

## *Retraction*

# **Retracted: Artificial Intelligence Teaching System and Data Processing Method Based on Big Data**

### **Complexity**

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Complexity. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] B. Xu, "Artificial Intelligence Teaching System and Data Processing Method Based on Big Data," *Complexity*, vol. 2021, Article ID 9919401, 11 pages, 2021.

## Research Article

# Artificial Intelligence Teaching System and Data Processing Method Based on Big Data

Bo Xu 

*School of Educational Science, Xianyang Normal University, Xianyang, Shaanxi 712000, China*

Correspondence should be addressed to Bo Xu; xubo@snnu.edu.cn

Received 4 April 2021; Revised 20 April 2021; Accepted 28 April 2021; Published 5 May 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Bo Xu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of big data, artificial intelligence teaching systems have gradually been developed extensively. The powerful artificial intelligence teaching systems have become a tool for teachers and students to learn independently in various universities. The characteristic of artificial intelligence teaching system is to get rid of the constraints of traditional teaching time and space and build a brand-new learning environment, which is the mainstream trend of future learning. As the carrier of students' autonomous learning, the artificial intelligence teaching system provides a wealth of learning resources and learning tools on the one hand, and on the other hand, it gradually accumulates more and more learning behaviors, learning status, and other large amounts of data, which is an in-depth study of online learning and provides valuable and generative dynamic resources. Based on relevant researches on domestic and foreign related learning analysis and common big data analysis methods, combined with actual learning evaluation goals, this paper proposes an artificial intelligence teaching system using big data analysis methods and a modeling process framework for online learning evaluation and uses student data to carry out predictive evaluation modeling to evaluate student learning outcomes. The evaluation results can enable teachers to predict whether students can successfully complete the course of learning after a period of teaching. Through the final evaluation, students' learning problems can be discovered in time based on the evaluation results, and targeted interventions can be made for students who are at risk. The scientific and objective learning evaluation obtained in this study through data analysis can not only provide teachers with relevant information and provide personalized guidance to students, but also improve the adaptive and personalized service functions of the learning platform of the artificial intelligence teaching system, greatly reducing teachers teaching burden. Artificial intelligence teaching evaluation can help educators understand the problems in teaching, adjust teaching strategies in time, and improve teaching results.

## 1. Introduction

In recent years, big data technology has penetrated more and more into various fields, and big data technology has changed many industries, allowing us to better understand the deep meaning of data [1]. Due to the difference between online learning and traditional learning methods, students' online learning evaluation cannot be carried out using traditional educational evaluation methods. Online learning breaks the time and space limitations of traditional teaching. There are a large number of students, and the learning process and behaviors are complex. Traditional teachers can only judge students by the results of homework and examination results, but they do not understand the behavior

of students in the learning process and cannot make a comprehensive evaluation of students in time [2]. The learning evaluation using big data analysis methods can track and utilize the learning data of students and calculate the data in a scientific way, so that students can be objectively evaluated in a timely and accurate manner. For educators, it can reduce the burden and improve the teaching effect; for students, they can understand their learning situation in time and encourage learning behavior [3]. At present, big data is developing rapidly, and in the field of education, analyzing and researching teaching and learning process data, that is, "big data and learning analysis technology," has gradually become a new research direction. Using big data to analyze the learning behavior of students can make learning

evaluations based on objective facts from multiple perspectives and diversity. Artificial intelligence teaching evaluation is an indispensable link in the process of online learning. Through the evaluation of learners, the monitoring and feedback of the learning process of learners can be enhanced, the relative separation of online learning can be improved, and the learning of learners can be strengthened. Guidance helps learners maintain long-term learning motivation, guides and promotes learners' learning, and improves learners' ability to learn independently, thereby improving the quality of learning and making learning a spontaneous behavior. At the same time, it is helpful to teachers and artificial intelligence teaching.

Big data analysis can analyze the data more quickly and accurately in the huge data. Therefore, the introduction of big data technology into artificial intelligence teaching is a major leap in artificial intelligence teaching research [4]. Under the current artificial intelligence teaching model, the personalized education of students has not been fully realized, and students cannot be taught in accordance with their aptitude. On the one hand, the reason is the large amount of student data in artificial intelligence teaching, the limited storage capacity, the lack of effective analysis tools, and the inability to teach students to carry out real-time evaluation and feedback. On the other hand, in the process of tracking students' learning behaviors, most teaching platforms lack the ability to record students' learning behaviors in an all-round way. They can only record and analyze certain specific action data, and they do not make good use of these data [5]. There are types of domestic research on using big data for learning behavior analysis; one is data index research, and the other is technical means research. The main research fields abroad are divided into three areas: use tool software to track and record learning behavior; pay attention to learner's needs and online learning environment; find the relationship between learning behavior and learning performance. Through the relevant research on big data learning and analysis at home and abroad in recent years, understand the research purpose, research content, research process, and current results and progress of the relevant research of big data learning and analysis and have an understanding of the relevant research of this research. And draw the research methods that are conducive to this research from the research at home and abroad in recent years, and refer to it for reference.

Although most experts and scholars at home and abroad have adopted a positive attitude towards artificial intelligence in education, the application of artificial intelligence in education is far from mature because artificial intelligence technology itself is still evolving and developing. Based on relevant researches on domestic and foreign related learning analysis and common big data analysis methods, combined with actual learning evaluation goals, this paper proposes a student artificial intelligence teaching system using big data analysis methods and a modeling process framework for online learning evaluation and uses student data to carry out predictive evaluation modeling to evaluate student learning outcomes. The evaluation results can enable teachers to predict whether students can successfully complete the

course of learning after a period of teaching. Through the final evaluation, students' learning problems can be discovered in time based on the evaluation results, and targeted interventions can be made for students who are at risk. The scientific and objective learning evaluation obtained in this study through data analysis can not only provide teachers with relevant information and provide personalized guidance to students, but also improve the adaptive and personalized service functions of the learning platform of the artificial intelligence teaching system, greatly reducing teachers teaching burden. Combining the actual learning evaluation goals, this paper proposes a student artificial intelligence teaching system using big data analysis methods and a modeling process framework for online learning evaluation.

## 2. Related Work

Although artificial intelligence has begun to enter the stage of human application development, it is still in the preliminary stage of integration of artificial intelligence and education and teaching practice. Through the research on artificial intelligence teaching and application at home and abroad, it can be seen that, in recent years, more attention has been paid to the technology development and application research of artificial intelligence education and teaching in education, and the research scope also involves many areas of education and teaching activities. "Intelligent learning guidance system, automated evaluation system, educational game, and educational robot," as the main four forms of artificial intelligence education applications, have attracted the attention of many scholars [6]. The intelligent learning guidance system can create personalized courses based on learners' language, learning style, knowledge structure and emotional state, and other personal characteristics to achieve "one person, one lesson schedule" and provide personalized learning guidance based on data feedback to satisfy different learners demand. And automatic evaluation systems based on natural language processing technologies such as E-rater, ProjectEssay Grade, IntelliMetric, My Access, Criterion, and Pianjiao.com have also been developed rapidly, which can realize automatic correction and grading of homework [7]. As a powerful learning tool, educational robots are becoming more and more common in the field of education. Chinese scholars also conduct analysis and research on the current status and future development trends of artificial intelligence education and teaching applications. Ikedinachi AP [8] and others introduced the four current combined results of artificial intelligence technology and the education industry, such as automatic homework correction, photo search for online question answering, intelligent assessment and personalized learning, and the future of artificial intelligence and education in the field of learner model, in addition to discussion and prospects for the integration and application of the teaching model field and the specific learning field. Williams [9] conducted research on the 6 fields of expert systems, robotics, machine learning, natural language understanding, artificial neural networks, and distributed artificial

intelligence, which are widely used and active in education, and explored their applications in education. Burton [10] also pointed out in the research that the current “artificial intelligence + education industry” approach mainly follows the separation of the main links of education and provides technical support in the links of practice, evaluation, learning process, teaching process, management process, etc. and application. Lu [11] and others combed and summarized the results of case studies of machine learning education applications based on real data abroad in recent years and found that the current machine learning education applications mainly focus on student modeling, student behavior modeling, predictive learning behavior, and early warning six aspects: risk of dropping out, learning support and evaluation, and resource recommendation. Schiff [12] analyzes the relationship between technology and education time and space and proposes personalized learning, appropriate services, academic evaluation, role changes, the application potential of the five major artificial intelligence education interdisciplinary studies, and the educational value of artificial intelligence, under the environment of human-computer coexistence. The application challenges of artificial intelligence education are based on advanced teaching experience, safety ethics of smart technology, effective collaboration between government enterprises and schools, and technological governance of the harmonious development of man-machines. At the same time, artificial intelligence has also been successfully applied in the field of special education. It can extend the functions of special people’s organs, use technical means to make up for their intellectual or physical deficiencies, meet the needs of different special people to the greatest extent, and promote their personalized learning. With the increasing development of artificial intelligence technology, the impact of artificial intelligence on education is becoming more and more profound. Therefore, some scholars have conducted research on the future trend of the integration of artificial intelligence and education, pointing out that the further development of artificial intelligence in the future will have a huge impact on educational goals, learning methods, educational content and educational models, educational environment and educational resources, and the role of teachers.

Although most experts and scholars at home and abroad have adopted a positive attitude towards artificial intelligence in education, the application of artificial intelligence in education is far from mature because artificial intelligence technology itself is still evolving and developing. Goel [13] believes that the integration of artificial intelligence and education is only limited to the teaching field and pays more attention to exploring the application of artificial intelligence technology in the education field, which is a superficial and shallow integration. Secondly, artificial intelligence education assumes part of the role of auxiliary teaching and auxiliary evaluation but has not achieved the integration with the whole process of education, and artificial intelligence education in this period is difficult to achieve in the cultivation of students’ emotions, attitudes, and values. The U.S. “Education Communication and Technology Research

Manual (Fourth Edition)” proposes that the application of artificial intelligence education currently faces three major challenges. The first is the research and development of the intelligent tutoring teaching system; the second is how to implement the intelligent tutoring teaching system on a regular basis; it is to make the application of artificial intelligence in education truly return to the original goal, that is, to provide learners with a highly “individualized” learning environment by truly implementing an “adaptive teaching system.” Dyachenko NN [14] analyzed the problems in the application of intelligent teaching system design. For example, the intelligent teaching system cannot realize self-renewal and self-improvement at present, the intelligent teaching system cannot be widely promoted, and the intelligent teaching system cannot realize emotional interaction. The application of artificial intelligence in the field of education has received more and more attention. The research results of artificial intelligence are gradually applied to education and teaching, which has a profound impact on teaching practice. Teachers themselves are also facing huge challenges. Therefore, the development of teachers has become a hot spot for scholars in the process of artificial intelligence education application. The existing research is mainly conducted from the three perspectives of teacher role development, teacher-student relationship, and teacher professional development in the era of artificial intelligence. The research on the role of teachers in the era of artificial intelligence mainly answers two questions: one is whether teachers can be replaced by artificial intelligence, and the other is the transformation of the role of teachers in the era of artificial intelligence.

### 3. Establishment of Artificial Intelligence Teaching System Based on Big Data

*3.1. System Functional Structure Design.* The artificial intelligence teaching system is very goal-oriented at the content and function levels. Therefore, the overall system needs to have good stability, low response time, stable and reliable user data, simple and convenient system operation, and friendly interface and main functions [15]. The system needs section analyzes and organizes the differentiated needs of multiple disciplines. For different disciplines, the refinement and labeling of different content require good support for text, pictures, formulas, audio, video, rich media, and rich text. Therefore, scalability needs to be considered and as much as possible to ensure smooth transition and upgrade when adding new subjects or content bearing forms, without affecting the original system business and other functional modules. In the requirements analysis stage, we determined the main functions that the system needs to implement. The main functions of the system are organized as shown in Figure 1 below.

Based on the previous demand analysis, the main functions of the system are divided into eight items. They are authority management, course management, knowledge point management, topic management, interactive learning, interactive exercises, interactive assessment, and data reporting. The authority management executive is the

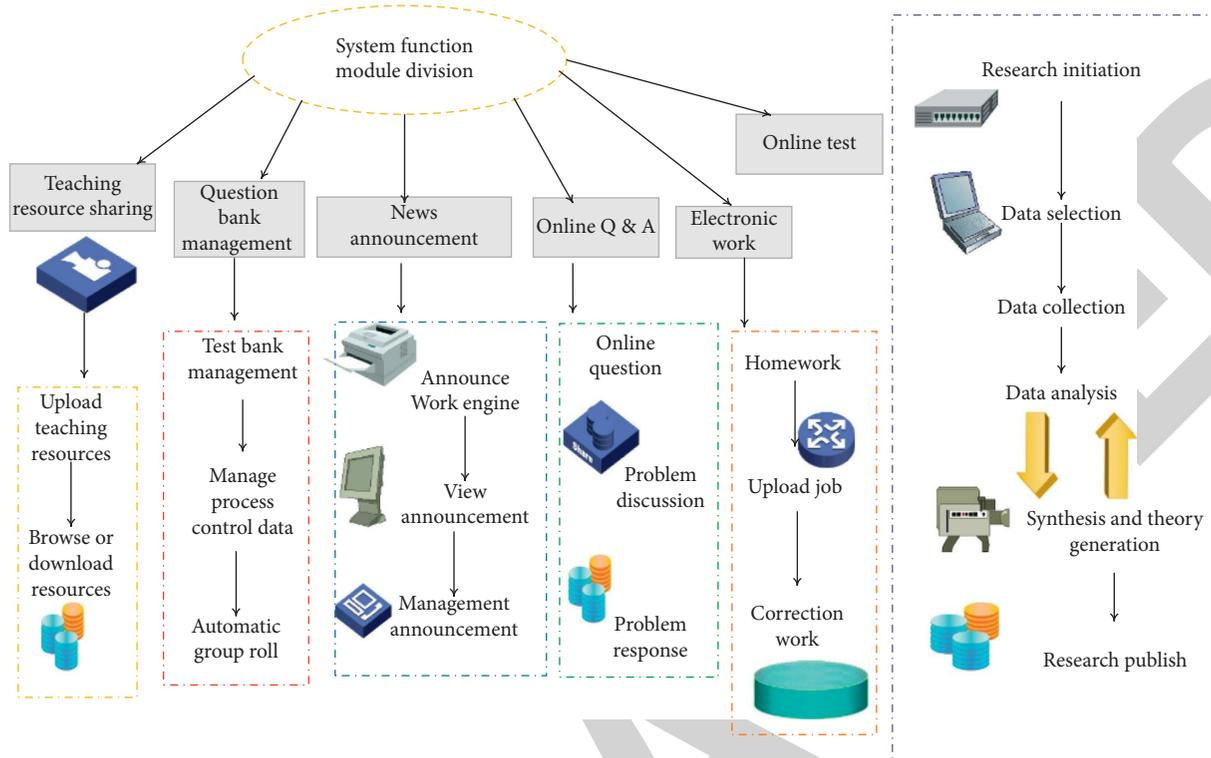


FIGURE 1: System functional structure diagram.

system administrator, who manages course administrators and users in the system. For system users, the basic information can be collected and sorted out, and it can be classified based on the purchase of courses and the usage of the system. Based on consumption records and content access rights, it can be divided into ordinary users, VIP users, and partner experience users, etc.; based on platform participation, it can be divided into normal users, service suspended users, and disabled users. It can realize the distribution and recovery of user access rights through the background and realize the user logout operation. For course administrators, they can divide operating permissions according to business modules and specific business lines, implement modularization for assembly line operations, refine permissions management, improve the fine-grained permissions setting, control the number of super administrators, and update administrator secret keys regularly.

Curriculum management, knowledge point management, and topic management executives are course administrators and mainly complete the maintenance, integration, and release of teaching materials in the system. Curriculum management is based on the teaching system to set up a hierarchical directory for accessing teaching resources. Knowledge point management realizes adding, deleting, and updating the content of knowledge points, setting the association between knowledge points, and setting the association between knowledge points and supporting materials [16]. The topic management realizes the basic operation permissions of adding, editing, deleting, updating, reviewing, returning, going online, offline,

combining questions, entering different types of questions, adding topic types, setting classification labels for topics, and setting topics for binding related knowledge points and other functions. For the maintenance work order of a single topic, the business can be disassembled in stages and the attributes can be configured. In the event of failure or temporary failure of the question, the system can make a judgment and give a premade error prompt page. Thinking label management can be carried out in the topic, including the creation of thinking label catalog and category and the addition, deletion, and editing of thinking labels, and complete the establishment of thinking tags in the question and be able to associate knowledge points and save the learning process and interaction process of the user in various ways in the question.

The user is the executor of the interactive learning, practice, evaluation, and data reporting functions. This part of the function carries the complete learning process provided by the system to users. Interactive learning realizes audio and video access in specific courses, participates in interactive questions, takes notes, asks questions, and jumps to practice questions and other functions. Interactive exercises and assessments realize the interaction with the system during the question-making process and establish contact with the teaching system by clicking on the label, clicking on the answer, and inputting feedback information, so that the system can more comprehensively obtain the user's information and establish a more complete user model and push content [17]. The data report implements functions such as scores, correctness rates, and weak point analysis, generates a general evaluation report of the learning

situation within a certain time frame, counts the indicators of learning participation, progress, and overall mastery of the course, and scientifically arranges the learning plan.

*3.2. System Database Design.* Based on the actual needs of the teaching system, the following basic principles need to be followed during the database design stage of the subject development:

- (1) The basic table maintains a one-to-one, one-to-many relationship. The relationship table is a many-to-many relationship.
- (2) Basic table characteristics: Atomicity, primitiveness, deduction, and stability.
- (3) Paradigm principle: The relationship between the basic table and its fields should meet the third paradigm as much as possible, that is, ensure that each column is directly related to the primary key column, rather than indirectly related. In actual operation, in order to improve the operating efficiency of the database, the paradigm standard will be reduced in some tables, redundancy will be appropriately increased, operating efficiency will be improved, and space will be exchanged for time.
- (4) Primary key value rule: The involved forms and records are mainly filtered and sorted based on time or subkey weight, so the primary key value is mainly generated in an automatic form.
- (5) Data redundancy: Data redundancy refers to the repetition of nonkey fields and does not include the repetition of primary and foreign keys in multiple tables. Data redundancy is very necessary in improving efficiency and ensuring data security and will be appropriately retained in this design.
- (6) View: The essence of a view is different from a table. It is a virtual table derived from one or more basic tables. It has a great effect in improving the speed of calculation, simplifying complex data operations, and saving storage space. In the teaching system, a large number of real-time table data operations are required, and full use of views can ensure operating efficiency.
- (7) Integrity constraints: follow entity integrity, domain integrity, referential integrity, and user-defined integrity.

### *3.3. Data Synchronization Design*

*3.3.1. Synchronization of Teaching Content.* When adding, deleting, and updating the reading content of the teaching content database, a switch is designed to realize the manual update of the online data on the browser side to keep the learning content synchronized. For error content or offline content, when the user accesses it through history records or other forms, a prompt page will be given, and the history records will be deleted synchronously.

*3.3.2. User Record Synchronization.* After the user completes the learning path in a specific module unit in accordance with the rules, the learning path and specific learning operation records are synchronized to the user's personal log. The conceptual model of the teaching system database is shown in an ER diagram, as shown in Figure 2.

*3.4. Database Table Design.* The teaching system has established databases and forms according to specific businesses to facilitate statistical calculations, set up system management log table, system administrator table, article table, audio table, collection question table, set of papers table, practice table, order table, question table, question type management table, question mode table, teaching table, subcourses main forms such as table, knowledge point table, etc., carries out the main business such as user management, course management, teaching content management, topic management, course catalog management, knowledge point map management, and so on.

The core function of the teaching system is to serve the closed-loop teaching logic of "learning, practicing, and testing" [18]. Therefore, it is necessary to link teaching, practice, and evaluation data and use the knowledge point map as the cornerstone of the system to continuously provide users with the optimal learning path to achieve predetermined goals based on the analysis of user data. As the basis for the establishment of the system, the knowledge point map needs to carry a large number of basic attributes, such as name, content, difficulty, and frequency of investigation. Knowledge points need to be related, and based on this, a basic model of content push is established. The teaching content associated with the knowledge points includes courses and topics. Each course can contain multiple knowledge points, and each topic can also contain multiple knowledge points. On this basis, the system needs to judge the user's mastery of related knowledge points through the performance of courses or topics and make targeted and divergent content recommendations based on the mastery situation and the connections established between the knowledge points. If you combine user real-time learning data in the content recommendation process, you can achieve preliminary adaptive content push based on sample statistics. When you have a large amount of associated data, you can try to use the probability graph model for data mining to further optimize the existing push model. The adaptive system is a system that is continuously optimized and advances based on data accumulation. In the early design, the database form and its associated establishment logic directly determine the system logic and the performance that can be achieved.

*3.5. Model Performance Evaluation.* For the prediction performance evaluation of the two-class training model in machine learning, some parameters are usually used: true positive (TP), false positive (FP), true negative (TN), false negative (FN). True positive, that is, an instance itself, is a positive class and is predicted to be a positive class; a false positive class is an instance itself and is a negative class but is

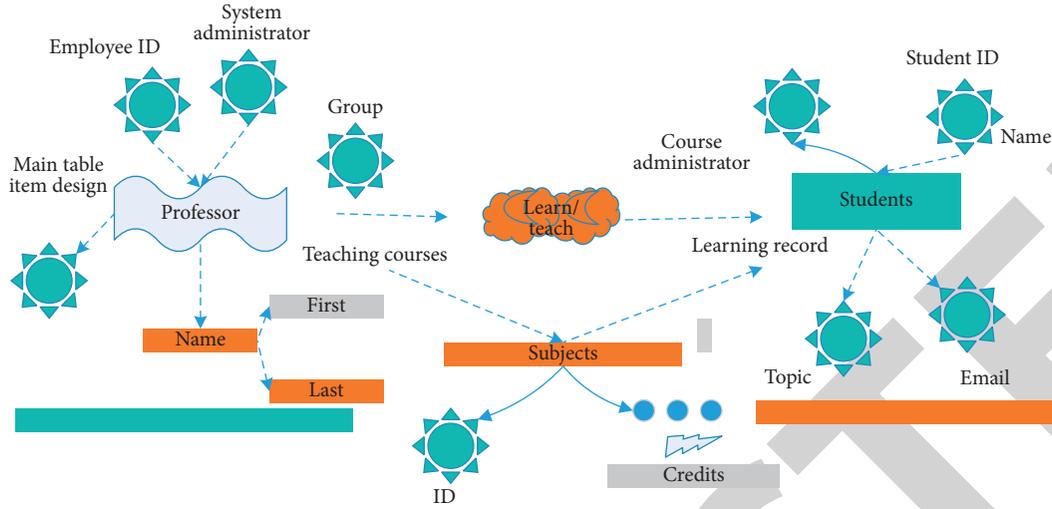


FIGURE 2: Conceptual ER diagram of teaching system database.

predicted to be a positive class; a true negative class is an instance itself and is a negative class and is predicted to be a negative class; a false negative class, an instance itself, is a positive class and is predicted to be a negative class [19–24]. These indicators are used to judge a classifier's ability to classify the samples used in different situations. It is inaccurate to judge the performance of a model simply by accuracy, especially when the deviation of the number of categories in the data is particularly large; if only the accuracy is considered, no training is required. All the test samples are divided into the category with a large number. The accuracy is also quite high, so it cannot reflect the ability of the model. Therefore, reference evaluation will be made in different applications based on several different indicators. In order to reflect the accuracy of the research in the research results, we will pay attention to the following data in results report:

$$\text{Precision} = \frac{\text{TruePos}}{(\text{TruePos} + \text{FalsePos})}, \quad (1)$$

which reflects the proportion of true positive samples among the positive examples judged by the classifier;

$$\begin{aligned} \text{Accuracy rate} &= \frac{(\text{TruePos} + \text{FalsePos})}{(\text{True} + \text{False})} \\ &= \frac{(\text{TruePos} + \text{FalsePos})}{(\text{TruePos} + \text{FalsePos} + \text{TrueNeg} + \text{FalseNeg})}, \end{aligned} \quad (2)$$

which reflects the ability of the classifier to judge the entire sample and can judge positive as positive and negative as negative.

The Recall formula is

$$\begin{aligned} \text{Recall} &= \frac{\text{TruePos}}{(\text{TruePos} + \text{FalseNeg})} \\ &= \frac{1 - \text{FalseNeg}}{\text{Total}}, \end{aligned} \quad (3)$$

which reflects the proportion of positive cases that have been correctly determined to the total positive cases.

False positive rate (specificity) is also known as NegativeRate.

The formula is

$$\begin{aligned} \text{Specificity} &= \frac{\text{TrueNeg}}{(\text{TrueNeg} + \text{FalsePos})} \\ &= \frac{1 - \text{FalsePos}}{\text{Total}}. \end{aligned} \quad (4)$$

## 4. Research on Machine Learning Algorithms for Predictive Models

**4.1. Algorithm Introduction.** This research will use big data analysis methods to evaluate the learning process of students. Therefore, it is necessary to select several big data analysis machine learning algorithms for research. There are many types of machine learning algorithms, and different algorithms can solve different problems. The appropriate algorithm needs to be selected according to the research objectives. This research hopes to evaluate the current learning situation of students through training on the learning data of students and predict whether the course can be successfully completed through the current learning situation. According to the actual situation of this research, through the evaluation of the algorithm, we selected four more common classification algorithms for comparison: logistic regression, support vector machine (SVM/SMO), J48 decision tree, and Bayesian; these classification algorithms can be classified, and the relevant calculation program code is provided in the system.

**4.2. Logistic Regression.** Logistic regression is a very commonly used machine learning algorithm, which can solve classification problems. Logistic regression is based on linear regression and can predict a variety of variables, but the most

commonly used one is to predict dichotomous variables [25]. The focus of this research is to find out whether students are in danger of failing the course at the end of the semester. Therefore, we are concerned with a binary classification problem. Logistic regression in this article also refers to binary logistic regression. Logistic regression does not have a hypothetical predictive distribution. Users choose to determine the predictive variables of the model, including qualitative and quantitative processing of these variables. In linear regression, the condition of multicollinearity will negatively affect the parameters, amplify their variance, and thus affect the fit of the model [14]. Logistic regression refers to gradually adapting the probability of an event to a logistic curve. The value range of this logic curve is between 0 and 1. The logic curve is an s-shaped curve. Its characteristic is that it starts to change quickly, gradually slows down, and finally saturates.

The advantage of logistic regression is that its variables range from negative infinity to positive infinity, and the value range is limited between 0 and 1. The function used between 0 and 1 can be a probability function, so that the logistic regression function is related to a probability distribution. And the advantage of the value of the independent variable between negative infinity and positive infinity is that it can combine such signals; no matter how large or small the combination is, a probability distribution can still be obtained in the end.

**4.3. J48 Algorithm.** J48 is a decision tree algorithm implemented based on C4.5, which is a commonly used algorithm for classification problems in machine learning and data mining. The goal of this algorithm is supervised learning. In a set of data sets, each element can be represented by a set of attribute values, and each element belongs to one of several mutually exclusive categories. The C4.5 algorithm can find the mapping relationship between the attribute value and the category through the learning of the data set, and this mapping can classify the elements of the new unknown category. J48 is a nonparametric algorithm that learns rules from data. The decision tree model uses a recursive method to distinguish dependent variables of different natures in each group and gradually divide the training data into groups. In each step of this process, the partition rule selects the preend variable, divide the data file into groups, and stop when the predetermined conditions are met. The result of the learning process is a set of rules (or a tree-like representation associated with it) describing the characteristics of the predictor and the range of values that specify a given class value. This makes the decision tree highly expressive: it is very good at predicting and describing the nature of the prediction (prediction is not the result of a black box).

**4.4. SVM/SMO.** Support vector machine (or SVM) is the state of supervised learning model proposed by Vladimir Vapnik, which has become more and more popular for classification, regression, and detection problem processing [16]. SVM is particularly suitable for analyzing data with a large number of predictive attributes, so it has a considerable

influence in text classification and bioinformatics [26]. Basic SVM is based on classifying data into two categories by finding the optimal decision boundary (minimum boundary hyperplane) that is as far away as possible from the data in each category. The vector near the hyperplane is the support vector. Therefore, the basic SVM is a nonprobabilistic binary linear classifier. In order to deal with nonlinear boundaries, SVM maps data into the dimensional feature space, even if there is no easy way to separate the points in the original dimensional space, where the data points can be accurately classified or predicted. This involves using kernel functions to map data from the original space to the new feature space. SVM, which is very similar to the multilayer perceptron neural network model, does not provide the output of its predictor in the form of a function. Therefore, like neural networks, they are less expressive than other machine learning algorithms (black box methods that use more predictions).

**4.5. Bayesian.** The naive Bayes classifier is a simplified Bayesian network, a graphical model based on the concept of conditional independence, which uses a directed graph to encode the joint probability distribution of a set of variables in a concise way to describe the dependence between probability variables. The naive Bayes classifier assumes that all predictors are conditionally independent for a given class of variables. This very strong independence assumption simplifies the calculation of the likelihood of the data, reducing it to the product of the likelihood of each attribute of a given class and therefore reducing the amount of training data required to estimate model parameters. The classifier can evaluate the category of the variable and the likelihood of this category (through the conditional probability of the data of a given class). The new input instance can be assigned to the class value with the highest probability. The probability distribution of this type of classifier formed on the data of a given category can be regarded as a random generator for the data samples of the given category value. When the attributes of the predictor variables are discrete, or the variance is Gaussian distribution independent of the class, the naive Bayesian learner can be regarded as a linear classifier; that is, each such naive Bayesian corresponds to a super in the attribute space of the predictor. Plane decision boundary: although the name of the algorithm feels that the algorithm is very simple, the naive Bayes classifier shows excellent performance in many complex real-world situations.

## 5. Teaching System Data Processing Evaluation Model

The data processing and evaluation process of the teaching system are shown in Figure 3. Four classification algorithms are used to perform modeling experiments using the training sets of different sizes and the original training sets proposed in the above research, respectively, and the performance of each model is measured through the analysis results. Four classification algorithms are used to perform machine learning on the above five samples, and a total of 20

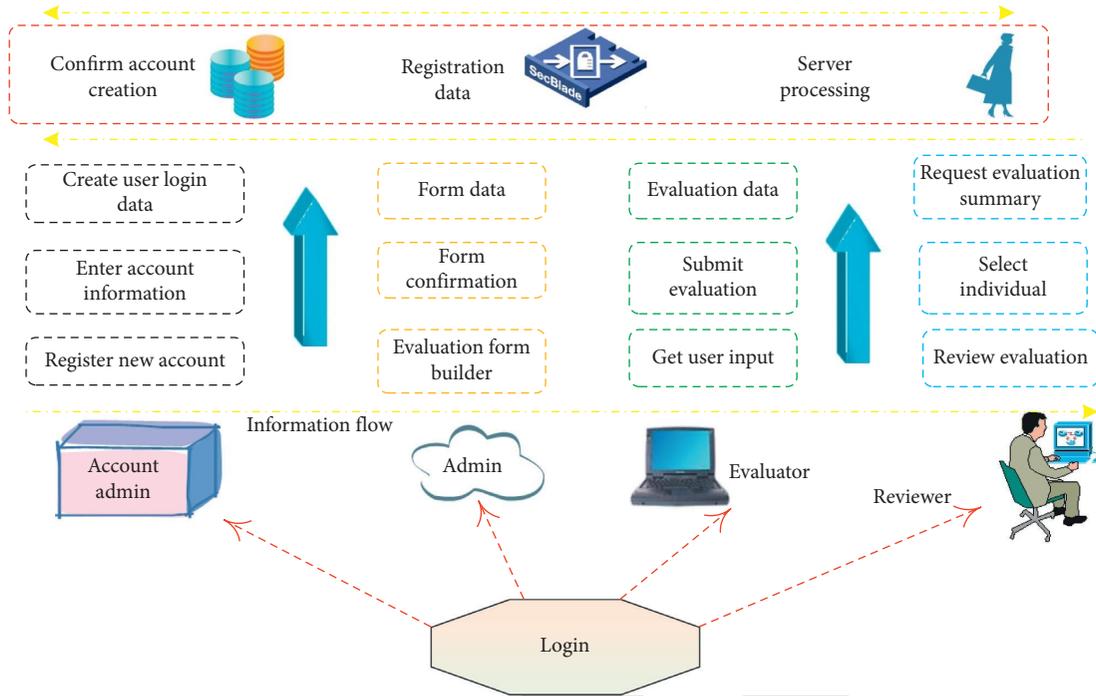


FIGURE 3: Evaluation model application process.

model results are obtained. The accuracy, FP rate, precision, and recall rate of each model are counted to obtain the following Figures. By comparing the machine learning results of the original data and the machine learning results of the sampled balanced data, the accuracy and recall rate have been greatly improved. It can be clearly seen that the sampled balanced data is compared with the unbalanced original model trained on the data which has higher accuracy and better effect. Therefore, balancing training data can help classification algorithms improve their ability to detect at-risk students.

As shown in Figure 4, comparing the accuracy, FP rate, precision, and recall rate of the four classification algorithms calculated by the model, it is found that when the overall sample size is different, including logistic regression, Bayesian, and SVM. The performance of the three classification algorithms/SMO is very stable, while the performance of J48 is very unstable, as shown in Figure 5. The reason is that the error in machine learning = deviation + variance. The deviation describes the gap between the expected value of the predicted value and the true value. The larger the deviation, the more it deviates from the real data. Variance describes the variation range of the predicted value, the degree of dispersion, that is, the distance from its expected value. The larger the variance, the more scattered the data.

Using a limited sample to estimate the infinite real data results in the bias and variance of the model not being able to have both. The reason is that if the accuracy of the model on the training samples is ensured as much as possible, the deviation of the model will be reduced. However, the flexibility of the model learned in this way is greatly impaired, resulting in overfitting, reducing the performance of

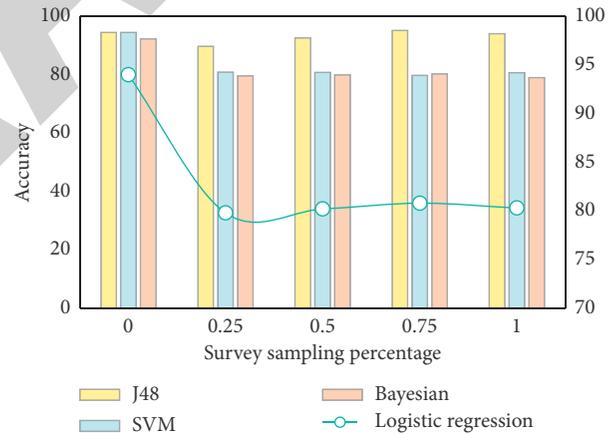


FIGURE 4: Comparison of accuracy rates.

the model on real data, and increasing the uncertainty of the model. On the contrary, if you add more restrictions to the model in the process of learning the model, so that the fluctuation of the training data will have less impact on its predictive ability, which can reduce the size of the variance, the model is more flexible, and it can fit the training well. Data improves the stability of the model, but it means that the deviation of the model becomes larger, which reduces the ability to predict new data. Therefore, low-bias learning algorithms usually have high variance and are more affected by the fluctuation of the training data set; low-variance calculations often have high bias and are affected by predicted data.

It can be seen from the performance of the four classification algorithms that logistic regression, Bayesian, and

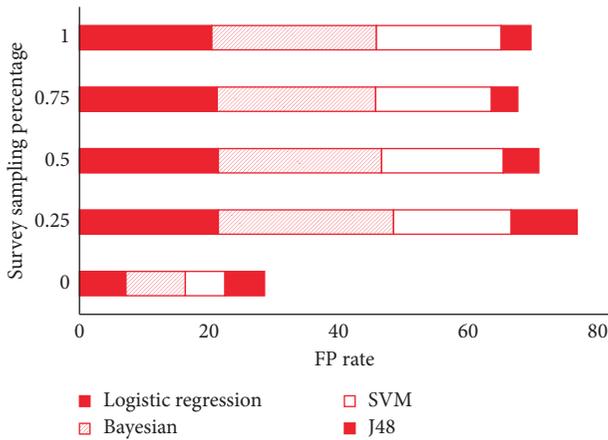


FIGURE 5: False positive rate comparison.

SVM/SMO are all types of high deviation and low variance, as shown in Figures 6 and 7. They are all linear models, and their performance is very stable under different data sets. The decision tree is a type with low deviation and high variance, which has stronger performance ability but will produce different trees according to the changes of training data.

Although the two categories of data in the four data sets (25%, 50%, 75%, and 100%) are almost balanced by sampling, when the total number of data changes, different trees will be produced, and the larger the data set, the bigger the tree. Through resampling, not only the total amount of the data set is changed, but the characteristics of the data set will be changed through oversampling of the minority class and subsampling of the majority class. Therefore, the lower the error of the generated tree, the stronger the expressiveness. The lower the predictive power, the worse the predictive power for new samples.

Figure 8 clearly shows the performance of different classification algorithms after using the student data experiment of the network school. Logistic regression, Bayesian, and SVM/SMO are stable when the total data set changes, and the index changes are very gentle. As the sample size decreases, the accuracy of the J48 algorithm decreases, and the FP rate increases. It can be seen from the information in the table that the recall rate of all classification algorithms after using the balanced data set experiment is very high, almost exceeding 80%, which is significantly higher than the ratio of the original unbalanced data after the experiment, but the FP rate is higher than the original unbalanced data. The results after the data set experiment have all increased and are relatively high. Therefore, it shows that the balanced data set generated by resampling is helpful for the classifier to detect at-risk students, but it increases the false predictions of risk-free students. In the actual operation process, it is likely that part of the risk-free students will be mistaken. It is predicted to be risky, but the ultimate goal of the experiment is to evaluate students' online learning. It is necessary to make a prediction on whether students can learn through the course, especially to monitor those students who cannot learn through the course, to discover their existing risk status in time, and to

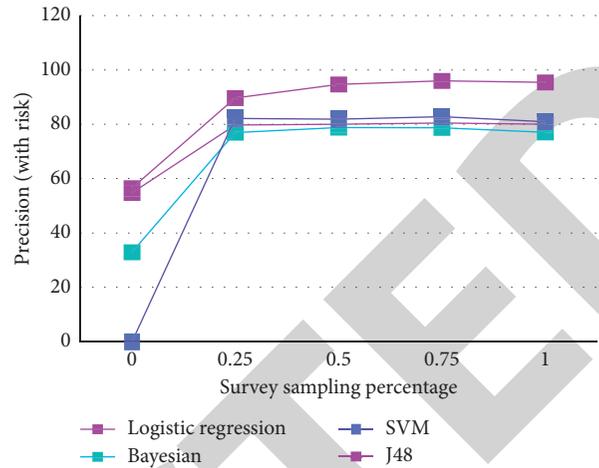


FIGURE 6: Comparison of accuracy (forecast with risk).

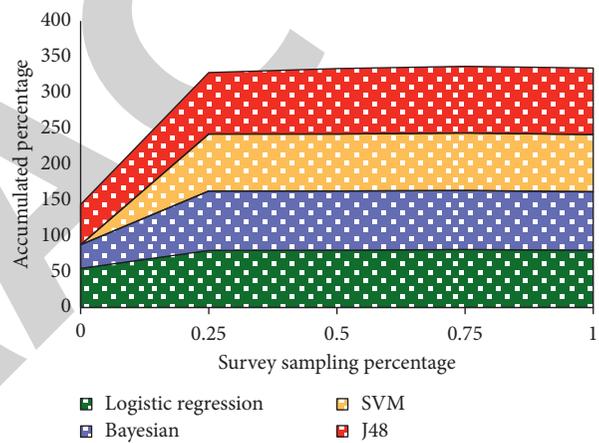


FIGURE 7: Comparison of accuracy (prediction without risk).

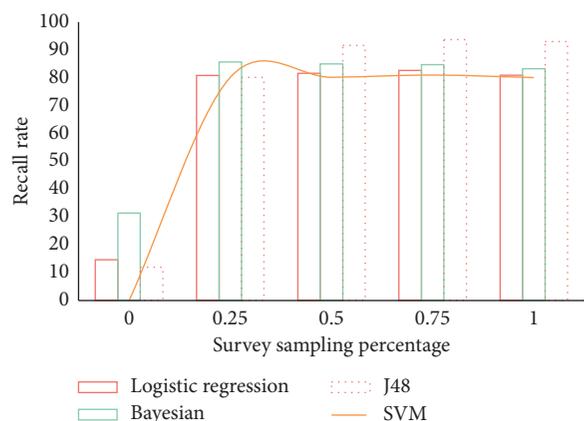


FIGURE 8: Recall ratio.

assess their risk. Make reasonable evaluations of the learning process and give timely interventions so that students can learn more efficiently. Therefore, if you cannot make good predictions for at-risk students, the effect of the model will be greatly reduced. By comparing the four learning

algorithms, logistic regression and SVM/SMO have relatively good performance; both have relatively good stability, and high recall rate, and low FP rate. Logistic regression has relatively high stability and relatively accurate predictions for at-risk students.

## 6. Conclusion

It can be seen from the performance of the four classification algorithms that logistic regression, Bayesian, and SVM/SMO are all types of high deviation and low variance. The decision tree is a type with low deviation and high variance, which has stronger performance ability but will produce different trees according to the changes of training data. The experimental results are more obvious. The performance of different classification algorithms can be seen after using student data. Logistic regression, Bayesian, and SVM/SMO are stable when the total amount of data set changes, and the index changes very smoothly; the J48 algorithm follows the smaller the sample size, the lower the accuracy rate, and the higher the FP rate.

With the advent of the era of big data, education is gradually applying big data to the industry. In the educational revolution in the era of big data, researchers and educators try to use big data to optimize teaching and apply it to all aspects of teaching. The artificial intelligence teaching system based on big data provides a learning environment for students to learn independently. The characteristic of the artificial intelligence teaching system is that it can allow students to study without time and space. However, online learning cannot make teachers better understand students. The characteristics of artificial intelligence teaching lead to the separation of teachers and students, and a teacher faces a large number of students. There is no way to pay attention to the learning progress of each student. Teachers cannot be like traditional teaching. Observe the student's learning process in the same way. Teachers hope to understand the learning situation of each student and give them targeted guidance so that students can improve their learning efficiency. Therefore, to solve this problem, it is necessary to use big data analysis methods in online learning. By collecting students' learning data, the students' learning situation can be analyzed and evaluated. Teachers can understand the students' learning situation through evaluation and give targeted guidance to find problems in teaching in time, improve teaching, and improve student learning efficiency. Artificial intelligence teaching evaluation can help educators understand the problems in teaching, adjust teaching strategies in time, and improve teaching effects. Online learning evaluation is the basis for providing students with adaptive teaching and can provide teachers with a basis for personalized teaching. However, online learning is a continuous, complex, and diverse behavior which requires a more in-depth study of online learning behavior in order to make an evaluation of artificial intelligence teaching suitable for teaching. At present, the research scope and applicable scope of artificial intelligence teaching evaluation are very wide. In order to optimize the teaching process, it is necessary to do more in-depth research on online learning evaluation.

## Data Availability

Data sharing does not apply to this article because no data set was generated or analyzed during the current research period.

## Consent

Informed consent was obtained from all individual participants included in the study references.

## Conflicts of Interest

The authors declare that there re no conflicts of interest.

## Acknowledgments

The study was supported by "Research on the Integration and Construction of Teaching Theory and Academic Discourse System in the Development of Socialism Education with Chinese Characteristics, Shaanxi Social Science Foundation, China (Grant no. 2019Q038)".

## References

- [1] Z. Allam and Z. A. Dhunny, "On big data, artificial intelligence and smart cities," *Cities*, vol. 89, no. 8, pp. 80–91, 2019.
- [2] Y. Lu, "Artificial intelligence: a survey on evolution, models, applications and future trends," *Journal of Management Analytics*, vol. 6, no. 1, pp. 1–29, 2019.
- [3] B. Sharma, "Processing of data and analysis," *Biostatistics and Epidemiology International Journal*, vol. 1, no. 1, pp. 3–5, 2018.
- [4] S. A. Popenici and S. Kerr, "Exploring the impact of artificial intelligence on teaching and learning in higher education," *Research and Practice in Technology Enhanced Learning*, vol. 12, no. 1, pp. 1–3, 2017.
- [5] K. Colchester, H. Hagra, D. Alghazzawi, and G. Aldabbagh, "A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 7, no. 1, pp. 47–64, 2017.
- [6] J. Keating and I. Nourbakhsh, "Teaching artificial intelligence and humanity," *Communications of the ACM*, vol. 61, no. 2, pp. 29–32, 2018.
- [7] K. Masters, "Artificial intelligence in medical education," *Medical Teacher*, vol. 41, no. 9, pp. 976–980, 2019.
- [8] A. P. Ikedinachi, S. Misra, P. A. Assibong, E. F. Olu-Owolabi, R. Maskeliūnas, and R. Damasevicius, "Artificial intelligence, smart classrooms and online education in the 21<sup>st</sup> century: implications for human development," *Journal of Cases on Information Technology (JCIT)*, vol. 21, no. 3, pp. 66–79, 2019.
- [9] R. Williams, H. W. Park, L. Oh, and C. Breazeal, "PopBots: designing an artificial intelligence curriculum for early childhood education," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 9729–9736, 2019.
- [10] E. Burton, J. Goldsmith, S. Koenig, B. Kuipers, N. Mattei, and T. Walsh, "Ethical considerations in artificial intelligence courses," *AI Magazine*, vol. 38, no. 2, pp. 22–34, 2017.
- [11] Y. Lu, "Artificial intelligence: a survey on evolution, models, applications and future trends," *Journal of Management Analytics*, vol. 6, no. 1, pp. 1–29, 2019.

- [12] D. Schiff, "Out of the laboratory and into the classroom: the future of artificial intelligence in education," *Ai & Society*, vol. 36, no. 1, pp. 331–348, 2021.
- [13] A. K. Goel and D. A. Joyner, "Using AI to teach AI: lessons from an online AI class," *AI Magazine*, vol. 38, no. 2, pp. 48–59, 2017.
- [14] N. N. Dyachenko, M. P. Bondarenko, and V. Pustovit, "Intelligent and cognitive technologies in education of International economic relations students and human resource development: methodology in language teaching and distance learning," *European Journal of Sustainable Development*, vol. 6, no. 4, pp. 353–355, 2017.
- [15] E. V. Polyakov, M. S. Mazhanov, A. Y. Rolich, L. S. Voskov, M. V. Kachalova, and S. V. Polyakov, "Investigation and development of the intelligent voice assistant for the internet of things using machine learning," *Electronic and Networking Technologies (MWENT)*, vol. 18, no. 4, pp. 1–5, 2018.
- [16] M. Syafrudin, G. Alfian, N. Fitriyani, and J. Rhee, "Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing," *Sensors*, vol. 18, no. 9, p. 2946, 2018.
- [17] I. Kotenko, I. Saenko, and A. Branitskiy, "Framework for mobile internet of things security monitoring based on big data processing and machine learning," *IEEE Access*, vol. 6, no. 6, pp. 72714–72723, 2018.
- [18] G. Manogaran, V. Vijayakumar, R. Varatharajan, P. Malarvizhi Kumar, R. Sundarasekar, and C.-H. Hsu, "Machine learning based big data processing framework for cancer diagnosis using hidden Markov model and GM clustering," *Wireless Personal Communications*, vol. 102, no. 3, pp. 2099–2116, 2018.
- [19] H. Wang, Z. Xu, and W. Pedrycz, "An overview on the roles of fuzzy set techniques in big data processing: trends, challenges and opportunities," *Knowledge-Based Systems*, vol. 118, no. 15, pp. 15–30, 2017.
- [20] N. Gao, D. Luo, B. Cheng, and H. Hou, "Teaching-learning-based optimization of a composite metastructure in the 0–10 kHz broadband sound absorption range," *The Journal of the Acoustical Society of America*, vol. 148, no. 2, pp. EL125–EL129, 2020.
- [21] Y. Zhang, L. Sun, H. Song, and X. Cao, "Ubiquitous WSN for healthcare: recent advances and future prospects," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 311–318, 2014.
- [22] W. Li and H. Song, "ART: an attack-resistant trust management scheme for securing vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 960–969, 2015.
- [23] J. Yang, C. Wang, B. Jiang et al., "Visual perception enabled industry intelligence: state of the art, challenges and prospects," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 2204–2219, 2020.
- [24] K. Sim, J. Yang, W. Lu, and X. Gao, "MaD-DLS: mean and deviation of deep and local similarity for image quality assessment," *IEEE Transactions on Multimedia*, p. 1, 2020.
- [25] M. Chassignol, A. Khoroshavin, A. Klimova, and A. Bilyatdinova, "Artificial Intelligence trends in education: a narrative overview," *Procedia Computer Science*, vol. 136, no. 36, pp. 16–24, 2018.
- [26] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education—where are the educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1–27, 2019.