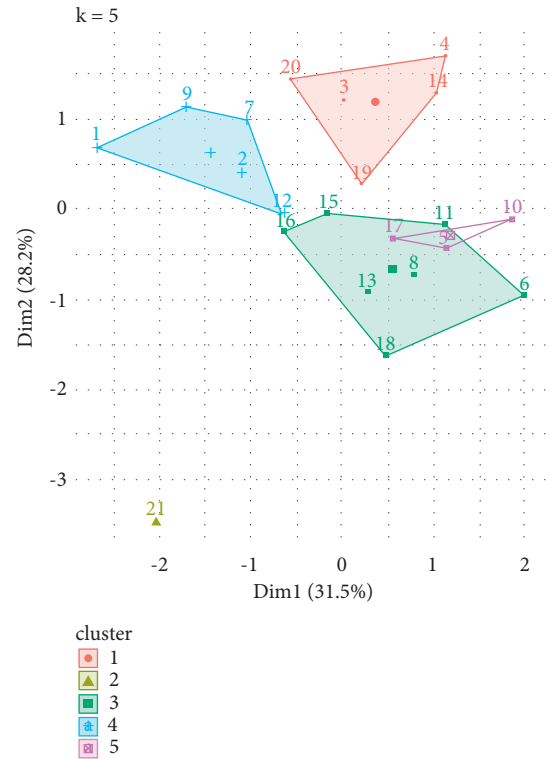| Parameters	Worst	Mean	Best
10	21	47	73
20	27	48	69
30	29	53	76
40	18	50	82
50	15	47	79
60	21	48	74
70	13	48	83
80	17	49	81
90	22	54	85
100	11	51	91
110	36	65	93
120	40	68	95
130	21	55	89
140	17	52	86
150	13	46	79

Figure 10: Clusters distribution.

The number of participants in a collaborative student group should be between four and six, as this allows good cooperation and communication among the members.

As a result, k was calculated as an integer consistently greater than or equal to the quotient of dividing the number of students by the number four. Following the completion of the group separation using the suggested algorithm, the solutions were assessed intragroup for homogeneity using the coefficient of variation (CV).

The CV is an index of relative variance or dispersion, which expresses the homogeneity of a set of measurements of values of a random quantitative variable and the accuracy of an experimental design.
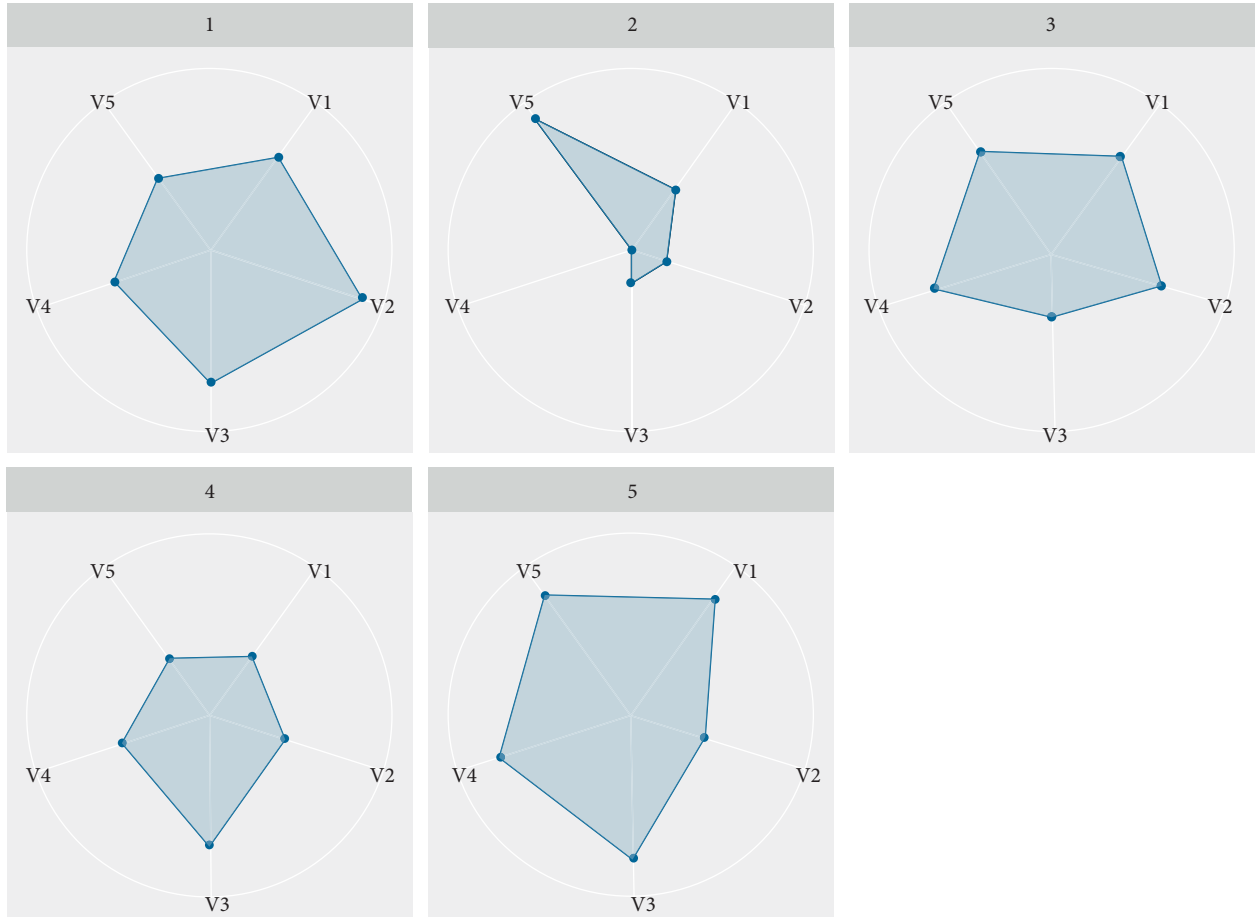
FIGURE 11: Differentiated learning attributes.

The following ratio was used to calculate the index for sample data:

$$CV = \frac{S}{\overline{Y}} 100, \tag{23}$$

where S is the standard deviation and Y is the arithmetic mean of the sample measurements.

Values close to zero indicate homogeneity in terms of characteristics, while values close to 1 indicate inhomogeneity. In general, the values of the CV index show the level of homogeneity as follows:

(1) High $(0.00 < CV \leq 0.25)$

(2) Medium $(0.25 < CV \leq 0.40)$

(3) Low $(CV > 0.40)$

The Kruskal–Wallis test [32] was used to determine if the resulting groups differed at different levels of statistical significance (0.01, 0.001).

The nonparametric Kruskal–Wallis test was used for the nonparametric analysis of variance in independent samples and was selected as in this case, the condition of normality of the populations examined is not met, the sample is small (<20 for each cluster), and the values of the dependent variables do not express quantity but are ranks where individuals are simply ordered according to some criterion.

The way to calculate the acceptance or rejection of the null hypothesis that the random samples are homogeneous is to examine whether the quantities R_i/n_i, $i = 1, \ldots, k$, $k > 3$ are approximately equal to each other and equal to $(n + 1)/2$ or if the following ratio is close to zero:

$$\sum_{i=1}^{k} \left(\frac{R_i}{n_i} - \frac{(n + 1)}{2} \right)^2. \tag{24}$$

The magnitude of the size effects was calculated using the Eta Squared η^2 and Cohen's d indicators [33]. Implementing the clustering process, initially, the parameters of the proposed algorithm were randomly initialized, based on the description performed above.

To find the optimal values for which the algorithm performs best, extensive trial and error tests were performed for different hyperparameters of the optimization algorithm. Initially, the population of optimal solutions was tested by testing values from 10 to 150 with a progressive increase of 10 units.

The algorithm performed 10 iterations for each value, whose diagrammatic representation of the results is presented in Figure 9, where the best, worst, and average values of their results are stated.

As it is understood, the optimal parameters that the algorithm shows greater convergence are for 120 particles,

TABLE 3: Mean and standard deviation of clusters.

Cluster		0 (6 studs)	1 (5 studs)	2 (4 studs)	3 (6 studs)	4 (6 studs)
V1	Mean	18.86	15.711	25.23	21.47	21.39
	StDev	0.77	1.18	0.94	1.08	1.07
V2	Mean	30.74	27.69	37.81	32.94	32.96
	StDev	1.02	0.44	0.98	1.35	1.34
V3	Mean	8.42	6.78	9.99	9.43	9.51
	StDev	0.53	0.47	0.03	0.56	0.57
V4	Mean	2.01	1.03	2.98	3.02	3.04
	StDev	0.02	0.46	0.01	0.01	0.01
V5	Mean	3.84	2.79	5.05	4.96	4.89
	StDev	0.77	0.48	0.02	0.04	0.05

TABLE 4: Coefficient of variation of each cluster.

Cluster	V1	V2	V3	V4	V5
0 (6 students)	0.04	0.03	0.06	0.01	0.02
1 (5 students)	0.06	0.02	0.05	0.03	0.01
2 (4 students)	0.04	0.03	0.01	0.01	0.01
3 (6 students)	0.05	0.04	0.05	0.01	0.02
4 (6 students)	0.06	0.05	0.05	0.01	0.02

which were selected for the further clustering process. Also, Table 2 presents in detail the above values.

The resulting clusters are shown in Figure 10.

Also, Figure 11 shows the exact distribution of the differentiated learning attributes.

The clusters created by the proposed algorithm with the average values per rating scale are presented in Table 3.

The values of the CVs, for the evaluation of the homogeneity within each cluster formed through the proposed algorithm, yielded values from 0.00 to 0.06, which are presented in Table 4.

Finally, the results of the Kruskal–Wallis test showed that the five groups differ significantly in [21]:

(1) The general mental capacity (V1), $[\chi^2(3) = 18, 67, p < 0, 001, d = 6, 9, \eta^2 = 0, 92]$

(2) Mathematical performance detection test (V2), $[\chi^2(3) = 18, 23, p < 0, 001, d = 5, 9, \eta^2 = 0, 90]$

(3) The quarter points that are an estimate of their performance (V3), $[\chi^2(3) = 17, 62, p < 0, 01, d = 4, 9, \eta2 = 0, 86]$

(4) The scale of detection of learning difficulties (V4), $[\chi^2(3) = 19, 26, p < 0, 001, d = 9, 4, \eta^2 = 0, 96]$

(5) Their psychosocial adaptation characteristics (V5), $[\chi^2(3) = 16, 92, p < 0, 01, d = 4, 3, \eta^2 = 0, 82]$

5. Discussion and Conclusions

Based on the results obtained and presented in detail above, it is obvious that the utilization of the proposed algorithm can find a reliable solution to the extremely difficult problem of creating and forming student groups for the implementation of individualized teaching programs. The methodology proved that through the widespread use of intelligent methods, small and heterogeneous groups of

students can emerge with the members of each group sharing similar features in terms of student ability [34], learning challenges [35], and psychosocial and cognitive profile [36]. In this way, in addition to being able to quickly manage the student potential in their class and knowing the individual characteristics of each group, the teacher can easily manage the student potential of their class [37]; he can offer high-quality education, through differentiated approaches that take into account the special educational needs and capabilities of each group, their particular interests, their unique experiences, their learning rhythms, their learning style, their cultural background, and their self-perception [38]. Also, as a clustering approach, the algorithm can be used in both traditional classrooms and digital or e-learning programs, facilitating the educational role [39], as it can compensate for difficulties in multicriteria grouping and differentiation of students in a variety of subjects [40]. Another significant benefit of the method is that it may be used with a large number of students and deliver results in a short period of time, provided of course there is the appropriate data for processing [41]. Another supporting presumption is that there is no limit to the data that can be accepted as quantitative data or to the evaluable factors that result from the multifaceted and holistic assessment of the student [42].

In this study, one limitation was the small number of participants, which may raise validity issues. However, the algorithmic approach used was weighted to compensate for any psychometric issues, and it is important to emphasize—which gives the method applicability in real conditions—that the application of computational methods was done in real order contexts, where there are physical limitations to a maximum number of students attending them.

From a technical point of view, the algorithm presented, which is proposed for the first time in the literature, shows a very high degree of convergence, which is evidenced by the very high clustering results that were achieved and confirmed experimentally. A very important observation also concerns the fact that the optimization method used converges very quickly, while, in all the tests, it was not observed to be trapped in local optimal, thereby avoiding incorrect cluster formations. The principle of differentiated learning is a modern educational method that aims to offer high-quality education, through differentiated approaches that take into account the special educational needs and capabilities of each student, their special interests, their unique experiences, their learning rhythms, their learning style, their cultural background, and their self-perception. Even, in this case, however, the level of students is never the same, resulting in the adaptation of teaching to the different levels of learning ability that exist within a classroom. The internal differentiation required in these cases should include a wide variety of practices and individualized forms of organizing the learning process. In that vein, this paper presented an innovative and fully efficient differentiated learning framework. It is an intelligent system that can classify students into similar homogeneous groups, based on their general mental ability, the performance of their student skills, grade points, learning difficulties they may face, and

finally the criteria of psychosocial adjustment for the assessment of skills and their school adaptation to the school environment. It is based on advanced engineering learning techniques for performing high-level analyzes for the effective reorganization of educational learning systems based on evaluation criteria. The implementation of the proposed algorithm is based on the ideal use and combination, for the first time in the machine learning literature, of the two well-known clustering methodologies (cooperation and competition) in order to produce an extremely efficient and fast neural system. The proposals for the continuation of this research focus mainly on the investigation and extension of the model with inherent capabilities of natural language processing, for the automated system to fully utilize the capabilities of the wider dependencies of modeling learning systems, with greater accuracy and efficiency. The future study of the effect of such a grouping methodology on the student's learning development in comparison to traditional methods of separation is also intriguing, as is the realization of such research using nonparametric machine learning methods.

Data Availability

The data used in the paper are available upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] H. J. Cha and M. L. Ahn, "Development of design guidelines for tools to promote differentiated instruction in classroom teaching," *Asia Pacific Education Review*, vol. 15, no. 4, pp. 511–523, 2014.

[2] X. Zhu, C. M. Cheong, G. Y. Li, and J. Wu, "Primary school teachers' conceptions of reading comprehension processes and its formulation," *Frontiers in Psychology*, vol. 11, p. 615, 2020.

[3] V. Demertzi and K. Demertzis, *A Hybrid Adaptive Educational eLearning Project Based on Ontologies Matching and Recommendation System*, http://arxiv.org/abs/2007.14771, 2021.

[4] A. Demetriou, G. Spanoudis, and A. Mouyi, "Educating the developing mind: towards an overarching paradigm," *Educational Psychology Review*, vol. 23, no. 4, pp. 601–663, 2011.

[5] M. Nedashkovskyy, G. Śmigielski, O. Sokolov, and A. Mreła, "Visual methods of processing survey data in social disciplines based on fuzzy logic," *Bulletin of the Polish Academy of Sciences*, vol. 69, no. 5, Article ID e138812, 2021, https://journals.pan.pl/dlibra/publication/138812/edition/120730.

[6] F. Essalmi, L. Jemni Ben Ayed, M. Jemni, S. Graf, and Kinshuk, "Generalized metrics for the analysis of E-learning personalization strategies," *Computers in Human Behavior*, vol. 48, pp. 310–322, 2015.

[7] A. Klašnja-Milićević and M. Ivanović, "E-learning personalization systems and sustainable education," *Sustainability*, vol. 13, no. 12, p. 6713, 2021.

[8] S. Monemian, *A Neuroevolutionary Neural Network-Based Collaborative Filtering Recommendation System,"* *Thesis*Laurentian University of Sudbury, Sawston, UK, 2020, https://zone.biblio.laurentian.ca/jspui/handle/10219/3503.

[9] I. Rodríguez-Ardura and A. Meseguer-Artola, "What leads people to keep on e-learning? An empirical analysis of users' experiences and their effects on continuance intention," *Interactive Learning Environments*, vol. 24, no. 6, pp. 1030–1053, Aug. 2016.

[10] I. Rodríguez-Ardura and A. Meseguer-Artola, "Flow experiences in personalised e-learning environments and the role of gender and academic performance," *Interactive Learning Environments*, vol. 29, no. 1, pp. 59–82, Jan. 2021.

[11] J. T. Nganji, "Towards learner-constructed e-learning environments for effective personal learning experiences," *Behaviour & Information Technology*, vol. 37, no. 7, pp. 647–657, Jul. 2018.

[12] S. Maghsudi, A. Lan, J. Xu, and M. van der Schaar, "Personalized education in the artificial intelligence era: what to expect next," *IEEE Signal Processing Magazine*, vol. 38, no. 3, pp. 37–50, May 2021.

[13] K. Demertzis, D. Tsiotas, and L. Magafas, "Modeling and forecasting the COVID-19 temporal spread in Greece: an exploratory approach based on complex network defined splines," *International Journal of Environmental Research and Public Health*, vol. 17, no. 13, p. 4693, Jan. 2020 Art. no. 13.

[14] Y. Contoyiannis, S. G. Stavrinides, M. P. Hanias, M. Kampitakis, P. Papadopoulos, and S. Potirakis, "On the effectiveness of imposing restrictive measures in a graded Self-Organized Criticality epidemic spread model the case of COVID-19," May 2020, http://arxiv.org/abs/2004.00682.

[15] K. Demertzis, D. Taketzis, D. Tsiotas, L. Magafas, L. Iliadis, and P. Kikiras, "Pandemic analytics by advanced machine learning for improved decision making of COVID-19 crisis," *Processes*, vol. 9, no. 8, p. 1267, Aug. 2021.

[16] S. Guzdek, "The importance of clusters for the region's economy," *Biul. KPZK*, vol. 2018, no. 272, pp. 54–64, 2018, https://journals.pan.pl/dlibra/publication/128619/edition/112206.

[17] S. Theodoridis and K. Koutroumbas, "Clustering algorithms III: schemes based on function optimization," in *Pattern Recognition*, S. Theodoridis and K. Koutroumbas, Eds., pp. 701–763, Academic Press, Boston, Fourth Edition, 2009.

[18] K. M. R. Alam, N. Siddique, and H. Adeli, "A dynamic ensemble learning algorithm for neural networks," *Neural Computing & Applications*, vol. 32, no. 12, pp. 8675–8690, 2020.

[19] L. Starkel, "Clustering of extreme rainfalls and evolution of fluvial systems in the Holocene," *Studia Quaternaria*, vol. 23, pp. 23–28, 2006, https://journals.pan.pl/dlibra/publication/136653/edition/119393.

[20] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, Dec. 2006.

[21] I. Kononenko and M. Kukar, "Chapter 12 - Cluster Analysis," in *Machine Learning and Data Mining*, I. Kononenko and M. Kukar, Eds., Woodhead Publishing, Sawston, United Kingdom, pp. 321–358, 2007.

[22] G. Peters, M. Lampart, and R. Weber, "Evolutionary rough k-medoid clustering," in *Transactions on Rough Sets VIII*, J. F. Peters and A. Skowron, Eds., Springer, Berlin, Heidelberg, pp. 289–306, 2008.

[23] M. Ragni and S. Neubert, "Analyzing Raven's intelligence test: cognitive model, demand, and complexity," in *Computational Approaches to Analogical Reasoning: Current Trends*, H. Prade

and G. Richard, Eds., Springer, Berlin, Heidelberg, pp. 351–370, 2014.

[24] P. A. Wilson and B. Lewandowska-Tomaszczyk, "Cognitive structure and conceptual clusters of emotion terms," *Filoz. Nauka*, vol. 2019, pp. 91–123, 2019, https://journals.pan.pl/dlibra/publication/131193/edition/114592.

[25] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, Berlin Heidelberg, 3rd ed edition, 2001.

[26] *A New Growing Neural Gas for Clustering Data Streams | Neural Networks*, vol. 78, 2021, https://dl.acm.org/doi/abs/10.1016/j.neunet.2016.02.003.

[27] W. Drozd, "Cluster analysis in research of accident rate in construction sector," *Archives of Civil Engineering*, vol. 64, no. 3, pp. 159–172, 2018, https://journals.pan.pl/dlibra/publication/132319/edition/115603.

[28] P. Golińska, "Information management supporting multi-modal transport utilization in virtual clusters," *Management and Production Engineering Review*, vol. 2013, no. 1, 2013, https://journals.pan.pl/dlibra/publication/103319/edition/89327.

[29] J. C. Bansal, H. Sharma, K. Deep, K. N. Das, and A. Nagar, "Special issue on swarm intelligence and its applications to engineering," *International Journal of System Assurance Engineering and Management*, vol. 9, no. 4, pp. 739-740, Aug. 2018.

[30] O. O. Koyejo, N. Natarajan, P. K. Ravikumar, and I. S. Dhillon, "Consistent binary classification with generalized performance metrics," *Advances in Neural Information Processing Systems*, vol. 27, 2014, https://papers.nips.cc/paper/2014/hash/30c8e1ca872524fbf7ea5c519ca397ee-Abstract.html.

[31] M. Bhaskar and C. Nick, *An Introduction to Neural Information Retrieval*, https://ieeexplore.ieee.org/document/8620670/, 2018.

[32] J. C. W. Rayner and G. Livingston, "The Kruskal-Wallis tests are Cochran-Mantel-Haenszel mean score tests," *Metron*, vol. 78, no. 3, pp. 353–360, Dec. 2020.

[33] M. A. Adams and T. L. Conway, "Eta squared," in *Encyclopedia of Quality of Life and Well-Being Research*, A. C. Michalos, Ed., Springer Netherlands, Dordrecht, pp. 1965-1966, 2014.

[34] J. S. Tahiri, S. Bennani, and M. K. Idrissi, "An assessment system adapted to differentiated learning within Massive Open Online Courses using psychometric testing," in *Proceedings of the 2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET)*, pp. 1–7, Istanbul, Turkey, 8-10, Sep. 2016.

[35] P. Wu and L. Sun, "Research on the cultivation of autonomous learning ability based on SPOC differentiated teaching mode: take "engineering contract management" as an example," in *Proceedings of the 2020 International Conference on Wireless Communications and Smart Grid (ICWCSG)*, pp. 408–411, Qingdao, China, 12-14, Jun. 2020.

[36] A. M. F. Yousef and A. R. Khatiry, "Cognitive versus behavioral learning analytics dashboards for supporting learner's awareness, reflection, and learning process," *Interactive Learning Environments*, vol. 0, no. 0, pp. 1–17, Dec. 2021.

[37] V. S. Varanasi and S. Chilukuri, "Adaptive differentiated edge caching with machine learning for V2X communication," in *Proceedings of the 2019 11th International Conference on Communication Systems Networks (COMSNETS)*, pp. 481–484, Bengaluru, India, 7-11 Jan. 2019.

[38] A. S. Zoolfakar and F. Sulaiman, "Providing differentiated learning experiences through e-Learning on Solid State Devices course," in *Proceedings of the 2009 International Conference on Engineering Education (ICEED)*, pp. 138–142, Kuala Lumpur, Malaysia, 7-8 Sep. 2009.

[39] A. Qodad, Y. Z. Seghrroucheni, M. Al Achhab, M. El Yadari, A. El Kenz, and A. Benyoussef, "An adaptive learning system based on a Job model, the differentiated instruction and Felder and Silverman's learning styles model," in *Proceedings of the 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt)*, pp. 506–510, Tangier, Morocco, 24-26 Jul. 2016.

[40] N. Aljowaysir, T. O. Ozdemir, and T. Kim, "Differentiated learning patterns with mixed reality," in *Proceedings of the 2019 IEEE Games, Entertainment, Media Conference (GEM)*, pp. 1–4, New Haven, CT, USA, 18-21, Jun. 2019.

[41] M. Andergassen, G. Ernst, V. Guerra et al., "The evolution of e-learning platforms from content to activity based learning: the case of Learn@Wu," in *Proceedings of the 2015 International Conference on Interactive Collaborative Learning (ICL)*, pp. 779–784, Firenze, Italy, 20-24, Sep. 2015.

[42] T. Keuning and M. van Geel, "Differentiated teaching with adaptive learning systems and teacher dashboards: the teacher still matters most," *IEEE Transactions on Learning Technologies*, vol. 14, no. 2, pp. 201–210, Apr. 2021.