Computer Science Journal of Moldova, vol.**34**, no.1(100), 2026

A new dynamic approach for grouping learners using a genetic algorithm in social collaborative learning environments

Noureddine Gouasmi, Mahnane Lamia, Yacine Lafifi

Abstract

Social media and social networking have spread widely in everyday life, so it is important to use them in collaborative learning. Forming appropriate learning groups is, therefore, an important objective. This paper presents a novel approach based on a Genetic Algorithm (GA) for dynamic learners grouping in a Social Network Learning system (SNL). It offers some improved attributes applied for grouping learners and new genetic operators applied in the GA. The efficiency of the proposed approach was evaluated by comparing the groups formed using the proposed GA with randomly formed groups, resulting in the conclusion that the proposed GA is more effective and that the groups formed are more efficient.

Keywords: Collaborative Learning, Genetic Algorithm, Group Formation, Social Learning Network, Modified Order Crossover, Modified One-Point Cut Crossover.

MSC 2020: 68T05, 68T20, 91D15.

1 Introduction

With the development of social networks, their use in human learning has become increasingly wide. Learners primarily use them to search for documents, communicate and collaborate with other learners, and share available information [1].

doi:10.56415/csjm.v34.02

Computer-Supported Collaborative Learning (CSCL) is an interdisciplinary field associating computer technology and educational sciences [2]. The aim is to promote collaborative learning through the incitement to cooperation and communication between learners by adopting electronic means, such as computers or smartphones connected to networks.

In the CSCL systems, group formation is an essential and challenging task, particularly the choice of learners to be grouped. Thus, the relationships between learners have an impact on the interactions among them [3], while learners' interactions lead to active learning, and to the acquisition of skills such as teamwork aptitudes and creative thinking, which are helpful competencies for employment [4].

The attributes used for grouping learners can be of very different natures, such as demographic variables (gender, age), grades, learning styles, learning activity, and personality [5]. Another crucial factor to consider is the interpersonal connections between learners. Therefore, analysing social networks in the CSCL systems enables the evaluation of these associations [6].

Group composition is another aspect to consider for their formation. There are three possible compositions: homogeneous, heterogeneous, or mixed groups [7]. Homogeneous groups are composed of learners with the same profile characteristics, whereas heterogeneous groups associate learners with different profiles. Mixed groups bring together learners whose profiles are similar in specific characteristics and different from others. Group composition has an impact on learning outcomes. At the same time, it may be augmented when heterogeneous members are in a group, but more effective when members have the same backgrounds [8].

Because of the above, this paper introduces a new genetic algorithm for forming improved groups of learners in an SNL environment, improving collaboration efficiency. Both SNL and genetic algorithms were used in higher education. The proposed approach offers new characteristics and operators used for grouping learners. These characteristics refer to the four main dimensions of learning in an SNL environment: Communication, Access, Activities, and Sociability. Concerning genetic operators, the algorithm uses two new operators: a Modified Order Crossover (MOC) as a mutation operator and a modified one-cut point crossover operator.

The paper is structured in 4 sections. After the Introduction, Section 2 contains several relevant works. Section 3 provides a comprehensive explanation of the approach, including which attributes are used to model the learners and what fitness functions allow for selecting the best solutions provided by the genetic algorithm. Section 4 describes the new genetic algorithm and its different stages for grouping learners in SNL environments. Sections 5 and 6 propose a detailed description of the experimentation of the proposed method, followed by an analysis of the results. Lastly, the Conclusion recapitulates the study's findings and proposes future work to expand upon this research.

2 Related works

Evolutionary algorithms are widely used for group formation problems, while Genetic Algorithms (GA) are the dominant solution in this set of techniques [7]. Since Professor John Holland proposed to resolve complex optimisation problems using methods and models from natural evolution in the late *60s*, and his introduction of GA in 1975 [9], GA has been extensively used to find optimal solutions for problems as different as the travelling salesman problem, combinatorial optimisation, etc. [10].

In recent years, many works have focused on forming collaborative groups using genetic algorithms for learning purposes. Most papers have focused on groups having intra-heterogeneity and inter-homogeneity [11], [12].

There are various attributes used to form groups, such as learning styles [11], personality traits [13], and in [12], Miranda *et al.* introduced the *maximization of empathy* as a way to increase the affinity between group members.

Sukstrienwong [11] proposed a particular chromosome representation called *circular chromosomes*, where each chromosome represents students with different learning styles divided into groups. The group assignment can be slid along the chromosome to change its allocation.

For GA operators, [13] uses a roulette wheel for selection, a one-point crossover operator, and a swap mutation, while [12] adopts the Mapped Crossover (PMX) operator and the Swap mutation with Shuffle Indexes mutation. And finally, [14] also uses PMX crossover with Tournament selection.

In this context, we present a genetic algorithm, offering innovative operators for group formation in the collaborative learning environment. The proposed solution for grouping learners considers all aspects of collaborative work: communication, access, and activities of the learner, as well as their sociability. This solution is integrated into a social network-based learning system called: CLISON (Collaborative Learning In SOcial Networks).

3 Problem formulation

The problem can be formulated in the following manner: let us consider a class S consisting of N learners. The goal is to assign these learners into L groups, each comprising M learners.

Let $S = \{s_1, s_2, \ldots, s_N\}$ be the class of N learners. Each learner is represented with K attributes, while $S_i = \{c_1(i), c_2(i), \ldots, c_K(i)\}$, where $c_k(i)$ is the attribute k for the learner i. Learners are assigned to a set of L groups $G = \{g_1, g_2, \ldots, g_L\}$.

To form the previous groups, the following criteria must be satisfied:

1. assigning each learner to only one group,
2. reducing the dissimilarity between the learners, to create heterogeneity between group members,
3. enhancing the similarity between the formed groups, so that they must be as homogeneous as possible.

This involves optimizing more than one objective function, which constitutes a problem of multi-objective optimization.

In heterogeneous grouping, learners from the same group must be different regarding some characteristics. This heterogeneity can improve learning outcomes, as the diversity leads to more dynamic and active groups [15].

While homogeneous groups give each member the same chance to participate in the group, heterogeneous groups allow roles distribution between members and increase creativity [15], [16].

3.1 Learner modelling

As presented above, each learner i is represented by a vector composed of the attributes used to form the groups:

$$S_i = \{c_1(i), c_2(i), \ldots, c_K(i)\}$$

assuming $c_j(i)$ is the value for the attribute j of the learner i.

In our approach, we used four main dimensions, from which we grouped attributes. These dimensions are: *communication*, *access*, *activities*, and *sociability*.

The attributes include:

- **Communication**: refers to messages or posts exchanged in an SNL environment, representing the learner's degree of involvement in the learning system.
- **Access**: refers to the number of accesses to the SNL environment, because it is important to know when learners were online and which resources they were accessing.
- **Activities**: measures the learner's performance and learning while doing learning activities on the platform.
- **Sociability**: refers to data related to describing how the learner is inclined to interact with others.

Communication is an essential element that must be analysed, while it is the primary purpose of SNL. According to Yassine [17], to assess learners' engagement, we need to understand their patterns of interactions with each other and with the content of the platform. Multiple variables can be used to measure communication, but the most important is *Total amount of messages* written and sent in the platform [18].

From an *access* perspective, two sets of variables are important to consider: (a) learners' access to the platform, which allows them to know when learners were online and what content they were accessing, and (b) information about the learner's study time and how engaged the learner is in his studies [18]. For this purpose, the variable *maximum number of accesses* is a good measure of the learner's involvement on the platform.

Considering *activities* dimension, they are a way to measure the learner's interest, performance, and outcomes. The SNL system provides information about the activities done by learners, the period in which the learner has delivered them, and whether they were delayed [19]. Thus, to measure the accomplishment of the activities provided to the learners, two indicators are used: *grade* and the *Number of complete activities*.

For the sociability aspect, social interactions are important, and learners' engagement depends on psychological factors such as sociability, extroversion, self-confidence, etc. Hence, *sociability* is a good indicator of social engagement in a group. It is defined as the inclination to be sociable and associate with others [20]. We therefore propose to measure this indicator for a learner by the number of learners with whom he interacts, based on the assumption that the more connections a learner has, the more sociable he is.

Therefore, five attributes are selected to represent the learners' profile. These attributes cover the main dimensions of learning through an SNL. The goal is to create learning groups that offer diversity in knowledge levels and skills among learners belonging to the same group.

While using the SNL, the learner's interactions with the system are captured through his connections and access to the system, the activities achieved, the messages and posts sent, and his interaction with his classmates on the system. Afterward, the captured data are used to determine the grouping attributes.

Table 1 summarizes the aspects and attributes used to form groups in our proposed approach.

3.2 Fitness function

To measure the quality of GA solutions, a fitness function is determined. A value is assigned to each chromosome to evaluate the best

Table 1. Group Formation Attributes

Dimension	Attribute	Formula
Communication skills	Total amount of messages sent by the learner i	$count(m_i)$
Access amount	platform access for learner i	$mode(access_i)$
Activities rating	grade	$average(grade_i)$
	Number of ended activities	$count(complete_i)$
Sociability level	degree of sociability	$count(connect_i)$

group formation.

Since the aim is to obtain similar groups having learners with different profiles, we calculate two fitness functions.

1. **Intra-groups fitness function** f_{intra}: evaluates the similarity between the learners' profiles in a group.

2. **Inter-groups fitness function** f_{inter}: evaluates the similarity between the profiles of the groups.

3.2.1 Intra-groups fitness function

The intra-groups fitness function for L groups of learners (f_{intra}) is calculated as in Eq. (1):

$$f_{intra} = \sum_{i=1}^{L} S_{intra}^{i}, \tag{1}$$

where S_{intra}^{i} is the similarity between profiles of learners for the group i that has M learners. It is calculated as in Eq. (2):

$$S_{intra}^{i} = \sum_{a=1}^{M-1} \sum_{b=a+1}^{M} DIST(S_a, S_b), \tag{2}$$

which is the sum of the Euclidean distances between the profiles of learners S_a and S_b for the group i.

This distance is calculated between K characteristics for two learners, as in Eq. (3):

$$DIST(S_a, S_b) = \sqrt{\sum_{i=1}^{K}(c_i(a) - c_i(b))^2}, \tag{3}$$

assuming $c_k(a)$ (respectively $c_k(b)$) is the attribute k for the learner a (respectively b).

To obtain heterogeneous groups, f_{intra} should be **maximized**.

3.2.2 Inter-groups fitness function

Inter-groups fitness function f_{inter} is calculated as in Eq. (4), for L groups:

$$f_{inter} = \sum_{a=1}^{L-1} \sum_{b=a+1}^{L} DIST(G_a, G_b), \tag{4}$$

which is the sum of Euclidean distances between profiles for L groups.

The profile of the group a is $G_a = \{p_a(1), p_a(2), \ldots, p_a(K)\}$, where each attribute of the group profile is the average value of member's attributes. Eq. (5) shows how to calculate $p_a(i)$.

$$p_a(i) = \frac{\sum_{j=1}^{M} c_j(i)}{M}, \tag{5}$$

assuming $p_a(i)$ is the value of the attribute i for the group a, and $c_j(i)$ is the value of the attribute j for the learner i.

To get similar groups, f_{inter} should be **minimized**.

4 Proposed genetic algorithm

Our system aims to form groups for collaborative learning where learners in a group are heterogeneous, while groups are homogeneous between them. To resolve this problem, a Genetic Algorithm (GA) with some characteristics is defined in the sections below.

Typically, a genetic algorithm adopts the steps described by the following stages.

4.1 Chromosome encoding stage

In our GA, a chromosome represents a possible solution to the question of assigning learners to groups. While this question is a combinatorial problem, the best solution representation consists of a set of symbols ordered in a list [21].

For a class containing N learners who must be assigned to L groups, the chromosome in our GA is an array of size N, where each element of the array has the group index value where the learner is assigned. Fig.1 shows an example of assigning 10 learners to 3 groups.

Student idx	1	2	3	4	5	6	7	8	9	10
Group idx	1	3	1	1	2	3	3	1	2	2

Figure 1. Example of assigning 10 learners to 3 groups

Indeed, a group must have (N/L) learners, if N is a multiple of L. If not, $(N \quad modulo \quad L)$ groups must have $(N/L)+1$ learners, whereas other groups must have (N/L) learners.

4.2 Population initialization stage

Two principal methods to initialize a population are randomly or using a heuristic [22].

A population initialized randomly drives mainly to optimality, and, on the other hand, the heuristic approach may, in some cases, make it difficult to find global optimal solutions because of the lack of diversification in the initial population [23], hence the choice of the random initialization method in our GA.

4.3 Crossover stage

The one-point crossover operator is easy to implement [24], chosen for our GA despite offering less diversity in solutions. To apply the operator, firstly, a crossover point is selected randomly; then, beyond the selected point, the genetic information of the two chromosomes is interchanged to obtain new offspring [25].

However, swapping genetic information may create invalid offspring; so, a modified version of the one-cut point crossover operator is proposed. In this version, after exchanging information, each offspring is analysed, and every possible value for the group index is counted. When a group index count exceeds M (number of learners per group), the last exceeding index is replaced by a missing one (*cf.* algorithm 1).

Algorithm 1.

Input: individuals s1, s2
Output: offspring o1, o2
Choose crossover point, randomly
Select the left substrings beyond the crossover point of chromosomes s1 and s2
Swap elements of left substring of s1 with s2
Count group index value in 1, .., M appearing in s1
For each element x in 1, .., M
If count(x)>L
swap x in s1 with the element in the same position in s2
End

We illustrate the previous algorithm with the following example:

Let there be two individuals ($s1$ and $s2$) used to create new offspring.

First, a crossover point is selected randomly. In the following example, the selected point is beyond the second element of $s1$ and $s2$.

$s1 = (12|2313)$ and $s2 = (23|3121)$

After crossover point selection, we swap left side substrings:

$s1 = (23|2313)$ and $s2 = (12|3121)$

For $s1$, index 1 appears 1 time, index 2, 2 times, and 3, 3 times. The last index, 3, is changed to 1.

For $s2$, index 1 appears 3 times, index 2, 2 times, and 3, 1 time. The last index 1 is changed to 3.

The generated offspring are:

$o1 = (232311)$ and $o2 = (123123)$

$o1$ and $o2$ are valid.

4.4 Mutation stage

Swap mutation is a broadly used mutation operator. However, our chromosome representation transforms the problem of assigning learners to groups into a permutation problem. So, the swap operator cannot be applied in this case because it creates invalid offspring. We therefore choose to use the *Modified Order Crossover (MOC)* mutation operator.

Instead of selecting only random positions in a parent chromosome, the elements on the left of the individual, regarding the crossover point, are kept in the same position, while all the positions to the right of the crossover point are chosen. The offspring is therefore valid [26].

Algorithm 2 shows the pseudo-code for this operator.

Algorithm 2.

Input: individuals s1, s2
Output: offspring o1, o2
Choose a crossover point, randomly
Keep elements of the left substring of s1 in their same positions in s2
Keep elements of the left substring of s2 in their same positions in s1
Fill up s1 with the right substring elements of s2 to form o1
Fill up s2 with the right substring elements of s1 to form o2

We illustrate the previous algorithm with the following example:

$s1 = (12|2313)$ and $s2 = (23|3121)$

After crossover point selection:

$s1 = (*2| * 3 * *)$ and $s2 = (2 * | * 1 * *)$

The resulted offspring are:

$o1 = (321321)$ and $o2 = (223113)$

The generated offspring ($o1$ and $o2$) are valid.

The mutation probability regulates the introduction of new genes in the population. If the probability is too low, some functional genes may not be reached, while a high probability leads to a great dissemblance between parents and descendant offspring.

4.5 Selection stage

The selection operator chooses which chromosomes will reproduce. A famous selection operator is *tournament selection.*

Tournament selection starts by extracting a competition subset by randomly selecting solutions; then, from the subset, the best solutions are chosen according to their fitness value as new parents [21]. Algorithm 3 shows a description of the technique.

The *tournament technique* offers a reasonable convergence rate in most problems, especially when the tournament size is binary [27].

Algorithm 3.

Input: the tournament size k
Output: selected individuals
Repeat
Select k individuals
Choose the individual from the k selected individuals with 2 criteria: lowest inter-fitness value and highest intra-fitness value
Until Expected number of best individuals selected

For example, with the following population, where, for each chromosome Ci, two fitness values $(intra, inter)$ are assigned:

C1	C2	C3	C4	C5	C6	C7	C8
(2, 3)	(3, 7)	(6, 2)	(7, 4)	(2, 1)	(3, 8)	(2, 5)	(9, 10)

With $k = 3$, three chromosomes are selected randomly: $C3, C5, C8$.

For each selected chromosome, the absolute value of the difference between intra-fitness and inter-fitness is calculated as $diff_{Ci} = |f_{intra} - f_{inter}|$. Then, we select the chromosome with the highest value of $diff_{Ci}$.

The differences between the three selected chromosomes are as follows:

$diff_{C3}$	$diff_{C5}$	$diff_{C8}$
4	1	1

Therefore, the selected chromosome is $C3$.

4.6 Elitism stage

Elitism consists of copying some of the best solutions into the next generation while these solutions remain unchanged, which allows maintaining the best solutions for the reproduction of a new generation [28]. The aim is to keep the best individuals in the final result. However, while propagating the same individuals to new generations, the possibility of finding new individuals better fitted to solve the problem is reduced. Therefore, it is associated with some additional operators [29].

Elites are generally small groups of individuals whose size is determined as a percentage of the population. Typically, the best 10% individuals of the population in a generation will form the elite group.

4.7 Evaluate termination stage

In our proposed GA, the number of iterations is the termination condition adopted. The GA must be tested with multiple values for this number to select an optimal one. This is important because when the number of iterations is reduced, it may not be sufficient to reach the optimal solution. But, when it is significant, the probability of achieving the optimal solution will be high, but the calculation time will be too long.

4.8 CLISON: the new social network learning system adopted for the proposed GA

The GA has been integrated into a learning system offering a social network and providing courses, as well as collaborative activities.

The major features proposed by the system are to provide lessons as posts, to allow students to comment and like posts, to offer instant messaging, collaborative learning tools, and an assessment sub-system, used by the students to answer MCQ individually and solve exercises (individually or collaboratively).

The social network-based learning system CLISON consists of four important modules (*cf.* figure 2):

- The **social network**, which is the main module of the system. It offers the features of a social media, such as publications, comments, messages, etc. It allows students to communicate with each other and with the teacher, and it allows the teacher with the capability to provide the students with lessons and activities.
- The **group formation module** is used to form groups of students whose aim is to carry out collaborative activities. In this module, there are two supported grouping methods: using the proposed GA and randomly.
- The **collaborative activities module** allows students to complete the activities, suggested by the teacher, in groups (groups formed by the previous module).
- The **learning tools**, which are tools proposed to help students solve the exercises and carry out the activities proposed on the system.

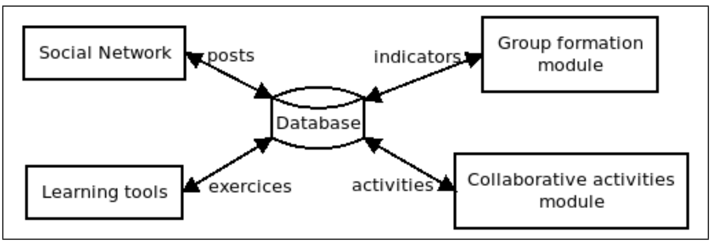

Figure 2. CLISON system architecture

5 Experimentation

The proposed system was assessed in an experimentation led at the University of Guelma (Algeria). The experiment focused on the use of the social learning network by the third-year computer science degree students for the Semi-Structured Data course, whose main purpose is to learn the XML language.

The purpose is to address the following research question: **Does the proposed grouping method improve students' learning?**

The experiment process is detailed in the following section.

5.1 Participants

Thirty-two third-year students participated in the experiment. They had access to the learning system, the lessons, and the proposed tools. However, for the collaborative exams, the class was divided into two groups of 16 students (*cf.* Section 5.2).

To test the efficiency of the students' grouping using the proposed GA, the students belonging to the experimental group ($N = 16$) were assigned to four heterogeneous groups ($L = 4$) of four members ($M = 4$), considering the 5 students' characteristics ($K = 5$) described in section 3.1. While for the control group, the 16 students were randomly divided into 4 groups of 4 members.

In the next section, to validate the proposed approach, we present the results of our experiment of creating heterogeneous groups of students using our GA. Then, we compare these results with those obtained by the random grouping of students, and, finally, we analyse and discuss the data collected.

5.2 Methodology

The experimental process lasted for one academic semester, conforming to the following steps:

1. **Familiarization phase:** during which the students received explanations about the platform, its objectives, and the tools it offers. They also answered a pre-test to determine their basic level.

2. **Learning phase**: in which lessons and exercises are regularly posted on the platform. In this phase, students have the opportunity to contact the teacher or other students for clarification

regarding the concepts being taught. In addition, they can solve the exercises using the available tools.

3. **Examination phase**: In this step, the students were associated, in groups of 4 members, to respond collaboratively to graded exercises.

4. **Assessment phase**: before the closure of the course, an individual MCQ test was submitted to students, in order to establish their final level regarding the concepts learned.

Regarding the grouping of learners held in the third phase, we choose to associate the students in heterogeneous groups of 4 members, because a group composed of a pair of students is too small and, therefore, lacks diversity. For larger groups, it is hard to ensure equality between members in activities and learning. Whereas, an odd number of group members can lead to one of the students remaining apart during collaborative activities.

Additionally, heterogeneous grouping is adopted for two reasons:

- Members with different academic and cognitive characteristics enable members with misconceptions to benefit from others more familiar with the topic.

- Members with different social skills enable students with good ones to encourage others who are less sociable, facilitating communication within the group.

Finally, regarding the proposed GA, we conducted a series of experiments on the same computer, varying each time only one of the five parameters: generations, population size, selection rate, crossover rate, and mutation rate.

While GA is a stochastic method, every run can give different results. Therefore, the experiments are performed several times, mainly around 10 to 30 times. In our case, the number of 20 runs was chosen because it is the most feasible, considering the number of genetic algorithm parameters to evaluate. Each test scenario was executed 20 times and was compared to the other scenarios according to two criteria: the average value of the highest intra-groups fitness and the computing time.

Table 2 shows the best parameters resulting from the experiments.

Table 2. Proposed genetic Algorithm's best parameters

Settings	Values
Number of Generations	70
Size of Population	20
Crossover rate	50%
Mutation rate	20%
Selection rate	30%

Comparative analysis of the proposed GA

To demonstrate the effectiveness of our new GA, we followed the experimentation methodology as presented above (*cf.* Section 5.2). Thus, in the third phase of the experimentation process, where students are associated to solve exercises, the grouping was carried out with two methods: grouping using GA and random grouping. Therefore, to validate the efficiency of the proposed GA at student grouping, we performed the process described below:

(I) The third degree class involved in the experiment was decomposed into two subclasses:

- an **experimental group** in which the grouping of students is done using the proposed GA, and
- a **control group** where the groups are randomly formed.

(II) All groups participated in performing collaborative activities of the third phase of the methodology.

(III) Finally, to validate the effectiveness of the grouping performed by our GA, we conducted a comparison study between the results obtained by the students of the experimental group and those of the control group.

The comparative analysis is presented and discussed in Section 6.

6 Results and discussion

As stated in the previous section, the experiment involved an experimental group and a control group, each composed of 16 students.

6.1 Pre-test results

In this first step, in the experiment process, the students were asked to fill out a questionnaire regarding the course. In the experimental group, one student, out of 16, didn't answer the questionnaire. Whereas, for the control group, 3 students did not respond to the questionnaire. Therefore, the participation rate for the experimental group is 94%, and 81% – for the control group.

The pre-test consisted of a multiple-choice questionnaire of 20 questions, each one introducing five possible answers, of which one or more may be correct. The maximum score for the questionnaire is 100 points. So, 28 students fulfilled the test, getting grades between 0 and 100.

To compare the obtained pre-test grades of the two groups of students, a T-test (also called Student's T-test) is used. It is a widely used statistical hypothesis test whose aim is to compare the means of two groups to determine if two groups of data are significantly different or not.

Table 3 presents a summary of the T-test for the two groups that took part in the experiment.

There was no statistically significant difference in grades, t(26) = 1.94, p = 0.063 ($p > 0.05$), despite experimental group ($M = 69.40$, $SD = 5.04$) attaining higher grades than control group ($M = 61.08$, $SD = 15.74$).

6.2 Post-test results

To complete the experiment and to close the course, the students were asked to take a post-test.

Table 3. T-Test: Pre-test results (experimental group students vs. control group students)

Parameters	Values
Null hypothesis	Both groups are equal regarding the students' grades in the pre-test
Alternative hypothesis	One group is superior than the other
Mean (experimental group)	69.40
Mean (control group)	61.08
Standard deviation (experimental group)	5.04
Standard deviation (control group)	15.74
Degree of Freedom	26
t	1.94126
p	0.063139

As for the pre-test, the post-test is a multiple-choice questionnaire composed of 20 questions. For this test, 25 students fulfilled the questionnaire, 13 from the experimental group and 12 from the control group. We noticed that the dropout rate is 19% for the experimental group, and 25% for the control group.

We compared the results obtained for the two groups involved in the experiment, using the T-test. A summary of the study is shown in Table 4.

The students in the experimental group ($M = 79.31, SD = 4.31$) compared to those in the control group ($M = 73.75, SD = 7.17$) demonstrated significantly better grades, $t(23) = 2.37032, p = 0.0265\,(p < 0.05)$.

We note, from these results, that there was a raise in grades between the pre-test and the post-test, for both groups, even if this improvement is more characterized in the experimental group than in the control group. These results suggest that the groups formed using our proposed method bring great benefits to their members.

Table 4. T-Test: Post-test results (experimental group students vs. control group students)

Parameters	Values
Null hypothesis	Both groups are equal regarding the students' grades in the pre-test
Alternative hypothesis	One group is superior than the other
Mean (experimental group)	79.31
Mean (control group)	73.75
Standard deviation (experimental group)	4.31
Standard deviation (control group)	7.17
Degree of Freedom	23
t	2.37032
p	0.0265

7 Conclusion and future work

The main goal of this paper is to form optimal collaborative groups using a Genetic Algorithm. Groups are formed from different attributes describing the learner profile according to four dimensions: *Communication*, *Access*, *Activities*, and *Sociability*. The use of these attributes makes it possible to have more efficient groups regarding collaborative learning activities in an SNL environment. To form these groups, we have defined two objectives to be respected: (a) each group must be made up of heterogeneous learner profiles, and (b) the groups must be homogeneous. The proposed GA uses two fitness functions (intra- and inter-group fitness functions) to respect grouping objectives. In contrast, a modified one-point crossover is used to create new valid individuals, and an elitism strategy is applied to preserve the best solutions.

The experimentation results show that the GA formed groups have better grades than randomly formed ones, which leads us to assume that they are more efficient in improving the learning of students.

As future work, we will try to study the improvements produced with our approach by comparing it to other group formation algorithms, and for a larger number of students.

References

[1] W. M. Al-Rahmi, N. Yahaya, U. Alturki, A. Alrobai, A. A. Aldraiweesh, A. Omar Alsayed, and Y. B. Kamin, "Social media–based collaborative learning: The effect on learning success with the moderating role of cyberstalking and cyberbullying," *Interactive Learning Environments*, vol. 30, no. 8, pp. 1434–1447, 2022, DOI: https://doi.org/10.1080/10494820.2020.1728342.

[2] C. E. Hmelo-Silver and H. Jeong, "Benefits and challenges of interdisciplinarity in cscl research: A view from the literature," *Frontiers in Psychology*, vol. 11, 2021, DOI: https://doi.org/10.3389/fpsyg.2020.579986.

[3] N. Hernández-Sellés, P.-C. Muñoz-Carril, and M. González-Sanmamed, "Computer-supported collaborative learning: An analysis of the relationship between interaction, emotional support and online collaborative tools," *Computers & Education*, vol. 138, pp. 1–12, 2019, DOI: https://doi.org/10.1016/j.compedu.2019.04.012.

[4] J. Hilliard, K. Kear, H. Donelan, and C. Heaney, "Students' experiences of anxiety in an assessed, online, collaborative project," *Computers & Education*, vol. 143, 2020, DOI: https://doi.org/10.1016/j.compedu.2019.103675.

[5] J. Moreno, J. D. Sánchez, and A. F. Pineda, "A hybrid approach for composing groups in collaborative learning contexts," *Heliyon*, vol. 7, no. 6, 2021, DOI: https://doi.org/10.1016/j.heliyon.2021.e07249.

[6] M. Saqr, U. Fors, M. Tedre, and J. Nouri, "How social network analysis can be used to monitor online collaborative learning and

guide an informed intervention," *PloS one*, vol. 13, no. 3, 2018, DOI: https://doi.org/10.1371/journal.pone.0194777.

[7] N. Maqtary, A. Mohsen, and K. Bechkoum, "Group formation techniques in computer-supported collaborative learning: A systematic literature review," *Technology, Knowledge and Learning*, vol. 24, no. 2, pp. 169–190, 2019, DOI: https://doi.org/10.1007/s10758-017-9332-1.

[8] J. Wu, "Online collaborative learning of a mixed team: A case study of hotel and tourism students from varied education backgrounds," in *Proceedings of the 2019 3rd International Conference on Education and Multimedia Technology*, 2019, pp. 12–17, DOI: https://doi.org/10.1145/3345120.3345137.

[9] J. Tao, R. Zhang, and Y. Zhu, *DNA Computing Based Genetic Algorithm.* Springer, 2020, DOI: https://doi.org/10.1007/978-981-15-5403-2.

[10] I. Gridin, *Learning Genetic Algorithms with Python.* BPB Publications, 2021.

[11] A. Sukstrienwong, "Group formation based on students' learning styles by circular genetic algorithm," *TEM Journal*, vol. 10, no. 3, pp. 1016–1021, 2021, DOI: https://doi.org/10.18421/TEM103-02.

[12] P. B. Miranda, R. F. Mello, and A. C. Nascimento, "A multi-objective optimization approach for the group formation problem," *Expert Systems with Applications*, vol. 162, 2020, DOI: https://doi.org/10.1016/j.eswa.2020.113828.

[13] O. Revelo-Sánchez, C. A. C. Ordóñez, and M. Á. R. Duque, "Group formation in collaborative learning contexts based on personality traits: An empirical study in initial programming courses," *IxD&A*, vol. 49, pp. 29–45, 2021, DOI: https://doi.org/10.55612/s-5002-049-002.

[14] G. Lescano, R. Costaguta, and A. Amandi, "Genetic algorithm for automatic group formation considering student's learning styles,"

in *2016 8th Euro American Conference on Telematics and Information Systems (EATIS)*, 2016, pp. 1–8, DOI: https://doi.org/10.1109/EATIS.2016.7520110.

[15] M. Kurni and K. Saritha, "Applying collaborative learning for enhancing the teaching-learning process in online learning through social media," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 16, pp. 251–259, 2021, DOI: https://doi.org/10.3991/ijet.v16i16.23207.

[16] I. Poort, E. Jansen, and A. Hofman, "Does the group matter? effects of trust, cultural diversity, and group formation on engagement in group work in higher education," *Higher Education Research & Development*, vol. 41, no. 2, pp. 511–526, 2022, DOI: https://doi.org/10.1080/07294360.2020.1839024.

[17] S. Yassine, S. Kadry, and M.-A. Sicilia, "Detecting communities using social network analysis in online learning environments: Systematic literature review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, no. 1, 2022, DOI: https://doi.org/10.1002/widm.1431.

[18] N. Gouasmi, M. Lamia, and Y. Lafifi, "Extracting learner's model variables for dynamic grouping system," in *Networking, Intelligent Systems and Security*. Springer, 2022, pp. 479–493, DOI: https://doi.org/10.1007/978-981-16-3637-0_34.

[19] K. D. Strang, "Beyond engagement analytics: Which online mixed-data factors predict student learning outcomes?" *Education and information technologies*, vol. 22, no. 3, pp. 917–937, 2017, DOI: https://doi.org/10.1007/s10639-016-9464-2.

[20] G. M. Harari, S. R. Müller, C. Stachl, R. Wang, W. Wang, M. Bühner, P. J. Rentfrow, A. T. Campbell, and S. D. Gosling, "Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life," *Journal of personality and social psychology*, vol. 119, no. 1, pp. 204–228, 2020, DOI: https://doi.org/10.1037/pspp0000245.

[21] O. Kramer, *Genetic algorithm essentials.* Springer, 2017, DOI: https://doi.org/10.1007/978-3-319-52156-5.

[22] J. K. Mandal, S. Mukhopadhyay, and P. Dutta, *Multi objective optimization.* Springer, 2018, DOI: https://doi.org/10.1007/978-981-13-1471-1.

[23] M. Gen, R. Cheng, and L. Lin, *Network models and optimization: Multiobjective genetic algorithm approach.* Springer Science & Business Media, 2008, DOI: https://doi.org/10.1007/978-1-84800-181-7.

[24] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, 2021, DOI: https://doi.org/10.1007/s11042-020-10139-6.

[25] K. Sastry, D. Goldberg, and G. Kendall, "Genetic algorithms," in *Search methodologies.* Springer, 2014, pp. 93–117, DOI: https://doi.org/10.1007/978-1-4614-6940-7_4.

[26] A. J. Umbarkar and P. D. Sheth, "Crossover operators in genetic algorithms: a review," *ICTACT journal on soft computing*, vol. 6, no. 1, pp. 1083–1092, 2015, DOI: https://doi.org/10.21917/ijsc.2015.0150.

[27] B. Rawat, D. Duwal, S. Phuyal, and A. Pant, "A comparative review between various selection techniques in genetic algorithm for finding optimal solutions," *International Journal of Computer Sciences and Engineering*, vol. 10, no. 10, pp. 15–22, 10 2022, Available: https://ijcseonline.org/index.php/j/article/view/6437/6428.

[28] G. Guariso and M. Sangiorgio, "Improving the performance of multiobjective genetic algorithms: An elitism-based approach," *Information*, vol. 11, no. 12, 2020, DOI: https://doi.org/10.3390/info11120587.

[29] C. Liu, A. Liu, R. Wang, H. Zhao, and Z. Lu, “Path planning algorithm for multi-locomotion robot based on multi-objective genetic algorithm with elitist strategy,” *Micromachines*, vol. 13, no. 4, 2022, DOI: https://doi.org/10.3390/mi13040616.

Noureddine Gouasmi,
Mahnane Lamia, Yacine Lafifi

Received September 17, 2025
Revised February 18, 2026
Accepted February 24, 2026

Noureddine Gouasmi
ORCID: https://orcid.org/0000-0002-7882-2835
Badji Mokhtar - Annaba University,
P.O. Box 12, Annaba, 23000, Algeria
E–mail: noureddine.gouasmi@univ-annaba.dz

Mahnane Lamia
ORCID: https://orcid.org/0000-0003-4050-7183
LRS Laboratory, Badji Mokhtar - Annaba University,
P.O. Box 12, Annaba, 23000, Algeria
E–mail: lamia.mahnane@univ-annaba.dz

Yacine Lafifi
ORCID: https://orcid.org/0000-0001-8232-4196
Labstic, 8 Mai 1945 University, Guelma
P.O. Box 401, Guelma 24000, Algeria
E–mail: lafifi.yacine@univ-guelma.dz