International Journal for Housing Science and Its Applications
Publish August 3, 2025. Volume 46, Issue 3 Pages 318-333

https://doi.org/10.70517/ijhsa46323

The use of digital twin technology in distance learning of Russian language

Yaowen Zhao[1,*] and Junping Li[2]
[1] Hohhot Vocational College, Hohhot, Inner Mongolia, 010051, China
[2] Inner Mongolia Open University, Hohhot, Inner Mongolia, 010010, China
Corresponding authors: (e-mail: wenzi7677@163.com).

Abstract The digital twin virtual simulation platform has strong visualization and interactivity, which can transform abstract theoretical knowledge into vivid and imaginative virtual scenes, helping to improve students' learning interest and enthusiasm. This paper introduces the digital twin education application framework, discusses the specific application methods of digital twin technology in Russian distance teaching, and constructs the corresponding Russian language teaching mode. For the examination management module, an improved genetic algorithm is used to realize the intelligent grouping of papers for remote Russian language teaching. The comparison results of the genetic algorithm before and after the improvement show that the fastest time of the genetic algorithm before the improvement is not less than 7500ms, and the difficulty of the generated test paper is not in the moderate [0.4~0.6) range. While the improved genetic algorithm is in the moderate range in terms of grouping difficulty, and the time can be reduced by at least 2 s. The article also proposes to introduce the improved fuzzy clustering algorithm incorporating AFSA into learning behavior analysis to portray learners. The final experiment confirms that the AFSA-FCM based model can accurately cluster learners, thus portraying learner profiles in a deeper and more comprehensive way. It helps teachers to carve out the characteristics of different learner groups and provides a key basis for decision-making in teaching and learning support services.

Index Terms digital twin, genetic algorithm, AFSA, fuzzy clustering algorithm, intelligent grouping, teaching Russian language

I. Introduction

Russian, as an international language, has a wide range of applications worldwide [1]. Therefore, Russian language teaching has always been of concern to many language learners [2]. Traditional face-to-face teaching methods show limitations in terms of location, content, and methods, which make it difficult to effectively improve the teaching of Russian in limited time. In order to improve the effectiveness and quality of Russian language teaching, the application of digital twin technology in Russian language distance teaching has received widespread attention, and has brought a lot of convenience and innovation to students and teachers [3]-[6].

Digital twin technology, i.e., digital-physical fusion technology, is the modeling and simulation of real-world physical entities or systems by digital means to achieve accurate simulation and analysis [7], [8]. Digital twin technology digitizes real-world physical entities or systems by mathematically modeling and physically simulating them and interacting them in real time with the actual physical system in order to achieve real-time monitoring, analysis, and optimization of the entities or systems [9]-[12]. In education, digital twin technology can enable students to understand and master knowledge more intuitively through simulations and virtual experiments, etc., and meet their learning needs through personalized and autonomous learning [13]-[15]. Through digital data collection and analysis, it helps educators to better assess teaching effectiveness and students' learning [16], [17]. In Russian distance learning, digital twin technology can provide students with an immersive distance learning experience [18]. Through digital twin technology, teachers can demonstrate and teach in virtual environments, while students can participate in real-time learning through virtual reality (VR) devices or other terminals, and this kind of teaching can break the geographic limitations and enable wider sharing of educational resources [19]-[22].

Literature [23] proposed a model for the application of digital twin technology and tested the model using decision tree and K-means algorithms, showing that digital twin technology has rich results and prospects in effective future planning of educational processes. Literature [24] points out the necessity of updating the educational environment in the context of digital prototyping of educational institutions in order to provide quality knowledge and competencies and proposes the use of state-of-the-art digital twin technology to build digital prototypes of educational institutions, showing that this technology is widely used. Literature [25] proposed the Digital Twin Approach (DTA) to reconstruct the learning area, stating that the application of digital twin technology provides a new path of technological integration and development for the management of smart universities, revealing that the effective use of DTA technology can motivate students to learn and improve learning outcomes. Literature [26] proposed DeepClass-Rooms, a digital twin framework for attendance and course content monitoring, which is cost-effective and requires RFID readers and high edge computing devices in Fog layer for attendance monitoring and content matching, and applies convolutional neural networks for online courses. Literature [27] identifies the current challenges in online education by highlighting the available innovative educational technologies and proposes a Unity3d based framework for optimal digital transformation of education by integrating these technologies into a learning management system to improve student engagement. Literature [28] describes Digital Twin Technology (DTT) and its applications in various industries including education and analyzes the potential of digital twins to enhance the learning experience and improve educational outcomes including immersive hands-on learning and personalization of the learning process, as well as the challenges of implementing DTT. Literature [29] examined the impact of digital twin (DT) technology on ubiquitous learning outcomes, and analyzed the impact of subsystems such as virtual reality symbiosis system and information transmission system of DT technology on ubiquitous learning outcomes, showing that all these subsystems have a significant positive impact on the ubiquitous learning outcomes, which is important for the construction of personalized DT learning spaces. Literature [30] proposes a systematic approach to the problem of creating digital twins for educational purposes, emphasizing the various ways in which digital twins created using multimedia and immersive technologies will expand opportunities and enhance learning. Literature [31] points out that current remotes have not yet fully realized the full potential of digital twins, emphasizing that digital twins can encompass all technologies used to design and deploy remote, virtual or online learning, and that the use of still-self twins is important for the education sector to achieve educational purposes. Literature [32] explores the application of existing solutions based on digital technologies to the practical needs in the field of education and personnel training at the International Center for Neutron Research (ICNR), proposes the use of digital twins of existing unique scientific facilities as a tool for contemporary education and discusses the main reasons for adopting this approach. The above study analyzes the application of digital twin technology in education, distance learning and other fields, and emphasizes the fact that the current digital twin technology is widely used with various fields and has a positive impact, which not only changes the traditional way of teaching and learning in education, but also personalizes the teaching and learning, and improves the interest of students and the effectiveness of teaching and learning.

The aim of the article is to apply the digital twin technology and carry out applied research on teaching Russian at a distance. It proposes a Russian language distance teaching model with the framework of teaching preparation, teaching implementation and teaching summary, including the functions of exam management and learning behavior analysis. Genetic algorithm is adopted as the basic algorithm of intelligent grouping in the examination management module, and according to the intelligent grouping constraints, the coding scheme, fitness function and cross-variable selection operator of genetic algorithm are improved to realize intelligent grouping of remote Russian language teaching, so as to improve the ability of online examination and teaching materials management. In the establishment of learning behavior analysis model, this paper uses AFSA to optimize FCA to get AFSA-FCM algorithm, which realizes the multifaceted analysis of learners' learning behavior and classifies them.

II. Research on Russian language teaching model based on digital twin technology

II. A.Framework for Digital Twin Educational Applications

Digital twin comprehensively uses communication technology, big data analysis, machine learning algorithms and other technologies to collect relevant and effective data of learners and physical entities through sensors, mobile learning platforms, etc., and then constructs digital twins of learners through simulation and simulation, and simulates and models the physical training equipment to form a learning space combining the real and the virtual [33], [34]. Finally, the dynamic writing (simulation, mirroring, anthropomorphism), integration, decision-making (coordination), supervision and assisted decision-making of Russian language education and teaching are accomplished through multidimensional interaction. The specific framework is shown in Figure 1.

II. B.Russian language teaching model driven by digital twin technology

The model of teaching Russian language driven by digital twin technology is mainly divided into teaching preparation, teaching implementation and teaching summarization. First of all, teachers design Russian language teaching scenarios, subdivided tasks and main practical activities according to the learning tasks and objectives, taking into account students' characteristics. At the same time, the corresponding digital twin model is generated based on the collected student data, and the tasks are pushed to the students through smart learning companions and so on. After the creation of the teaching project, centered on the construction of knowledge, students carry out learning, embodied experience, collaborative inquiry, and program verification. In this process, the digital twin monitors students' learning status in real time, analyzes students' learning behaviors through the AFSA-FCM algorithm, iteratively optimizes students' learning solutions and guides them in project implementation. Finally, after the completion of the project, students present the project, teachers and enterprises evaluate the students' programs, and the digital twin improves the teaching process and gives adjustment opinions according to the evaluation results. The specific model design is shown in Figure 2.

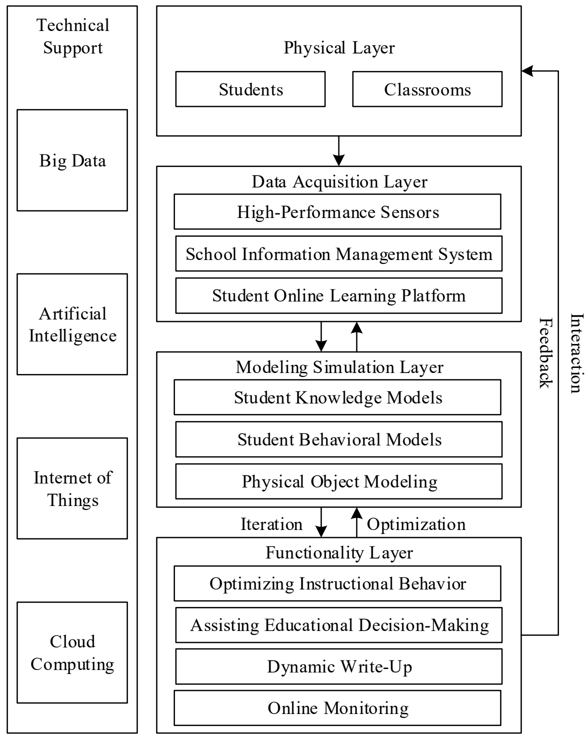

Figure 1: Digital twin education application framework

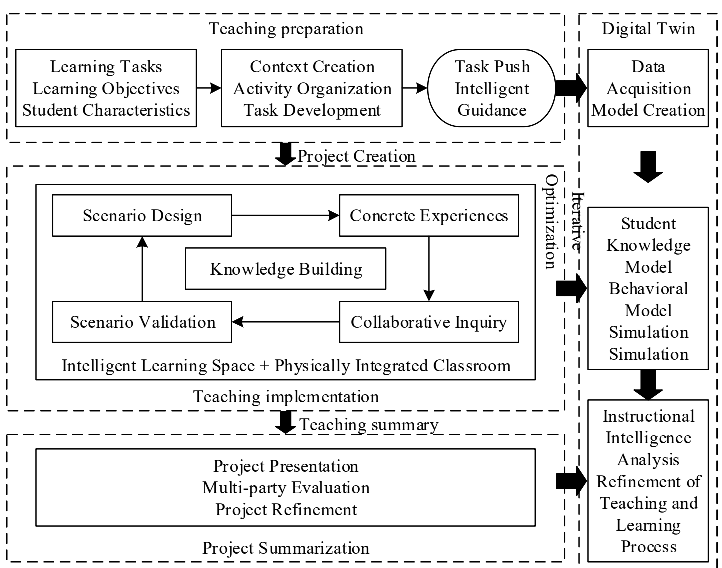

Figure 2: Is based on the Russian teaching pattern framework of digital twin technology

II. C.Exam Management Module Design

As the technology involved in distance education systems continues to evolve, so does the exam technology carried by distance education systems. The examination module is the core part of the system, through which teachers can freely change, add, delete, print and other operations on Russian examination questions and contents. Scroll formation is a basic sub-function of the online examination function, in which the core of realizing scroll formation is to realize automatic intelligent scroll formation algorithm. In the examination module, the manual grouping of papers by teachers will consume a lot of manpower and material resources, the work efficiency is not high, and the results of grouping papers may be unreasonable. To address this problem, it is proposed to use the genetic algorithm, which has the advantages of scalability, parallelism and robustness, has strong global search capability, and can effectively improve the search speed of various algorithms, to be applied to intelligent paper formation, and to screen and generate test papers through the characteristics of the algorithm, so as to improve the efficiency and quality of intelligent paper formation.

The specific process of grouping is: first, the grouped papers are divided into a number of question blocks according to the type of questions, and then the related operations of grouping are carried out separately according to the type of question blocks, and finally, when the grouping of all question blocks is completed, the results are taken to be compared with the difficulty level of the grouping module and the previously set conditions, if the grouping results of the module meet the set conditions, then it means that the grouping of this module is completed; otherwise, the system has to continuously adjust the question paper question types, so as to enhance the efficiency and quality of intelligent grouping. Otherwise, the system has to keep adjusting the difficulty level of the question blocks until the conditions are met.

II. C. 1) Group Volume Constraints

Generating test papers based on genetic algorithms allows intelligent grouping of test papers based on constraints such as overall number of test papers, number of question types, distribution of knowledge points and difficulty of test questions.

(a) Knowledge point constraints. That is, the scope of the examination content constraints, the Russian language examination paper needs to be knowledge points need to cover the content of the examination course. The formula is:

$$S_k = \sum_{i=1}^{n} s_i T(k) \tag{1}$$

Where, s_i denotes the i th question; s_k denotes whether it is taken.

$$T(k) = \begin{cases} 1 & \text{Value equal to 1 indicates that the section is selected} \\ 0 & \text{Value equal to 0 indicates that the section is not selected} \end{cases} \tag{2}$$

(b) Question type constraints: question type constraints that is, the need to ensure the diversity of the question paper to avoid a single question type. The specific constraint expression is:

$$S_j = \sum_{i=1}^{n} s_i T(j) \tag{3}$$

Where, s_i denotes the i th question.

$$T(j) = \begin{cases} 1 & \text{Value equal to 1 indicates selection to question type } j \\ 0 & \text{Value equal to 0 indicates no selection to question type } j \end{cases} \tag{4}$$

(c) Total score constraint. Can be expressed as:

$$Z = \sum_{i=1}^{n} f_i \tag{5}$$

f_i denotes the score for each sub-question. Z and n denote the total score of the paper and the number of questions in the paper, respectively, and the scores of each sub-question are added up to the total score of the paper.

(d) Difficulty constraints. High difficulty and low difficulty set a uniform ratio, can ensure the quality of test questions, improve the efficiency of the group paper. The difficulty constraint expression is:

$$N = \frac{\sum_{i=1}^{n} N_i \cdot f_i}{Z} \tag{6}$$

n denotes the total number of questions in the paper; N_i and f_i denote the difficulty and score of the i th question, respectively; and Z denotes the overall score of the paper.

(e) Exposure constraint. The exposure degree is the number of times an exam question appears in the paper. The specific expression is:

$$b_t = \frac{c}{s} \tag{7}$$

In Eq. (7), c and s denote the number of times a test question is selected and the interval between test question appearances, respectively.

The above constraints can be influenced according to the shape, content and name of the exam. Test paper generation requires comprehensive consideration of grouping constraints. Therefore, when the genetic algorithm is used to group papers, it is necessary to ensure that the constraints are established. If the comprehensive constraints are expressed as a 5-dimensional vector $(a_1, a_2, a_3, a_4, a_5)$, a_i is the i th indicator in the feature vector; when the number of test questions in the test paper is m, the test paper is represented by an $m*5$-dimensional feature matrix as:

$$s = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{15} \\ a_{21} & a_{22} & \cdots & a_{25} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{m5} \end{pmatrix} \tag{8}$$

In Eq. (8), the horizontal parameter represents the 5 constraints and the vertical parameter represents the number of test questions.

II. C. 2) Improved genetic algorithms

Genetic algorithm is simple and operational, but it may have slow convergence and easy to fall into local optimal problems in multiple constraints of intelligent paper grouping [35], [36]. To solve this problem, it is proposed to optimize the encoding method, fitness function, selection algorithm, crossover operator and variation operator of the genetic algorithm into in order to enhance the overall evaluation performance of the intelligent grouping of papers in order to screen the papers with the highest rationality. The specific improvement methods are.

(1) Coding scheme design

Real number coding method based on question type grouping method is used to realize coding method improvement. Based on the type of test questions, the question types are categorized into single-choice, multiple-choice, judgment, fill-in-the-blank, and short-answer questions. The corresponding coding ranges are respectively: $d_1 \ldots d_n, D_1 \ldots D_n, P_1 \ldots P_n, T_1 \ldots T_n, J_1 \ldots J_n$. By this way, each different question type can be divided into an independent group, which can be decoded quickly, thus improving the efficiency of the genetic algorithm operation during the test question search.

(2) Adaptation function design

Set the coefficients before the constraints of the test paper expressed as $\omega_1, \omega_2, \cdots, \omega_n$, then the sum of the coefficients before the constraints is controlled in the range of 0-1, and summed up to 1. Based on the above five constraints and mathematical models, choose the difficulty limit of the test questions, the limit of the test paper score, the limit of test question exposure, the limit of the distribution of the knowledge chapter, and the limit of the question type of the test paper. Distribution limit and the question type limit of the test paper to determine the objective function. The specific expression is:

$$f(x) = \omega_1 Z + \omega_2 S_k + \omega_3 b_t + \omega_4 N + \omega_5 S_j \tag{9}$$

The ω is the size of the weight occupied by each of the five constraints. When a certain question type is selected as S_j, it means that this constraint is selected and the sum of the coefficients in front of its constraints has taken the value of 1. Therefore, it is necessary to minimize the weight value of this constraint to ensure that other constraints meet the setting requirements.

(3) Selection operator design

The roulette selection method is chosen for selecting the operation operator, which is also an adaptation scale method. The specific expression of the selection operator is:

$$W_i = \frac{f(n_i)}{\sum_{i=1}^{n} f(n_i)} \tag{10}$$

The numerator and denominator denote the fitness value of the test question n_i and the total fitness value of the test question population, respectively; the fitness value of the n th test question corresponds to the proportion value of the whole test question as W_i . The larger the proportion value indicates that the better the feedback of the test question to the various constraints of the test paper, the higher the probability that this test question will be inherited into the next generation of the excellent population.

(4) Crossover operator design

A single-point crossover is selected for crossover. Its crossover replacement operation means replacing the real values of genes in each chromosome segment, thus realizing the corresponding test substitution, which generates the next generation of test populations and completes chromosome reorganization.

(5) Mutation operator design

According to the design of group coding and group crossover, a single-point mutation operation of the question type is used. This method can use variation operation independently in each of the test questions. The basic process of the variation operator is: First, the algorithm satisfies the set probability of variation Pm, then the variation operation can be carried out. Then the mutation point is determined. That is, based on the mutation probability to select the individual that needs to be mutated. That is, comparing the fitness between two chromosomes after the crossover operation, the chromosome with a low overall average fitness value is selected, and the mutation operation is carried out from a chromosome with a lower fitness value. The test subject with low fitness in the chromosome is thus selected as the point of mutation for mutation. Afterwards, the direction of mutation is determined, i.e., after adding or subtracting the number of the test question with the highest fitness of the current chromosome to obtain a certain value, a fitness range is generated, and the direction of mutation is randomly selected within this range. Finally, from the list of test questions of the same question type, a test question with the load variation operator requirement is selected as the replacement test question, and this test question replaces the selected question type variation point test question.

Based on the above improved methods and constraints, the intelligent paper grouping process based on improved genetic algorithm can be obtained as shown in Fig. 3.

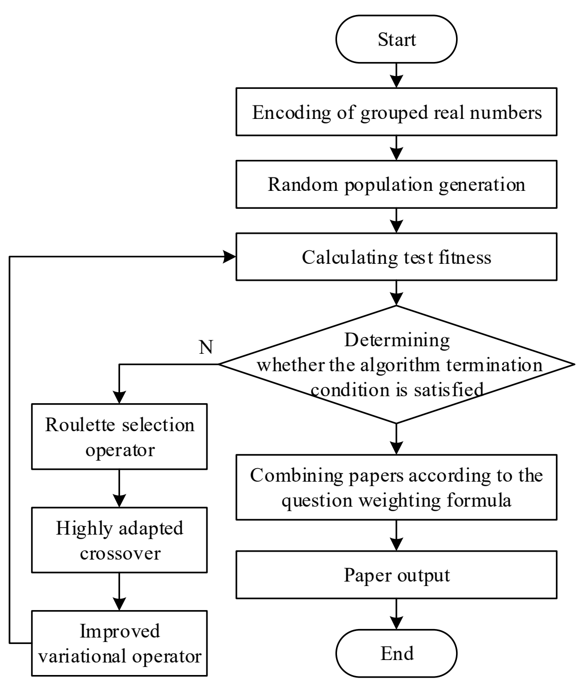

Figure 3: Is based on an intelligent group volume process based on improved genetic algorithms

At the beginning of the improved algorithm, a grouped real number coding method is required to encode the parameters. Then the population is randomly generated and the fitness function of the test questions is calculated and the fitness is compared separately. It is judged whether it reaches the termination condition of the algorithm or not. If yes, then the combination of test papers can be performed and the test papers can be output. On the contrary, the improved variation algorithm, high adaptability crossover algorithm and roulette wheel selection algorithm are used to calculate the fitness, and thus determine whether the standard is reached. And based on the question weights, intelligent paper combination and generation is carried out.

III. Intelligent grouping algorithm validation

III. A. Question paper question percentage

In order to verify whether the proposed optimized genetic algorithm is effective, the experiment will use 100 Russian language test paper question types as experimental data. The question types mainly include five types of single choice, multiple choice, judgment, fill in the blank and short answer questions. The number of questions in each type is 30, 10, 30, 20 and 10, respectively.

III. B. Algorithm runtime comparison

In order to verify the optimization performance of the proposed improved genetic algorithm, the experiment will be before and after the improvement of the genetic algorithm for intelligent grouping volume comparison and analysis, for 20 times of grouping volume to draw the corresponding time comparison shown in Figure 4.

From the graph analysis, it can be seen that the pre-improved genetic algorithm grouping time is in the range of 7500~9500ms. The optimized genetic algorithm's winding time is lower than 5500ms, which is significantly lower than that of the pre-improved genetic algorithm. This shows that the optimized genetic algorithm can improve the iteration speed, shorten the grouping time and improve the grouping efficiency in the iterative optimization process.

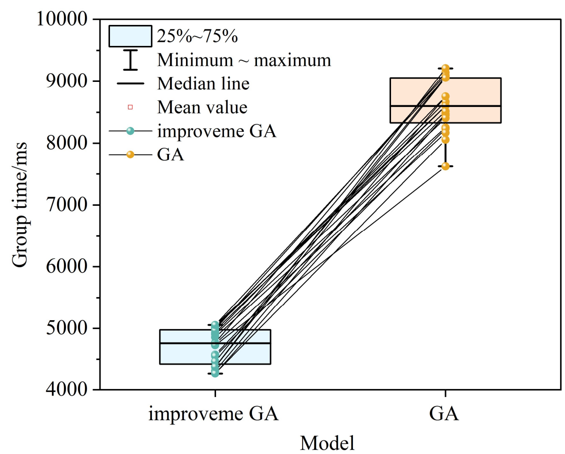

Figure 4: The time comparison diagram of the genetic algorithm

III. C. Test paper difficulty test

The experiment will be conducted by selecting 1 set of test papers each from the test papers generated by the genetic algorithm before and after optimization for difficulty comparison. Among them, test paper 1 is generated by the original genetic algorithm and test paper 2 is generated by the post-optimization genetic algorithm. In order to better distinguish the difficulty level of the test papers, the difficulty level of the test papers is categorized as hard [0~0.2), harder [0.2~0.4), moderate [0.4~0.6), easier [0.6~0.8), and easy [0.8~1.0]. The difficulty of multiple choice questions obtained for paper 1 and paper 2 is shown in Table 1. The results of the comparison of the difficulty of each question type in the two question papers are shown in Table 2. From the two tables, it can be seen that the 30 questions of paper 1 contained 4 difficult questions. Paper 2 contained only one difficult question out of 30 questions. The difficulty values of multiple-choice questions in Paper 1 and Paper 2 are 0.385 and 0.572, respectively, which shows that the multiple-choice questions generated by the original genetic algorithm are not reasonably set and the difficulty is high. Among the multiple choice, judgment, fill-in-the-blank, and short answer question types, the difficulty of the questions in Paper 2 are all in the moderate [0.4~0.6) range, while only the judgment questions in Paper 1 are in the moderate difficulty. It continues to prove the excellent performance of the optimized genetic algorithm in this paper in grouping papers.

Table 1: Choice difficulty comparison table

Selection problem			
Numbering	Test 1 difficulty	Numbering	Test 2 difficulty
1	0.408	1	0.699
2	0.358	2	0.91
3	0.724	3	0.547
4	0.449	4	0.532
5	0.367	5	0.791
6	0.496	6	0.877
7	0.526	7	0.348
8	0.232	8	0.672
9	0.565	9	0.601
10	0.441	10	0.839
11	0.249	11	0.606
12	0.479	12	0.32
13	0.111	13	0.589
14	0.256	14	0.55
15	0.114	15	0.374
16	0.556	16	0.234
17	0.376	17	0.777
18	0.521	18	0.514
19	0.452	19	0.518
20	0.583	20	0.433
21	0.451	21	0.178
22	0.229	22	0.622
23	0.509	23	0.54
24	0.091	24	0.773
25	0.208	25	0.448
26	0.218	26	0.536
27	0.548	27	0.441
28	0.393	28	0.725
29	0.491	29	0.357
30	0.151	30	0.795

Table 2: Difficulty comparison table of test papers

	Test 1	Test 2
Selection problem	0.385	0.572
Multiple-choice	0.323	0.534
Judgement	0.491	0.511
Fill in	0.355	0.569
Short answer	0.388	0.531
Whole	0.389	0.543

III. D. Percentage of knowledge points in the question paper

Based on Test Paper 2, the calculation of the coverage rate of knowledge points of Russian language teaching in the test paper is carried out, and the calculation results are obtained as shown in Figure 5. The 24 knowledge points involved in the Russian language examination, test paper 2 can basically cover most of the knowledge points, and the overall knowledge point coverage rate is maintained at about 95.83%, and its knowledge point distribution is more reasonable.

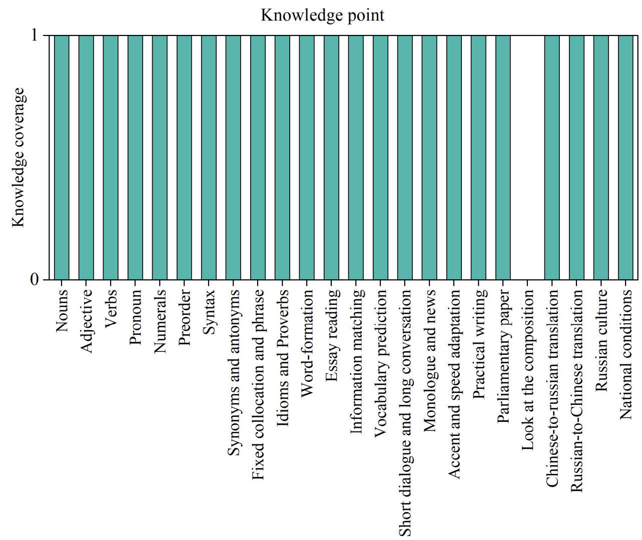

Figure 5: knowledge section coverage

IV. Learning behavior analysis method based on AFSA-FCM

Learning behavior of students in Russian distance learning refers to the sum of learning attitudes, learning styles, learning styles, and learning processes expressed by learners in the process of independent learning. Specific content includes the planning of learning, resource use, time allocation, communication, self-regulation, task completion and other aspects. Learning behavior has a close relationship with learning effects, and good learning behavior is an important guarantee for completing learning tasks and improving the quality of Russian language learning.

In the distance learning environment, the time and space separation of teaching and learning makes it difficult for teachers to directly observe the behavior of learners, effectively monitor and evaluate the learning process of learners, and give targeted learning feedback and guidance. Therefore, it is of great significance to establish a learning behavior analysis model in the distance education environment and conduct scientific and intelligent learning behavior analysis.

The basic idea of FCA is to assign data points to different clusters and allow the data points to belong to multiple clusters with different degrees of affiliation. Compared with traditional hard clustering algorithms, FCM is more flexible and can better reflect the actual situation of the data [37].The first step in the operation of FCM is to input the initial data, and then initialize the parameters and get the affiliation matrix U , the expression of affiliation matrix U is shown in equation (11).

$$U = [uk_i] \tag{11}$$

In Eq. (11), $U = [uk_i]$ denotes the affiliation degree of the k th data point belonging to the i th clustering center. After that, the spatial correlation $SC_{ij}(k)$ is calculated and its computational expression is shown in Eq. (12).

$$SC_{ij}(k) = \frac{1}{1 + \frac{d(x_k, c_i)^2}{a}} \tag{12}$$

In Eq. (12), $d(x_k, c_i)$ denotes the distance between the data point x_k and the clustering center c_i; a denotes the fuzzy degree constant. The next step is to calculate the clustering center c_i , and the formula for c_i is shown in equation (13).

$$c_i = \frac{\sum_{k=1}^{n} u_{k_i}^m x_k}{\sum_{k=1}^{n} u_{k_i}^m} \tag{13}$$

In Eq. (13), m is a fuzzification parameter, which is used to control the degree of fuzziness. Then the subordination degree $U = [uk_i]$ is calculated and its formula is shown in Eq. (14).

$$uk_i = \frac{1}{\sum_{i=1}^{c}\left(\frac{d(x_k, c_i)}{d(x_k, c_i)}\right)^{\frac{2}{m-1}}} \tag{14}$$

After obtaining the degree of affiliation, the degree of affiliation is judged, and if it meets the requirements then the clustering results are obtained through equation (15).

$$C_i = \arg_{k\in[1,c]}\{\max(u_{c\times N})\} \tag{15}$$

In Eq. (15), c is the number of clusters; N denotes the total number of data volume. If it does not meet the requirements, it returns to recalculate the spatial correlation and proceeds to the subsequent steps.FCM suffers from the deficiencies of sensitive initialization, easy to fall into local optimum, and high computational complexity. In order to improve the above deficiencies of FCM, the study introduces the AFSA which simulates the foraging behavior of fish schools, and improves the overall performance of FCM through the good global search capability of AFSA. Artificial fish schools have foraging behavior, schooling behavior, tail chasing behavior and random behavior, and AFSA mimics the above behaviors of fish schools by constructing artificial fish to achieve the optimization search [38].

The foraging behavior of AFSA is that the artificial fish randomly selects a location within its field of view and moves one step towards that location if the value of the objective function at that location is better than the current location. The behavior can be represented by equation (16).

$$X_o(t+1) = X_o(t) + Step \times Rand(\ ) \times \frac{X_p(t) - X_o(t)}{|X_p(t) - X_o(t)|} \tag{16}$$

In Eq. (16), $X_o(t)$ denotes the position of the artificial fish o at moment t, $X_p(t)$ denotes a random position within its field of view, and $Step$ denotes the step size. The clustering behavior is defined as the movement of the artificial fish towards the central position of other artificial fish within its field of view, taking into account the value of the objective function of the central position. The behavior can be represented by equation (17).

$$X_o(t+1) = X_o(t) + Step \times Rand(\ ) \times \frac{\bar{X}(t) - X_o(t)}{|\bar{X}(t) - X_o(t)|} \tag{17}$$

In Eq. (17), $\bar{X}(t)$ denotes the center position of other artificial fish within the field of view. The tail chasing behavior means that the artificial fish tracks the artificial fish with the optimal objective function value within its field of view. This behavior can be represented by equation (18).

$$X_o(t+1) = X_o(t) + Step \times Rand(\ ) \times \frac{X_{best}(t) - X_o(t)}{|X_{best}(t) - X_o(t)|} \tag{18}$$

In Eq. (18), $X_{best}(t)$ denotes the position of the artificial fish with the optimal objective function value within the field of view. Through the above behaviors, the artificial fish is able to find the optimal solution in the search space while avoiding falling into the local optimum.The specific operation process of AFSA is divided into six steps, firstly, the parameters of the algorithm are initialized and set up, and secondly, the adaptive value of the initial fish population is calculated and the optimal artificial fish state is publicized. Then individual fish are evaluated according to the current position and adaptation value, and behaviors are executed according to the results. After that the corresponding behavior is executed based on the behavior selection and the position is updated. Then it is to evaluate the artificial fish after iteration and compare it with the optimal individual. If a better individual is found, the publicized individual is replaced. Finally, the algorithm ends when the optimal solution of the public individual reaches a satisfactory value. To study the optimization of FCM using AFSA, the idea is to use the global search ability of AFSA to obtain the initial clustering center in the global range, and then this initial clustering center is operated in FCM to find the global optimal solution, and the specific process is shown in Fig. 6.

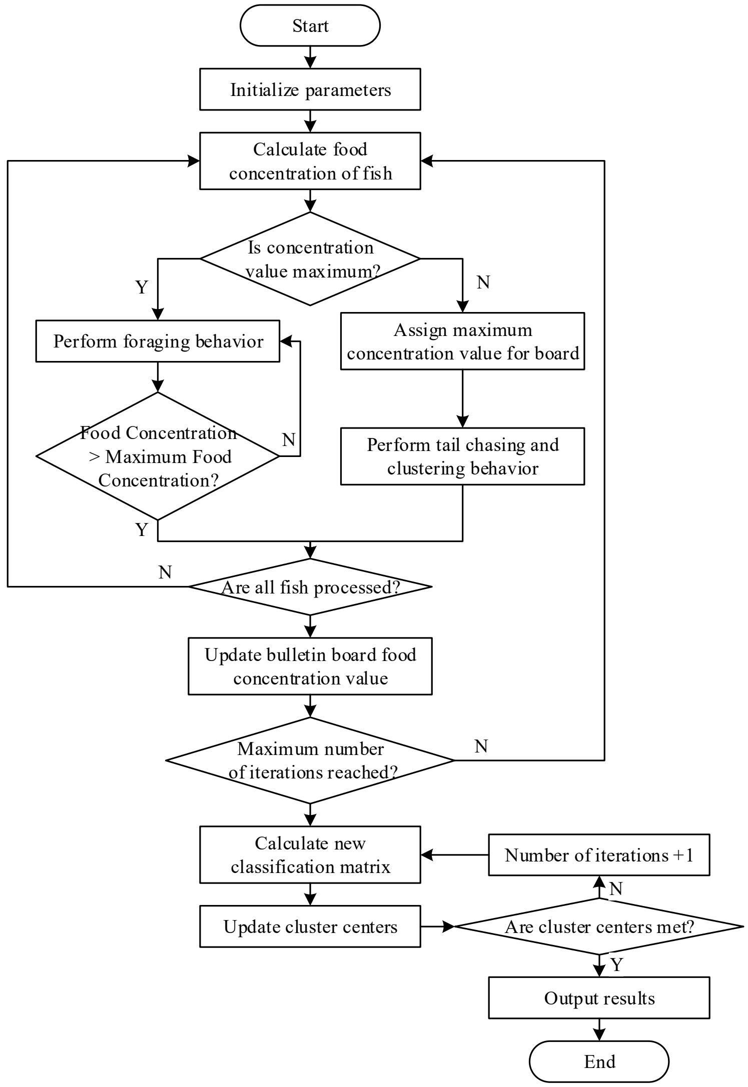

Figure 6: AFSA-FCM algorithm running process

As can be seen in Fig. 6, the AFSA-FCM algorithm firstly initializes the parameters, and then calculates the fitness value of each artificial fish by simulating its behaviors such as foraging, clustering and tail chasing. After that, according to the fitness value and behavioral rules, the optimal behavior is selected and executed, and then the optimal result is retained and the bulletin board data is updated. The process is repeated until the termination condition is satisfied. At the end of the search, the optimal data on the bulletin board is recorded as the initial clustering center, and the clustering center is substituted into the FCM algorithm for the classification operation.

V. Clustering results and analysis

The sample was selected from actual teaching and surveys of Russian language courses in the College of Foreign Languages at the University of W in the fall of 2023. And the related data were screened more necessary, mainly selecting representative data. Initialization process was carried out. In this paper, the following eight data features are all constructed for the samples: the number of video on-demand (PlayCount), the total length of video viewing (PlayTime), the number of knowledge points (KCount), the number of participation in discussions (DiscussCount), the amount of discussion speeches (DiscussAmount) and three fine-grained features (the value of learning attitude (SAttitude), Knowledge Entropy (KEntropy) and Knowledge Pass Rate (KPassPercent)).

In this paper, we use the AFSA-FCM algorithm to categorize 11,465 Russian language learners into five categories. To further test the clustering effect, this paper uses the PCA dimensionality reduction algorithm to visualize the clustering effect. The results are shown in Figure 7, where the learners are clearly clustered into 5 categories. According to the clustering results, Cluster1, Cluster2, Cluster3, Cluster4, and Cluster5 contain 4862, 2514, 1839, 1581, and 669 students, respectively.

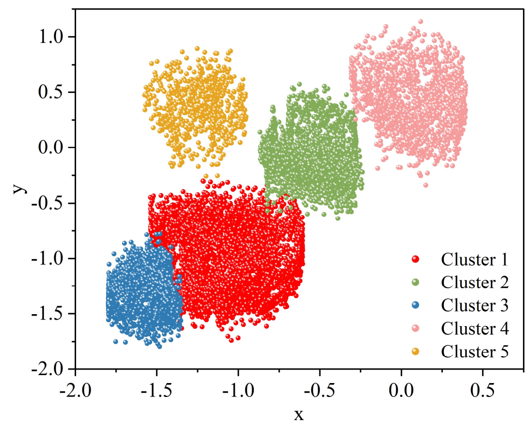

Figure 7: The visualization of clustering results

In order to better show the difference between different categories of learners, the statistical characteristics of learners in each category are calculated as shown in Table 3, such as the minimum value, maximum value, and mean.

The following analysis of the behavior of these 5 categories of learners, in terms of learning attitude, Cluster4 has the best learning attitude, with a learning attitude mean value of 17.52, which is far more than the other 4 categories. In contrast, the attitudes of the other categories of learners are more sluggish, with the values of learning attitudes lower than 2.5. From the perspective of video viewing, Cluster4 is still in a prominent position, with the average value of the number of times of clicking close to 300 times, and the average value of the playback time is more than 67,000s, which is much higher than that of the other categories of learners, indicating that Cluster4 watches the learning videos more frequently, and also confirms that Cluster4 has the best learning attitude.Cluster5 is outstanding in participating in discussions, far more than the other categories of learners, indicating that Cluster5 is very active in the classroom and often uses the forum to communicate with classmates and teachers.

Further analyzing the performance of learners in each category in terms of academic characteristics, Cluster4 learns the most knowledge points, has the highest entropy of knowledge points and the highest pass rate of knowledge points, which indicates that Cluster4 learns knowledge comprehensively but with focus and learns well. Cluster2 did not learn many points, but had the second highest knowledge pass rate after Cluster4, suggesting that Cluster2 was targeted for selective learning, and that these students tended to be the ones with the foundation. Cluster3 is in a low state for all academic characteristics, indicating that the learning state is in dire need of adjustment. Cluster1 learns more than Cluster3, but is still in a suboptimal state with a pass rate of only 25%, which represents a somewhat one-sided and insufficiently in-depth study of Cluster1. Cluster5 watched more courses than Cluster1, but had a pass rate of only 0.16, a slightly worse learning status than Cluster1.

The enrollment ratio by different categories of learners at different times is shown in Figure 8. Cluster2 has a high enrollment ratio of 75% at the beginning of the course, then drops to about 15%, and then gradually increases again in the middle and late semester, which suggests that students with a basic knowledge will be more active in enrolling in the beginning of the course, and then start a wave of learning again towards the end of the course.

Cluster1 accounts for about half of the total learners, gradually increasing its share after the beginning of the semester and finally stabilizing at about 60%.

The curves for Cluster3 and Cluster4 are in a steady and slowly decreasing state, while the curve for Cluster5 has remained stable.

In the period after December 20, the curves of Cluster1, Cluster2, and Cluster4 show large oscillations, this is because this period is at the end of the semester, there are fewer students joining the course, and some small increase or decrease in the data can cause a relatively large oscillation.

Table 3: The overview of the statistical attributes of learners in different clusters

		Play Count	Play Time	Discuss Count	Discuss Amount	SAttitude	KCount	KEntropy	KPass Percent
Cluster 1	Min	6.00	652.15	0.00	0.00	0.64	3.00	0.93	0
	Mean	65.63	15470.46	6.27	175.89	2.06	8.64	1.88	0.25
	Max	267.00	46655.43	15.00	924.00	88.82	18.00	2.09	0.5
Cluster 2	Min	1.00	2.40	1.00	0.00	1.55	1.00	1.75	0
	Mean	11.33	1885.55	6.29	169.01	2.14	2.94	2.63	0.5
	Max	121.00	30924.21	18.00	874.00	64.37	12.00	1.34	1
Cluster 3	Min	1.00	33.94	1.00	4.00	1.01	0.00	0.66	0
	Mean	61.30	5176.21	5.54	175.96	2.90	7.33	0.83	0.25
	Max	292.00	58672.13	12.00	920.00	77.83	10.00	1.12	0.5
Cluster 4	Min	99.00	24984.87	0.00	5.00	0.35	13.00	0.99	0
	Mean	290.57	67178.22	6.45	214.09	17.52	22.12	1.86	0.6
	Max	1543.00	90494.96	22.00	1820.00	330.93	25.00	1.37	1
Cluster 5	Min	6.00	72.07	4.00	73.00	0.60	2.00	2.95	0
	Mean	67.68	16060.38	14.17	813.27	1.75	9.90	0.96	0.16
	Max	282.00	40163.53	85.00	4557.00	156.48	18.00	1.85	0.5

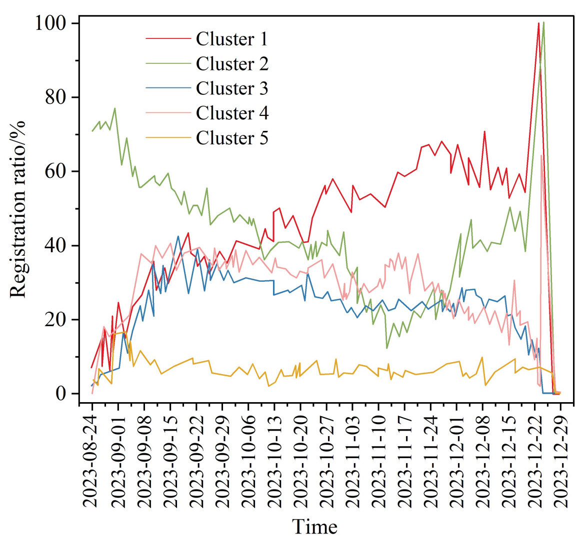

Figure 8: The registration ratio of different learners at different times

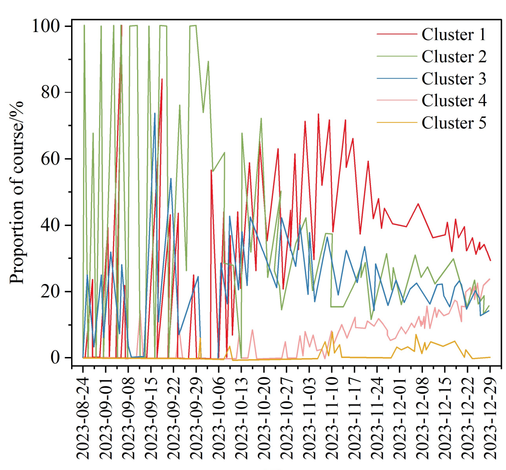

Figure 9: The proportion of different learners in different times

The distribution of the percentage of class closures by each type of learner is shown in Figure 9. In the middle and early part of the semester, the completion ratio of each type of learner is in a wide oscillation (the total number of completion in the early part of the semester is very small.) After November 24, Cluster4 gradually increases the completion ratio, which indicates that the more excellent learners have a later completion time and a longer duration of study. Most of the rest of the types of learners have a steady decreasing percentage of course closures from November 10th onwards.

Summarizing the results of the analysis, the characteristics of Cluster1, Cluster2, Cluster3, Cluster4 and Cluster5 are obtained as follows:

Cluster1: Learning one-sided type. This kind of students learn fewer knowledge points, which leads to insufficient learning hours, learning attitude value, knowledge point entropy and knowledge point pass rate and other characteristics of mediocre performance, and the large number of this kind of learners, indicating that the level of network education still has more room for improvement.

Cluster2: Key Learning Type. This type of learners are mostly basic students, similar to Cluster1 in the behavioral characteristics, and does not dominate, but should be distinguished from Cluster1, because Cluster2 students are selective to learn certain knowledge points in depth.

Cluster3: The soy sauce type. These students are on the verge of giving up learning, and all the characteristics are bad, showing a negative and slow state. Network education administrators should fully mobilize the learning enthusiasm of such students.

Cluster4: Comprehensive learning type. The characteristics of such students show that the students' learning state is very good, the knowledge point learning is more comprehensive, the duration of learning time is long, but also will distinguish the key knowledge, with a certain degree of subjective initiative.

Cluster 5: Discussion-oriented. This type of student actively participates in the discussion in the forum, is the main force of the active forum, but the learning situation is not as comprehensive as Cluster4, so it is presumed that this type of learner's learning method may be problematic, and need to be corrected by the teacher in a timely manner.

VI. Conclusion

This paper establishes a remote Russian language teaching model based on the idea of digital twin technology and combining the intelligent grouping technology of improved genetic algorithm and the learning behavior analysis technology based on AFSA-FCM.

As shown by the algorithm validation, the time taken by the original genetic algorithm to run is in the range of 7500~9500ms. The improved genetic algorithm does not exceed 5500ms.The difficulty scores of the test paper questions generated by the improved genetic algorithm in the multiple choice, judgment, fill-in-the-blank, and short answer question types are 0.572, 0.534, 0.511, 0.569, and 0.531, respectively, which are in the range of moderate [0.4~0.6].The difficulty of the test paper questions generated by the original genetic algorithm, only judgment questions are moderately difficult. It is confirmed that the improved genetic algorithm can not only improve the efficiency of organizing papers, but also meet the requirement of the reasonableness of the difficulty of the test paper questions.

Using the proposed AFSA-FCM algorithm to mine the learning data of 11,465 students at the University of W and analyze all types of learners in detail, five typical learner portraits are portrayed. It is confirmed that the model can effectively achieve learning behavior analysis.

Through the personalized analysis of 5 types of typical students and combining the advantages of distance education, the following suggestions are made.

(1) The learning effect of distance online learning can be further improved, teachers should pay more attention to the breadth and depth of students' learning, and some flexibility can be appropriately reduced in the design of the learning platform.

(2) An improvement class can be set up for students who are good at learning to learn knowledge that is more in line with their own level. At the same time, the learning platform can add a “traffic light” function to warn students, when a red light, it means that their learning status has been lower than the average level of students in the same batch. This kind of internal competition may be more effective than external pressure.

The application of digital twin technology in distance learning has important value and significance. It can enhance students' practical ability and innovative spirit and improve the quality of teaching. However, at present, the application of digital twin technology in distance teaching is still in the exploratory stage, and needs to be further improved and optimized. In the future, educators and related researchers can carry out in-depth research and practice in the following aspects: strengthen the research and development and innovation of virtual simulation technology, improve the simulation degree and real-time nature of the platform, so as to make it closer to the real-world IoT environment and system; enrich the content and form of the virtual simulation teaching resource base, to meet the learning needs of students of different specialties and levels. Explore the teaching mode of combining virtual simulation and physical equipment to realize the organic integration of online and offline teaching.

VII. References

[1] Aref'ev, A. L. (2014). The Russian language in the world: Past, present, and future. Herald of the Russian Academy of Sciences, 84(5), 357-364.

[2] Egorov, V. G. (2021). Russian language in the global linguocultural space. RUDN Journal of Language Studies, Semiotics and Semantics, 12(4), 1189-1215.

[3] Tekucheva, I. V., & Gromova, L. Y. (2016). On Present State of Teaching Russian Language in Russia. International Journal of Environmental and Science Education, 11(14), 6504-6511.

[4] Nuriddinovna, M. D. (2022). Methodology of Teaching the Russian Language. International Journal on Orange Technologies, 4(2), 57-60.

[5] Vikhman, V. V., & Romm, M. V. (2021). "Digital twins" in education: prospects and reality. Vysshee Obrazovanie v Rossii= Higher Education in Russia, 30(2), 22-32.

[6] Eriksson, K., Alsaleh, A., Behzad Far, S., & Stjern, D. (2022). Applying digital twin technology in higher education: An automation line case study. In SPS2022 (pp. 461-472). IOS Press.

[7] Omrany, H., Al-Obaidi, K. M., Ghaffarianhoseini, A., Chang, R. D., Park, C., & Rahimian, F. (2025). Digital twin technology for education, training and learning in construction industry: implications for research and practice. Engineering, Construction and Architectural Management.

[8] Wang, K. J., Lee, T. L., & Hsu, Y. (2020). Revolution on digital twin technology—a patent research approach. The International Journal of Advanced Manufacturing Technology, 107, 4687-4704.

[9] Zaballos, A., Briones, A., Massa, A., Centelles, P., & Caballero, V. (2020). A smart campus' digital twin for sustainable comfort monitoring. Sustainability, 12(21), 9196.

[10] Botín-Sanabria, D. M., Mihaita, A. S., Peimbert-García, R. E., Ramírez-Moreno, M. A., Ramírez-Mendoza, R. A., & Lozoya-Santos, J. D. J. (2022). Digital twin technology challenges and applications: A comprehensive review. Remote Sensing, 14(6), 1335.

[11] Lexman, R. R., & Baral, R. (2024). Digital twins in MOOCs: exploring ways to enhance interactivity. Development and Learning in Organizations: An International Journal, 38(4), 23-26.

[12] Siyan, C., Tinghuai, W., Xiaomei, L., Liu, Z., & Danying, W. (2021). Research on the improvement of teachers' teaching ability based on machine learning and digital twin technology. Journal of Intelligent & Fuzzy Systems, 40(4), 7323-7334.

[13] Zheng, Y., Yang, S., & Cheng, H. (2019). An application framework of digital twin and its case study. Journal of ambient intelligence and humanized computing, 10, 1141-1153.

[14] Demir, L. (2024). DIGITAL TWIN TECHNOLOGY IN HIGHER EDUCATION MANAGEMENT: THE CASE OF TURKEY. International Online Journal of Education & Teaching, 11(1).

[15] Chamorro-Atalaya, O., Flores-Velásquez, C. H., Flores-Cáceres, R., Arévalo-Tuesta, J. A., Zevallos-Castañeda, M., Tomás-Quispe, G., ... & Alarcón-Anco, R. (2024). Use of Digital Twin Technology in the Teaching-Learning Process, in the field of University Education: A Bibliometric Review. International Journal of Learning, Teaching and Educational Research, 23(2), 313-329.

[16] Guangyu, L. I. A. N. G., Xinyu, P. A. N. G., Shuojie, L. I., Bozhao, S. H. I., & Dengyu, L. I. U. (2024). Application of a digital twin in virtual simulation experiment teaching of the control course. Experimental Technology & Management, 41(3).

[17] Sepasgozar, S. M. (2020). Digital twin and web-based virtual gaming technologies for online education: A case of construction management and engineering. Applied Sciences, 10(13), 4678.

[18] Deniz, S., Müller, U. C., Steiner, I., & Sergi, T. (2022). Online (remote) teaching for laboratory based courses using "digital twins" of the experiments. Journal of Engineering for Gas Turbines and Power, 144(5), 051016.

[19] Guc, F., Viola, J., & Chen, Y. (2021, July). Digital twins enabled remote laboratory learning experience for mechatronics education. In 2021 IEEE 1st international conference on digital twins and parallel intelligence (DTPI) (pp. 242-245). IEEE.

[20] Khan, T. H., Noh, C., & Han, S. (2023). Correspondence measure: a review for the digital twin standardization. The International Journal of Advanced Manufacturing Technology, 128(5-6), 1907-1927.

[21] Kwok, P. K., Yan, M., Qu, T., & Lau, H. Y. (2021). User acceptance of virtual reality technology for practicing digital twin-based crisis management. International journal of computer integrated manufacturing, 34(7-8), 874-887.

[22] Tuhaise, V. V., Tah, J. H. M., & Abanda, F. H. (2023). Technologies for digital twin applications in construction. Automation in Construction, 152, 104931.

[23] Selim, A., Ali, I., Saracevic, M., & Ristevski, B. (2024). Application of the digital twin model in higher education. Multimedia Tools and Applications, 1-18.

[24] Kartashova, L. A., Gurzhii, A. M., Zaichuk, V. O., Sorochan, T. M., & Zhuravlev, F. M. (2020). Digital twin of an educational institution: an innovative concept of blended learning. In Proceedings of the symposium on advances in educational technology, aet.

[25] She, M., Xiao, M., & Zhao, Y. (2023). Technological implication of the digital twin approach on the intelligent education system. International journal of humanoid robotics, 20(02n03), 2250005.

[26] Razzaq, S., Shah, B., Iqbal, F., Ilyas, M., Maqbool, F., & Rocha, A. (2023). DeepClassRooms: a deep learning based digital twin framework for on-campus class rooms. Neural Computing and Applications, 1-10.

[27] Hlayel, M., & Jawad, M. S. (2022). Leverage Digital Transformation in Distance Learning for Maximum Students' Engagement by Utilization of the Integrated Technologies of Digital Twin, Extended Realities and Gamification. International Journal of Computing Sciences Research, 7, 1594-1620.

[28] Damaševičius, R., & Zailskaitė-Jakštė, L. (2024). Digital twin technology: Necessity of the future in education and beyond. Automated Secure Computing for Next-Generation Systems, 1-22.

[29] Zhang, Y., Pang, L., Wang, D., & Liu, S. (2022). Influences of Digital Twin Technology on Learning Effect. Journal of Engineering Science & Technology Review, 15(4).

[30] Hladký, V. (2023, May). A Digital Twin for Remote Learning: A Case Study. In Advances in Design, Simulation and Manufacturing VI: Proceedings of the 6th International Conference on Design, Simulation, Manufacturing: The Innovation Exchange, DSMIE-2023, June 6–9, 2023, High Tatras, Slovak Republic-Volume 1: Manufacturing Engineering (p. 379). Springer Nature.
[31] Ağca, R. K. (2023). Using digital twins in education from an innovative perspective: Potential and application areas. Education Mind, 2(2), 65-74.
[32] Balyakin, A. A., Nurakhov, N. N., & Nurbina, M. V. (2021). Digital twins in contemporary education: Virtual workshop. In Perspectives and Trends in Education and Technology: Selected Papers from ICITED 2021 (pp. 473-483). Singapore: Springer Singapore.
[33] Deniz Sabri,Müller Ulf Christian,Steiner Ivo & Sergi Thomas. (2022). Online (Remote) Teaching for Laboratory Based Courses Using "Digital Twins" of the Experiments. Journal of Engineering for Gas Turbines and Power,144(5),
[34] Maddahi Yaser & Chen Siqi. (2022). Applications of Digital Twins in the Healthcare Industry: Case Review of an IoT-Enabled Remote Technology in Dentistry. Virtual Worlds,1(1),20-41.
[35] HanYang. (2023). A Study on an Intelligent Algorithm for Automatic Test Paper Generation and Scoring in University English Exams. Journal of ICT Standardization,11(4),391-402.
[36] Yin Qilong. (2021). Research on Intelligent Test Paper Generating System Based on Improved Genetic Algorithm. Journal of Physics: Conference Series,1744(3),032203-.
[37] Ma Li,Li Yang,Fan Suohai & Fan Runzhu. (2015). A Hybrid Method for Image Segmentation Based on Artificial Fish Swarm Algorithm and Fuzzy c-Means Clustering. Computational and mathematical methods in medicine,2015,120495.
[38] Ma Li,Li Yang,Fan Suohai & Fan Runzhu. (2015). A Hybrid Method for Image Segmentation Based on Artificial Fish Swarm Algorithm and Fuzzy c-Means Clustering. Computational and mathematical methods in medicine,2015,120495.