IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 18, 2022, accepted February 10, 2022, date of publication February 15, 2022, date of current version February 24, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3151652

Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining

GUIYUN FENG, MUWEI FAN, AND YU CHEN
School of Management, Guizhou University, Guiyang 550025, China

Corresponding author: Muwei Fan (mwfan@gzu.edu.cn)

This work was supported in part by the Construction Project of 2017 First-Category Discipline in Guizhou Province: Big Data Discipline Group under Grant GNYL[2017]005.

ABSTRACT The development of intelligent technologies gains popularity in the education field. The rapid growth of educational data indicates traditional processing methods may have limitations and distortion. Therefore, reconstructing the research technology of data mining in the education field has become increasingly prominent. In order to avoid unreasonable evaluation results and monitor the students' future performance in advance, this paper comprehensively uses the relevant theories of clustering, discrimination and convolution neural network to analyze and predict students' academic performance. Firstly, this paper proposes that the clustering-number determination is optimized by using a statistic which has never been used in the algorithm of K-means. Then, the clustering effect of K-means algorithm is tested by discriminant analysis. The convolutional neural network is introduced for training and testing data that are labeled with categories. The generated model can be used to predict prospective performance. Finally, in order to validate the prediction results, the effectiveness of the generated model is evaluated by using two metrics in two cross-validation methods. The experimental result demonstrates that the statistic not only solves the difficulty to determine the clustering number in K-means algorithm from an objective and quantitative point of view, but also improves the reliability of prediction results.

INDEX TERMS Academic performance, clustering analysis, convolutional neural networks, discriminant analysis, educational data mining.

I. INTRODUCTION

Data mining (DM) can discover hidden information in a large amount of unordered data. Educational data mining (EDM) is a research field of data mining, which focuses on the application of data mining, machine learning and statistical methods. In the past few decades, the application of data mining technology in the educational environment has become an active research field. Due to the availability of online datasets and learning systems, it has gained wide popularity in recent years [1]. EDM involves the development and realization of data mining methods, which facilitate the analysis of massive data from various educational backgrounds. Higher education institutions regard students' academic performance as one of the most important criteria. Consequently, predicting learning process and analyzing students' performance are considered to be significant tasks in the field of EDM [2]. EDM is a continuously evolving subject which focuses on the improvement of self-learning and adaptive methods to reveal hidden patterns or internal relationship of educational data. In the field of education, heterogeneous data is participating and growing in the paradigm of big data. In order to adaptively extract meaningful information from massive educational data, some specific data mining techniques are needed [3]. Because technologies which are relevant to data mining make massive student information used to study valuable patterns of student learning behavior possible, the application research of EDM develops rapidly. Data mining methods have been applied in many aspects of educational data processing, such as student retention, dropout prediction, academic data analysis, and student behavior analysis [4]. The evaluation and prediction of student academic performance have always been important parts in EDM.

The final exam scores of college students reflect the students' learning effects to some extent, but the evaluation of

The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai.

learning effects cannot only be based on absolute scores. The traditional absolute score has certain limitations in reflecting the learning situation. The reasons are that difficulty of different courses is different, the marking standards of different teachers in the same course are different, and so on. In order to ensure the quality of talents, colleges and universities should not only judge the students by scores, but also analyze the learning effects of students, predict the academic performance of students in the future based on the analyzed results, and then set academic warnings in time. This work will not only help colleges and universities to improve the quality of education, but also help students improve their overall performance, thereby improving the management of educational resources.

The research problem of this paper is to objectively evaluate students' academic achievement from the perspective of clustering, and predict the future achievement based on the existing achievement. To achieve the above-mentioned goals, this study formulates the following two subdivided research problems:

1) How to determine the clustering number objectively in order to evaluate academic performance?
2) How to train unlabeled data with the implementation of constructing prediction model?

Therefore, the main contributions of this paper are:

1) Improve the traditional K-means algorithm. Using objective and quantitative analysis instead of subjective evaluation to determine the value of k makes the academic performance evaluation results obtained by clustering more convincing. Furthermore, the models and prediction results trained by deep learning algorithm are more reliable.
2) The deep learning algorithm trains data with category labels which the metadata do not have. Clustering is used to add category labels to the metadata, and then deep learning algorithm is used to train the data, which provides a new idea for deep learning without feature labels.

The remaining sections of this paper are organized as follows. Section II reviews the literature that are relevant to the research of analysis and prediction in EDM and the application of convolutional neural network (CNN). Section III demonstrates the research methods and theory. Section IV discusses the experimental design and results, and analyzes the reliability of the prediction. Finally, we summarize the full text, highlight our limitations and introduce the research directions of EDM in the future in Section V.

II. RELATED WORK

Many scholars pointed out the application of data mining technology in the field of education, and highlighted the importance of EDM. Angeli *et al.* [5] illustrated how educational technicians use data mining to guide and monitor school-based technical integration work. They discuss that the significance of the research lies in the necessity of developing EDM tools which can be used in meaningful ways and users can display results and suggestions. Javier *et al.* [6] pointed out that EDM combines data mining technology with educational data, and summarize the most commonly used data mining methods: factor analysis, regression, correlation mining and so on. Romero *et al.* [7], [8] believed that the goal of EDM is to better understand how students learn, and it is also a method developed to explore the unique data types from the educational environment, which can be defined as the application of data mining technology. In order to optimize the education system, Wang [9] used data mining algorithms to classify and summarize educational data. Starting from the process of EDM to the algorithm process and evaluation model, he proposes an optimization plan. Through effective evaluation of the development of smart education, a comprehensive evaluation model has been verified the high efficiency of intelligent education and the accuracy of data mining algorithms. In this section, we will review the literature on the application of data mining technology in the analysis and prediction of academic performance.

A. RESEARCH ON ANALYSIS IN EDM

Some scholars apply clustering and other data mining techniques to the analysis of academic performance.

James *et al.* [10] pointed out that currently some scholars are using clustering analysis to analyze student performance and distinguishing students' categories based on their performance. They add a K-means clustering algorithm combined with a deterministic model to analyze student performance. Ani *et al.* [11] used entropy to study the degree of confidence changes over time, and use student data from eight courses to determine the student category based on the degree of student confidence. Moises *et al.* [12] used a clustering algorithm to detect six different student groups, analyze the interaction patterns of each category, and find that these patterns would be repeated in the early stages of the course. John *et al.* [13] used log data to identify the online behavior patterns of a group of students, and determine the student's category by using clustering analysis. Karthikeyan *et al.* [14] developed a novel approach called hybrid educational data mining model for analyzing the student performance for effectively enhancing the educational quality for students. Crivei *et al.* [15] investigated the usefulness of unsupervised machine learning methods, particularly principal component analysis and relational association rule mining in analyzing students' academic performance data. Delgado *et al.* [16] used a new unsupervised clustering technology based on self-organizing mapping which performs accurate and diversified user clustering according to student behavior records. Okoye *et al.* [17] proposed an educational process and data mining plus machine learning (EPDM + ML) model applied to contextually analyze the teachers' performance and recommendations based on data derived from students' evaluation of teaching. Kumar *et al.* [18] developed a model called Multi-Tier Student Performance Evaluation Model (MTSPEM) using single and ensemble classifiers.

The student data from higher educational institutions are obtained and evaluated in this model based on significant factors that impacts greater manner in students' performance and results. Duhayyim *et al*. [19] presented an improved evolutionary algorithm based feature subsets election with neuro-fuzzy classification (IEAFSS-NFC) for data mining in the education sector. The presented IEAFSS-NFC model involves data pre-processing, feature selection, and classification.

The analysis in EDM is mostly reflected in the evaluation of students' performance.

B. RESEARCH ON PREDICTION IN EDM

Some scholars apply CNN and other data mining techniques to the prediction of academic performance.

Agaoglu [20] used four classification techniques to predict instructors' performance according to students' evaluation of courses. Qiu *et al*. [21] proposed an integrated framework with feature selection (FSPred) to predict the dropout in MOOCs, which includes feature generation, feature selection, and dropout prediction. Akram *et al*. [22] presented an algorithm called students' academic performance enhancement through homework late/non-submission detection (SAPE) for predicting students' academic performance. Based on the characteristics of MOOCs learning, Wen *et al*. [23] proposed a new simple feature matrix to maintain the information related to the local correlation of learning behavior, and propose a new CNN model to predict the dropout rate. Lin *et al*. [24] proposed a continuous facial emotion pattern recognition method based on deep learning to analyze students' learning emotions. This method combines CNNs and long-short memory networks for deep learning to identify and analyze students' continuous facial emotions and predict academic emotions. Farissi *et al*. [25] proposed a method based on genetic algorithm feature selection technique with classification method in order to predict student academic performance. Turabieh *et al*. [26] proposed a modified version of Harris Hawks Optimization (HHO) algorithm by controlling the population diversity. The proposed approach is employed as a feature selection algorithm to discover the most valuable features for student performance prediction problem. Ma *et al*. [27] proposed a new perspective called progressive imitation learning to train a lightweight CNN model by imitating the learning trajectory of the teacher model to construct a prediction model. Nabil *et al*. [28] explored the efficiency of deep learning in the field of EDM, especially in predicting students' academic performance in order to identify students at risk of failure. Gao *et al*. [29] proposed a deep cognitive diagnosis framework to obtain students' mastery of skills and solve problems by enhancing traditional cognitive diagnosis methods with deep learning.

The prediction in EDM is mostly reflected in students' achievement, dropout rate, psychological problems and so on.

C. SUMMARY

The ever-increasing amount of educational data makes educators, education-related personnel and even the general public focus on the important field of turning massive disorderly educational data into useful information. Whether it is to predict student performance and dropout rate, or evaluate student academic performance, or evaluate teacher performance, data mining technology is widely used in all aspects. EDM researchers usually use automated methods such as data mining and machine learning to explore educational data, which is confirmed on the basis of the previous literature review. Combining computer technology with educational big data, knowledge of subjects such as mathematics, engineering, pedagogy and psychology can also be applied to EDM.

From the literature review, it can be found that EDM models can be divided into two types: descriptive models and predictive models. Descriptive models are used to describe models and provide reference for decision-making, whereas predictive models are mainly used for data-based prediction. The former is mostly used to evaluate students' academic performance and provide a reference for teaching managers to make decisions, whereas the latter is mostly used to predict students' academic performance, help prevent the risk of dropout, and improve students' academic performance. This paper intends to use a combination of descriptive and predictive models. This study will use clustering to analyze students' academic performance and convolutional neural network to predict students' future academic performance. Table 1 presents a comparison of the study with the work mentioned in recent three years. The differences between our method and existing ones are as follows:

1) The method for determining clustering number is different from others.
2) We use the clustering results as the label of the data to train the data.
3) We combine analysis and prediction.
4) We combine clustering algorithm and classification algorithm.

III. METHODOLOGY

A. DATA PREPROCESSING

The study selects three datasets A, B, and C in a university in which students in the same dataset have the same courses for analysis. The global objective of data preprocessing is to remove the unwanted variability or effects so that the useful information related to the properties can be used for efficient modelling [30]. Data preprocessing serves as the foundation for valid data which is an indispensable step in building operational data analysis considering the intrinsic complexity of data quality [31]. Data preprocessing refers to a series of necessary cleaning, integration, transformation and reduction of the original data before data mining, so as to achieve the minimum specifications and standards required by algorithms for knowledge acquisition research.

TABLE 1. Comparison of the study with the work mentioned in recent three years.

Reference	Year	Dataset	Method	Objective
Karthikeyan *et al.* [14]	2020	the benchmark education dataset	naive bayes, J48	analysis
Crivei *et al.* [15]	2020	a real academic data set	principal component analysis, association rule mining	analysis
Farissi *et al.* [25]	2020	kaggle repository datasets	genetic algorithm	prediction
Okoye *et al.* [17]	2021	data collected from students' evaluation of teaching	text mining, *k*-nearest neighbors	analysis
Kumar *et al.* [18]	2021	a benchmark student data set	naive bayes, random forest	analysis
Turabieh *et al.* [26]	2021	a real dataset obtained from UCI	harris hawks optimization	prediction
Nabil *et al.* [28]	2021	a dataset collected from a public 4-year university	deep neural network	prediction
Duhayyim *et al.* [19]	2022	a benchmark student performance dataset from UCI	an improved evolutionary algorithm based feature subsets election with neuro-fuzzy classification, chaotic whale optimization algorithm	analysis
Gao *et al.* [29]	2022	two real-world data sets	a deep cognitive diagnosis framework	prediction
Present Study	2022	three datasets in a university	improved K-means, CNN	analysis and prediction

1) DATA CLEANING

Data cleaning refers to the elimination of incomplete, missing or duplicate data. There are many ways to fill in missing values for attributes, such as ignoring tuples, using a global constant to fill in missing values, using the mean of attributes to fill in missing values, etc. Delete the grade records of the courses with more missing courses, and fill in the grade records of the courses with fewer missing courses. This paper follows the following principles: Delete the score records with empty scores in more than two courses, and if there are still students whose course scores are empty, fill it with the average value of the course. It is understood that a course with a score of 0 is a student's absence from the exam, and the corresponding student's score record is deleted.

2) DATA INTEGRATION

In order to solve data redundancy, it is necessary to merge related courses. Since some courses are divided into several semesters, merging these courses and taking the average score of several semesters as the score of the course is conducive to reducing the characteristics in the process of subsequent analysis. After merging, datasets A, B, and C have 9, 13, and 13 courses respectively for analysis.

3) DATA TRANSFORMATION

The original score data are presented in the form of a percentile system, with no difference of order of magnitude, and no standardized operation is required. The K-means algorithm is only suitable for processing numerical data. When the data for analysis is combined into a table, in addition to setting the student number to character type, the data type of each subject score is converted to numerical type, and decimal places are set to 0.

4) DATA REDUCTION

In order to ensure the simplicity of the presentation results, irrelevant attributes such as credits and class time are omitted, and only the student number and the corresponding score of each course are retained. In order to ensure the statistical convenience of the final clustered cases, a student record corresponds to a serial number. Datasets A, B, and C have 52, 66, and 51 original score records, respectively. After preprocessing, there are 46, 61, and 51 score records, respectively.

B. CLUSTERING ANALYSIS AND K-MEANS ALGORITHM

Clustering is one of the most common unsupervised data mining methods. Objects with similar characteristic attributes are placed in a category, and the characteristic attributes of the objects in different categories are different. The purpose of clustering analysis is simply to find a convenient and efficient way to organize data, not to establish rules for classifying future data [32]. Classification is to classify samples of unknown categories into a certain category according to the classification criteria that have been determined in advance, whereas clustering is to find features from the data and then classify them according to the features. The clustering results are closer to the actual situation, making the clustering analysis widely used in various areas.

K-means algorithm, also known as a kind of fast clustering method, was proposed by Macqueen in 1967. K-means can maintain good scalability and efficiency when dealing with datasets, with strong local search ability and fast convergence speed. However, due to the sensitivity of K-means to outliers and initial clustering centers, the number of data points in different clusters may vary greatly, and it is often not easy to obtain ideal clustering results [33]. The pseudocode of K-means is provided in Algorithm 1. The procedures of the K-means algorithm:

1) Arbitrarily select k samples from n samples as the initial clustering centers, and the initial clustering center is randomly determined.
2) Assign all other sample to the nearest clustering center.
3) Calculate the clustering center of each cluster, and Euclidean distance is used as the formula for calculating distance.

Algorithm 1 Clustering for Analyzing Academic Performance

Require: parameter k, inputs $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$

initialize clustering centers $c_1, c_2, \ldots, c_k$
repeat
for $i = 1 : n$ **do**
  label the input x_i as belonging to the nearest cluster, $y_i := \underset{j=1,2,\ldots,k}{\arg\min} \left\|x_i - c_j\right\|^2$
end for
for $i = 1 : \sqrt{n}$ **do**
  compute clustering center c_j as the mean of all inputs of the jth cluster, $c_j := mean(\{x_i : y_i = j\})$
end for
until no change in clustering centers between subsequent iterations
return clustering centers $c_1, c_2, \ldots, c_k$

4) Repeat 2), 3), until the k clustering center no longer change.

Rezaee *et al.* [34] pointed out that the optimal value of k is within the range of 1 to $\sqrt{n}$. If the k value is selected appropriately, the sum of squared deviations between samples within the same category should be small, and the sum of squared deviations between categories should be large. In order to solve the problem that clustering number is difficult to determine, this paper proposes to use a statistic R_k^2 to determine the value of k, and the calculation equation is

$$R_k^2 = 1 - \frac{\sum_{p=1}^{k}\sum_{i=1}^{j} d_{ip}^2}{\sum_{p=1}^{k}\sum_{i=1}^{j} d_{ip}^2 + \sum_{p=1}^{k} d_p^2} \tag{1}$$

In (1), d_{ip} represents the distance between the object in the p-th category and the clustering center of the p-th category under the premise of clustering into k categories, and the numerator represents the sum of squared deviations within the category. The denominator represents the sum of the sum of squared deviations within the category and the sum of squared deviations between categories. If R_k^2 is relatively large, it means that the sum of squared deviations within the category is relatively small and the sum of squared deviations between categories is relatively large when divided into k categories, so it is appropriate to divide into k categories. However, the more categories there are, the smaller the sum of squared deviations within each category, and the larger R_k^2 is. Therefore, we can only choose an appropriate k to make R_k^2 large enough, and as the value of k increases, the increase in R_k^2 is relatively small.

We use the traditional K-means algorithm to cluster students. With the implementation of solving the problem that it is difficult to determine clustering number, the proposed statistic is used to determine the clustering number, and then evaluate and analyze the academic performance.

C. DISCRIMINANT ANALYSIS AND BAYES DISCRIMINATION

Discriminant analysis is a statistical analysis method to determine the category of samples. According to different discriminant criteria, discriminant analysis can be divided into distance discrimination, Fisher discrimination, Bayes discrimination and so on. Bayes discriminant method is based on the prior probability of the population to minimize the average loss of misjudgment [35]. As the clustering number increases, the number of Fisher discrimination functions will increase. Although the distance discrimination method is intuitive and easy to understand, it does not consider the prior probability of each category. In order to avoid the above problems, the Bayes discrimination is selected to test the clustering effect. The idea of Bayes discrimination is to obtain the posterior probability distribution according to the prior probability of the sample. At this time the sample is judged to be the category with the largest posterior probability.

Suppose k populations $G_1, G_2, \ldots, G_k$ have meta-probability density functions $f_1(x), f_2(x), \ldots, f_k(x)$, respectively. The prior probability of k populations is known to be $q_1, q_2, \ldots, q_k$, the loss of the sample from the population G_t is $L(j/i)(i, j = 1, 2, 3, \ldots, k)$, and it is specified $L(j/i) = 0$. For any sample x, it can be calculated in turn:

$$D_\alpha(x) = \sum_{j=i}^{k} q_j f_i(x) L(\alpha/j) \quad (\alpha = 1, 2, \ldots, k) \tag{2}$$

According to (2), it can be determined the Bayes criterion: if $D_t(x) = \min_{1\le\alpha\le k} D_\alpha(x)$, then judge $x \in G_k$. In particular, if the loss function $L(j/i)$ is the same when $i \neq j$, the Bayes criterion is: if $q_i f_i(x) = \max_{1\le\alpha\le k} q_\alpha f_\alpha(x)$, then judge $x \in G_k$.

D. DEEP LEARNING AND CNN

Hinton proposed a deep belief network in 2006. The neural network completed the transition from shallow to deep, and the concept of deep learning was born. It has turned out to be very good at discovering intricate structures in high-dimensional data [36]. Deep learning is a general term for a category of methods developed by machine learning, which refers to the laws in data, especially the classification laws. The filtering operation performed by the feature map is discrete convolution, hence the name of the convolutional neural network. CNN is a deep neural network learning algorithm proposed by LeCun *et al.* [36], which is a supervised deep model architecture. The characteristics of local connection, weight sharing and pooling operation of CNN can effectively reduce the complexity of the network, reduce the number of training parameters, have strong robustness and fault tolerance, and are easy to train. CNN is designed to process data that come in the form of multiple

TABLE 2. Confusion matrix.

Actual Category	Prediction Category	
	Positive	Negative
Positive	true positive	false negtive
Negative	false positive	true negative

arrays [36]. A trained CNN can reliably identify objects without requiring specific morphologic parameters [37]. Therefore, it is suitable for classifying student samples.

The basic structure of a convolutional neural network consists of five layers:

1) The number of input neurons in the input layer depends on the number of features, and each course is used as the feature input.
2) In the convolutional layer, some courses have little influence on the classification results of students, and some courses have a great degree of distinction among students. In short, the function of the layer is to extract more abstract features to make the classification results more accurate.
3) The role of the pooling layer is to reduce the size of the data space, thereby reducing the number of parameters in the entire neural network and preventing over-fitting, and the layer also has a certain anti-noise ability.
4) After convolution and pooling operations are superimposed layer by layer, the classification result is given through the fully connected layer.
5) The number of output neurons in the output layer depends on the classification result. The accuracy of the results and the loss caused by the error can be obtained by comparing the output classification results with the original labels.

E. METRICS FOR PERFORMANCE EVALUATION

The loss caused by error and accuracy are used as the metrics of evaluation results. Overall accuracy is often used to measure model performance. The confusion matrix is shown in Table 2. The calculation equation of accuracy is

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

where "positive" denotes a certain category, and "negative" means other categories. True positive (TP) denotes that the actual category of a student is this category, and the prediction category is also this category. True negative (TN) means that the actual category of a student is not this category, nor is the prediction category. False positive (FP) is that the actual category of a student is not this category, but it is predicted to be this category. False negative (FN) indicates that the actual category of a student is this category, but the prediction is not this category.

F. CROSS VALIDATION

Validation is an important phase in building the predictive model. It analyzes how realistic the predictive model is. [38]

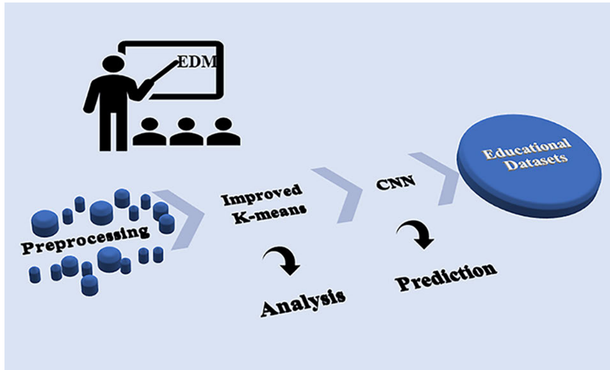


FIGURE 1. We use clustering algorithm named improved K-means which is applied to analyze students' academic performance and provide a reference for teaching managers to make decisions and classification algorithm named CNN which is used to predict students' future academic performance and help prevent the risk of dropout.

Cross-validation is a model validation technique applied to evaluate how the statistical analysis results are generalized into an independent dataset [39]. We prepare to use two methods for model validation. One is random hold-out method and the other is x-fold cross-validation method.

1) RANDOM HOLD-OUT METHOD

The original data are randomly divided into two groups, one as the training set and the other as the test set. The training set is used to train the classifier, and then the test set is used to verify the model, and the final classification accuracy is recorded as the performance index of the classifier. 70% for training and 30% for testing is used in [20], [40]. 80% for training and 20% for testing is used in [28], [39], [41].

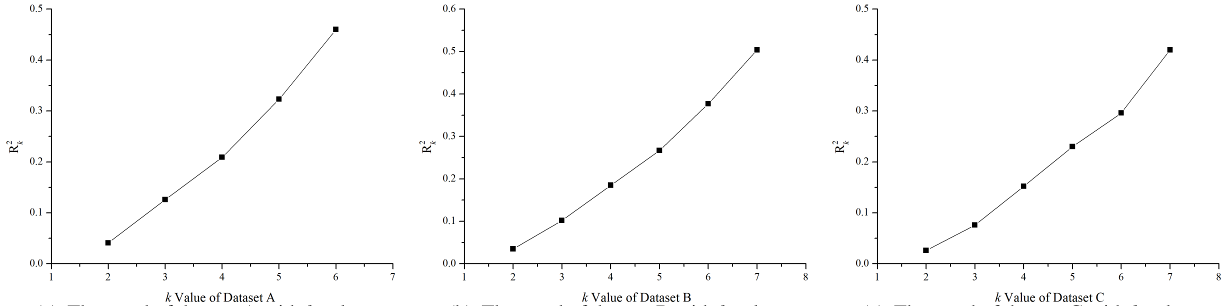
2) SHUFFLE X-FOLD CROSS-VALIDATION METHOD

The data are randomly divided into x parts, and x experiments are carried out respectively. (x-1) of them are used as training data each time, and the remaining one is used as test data. Finally, the average value of each index in the x experiments is taken as the final classification result. 5-fold cross validation is used in [28], [39]. 10-fold cross validation is used in [21], [22], [38], [42].

Using the above methods, the research content of this paper ranges from data preprocessing, clustering and discrimination to evaluating academic performance, and then to convolution neural network training data with category labels, and using metrics to evaluate the prediction results. The simple process of EDM is shown in Fig. 1.

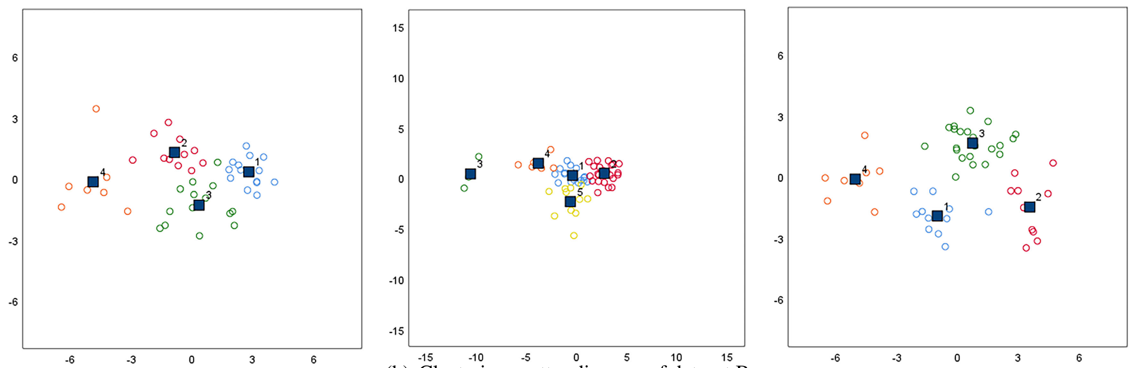
IV. EXPERIMENTAL RESULTS AND DISCUSSION

To ensure the universality of the analysis results, the scores of make-up exams and retakes were not included. There is a problem of missing marks in elective courses. If the results of elective courses are included in the model, it will seriously affect the performance of the algorithm and model. Therefore, only compulsory course data will be analyzed.



(a) The trend of dataset A with k value (b) The trend of dataset B with k value (c) The trend of dataset C with k value

FIGURE 2. The trend of each dataset with clustering number.



(a) Clustering scatter diagram of dataset A (b) Clustering scatter diagram of dataset B (c) Clustering scatter diagram of dataset C

FIGURE 3. Clustering scatter diagrams of three datasets.

A. DETERMINING k VALUE

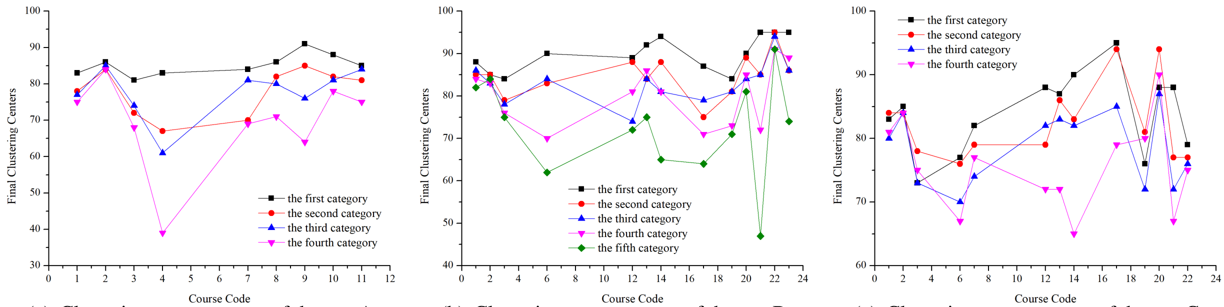
Use the K-means algorithm to cluster the preprocessed data. At this time, the intra-category distance and the inter-category distance corresponding to different clustering numbers can be obtained, and the corresponding statistics can be calculated. The trend of each dataset with clustering number is shown in Fig. 2. According to the rules for determining clustering number and comprehensive consideration of the actual situation, it is more appropriate to cluster the datasets A, B, and C into four, five, and four categories, respectively. Taking dataset A as an example, as the value of the clustering number increases, the value of the statistic also increases. However, when $k = 4$, the growth of the statistic decreases, so it is more appropriate for the k value in dataset A to be 4.

B. CLUSTERING ANALYSIS

The clustering scatter diagrams of datasets A, B, and C are shown in Fig. 3. The square box is the clustering center, and the objects in this category surround the clustering center. The overall clustering results of datasets A and C are relatively scattered, and the objects in the category and their corresponding clustering centers are also relatively scattered, whereas the overall clustering results of dataset B and the objects in the category and their corresponding clustering centers are relatively close, so the boundaries among categories are not very obvious. The clustering results of datasets A and C are relatively scattered, and the number of objects in each category is not much different, but the number of objects in each category of dataset B with closer clustering results is relatively large. Some clusters have only single digits, but there are more objects in other clusters.

The clustering center curves of datasets A, B, and C are shown in Fig. 4. Compared with other courses, the degree of difference in courses 1, 2, and 3 which three datasets all have is relatively small. The higher the curve position, the better the academic performance of the students in this category. The fourth category of students is obviously inferior to the other three categories of students in the overall academic performance of dataset A, the fifth category of students is obviously inferior to the other four categories of students in the overall academic performance of dataset B, and the fourth category of students is obviously inferior to the other three categories of students in the overall academic performance of dataset C. The current grades of students in these categories are lagging behind, which should arouse the focus of teachers' attention and should be set early warning.

The course 4 in dataset A, the course 21 in dataset B, and the course 14 in dataset C have the largest difference



(a) Clustering center curves of dataset A (b) Clustering center curves of dataset B (c) Clustering center curves of dataset C

FIGURE 4. **Clustering center curves of three datasets.**

among all the courses of the corresponding datasets. Course 4 is a basic course. Courses 21 and 14 are professional courses. Different students have different weak courses. The corresponding teacher should adjust the teaching plan in time to achieve the purpose of teaching students in accordance with their aptitude.

The values of clustering centers of three datasets are quite different. The clustering center curves range from 30 to 100, 40 to 100, and 60 to 100, respectively. Dataset A clustering center has the largest floating range. The floating range of the clustering center of dataset C is the smallest, indicating that students of different datasets have greater differences in the degree of mastery in the courses.

The traditional evaluation method of students' grade is based on specific relationship between scores and ranks. The clustering centers of five categories in course 1 are between 80 and 90 in Fig. 4 (b), and the ranks are all "good". The clustering centers of four categories in course 6 are below 80 in Fig. 4 (c), and there is no "excellent" or "good" in all ranks of the course.

Our method does not evaluate the student group with absolute score, but summarizes and reflects on the learning effect of students according to the actual situation. Taking Fig. 4 (b) as an example, according to our evaluation method, even if the clustering value is above 80, the fourth category of students in course 1 should be ranked as "poor". Taking Fig. 4 (c) as an example, according to our evaluation method, even if the clustering value is below 80, the first category of students in course 6 should be ranked as "excellent". Therefore, in the 100-point system, less than 90 points does not mean not excellent; similarly, reaching 80 points does not mean not poor.

C. DISCRIMINANT ANALYSIS

After determining clustering number, in order to improve the accuracy of the subsequent prediction results, it is necessary to test whether the clustering effect is good or not. Before performing discriminant analysis, it is necessary to check whether the sample is suitable for discriminant analysis. In the equality test of the group mean, the significance of all the characteristics of the three datasets is less than 0.05, indicating that there are significant differences in the means within different categories, so it is reasonable to do discriminant analysis. The clustering results of K-means algorithm are basically the same as those of Bayes discrimination. Each student has a higher posterior probability value in the category to which they belong. The clustering analysis results of datasets A, B, and C are correctly aligned 95.7%, 98.4%, and 98.0% of the original grouped cases which were clustered, and the clustering effect was good. The discriminant analysis verifies the validity and reliability of the clustering results, which is conducive to the next experiment.

The inconsistencies between the clustering and discrimination results of datasets A, B, and C are shown in Table 3, Table 4, and Table 5, respectively. The clustering results of two objects, one object, and one object are inconsistent with the discrimination results.

Judging from the number of inconsistent cases, there are fewer inconsistencies in the clustering and discrimination results of the three datasets. From the perspective of posterior probability, the discriminant result is the second category and the clustering result is the third category in the first inconsistency in dataset A. The posterior probabilities of the second category and the third category are 0.50075 and 0.34131 respectively. The difference is not large, and the posterior probability value of the result is relatively high. The objects whose clustering results of other datasets are inconsistent with the discrimination results are discriminated with slightly higher probability than the clustering probability.

Both angles show that the clustering effect is better. Even if the clustering result is inconsistent with the discrimination result, the clustering result at this time is acceptable for training in CNN. At the same time, it is verified from the side that the determination of clustering number by the statistic is helpful to ensure the reliability of the clustering results, and it is helpful to add category labels to the data for constructing the student performance prediction model in the next step.

D. CONVOLUTIONAL NEURAL NETWORK TRAINING

The data used in CNN must be data with category labels, so after ensuring a good clustering effect, it is necessary to add category labels to the clustering results.

TABLE 3. Cases and posteriori probability values of inconsistent clustering and discrimination in dataset A.

Dataset A	Clustering Category	Discrimination Category	Posteriori Probability of Category 1	Posteriori Probability of Category 2	Posteriori Probability of Category 3	Posteriori Probability of Category 4
1	1	1	0.99941	0.00001	0.00058	0.00000
2	1	1	0.99980	0.00001	0.00018	0.00000
3	1	1	0.97390	0.00607	0.02003	0.00000
4	1	1	0.98862	0.00008	0.01130	0.00000
5	1	1	0.96654	0.00012	0.03333	0.00000
6	1	1	0.95463	0.00539	0.03997	0.00000
7	1	1	0.99530	0.00260	0.00210	0.00000
8	1	1	0.99887	0.00000	0.00113	0.00000
9	1	1	0.99449	0.00136	0.00415	0.00000
10	1	1	0.96992	0.00004	0.03005	0.00000
11	1	1	0.99294	0.00004	0.00702	0.00000
12	1	1	0.94837	0.00194	0.04969	0.00000
13	1	1	0.97053	0.00439	0.02508	0.00000
14	3	3	0.26267	0.03975	0.69758	0.00000
15	1	1	0.90927	0.00889	0.08184	0.00000
16	2	2	0.00445	0.94552	0.05002	0.00001
17	3	3	0.06557	0.19409	0.74032	0.00001
18	3	3	0.04181	0.00015	0.95804	0.00000
19	2	2	0.04013	0.70051	0.25937	0.00000
20	3	2	0.15794	0.50075	0.34131	0.00000
21	3	3	0.06906	0.00083	0.93011	0.00000
22	3	3	0.00196	0.24270	0.75531	0.00003
23	3	1	0.52552	0.00027	0.47421	0.00000
24	2	2	0.02375	0.59229	0.38396	0.00000
25	3	3	0.11852	0.01322	0.86826	0.00000
26	2	2	0.00302	0.98973	0.00723	0.00002
27	3	3	0.00614	0.01150	0.98236	0.00000
28	2	2	0.00006	0.99957	0.00036	0.00002
29	3	3	0.00161	0.00022	0.99817	0.00000
30	2	2	0.00243	0.96125	0.03632	0.00000
31	2	2	0.00000	0.99908	0.00055	0.00037
32	3	3	0.00251	0.08609	0.91139	0.00000
33	2	2	0.00976	0.82063	0.16936	0.00025
34	3	3	0.00007	0.00609	0.99289	0.00096
35	2	2	0.00020	0.96500	0.03471	0.00009
36	3	3	0.00001	0.05546	0.94452	0.00001
37	4	4	0.00000	0.00010	0.00001	0.99989
38	2	2	0.00000	0.91814	0.00376	0.07810
39	2	2	0.00004	0.98003	0.01980	0.00013
40	3	3	0.00002	0.00649	0.99194	0.00156
41	4	4	0.00000	0.00000	0.00000	1.00000
42	4	4	0.00000	0.01553	0.04266	0.94181
43	4	4	0.00000	0.00395	0.00004	0.99601
44	4	4	0.00000	0.00000	0.00000	1.00000
45	4	4	0.00000	0.01131	0.00000	0.98869
46	4	4	0.00000	0.00000	0.00000	1.00000

According to the loss generated by the error, judge whether the network has converged, and use it as the criterion for stopping iteration. Use Python, deep learning framework Keras and TensorFlow, Spyder and other development tools to conduct experiments. The parameters are set to: batch_size=32, filters=32, kernel_size=3. The epoch value is determined according to whether the network has converged or not, and feature number is determined according to the feature value of each dataset. In order to avoid the saturation of the output with the gradual increase of the input, the Rectified Linear Unit (ReLU) is selected as the activation function of the convolutional layer. The five-layer convolutional layers are used for training when considering the characteristic values of three datasets, and the Softmax function is used for classification in the fully connected layer. Dropout makes the network structure of each round of training different. In the final classification, the entire network is used for classification, which is similar to taking the average of different classifiers. Dropout is a technique for addressing the problem of overfitting [43]. Therefore, in order to prevent the occurrence of over-fitting phenomenon, the value of dropout in the model is set to 0.2.

This paper uses two popular different cross-validation methods. Considering the size of the dataset, one is random hold-out method (80% of the data are randomly divided into training sets and 20% are test sets) and the other is shuffle 5-fold cross-validation method. The data of three

TABLE 4. Cases and posteriori probability values of inconsistent clustering and discrimination in dataset B.

Dataset B	Clustering Category	Discrimination Category	Posteriori Probability of Category 1	Posteriori Probability of Category 2	Posteriori Probability of Category 3	Posteriori Probability of Category 4	Posteriori Probability of Category 5
1	2	2	0.00002	0.99998	0.00000	0.00000	0.00000
2	2	2	0.00004	0.99996	0.00000	0.00000	0.00000
3	2	2	0.00002	0.99997	0.00000	0.00000	0.00000
4	2	2	0.00001	0.99999	0.00000	0.00000	0.00000
5	2	2	0.00001	0.99987	0.00000	0.00000	0.00011
6	2	2	0.00002	0.99998	0.00000	0.00000	0.00000
7	2	2	0.00014	0.99986	0.00000	0.00000	0.00001
8	2	2	0.00150	0.99848	0.00000	0.00000	0.00002
9	2	2	0.02463	0.97520	0.00000	0.00000	0.00017
10	2	2	0.00018	0.99981	0.00000	0.00000	0.00001
11	2	2	0.00016	0.99984	0.00000	0.00000	0.00000
12	2	2	0.00820	0.99175	0.00000	0.00000	0.00004
13	2	2	0.01135	0.98764	0.00000	0.00000	0.00101
14	2	2	0.00947	0.93677	0.00000	0.00000	0.05376
15	2	2	0.00010	0.99947	0.00000	0.00000	0.00043
16	2	2	0.00003	0.99997	0.00000	0.00000	0.00000
17	2	2	0.08304	0.90985	0.00000	0.00000	0.00711
18	5	5	0.07246	0.02162	0.00000	0.00004	0.90587
19	2	2	0.00129	0.99870	0.00000	0.00000	0.00001
20	1	1	0.71226	0.07475	0.00000	0.00002	0.21296
21	2	2	0.03030	0.96953	0.00000	0.00000	0.00017
22	2	2	0.00043	0.99919	0.00000	0.00000	0.00038
23	2	2	0.34985	0.64997	0.00000	0.00000	0.00017
24	1	1	0.55757	0.41209	0.00000	0.00000	0.03033
25	2	2	0.00234	0.99764	0.00000	0.00000	0.00002
26	1	1	0.99761	0.00135	0.00000	0.00000	0.00104
27	1	1	0.99633	0.00210	0.00000	0.00122	0.00035
28	2	2	0.00919	0.99043	0.00000	0.00000	0.00038
29	2	2	0.00486	0.99512	0.00000	0.00000	0.00001
30	1	1	0.99573	0.00035	0.00000	0.00019	0.00373
31	1	1	0.89062	0.09671	0.00000	0.00000	0.01267
32	1	1	0.99143	0.00655	0.00000	0.00000	0.00201
33	5	1	0.53887	0.00000	0.00000	0.00377	0.45736
34	1	1	0.99211	0.00013	0.00000	0.00519	0.00256
35	5	5	0.17905	0.07652	0.00000	0.00000	0.74443
36	1	1	0.97450	0.02036	0.00000	0.00000	0.00513
37	2	2	0.10469	0.88944	0.00000	0.00000	0.00587
38	5	5	0.35688	0.16565	0.00000	0.00004	0.47742
39	1	1	0.98698	0.00053	0.00000	0.00118	0.01130
40	5	5	0.03086	0.00066	0.00000	0.00020	0.96828
41	5	5	0.00485	0.00004	0.00000	0.00000	0.99511
42	1	1	0.99687	0.00082	0.00000	0.00000	0.00231
43	1	1	0.88938	0.09733	0.00000	0.00122	0.01207
44	1	1	0.99394	0.00360	0.00000	0.00117	0.00129
45	5	5	0.03404	0.00354	0.00000	0.00000	0.96243
46	1	1	0.98364	0.00001	0.00000	0.01031	0.00604
47	4	4	0.00376	0.00000	0.00000	0.99624	0.00000
48	5	5	0.00001	0.00000	0.00000	0.00000	0.99999
49	1	1	0.73697	0.00007	0.00000	0.02579	0.23718
50	4	4	0.00150	0.00000	0.00000	0.99835	0.00015
51	5	5	0.34299	0.00030	0.00000	0.00157	0.65514
52	4	4	0.00002	0.00000	0.00000	0.99998	0.00000
53	5	5	0.00436	0.00000	0.00000	0.00000	0.99564
54	5	5	0.00000	0.00000	0.00000	0.00002	0.99998
55	4	4	0.00640	0.00003	0.00000	0.98978	0.00379
56	1	1	0.99905	0.00002	0.00000	0.00000	0.00093
57	3	3	0.00000	0.00000	1.00000	0.00000	0.00000
58	4	4	0.00003	0.00000	0.00000	0.99997	0.00000
59	3	3	0.00000	0.00000	1.00000	0.00000	0.00000
60	3	3	0.00000	0.00000	1.00000	0.00000	0.00000
61	4	4	0.00000	0.00000	0.00000	0.99999	0.00000

datasets are basically converged after iterating 2000, 2000, and 1000 times, respectively, and the accuracy and loss are respectively are reaching better levels. In order to ensure the reliability of the result analysis, the accuracy

TABLE 5. Cases and posteriori probability values of inconsistent clustering and discrimination in dataset C.

Dataset C	Clustering Category	Discrimination Category	Posteriori Probability of Category 1	Posteriori Probability of Category 2	Posteriori Probability of Category 3	Posteriori Probability of Category 4
1	3	3	0.00014	0.00007	0.99979	0.00000
2	2	2	0.00000	0.99969	0.00031	0.00000
3	3	3	0.00009	0.00000	0.99990	0.00000
4	3	3	0.00003	0.00089	0.99908	0.00000
5	3	3	0.00416	0.00000	0.99584	0.00000
6	2	2	0.00002	0.88047	0.11951	0.00000
7	3	3	0.00003	0.00000	0.99997	0.00000
8	3	3	0.00000	0.00013	0.99987	0.00000
9	3	3	0.00734	0.00002	0.99264	0.00000
10	3	3	0.00000	0.00000	1.00000	0.00000
11	3	3	0.00007	0.00087	0.99905	0.00000
12	3	3	0.00000	0.00180	0.99820	0.00000
13	2	2	0.00000	0.99993	0.00007	0.00000
14	3	3	0.00311	0.00000	0.99689	0.00000
15	3	3	0.00209	0.00355	0.99436	0.00000
16	2	2	0.00000	1.00000	0.00000	0.00000
17	3	3	0.00060	0.01521	0.98419	0.00000
18	2	2	0.00002	0.99957	0.00041	0.00000
19	2	2	0.00444	0.84482	0.15074	0.00000
20	1	1	0.98327	0.00000	0.01670	0.00002
21	3	3	0.00000	0.00476	0.99524	0.00000
22	3	3	0.00009	0.00000	0.99991	0.00000
23	2	2	0.00068	0.94805	0.05127	0.00000
24	2	2	0.00001	0.99999	0.00000	0.00000
25	3	3	0.00009	0.00680	0.99312	0.00000
26	3	3	0.00003	0.00000	0.99997	0.00000
27	3	3	0.00001	0.00000	0.99998	0.00000
28	3	3	0.09689	0.00103	0.90207	0.00001
29	1	1	0.99840	0.00002	0.00080	0.00079
30	1	1	0.99227	0.00000	0.00495	0.00279
31	3	3	0.00826	0.00000	0.98871	0.00304
32	3	3	0.00013	0.00000	0.99987	0.00000
33	2	2	0.00000	1.00000	0.00000	0.00000
34	3	3	0.00235	0.00014	0.99752	0.00000
35	1	2	0.39656	0.54025	0.06319	0.00000
36	2	2	0.00000	1.00000	0.00000	0.00000
37	4	4	0.00000	0.00000	0.00000	1.00000
38	4	4	0.00001	0.00000	0.00003	0.99996
39	4	4	0.00008	0.00000	0.00000	0.99992
40	4	4	0.00865	0.00000	0.00000	0.99135
41	1	1	0.99994	0.00000	0.00004	0.00002
42	4	4	0.00114	0.00000	0.00006	0.99880
43	4	4	0.00000	0.00000	0.00000	1.00000
44	1	1	0.99989	0.00000	0.00011	0.00000
45	1	1	0.99668	0.00000	0.00033	0.00299
46	4	4	0.00000	0.00000	0.00000	1.00000
47	1	1	0.99979	0.00000	0.00018	0.00003
48	1	1	1.00000	0.00000	0.00000	0.00000
49	1	1	0.99998	0.00000	0.00002	0.00000
50	4	4	0.00000	0.00000	0.00000	1.00000
51	1	1	0.99705	0.00001	0.00293	0.00000

TABLE 6. Accuracy, loss, and time of training and testing in three datasets using two cross-validation methods.

		Accuracy	Loss	Val_accuracy	Val_loss	Time(ms)
Random Hold-out Method	Dataset A	94.59%	15.29%	88.89%	34.94%	432
80% for Training	Dataset B	94.29%	15.91%	90.00%	28.26%	448
20% for Testing	Dataset C	93.29%	18.51%	92.50%	21.76%	170
	Dataset A	91.22%	16.53%	88.89%	36.67%	430
Shuffle 5-fold Cross Validation	Dataset B	95.92%	13.68%	90.00%	30.75%	366
	Dataset C	93.90%	15.39%	92.50%	19.42%	170

and loss are used to evaluate the validity of the model. Accuracy and loss are the accuracy and loss of the three training sets respectively. Val_accuracy and val_loss are the accuracy and loss of the three test sets respectively.

Time means running time. Random hold-out method results and shuffle 5-fold cross-validation results are shown in Table 6.

In the random hold-out method results, the number of iterations of datasets A and B is 2000, but because the score records of dataset B are more than those of dataset A, the running time of dataset B is slightly longer than that of dataset A. Dataset C has the highest time efficiency and the best prediction performance. The accuracy of the test set is lower than that of the training set, and the loss value of the test set is higher than the loss value of the training set. In the training set, although the accuracy and loss of dataset A is the best among three datasets. However, in the test set, the accuracy and loss of dataset A is the worst among the three datasets. It can be concluded from random hold-out method results that the better the training set performs, the worse the test set performs. The fundamental reason is that the over-fitting of the training data leads to insufficient prediction accuracy of the model.

In shuffle 5-fold cross-validation results, the training set in dataset B has the highest accuracy and the lowest loss value. The training set in dataset C does not perform best in the training sets of three datasets, but the test set in dataset C performs best in the test sets of three data sets, and the running time is the shortest.

No matter random hold-out method or shuffle 5-fold cross-validation method, datasets A and B run longer than dataset C. The accuracy of the test set is correspondingly consistent in two methods. Comparing the two methods, the performance of training set and test set in dataset A is worse, but the performance of training set and test set in dataset C is better. There is little difference in the running time between datasets A and C. In the second method, the performance of the training set is better and the performance of the test set is worse, but the running time is significantly shortened in dataset B. In terms of accuracy, the model can guarantee a certain degree of reliability of the prediction results, and can be used to predict whether students' academic conditions are in an early warning state.

With the purpose of verifying whether a significant difference between the proposed method with the statistic and a method without the statistic, the results of both methods are compared. This study performs the statistical t-Test on student performance datasets. There is no statistically significant difference if P value > 0.05. The results show that there is a significant difference in the method of using the statistics to determine the clustering number and then using the clustering results as labels to construct a convolutional neural network classifier (P value < 0.05). There is no significant difference in the classifier without the statistics (P value > 0.05).

The high consistency between clustering results and discrimination results shows that the statistic proposed in this paper determines the clustering number from an objective and quantitative point of view. At the same time, the statistical test results also show that the classifier constructed by the labels added to the data by clustering has improvement in predicting students' academic performance.

V. CONCLUSION, LIMITATION, AND FUTURE WORK

The implications of the above analysis for the education sector are as follows:

1) Make good use of the excellent group to drive the overall development.
2) Modify the training program in a targeted manner to achieve the purpose of teaching students in accordance with their aptitude.
3) Explore more efficient instructional programs to accommodate student development.

Considering there is a certain degree of irrationality and subjectivity in the results of the school's evaluation, by using the K-means algorithm in unsupervised learning to perform clustering analysis on student performance, and then using the clustering results as the category label of CNN, the paper starts with data mining. It is ultimately found that the model has a more ideal prediction accuracy, which is beneficial to ensure the objective and fair evaluation of students by school. Moreover, it is available to promptly remind students who are in a state of academic warning. When considering data labels, the selection range of label value must be involved, and the selection range of label value is related to clustering number. The K-means algorithm has a well-known defect: the value of k is artificially determined. To improve the algorithm, the paper uses an objective statistic to optimize k-value selection, and replaces subjective evaluation with quantitative analysis, which makes the clustering results more powerful. The persuasiveness also makes the training and prediction results of CNN more reliable, and the effectiveness of the model is naturally guaranteed.

Although the clustering results are obtained after comprehensive consideration of the actual situation and the use of quantitative analysis, the selection of the initial clustering center is determined randomly, which may have a certain impact on the accuracy of the clustering results. Although the proposed statistic makes the results obtained by CNN better than those without the statistic, we do not compare it with other classifiers.

In the era of big data, EDM faces many opportunities in policies, resources, and technology. EDM research is beneficial to the development and innovation of education and the entire society. The complexity of educational issues and the nature of interdisciplinary make EDM show its uniqueness in terms of data sources, data characteristics, research methods and application purposes. Revealing and solving research problems in the education field has been aimed by EDM which uses a series of data mining techniques to analyze educational data, and uses existing data to discover new knowledge, thereby improving the quality of education and the learning process. The hybrid model that combines data mining techniques with the existing education data processing technology is applied to the analysis process of the student dataset. In the future, it can be further enhanced by combining association models or some integration-based technologies. In addition, EDM can also be extended to medical data processing, sports data processing and other fields. Applying educational data mining methods to finding suggestions to promote discipline development [44], learning analysis in a virtual learning environment, technology-assisted teaching methods, and monitoring student mental health [45] can all be used as future research content. The significant role of data mining technology in predicting academic performance and improving learning ability prompts us to carry out the next step of research.

REFERENCES

[1] H. B. Antonio, H. F. Boris, T. David, and N. C. Borja, "A systematic review of deep learning approaches to educational data mining," *Complexity*, vol. 2019, May 2019, Art. no. 1306039.

[2] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Implementing AutoML in educational data mining for prediction tasks," *Appl. Sci.*, vol. 10, no. 1, pp. 90–117, Jan. 2020.

[3] S. Kausar, X. Huahu, I. Hussain, W. Zhu, and M. Zahid, "Integration of data mining clustering approach in the personalized E-learning system," *IEEE Access*, vol. 6, pp. 72724–72734, 2018.

[4] D. Buenaño-Fernandez, W. Villegas, and S. Luján-Mora, "The use of tools of data mining to decision making in engineering education—A systematic mapping study," *Comput. Appl. Eng. Educ.*, vol. 27, no. 3, pp. 744–758, May 2019.

[5] C. Angeli, S. K. Howard, J. Ma, J. Yang, and P. A. Kirschner, "Data mining in educational technology classroom research: Can it make a contribution?" *Comput. Educ.*, vol. 113, pp. 226–242, Oct. 2017.

[6] B. A. Javier, F. B. Claire, and S. Isaac, "Data mining in foreign language learning," *WIREs Data Mining Knowl. Discov.*, vol. 10, no. 1, Jan./Feb. 2020, Art. no. e1287.

[7] C. Romero and S. Ventura, "Data mining in education," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 3, no. 1, pp. 12–27, 2013.

[8] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining Knowl. Discov.*, vol. 10, no. 1, May 2020, Art. no. e1355.

[9] S. Wang, "Smart data mining algorithm for intelligent education," *J. Intell. Fuzzy Syst.*, vol. 37, no. 1, pp. 9–16, Jul. 2019.

[10] M. J. James, S. H. Ganesh, M. L. P. Felciah, and A. K. Shafreenbanu, "Discovering students' academic performance based on GPA using K-means clustering algorithm," in *Proc. World Congr. Comput. Commun. Technol.*, Trichirappalli, India, 2014, pp. 200–202.

[11] A. Ani, L. Nicholas, and S. B. Ryan, "Enhancing the clustering of student performance using the variation in confidence," in *Proc. Int. Conf. Intell. Tutoring Syst.* Cham, Switzerland: Springer, 2018, pp. 274–279.

[12] R. G. Moises, D. P. P. R. Maria, and O. Francisco, "Massive LMS log data analysis for the early prediction of course-agnostic student performance," *Comput. Educ.*, vol. 163, Apr. 2020, Art. no. 104108.

[13] J. N. Walsh and A. Rísquez, "Using cluster analysis to explore the engagement with a flipped classroom of native and non-native English-speaking management students," *Int. J. Manage. Educ.*, vol. 18, no. 2, Jul. 2020, Art. no. 100381.

[14] V. G. Karthikeyan, P. Thangaraj, and S. Karthik, "Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation," *Soft Comput.*, vol. 24, no. 24, pp. 18477–18487, Dec. 2020.

[15] L. M. Crivei, G. Czibula, G. Ciubotariu, and M. Dindelegan, "Unsupervised learning based mining of academic data sets for students' performance analysis," in *Proc. IEEE 14th Int. Symp. Appl. Comput. Intell. Informat. (SACI)*, Timisoara, Romania, May 2020, pp. 11–16.

[16] S. Delgado, F. Moran, J. C. S. Jose, and D. Burgos, "Analysis of students' behavior through user clustering in online learning settings, based on self organizing maps neural networks," *IEEE Access*, vol. 9, pp. 132592–132608, 2021.

[17] K. Okoye, A. Arrona-Palacios, C. Camacho-Zuñiga, J. A. G. Achem, J. Escamilla, and S. Hosseini, "Towards teaching analytics: A contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification," *Educ. Inf. Technol.*, vol. 1, pp. 1–43, Oct. 2021.

[18] E. S. V. Kumar, S. A. A. Balamurugan, and S. Sasikala, "Multi-tier student performance evaluation model (MTSPEM) with integrated classification techniques for educational decision making," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 1796–1808, Mar. 2021.

[19] A. Siddiqa, S. A. Z. Naqvi, M. Ahsan, A. Ditta, H. Alquhayz, M. A. Khan, and M. A. Khan, "An improved evolutionary algorithm for data mining and knowledge discovery," *Comput., Mater. Continua*, vol. 71, no. 1, pp. 1233–1247, 2022.

[20] M. Agaoglu, "Predicting instructor performance using data mining techniques in higher education," *IEEE Access*, vol. 4, pp. 2379–2387, 2016.

[21] L. Qiu, Y. Liu, and Y. Liu, "An integrated framework with feature selection for Dropout prediction in massive open online courses," *IEEE Access*, vol. 6, pp. 71474–71484, 2018.

[22] A. Akram, C. Fu, Y. Li, M. Y. Javed, R. Lin, and Y. Jiang, "Predicting students' academic procrastination in blended learning course using homework submission data," *IEEE Access*, vol. 7, pp. 102487–102498, 2019.

[23] Y. Wen, Y. Tian, B. Wen, Q. Zhou, G. Cai, and S. Liu, "Consideration of the local correlation of learning behaviors to predict dropouts from MOOCs," *Tsinghua Sci. Technol.*, vol. 25, no. 3, pp. 336–347, Jun. 2020.

[24] S. Y. Lin, C. M. Wu, S. L. Chen, and T. L. Lin, "Continuous facial emotion recognition method based on deep learning of academic emotions," *Sens. Mater.*, vol. 32, no. 10, pp. 3243–3259, Oct. 2020.

[25] A. Farissi, H. Mohamed Dahlan, and Samsuryadi, "Genetic algorithm based feature selection with ensemble methods for student academic performance prediction," *J. Phys., Conf. Ser.*, vol. 1500, no. 1, Apr. 2020, Art. no. 012110.

[26] H. Turabieh, S. A. Azwari, M. Rokaya, W. Alosaimi, A. Alharbi, W. Alhakami, and M. Alnfiai, "Enhanced Harris hawks optimization as a feature selection for the prediction of student performance," *Computing*, vol. 7, pp. 1417–1438, Jan. 2021.

[27] H. B. Ma, S. Y. Yang, D. Z. Feng, and L. C. Jiao, "Progressive mimic learning: A new perspective to train lightweight CNN models," *Neurocomputing*, vol. 456, pp. 220–231, Oct. 2021.

[28] A. Nabil, M. Seyam, and A. Abou-Elfetouh, "Prediction of students' academic performance based on courses' grades using deep neural networks," *IEEE Access*, vol. 9, pp. 140731–140746, 2021.

[29] L. Gao, Z. Zhao, C. Li, J. Zhao, and Q. Zeng, "Deep cognitive diagnosis model for predicting students' performance," *Future Gener. Comput. Syst.*, vol. 126, pp. 252–262, Jan. 2022.

[30] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and D. N. Rutledge, "New data preprocessing trends based on ensemble of multiple preprocessing techniques," *TrAC Trends Anal. Chem.*, vol. 132, Nov. 2020, Art. no. 116045.

[31] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Frontiers Energy Res.*, vol. 9, Mar. 2021, Art. no. 652801.

[32] L. Cricelli, M. Greco, M. Grimaldi, and L. P. Llanes Dueñas, "Intellectual capital and university performance in emerging countries: Evidence from Colombian public universities," *J. Intellectual Capital*, vol. 19, no. 1, pp. 71–95, Jan. 2018.

[33] T. Zhang, C. Yin, and L. Pan, "Improved clustering and association rules mining for university student course scores," in *Proc. 12th Int. Conf. Intell. Syst. Knowl. Eng. (ISKE)*, Nanjing, China, Nov. 2017, pp. 1–6.

[34] M. Ramze Rezaee, B. P. F. Lelieveldt, and J. H. C. Reiber, "A new cluster validity index for the fuzzy c-mean," *Pattern Recognit. Lett.*, vol. 19, nos. 3–4, pp. 237–246, Mar. 1998.

[35] J. Shi, S. Zhu, Y. Zhou, and R. Li, "Bayes-extension discriminant method of two populations based on multivariate kernel density estimation," *Proc. Comput. Sci.*, vol. 122, pp. 780–787, Jan. 2017.

[36] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, Sep. 2015.

[37] P. Chea and J. C. Mandell, "Current applications and future directions of deep learning in musculoskeletal radiology," *Skeletal Radiol.*, vol. 49, no. 2, pp. 183–197, Feb. 2020.

[38] P. Kumari, P. K. Jain, and R. Pamula, "An efficient use of ensemble methods to predict students academic performance," in *Proc. 4th Int. Conf. Recent Adv. Inf. Technol. (RAIT)*, Dhanbad, India, Mar. 2018, pp. 1–6.

[39] R. Ghorbani and R. Ghousi, "Comparing different resampling methods in predicting students' performance using machine learning techniques," *IEEE Access*, vol. 8, pp. 67899–67911, 2020.

[40] N. Yanes, A. M. Mostafa, M. Ezz, and S. N. Almuayqil, "A machine learning-based recommender system for improving students learning experiences," *IEEE Access*, vol. 8, pp. 201218–201235, 2020.

[41] M. A. Prada, M. Dominguez, J. L. Vicario, P. A. V. Alves, M. Barbu, M. Podpora, U. Spagnolini, M. J. V. Pereira, and R. Vilanova, "Educational data mining for tutoring support in higher education: A web-based tool case study in engineering degrees," *IEEE Access*, vol. 8, pp. 212818–212836, 2020.

[42] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020.

[43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jun. 2014.

[44] Y. Long, "Research on art innovation teaching platform based on data mining algorithm," *Cluster Comput.*, vol. 22, no. S6, pp. 14943–14949, Nov. 2019.

[45] V. Murugananthan and B. L. S. Kumar, "Educational data mining life cycle model for student mental healthcare and education in Malaysia and India," *J. Med. Imag. Health Informat.*, vol. 7, no. 3, pp. 554–560, Jun. 2017.

GUIYUN FENG was born in Nantong, China, in 1996. She received the bachelor's degree in information management and information system from Nantong University, Nantong, in 2019. She is currently pursuing the master's degree with the School of Management, Guizhou University, Guiyang, China. Her research interests include educational data mining, data analysis, pattern recognition, and service science and innovation management.

MUWEI FAN was born in Guiyang, China, in 1987. He received the B.S. degree in electronic engineering from the Beijing Institute of Technology, China, in 2009, the M.S. degree in computer integrated manufacturing from Nanyang Technological University, Singapore, in 2012, and the Ph.D. degree in petroleum storage and transportation engineering from the China University of Petroleum, China, in 2017.

Since 2017, he was an Associate Professor with the Center of Big Data Creative Innovation and Technology, Guizhou University. His research interests include complex system modeling and simulation, reliability analysis and efficiency evaluation, big data analysis, and pattern recognition.

YU CHEN was born in Guiyang, China, in 1998. She received the bachelor's degree in engineering management from Guizhou University, Guiyang, in 2021, where she is currently pursuing the master's degree with the School of Management. Her research interests include industrial engineering (including logistics and supply chain), data analysis, and pattern recognition.