J. COMBIN. MATH. COMBIN. COMPUT. 127a (2025) 411–428

Journal of Combinatorial Mathematics and
Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

A study of educational evaluation models and their legal compliance based on big data analytics models

Ang Li[1,✉]

[1] Office of Educational Administration, Guangdong University of Finance & Economics, Guangzhou, Guangdong, 510320, China

ABSTRACT

Teaching evaluation is the feedback on the teaching effect of teachers and the learning effect of students. It has become a critical link in colleges and universities teaching management and teaching inspection. This paper proposes and applies an improved BT-SVM multi-classification algorithm to the education evaluation model. By calculating the relative distance between classes, the error accumulation phenomenon existing in the traditional SVM when dealing with multi-classification problems is solved. A classifier structure based on an incomplete binary tree is constructed to automatically classify teaching data by gradually dividing the data set and training the SVM classifier. By calculating the decision function value of the test sample in the binary tree, the category to which it belongs can be quickly determined. The education evaluation model follows the principle of legal compliance to improve the quality and efficiency of model evaluation and ensure the rule of law construction in colleges and universities. The research results show that the error rate of the BT-SVM algorithm in machine learning is below 0.1%, the fairness index is between 0.1-2, and the prediction accuracy is 96%. It shows that the machine learning algorithm can effectively improve the efficiency of education evaluation work and has the principle of fair legal compliance.

Keywords: educational evaluation; classifier structure; decision function; legal compliance; machine learning; BT-SVM

1. Introduction

Educational evaluation is an affirmation and basis for judging teachers' teaching. It is a value evaluation activity and can also judge whether students' learning outcomes meet expected requirements. Therefore, the causes of problems in educational evaluation are very complex. They are restricted by students' cognitive abilities, as well as the value standpoints, value standards and value ideals of the teaching subjects [1-2]. The classroom is the main battlefield of school education. Classroom teaching is an interactive process in which teachers and students conduct dialogue and collaboration in this main battlefield.

✉ Corresponding author.
E-mail address: leon340104496@163.com (A. Li).

Received 09 May 2024; Revised 11 August 2024; Accepted 21 October 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-025

Classroom teaching evaluation measures and evaluates the gap between the actual interactive effect and the desired design goal, and finds improvement measures to narrow this gap [3]. In the context of modern information technology accelerating the digital transformation of education, constructing new features and methods of teaching evaluation and formulating classroom teaching improvement plans more scientifically is a forward-looking research work [4]. With the rapid development of educational evaluation models and the deep integration of intelligent algorithms, while fully demonstrating the dividends of intelligent technology, the risks of privacy leakage, violation of property rights and ethical violations are also increasingly prominent, triggering widespread concern and worry among educational researchers [5].

To accurately judge educational effectiveness, this paper proposes an improved BT-SVM multi-classification algorithm to divide teaching samples in view of the error accumulation problem in traditional machine learning algorithms. The algorithm process is described in detail to ensure the working efficiency of the education evaluation model, and it is pointed out that the evaluation model should follow the compliance principle. The constructed teaching evaluation model is divided into evaluation and classification modules. All regulatory issues of the model are assumed to be independent. The degree of compliance of the education evaluation model is judged based on the integrity of the evaluation indicators contained in the privacy policy. Four first-level and eight second-level indicators are constructed, and the privacy protection port structure of privacy and non-privacy is designed. The education evaluation model can not only weigh the evaluation quality but also reduce information loss, ensure data privacy, etc., to ensure the comprehensive effect of the evaluation model while meeting the legal compliance requirements.

2. Related Words

With the changes in education in recent years, many scholars have conducted a lot of research on teaching methods as well as educational evaluation. Lu, L et al. conducted a user characterization test by obtaining user data from the web. Tian, Y et al. pointed out the publicizing of educational resources to realize an educational evaluation system which is equivalent to a learning system through which learning programs are improved and the user's characteristics are accurately evaluated [6-7]. He, H et al. concluded that automated scoring of physical tests for college students could be achieved with Deep BeliefNetworks with more desirable accuracy [8]. Gu, J et al. tracked learning related variables and analyzed the role of existing findings future related research and provided recommendations for the practice process [9]. Kawaguchi, O processed textual information on the basis of machine learning, constructed an educational model by extracting feature words, and obtained evaluation results [10]. Cresci, S et al. argued that personalized needs of all users can be achieved through online learning [11]. Li, C et al. argued that the online learning process can be extracted from the user's characteristic information data, through which the user's characteristics can be analyzed as a way to assist the reform of online teaching [12]. Wang, D et al. constructed an evaluation model for online learning in their study, through which the learning process can be evaluated effectively [13]. Sacco, M. A. et al. believe that online teaching needs the assistance of educational evaluation model, in which the class real-time mastery of the user's learning status, and based on this proposed corresponding educational strategies. All of the above studies belong to the theoretical research in the field of education, not down to the specific implementation level [14].

The field of education has a huge amount of data, but with the opening of educational data, data security and ethical issues have gradually come to the fore, and the fragmentation, siloing and privatization of data have become increasingly prominent. At present, the legal regulation of education evaluation models has not yet formed a completely clear understanding, so how to promote the legal normative opening of education evaluation models in a focused and step-by-step manner, identify the risks brought about by the opening of education governmental data, and carry out effective supervision and prevention has become a hot and difficult issue that needs to be solved urgently at the present time.

3. Machine Learning in Education

3.1. *Application of machine learning algorithms in educational evaluation*

In the process of constructing educational evaluation models, support vector machines in machine learning algorithms are mostly used in classification problems. However, due to the combination of multiple binary classifiers, data redundancy and many evaluation errors will occur. In this paper, an improved binary tree support vector machine multi-classification algorithm is proposed for the existing algorithm, which is referred to as BT-SVM in the following text [15].

3.1.1. Fundamentals of the BT-SVM classification algorithm. The BT-SVM multiclassification algorithm first divides all educational data into two subclasses. Further, it divides the subclasses into two sub-subclasses, and so on, until all nodes contain only one class [16]. Then, a binary SVM classifier is trained at each non-leaf node. The algorithm's training speed and classification accuracy will be optimal when the structure of the binary tree is close to normal. The BT-SVM algorithm is described as follows:

1) In the early stages of training, set the training education set as A, and remember the category number of A as n. If $n=1$, the category number of the sample data of the data set is used as the leaf node, and is added to the bifurcation classification. The training set A is randomly divided into B, C two subsets, so that the data subset B for the positive class, C for the negative class, construct the classification function SVM_1. construct the bifurcation classification tree with SVM_1 as the root node, repeat the steps of training for B, C two subsets to B as the classification function generated by the training set SVM_2 for the left sub-tree, C as the classification function generated by the training set SVM_3 for the right sub-tree. The category number of the sample data of this dataset is added to the binary classification tree using the category number of the sample data of this dataset as the leaf node.

2) In the testing phase, input the test education data into the root node of the binary classification tree constructed in the training phase. If the node is a leaf node, output the category number of the leaf node, i.e., the category to which the test evaluation results belong. Calculate the value of the decision function of the test data, if the value is +1, go to into the left subtree of that node, otherwise into the right subtree of that node [17].

3.1.2. Improved BT-SVM multi-classification algorithm. Due to the limitations of sample quantity and training time, using Euclidean distance to classify sample attributes is simple in educational evaluation. However, if there is a low or high overlap between categories, the traditional algorithm cannot make a good evaluation, which will lead to differences in evaluation results. Therefore, this paper uses inter-class distance to divide and evaluate samples, and the established evaluation model is more suitable for different situations and has higher applicability [18].

Let X be the sample set containing the k educational evaluation categories, and x_i be the training set of the first class, then the sample center of the i th class is:

$$C_i = \frac{1}{n_i}\sum_{x \in X_i} x \tag{1}$$

where $i = 1, 2 \cdots, k$, let n_i be the sample size of class i.

Assuming c_i, c_j is the cluster center of teacher i and student j, the Euclidean distance between teachers and students is:

$$d_{ij} = \left\| c_i - c_j \right\| \tag{2}$$

Based on formula (2), the minimum hypersphere radius between samples is calculated:

$$R_i = \max\left\{ \left\| c_i - x_i \right\| \right\} \tag{3}$$

where X_i is a sample of class i.

Assuming that the minimum hypersphere radius of i and j satisfies R_i, R_j and the Euclidean distance satisfies Dd_{ij} the relative distance between the teacher and the student is:

$$D_{ij} = \frac{d_{ij}}{R_i + R_j} \tag{4}$$

From formula (4), we can see that the relative distance of the education sample includes two indicator classes, namely the Euclidean distance and the minimum hypersphere radius, to divide the evaluation attributes and evaluation object clusters. In the evaluation model, the Euclidean distance between the two column centers reflects the attribute degree of the education data of the evaluation object in the model. The larger the Euclidean distance, the lower the correlation between the education object and the evaluation attribute, and it does not belong to the evaluation category. At the same time, the minimum hypersphere radius indicates the number of clusters of the evaluation object in the model [19]. So that the relative distance can simultaneously reflect the distance between two educational attributes and the degree of intersection of their sample distributions, i.e., the larger D_{ij} indicates that the degree of separability of class i and class j is better and easier to be divided, otherwise it indicates that the separability between the sets of educational samples is less.

3.1.3. Algorithm description. The educational samples in the training phase are normalized, and all evaluation criteria are stored in set C in ascending order. The flow of the algorithm is shown in Fig. 1, where the relative distance matrix between classes $i, j(i \neq j)$ is computed for an evaluation problem containing k classes D, the

$$D = \begin{bmatrix} 1 & 2 & D_{12} \\ 1 & 3 & D_{13} \\ \cdots & \cdots & \cdots \\ k-1 & k & D_{k(k-1)} \end{bmatrix} \tag{5}$$

The first two columns represent the category attributes of class i, j, and the third column is the evaluation index between class i, j. In D find the two classes i, j in set C that have the largest relative distance, and deposit them in set C_1, C_2 according to the size of the class labels, $C = C - (C_1 \cup C_2)$. If $C = \phi$, then use C_1, C_2 as the left and right subtrees of the binary tree, respectively, to complete the selection of a binary classification of positive and negative classes. Find the minimum distance D_{mc1} and D_{mc2} of class $m(m \in C)$ to each class in C_1, C_2, respectively, in D, and if $D_{mc1} \prec D_{mc2}$, add m to C_1, otherwise go back to C_2, and train an SVM classifier at each non-leaf node of the binary tree.

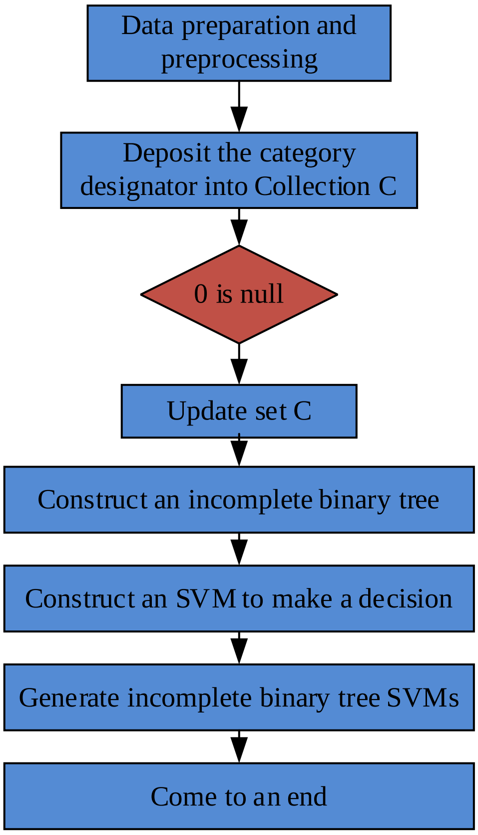

Fig. 1. Algorithm flow

In the test phase, the decision function $D_1(x)\succ 0$ of the unevaluated sample x at the root node of the binary tree is calculated. If $D_1(x)\succ 0$ is satisfied, the root node is entered to calculate the $D_2(x)$ value. When the result is greater than 0, it is still transferred to the left subtree for calculation. If $D_1(x)\prec 0$, go to the right subtree of the root node and calculate the value of $D_2(x)$. If the result is greater than 0, go to the left subtree for calculation. Otherwise, go to the right subtree and continue calculating until the leaf stops. At this time, output the category attribute of the unrated sample and divide it into the evaluation label [20]

This paper proposes a teaching quality evaluation model based on an incomplete binary tree support vector machine. In this model, the evaluator does not need to give an evaluation result for the teacher's teaching quality but gives a score for each evaluation indicator. Then, the incomplete binary tree support vector machine classifier is called to automatically classify the teacher's teaching quality and improve the efficiency of the evaluation work.

3.2. *Principles of laws and regulations involved in educational evaluation*

With the development of educational evaluation theory and the deepening of evaluation practice, the laws and regulations involved in educational evaluation are also being emphasized. The one-dimensional evaluation executed by the government and education administration is gradually developing into a joint three-dimensional evaluation by the government, society and schools. Based on the BT-SVM multi-classification algorithm, each type of evaluation has its own characteristics in terms of form, organizational settings and functions. However, among the different subjective views, should the ethical principles of evaluation be unified to make the educational evaluation system legally compliant [21]. Figure 2 shows the principles of evaluation model compliance, and justice and fairness are the bases of legal compliance of educational evaluation. Educational evaluation needs the principle of truthfulness, needs to be trusted and recognized, the subject of educational evaluation contains two main categories, one is the subject of evaluation system development, and the other is the subject of evaluation implementation, which is also the members of the expert group in the educational evaluation activities [22]. With the concept of the legal system of educational evaluation as the core, the six ethical principles of fairness, truthfulness, effectiveness, rationality of interests, plurality of subjects, and human-centeredness of equal educational evaluation are proposed.

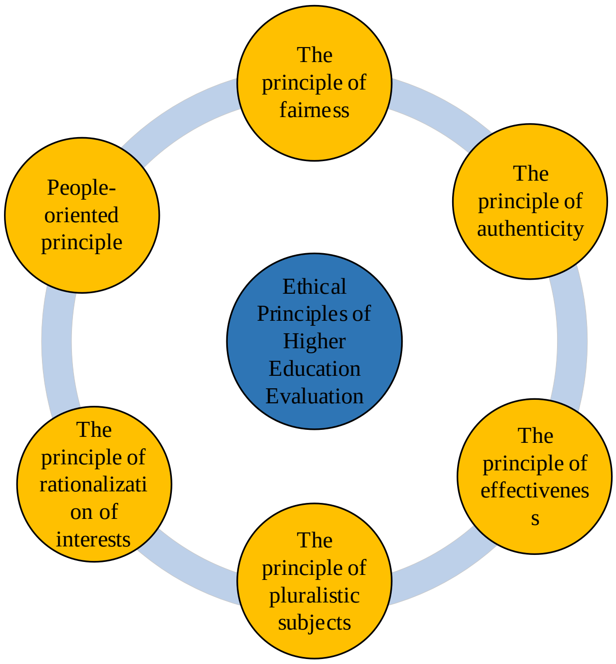

Fig. 2. The compliance principle of evaluation model

4. Legal compliance analysis

4.1. Evaluation modeling

Legal compliance of educational evaluation takes cultivating students' vocational ability and professional ethics as the teaching goal and purpose, which is an important model and method to promote the transformation of educational evaluation to the transformation, and itself has a very strong vocational education characteristics. Therefore, the evaluation model legal compliance will be used as a perspective to provide the corresponding standard basis for constructing a scientific and effective education teaching evaluation system. Figure 3 shows the model principle based on the BT-SVM multi-classification algorithm, the legal compliance detection model assumes that each base learner is a voter, and uses the training set and test set to train and test the corpus of the base learners respectively. Unclassified privacy policy statements are fed into the base learners as input variables, and the learners automatically make judgments about their compliance based on the measurement metrics to determine which metrics the statement meets [23-24]. The three learners output their respective judgment results, and then the classified values output by the base learner are sent to the integrated classifier, which votes on the classified values, and the indicator with more than half of the votes is the corresponding evaluation indicator of the unclassified policy statement, so as to determine the evaluation criteria one by one, and whether the legal compliance meets the specified requirements.

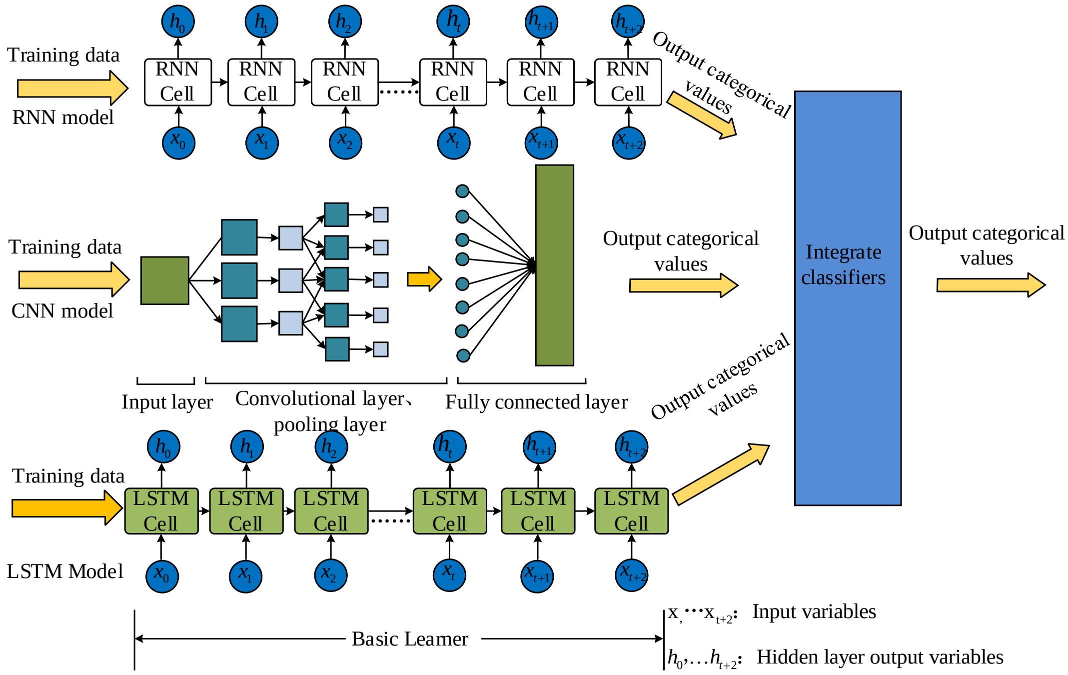

Fig. 3. Evaluation principle based on machine learning

In the above evaluation principle based on machine learning, the teaching quality evaluation model is shown in Fig. 4, which mainly includes the evaluation module and the classification module, and the figure shows the general workflow of the whole teaching quality evaluation system, in which the basic data set contains the basic information of the teacher, the course information, and the grade table, etc., and the evaluation result data set includes the output results of each module [25-26]. In the evaluation module, the BT-SVM algorithm is used to preprocess the data of the teaching subject, complete various evaluations, and obtain the final score. The evaluation of the teaching object is first divided into categories, and then evaluated from large to small according to the clustering results. In the classification model, the evaluation results of the teaching subject and teaching object are gradually output through the classifier. Considering that the evaluation model should be more in line with legal compliance requirements, the classifier established using the BT-SVM algorithm can better protect data privacy[27].

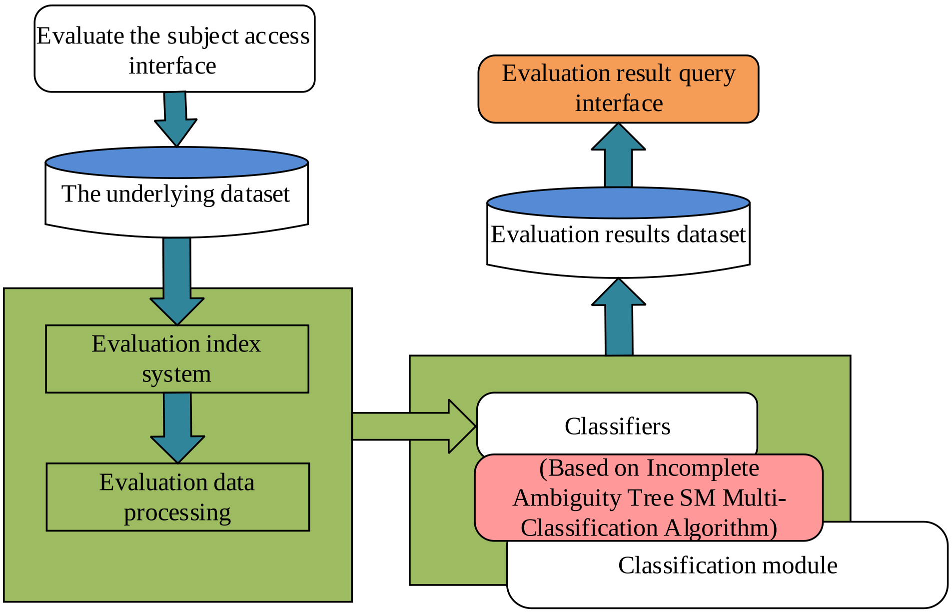

Fig. 4. Teaching evaluation model

In the model constructed in this paper, the classification module introduces relative distance under the premise of applying binary trees, giving full play to the advantages of binary support vector machines and solving the problem of limited classification conditions of traditional machine learning algorithms. It ensures the performance of the education evaluation model and data security by reducing redundant data generated by the model. At the same time, because the education evaluation model is applied in the campus network, it is less affected by the outside world, which not only improves the efficiency of the evaluation process but also ensures the legal compliance of the model [28-29].

4.2. *Model Legal Compliance Testing*

4.2.1. Evaluation Model Processing and Classification. In order to test the legal compliance of the educational evaluation model, the labeled values of the samples in the training set need to be converted into different category values. In this paper, the label y, which indicates the degree of compliance of the model after normalization, is divided into 4 categories according to $y \prec 0.6, 0.6 \le y \prec 7, 0.7 \le y \prec 0.8, y \ge 0.8$.

The category of test model x is predicted as the category value c that maximizes the conditional probability $P(c|x)$, i.e., the value of the said marker y category [V]. If model x has 4 possible violation markers, for simplicity, the numbers 1, 2, 3, and 4 can be used to denote the different categories in the algorithmic program and in the sample array. Where 1 i.e. $c=1$ corresponds to the aforementioned $y<0.6$, the model has a legal security problem.

$P(c|x)$ is the conditional probability, and here is the a posteriori probability, which refers to the probability that the value of its indicator is c, provided that x is known. For example, it is known that the normalized sample eigenvalue of evaluation model A is $x_A=(0.5,0.6,0.4,0.6,0.7,1)$, and in order to predict the category to which it belongs, the a posteriori probability of $P(c=1|x_A)$, $P(c=2|x_A)$, $P(c=3|x_A)$, and $P(c=4|x_A)$ is calculated using the data in the training set, in which the category corresponding to the maximum value is the predicted value of the index of model A. In the case of a limited number of samples in the training set, the above calculation is really an estimation. The value of $P(c|x)$ is usually not directly estimated and needs to be transformed into $P(x|c)$ as follows:

$$P(x|c)=\frac{P(x|c)P(c)}{P(x)} \tag{6}$$

where $P(c)$ is the probability that the indicator is c. If the total number of samples in the training set is m, and the number of samples in which the indicator value is c is n_c, then the estimate of $P(c)$ is:

$$P(c)=\frac{n_c}{m} \tag{7}$$

$P(x)$ is the probability of observing a sample x with the same $P(x)$ value corresponding to all category markers. Therefore, $P(x)$ can be disregarded in the process of predicting category values by comparing the $P(c|x)$ values corresponding to different categories.

For a limited number of training sets, $P(x|c)$ is also usually difficult to estimate directly. To make it estimable, assuming that all regulatory problems of the model are independent of each other, there are:

$$P(x|c)=\prod_{i=1}^{d}P(x_i|c) \tag{8}$$

where x_i denotes the i rd eigenvalue of the model book x. If the number of samples in which the i th eigenvalue takes the value of x_i is q_i out of the n_c class c samples, then the estimate of $P(x_i|c)$ is:

$$P(x_i|c)=\frac{q_i}{n_c} \tag{9}$$

Equation (4) corresponds to the case where the sample eigenvalues are discrete, e.g., the A1 indicator is categorized and converted to a category value so that the number of samples with an eigenvalue of x_i can be counted. In the case of continuous values of the characteristics, such as the above normalized scores, it is necessary to use the great likelihood estimation to obtain the value of the probability density function $p(x_i|c)$. Assuming that all Type c samples conform to a certain distribution, e.g., normal, i.e., there are $p(x_i|c)\sim N(\mu_i,\sigma_i^2)$, then:

$$p(x_i|c)=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right) \tag{10}$$

where μ_i and σ_i^2 are the mean and variance of the eigenvalues x_i of the sample cases of class c in the training set, in that order.

The features used for legal compliance are still continuous values after normalization, and the probability density of the output results of the evaluation model is usually consistent with a normal distribution. After the categorized level 1 metrics distinguish the full training samples into certain categories, it is likely that one of these categories will no longer conform to a normal distribution. Especially in the case of the classification of the second-level metrics, the distributional state of the model is uncertain and dynamically changing [31]. Therefore, it is not possible to estimate the value of $p(x_i|c)$ assuming that the samples of category c conform to a certain definite form of distribution, but only to estimate $P(x_i|c)$ by classifying the individual features of the model.

In the case of a limited number of samples in the training set, it is possible that the process of calculating $P(x|c)$ the number of samples for a certain metric is zero, i.e., a certain $P(x_i|c)=0$ situation where the feature classification result is zero, thus wiping out the information contained in the other non-zero $P(x_i|c)$ values as well. To avoid this situation, the Laplace correction formula can be used as:

$$P(c)=\frac{n_c+1}{m+N} \tag{11}$$

$$P(x_i|c)=\frac{q_i+1}{n_c+N_i} \tag{12}$$

where N describes the number of labeled compliance and N_i describes the number of categorizations for the i rd indicator.

4.2.2. Legal privacy compliance indicator system. In order to further quantify the degree of compliance of the constructed educational evaluation model, this study uses the completeness of the measurement indicators included in the privacy policy as a judgment standard to calculate the compliance score. Table 1 shows that this paper lists four first-level indicators and eight second-level indicators, and the more the evaluation model complies with the requirements of legal norms, the higher the compliance score.

Table 1. Evaluation index of compliance of educational evaluation model

Primary index	Secondary index	Description of compliance
A1 Privacy Policy Information	B1 Policy scope, time	Whether functional services and aging time are specified
	B2 Amendments and updates to the Privacy Policy	Whether to clearly look at the revision and update methods
A2 Educational information collection and use	B3 Purpose, method and scope of collecting educational information	Whether the purpose, method and scope of collecting educational information are clearly defined
	B4 Consent to collect educational information	Whether the collection and use of educational information has obtained the authorization and consent of the educational information subject
A3 Rules for storing educational information	B5 Location of educational information storage	Whether the server and geographic location for storing educational information are specified
	B6 Storage period of educational information	Whether the time limit of educational information storage and its basis, and the special circumstances for extending the storage period are specified
A4 Educational information sharing, transfer, public disclosure rules	B7 External sharing, transfer, public disclosure of personal information	Whether the purpose and scenario of external sharing, transfer and public disclosure of personal information are clear
	B8 Exceptions for sharing, transferring, and publicly disclosing personal information with prior authorized consent	Whether it is clear that the APP shares, transfers and publicly discloses personal information without seeking authorization according to relevant laws and regulations and regulatory requirements

In the education system evaluation model, there are strict regulations on data privacy protocols that do not conform to the norms, and normal access and querying are tampered with by human attacks and thus do not conform to the norms, thus preventing the correct evaluation system information from being delivered to the destination station. Non-compliant protocols are deleted, generating a large amount of garbage data in the entire educational evaluation model, making the model become congested resulting in packet loss and other phenomena. Therefore, it is crucial to establish a legal privacy compliance index system before constructing a privacy protection structure.

4.2.3. Privacy protection structure. Taking the constructed primary and secondary indicators into account, Figure 5 shows the data privacy protection structure of the evaluation model. For the query request made by the user, the BT-SVM algorithm is used in satisfying the user query service, which effectively protects the evaluation data privacy. Therefore, oriented to the legal compliance of the model, the goal of privacy protection structure design is to provide users with query results to achieve the coordination and unity of the three quality assurance, security and trustworthiness, and efficient service. Quality assurance means that the evaluation service results that can eventually be provided to users are accurate and meet the requirements of laws and regulations. Security and trustworthiness means that sufficient protection is provided for private data, and efficient service means that the system overhead, response time and time complexity are small. These three requirements are often contradictory in the real evaluation model, and the pursuit of balance and conciliation is the key to the realization of the algorithm.

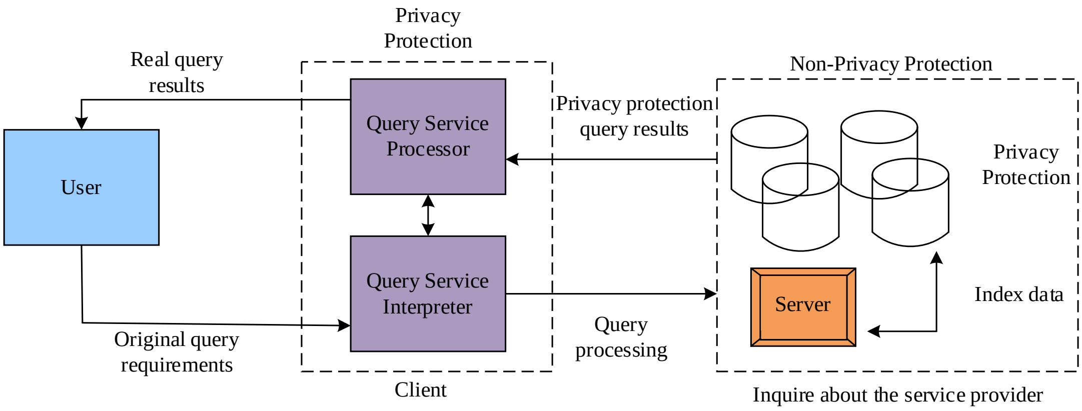

Fig. 5. Model privacy protection structure

In this paper, we establish a privacy-preserving structure that reflects the brand new utility matrix of different quasi-identifiers for different sensitive attribute values, and gives a relatively perfect guarantee metric that reflects the loss of information in anonymized data. A port structure that fully considers privacy and non-privacy is proposed. The educational evaluation model is made to weigh the evaluation quality, information loss, running time and data privacy protection in order to enhance the comprehensive effect of the evaluation model.

5. Educational evaluation modeling and legal compliance testing

5.1. Data sources

In this paper, it is assumed that all input data in the experiments are pre-processed canonical data and most of the training and datasets belong to online teaching resources, so the simulation experiments adopt the public education dataset as the training and testing dataset. The public education dataset includes 500 data records, 30 indicator evaluation contents and 100000 evaluation result outputs. The above feature labels and data volume are sufficient to support the experimental requirements of this paper. The specific parameters of hardware and software configurations of the experimental platform are shown in Table 2, and the BT-SVM algorithm is applied on an Inteli7-11700 desktop computer with 16GB of RAM to process the data relevantly in order to verify the performance of the evaluation model built in this paper and the legal compliance.

Table 2. Experimental hardware and software configuration parameters

Category	Disposition	Parameter value
Hardware configuration	CPU	Inteli7-11700
	GPU	GeForceRTX3060
	Internal memory	16GB
	Video memory	12GB
	Computing equipment	Desktop computer
Software configuration	Operating system	Win10x64
	Programming language	Python3.8.8
	Data set	TMDB
	Machine learning algorithm	BT-SVM

5.2. Evaluation of model performance in legal compliance

5.2.1. Data privacy protection. Data sharing is an unavoidable part of educational evaluation modeling, and many of these data contain users' private information. Although some data can guarantee that their private information will not be disclosed, in some cases, these data can be associated with other data, and the associated data may expose their private information. For example, there are two sets of data, student performance and school curriculum, the school curriculum is publicly available, and the student performance data is not publicly available. However, there are shared attributes between the data on student grades and the data on school course schedules, including student names and class markets. Through the records associated with attributes in these two batches of data, every piece of student personal information can be extracted, which destroys the privacy of the educational evaluation model. Figure 6 shows the comparison of the error rate under different evaluation samples, the error rate of BT-SVM algorithm in machine learning are all below 0.1%, and the four algorithms of decision tree, random forest, support vector machine, and neural network are all above 18%, which clearly shows that the proposed algorithm has the advantage of low data error rate, and it can ensure the privacy performance of the educational evaluation model better.

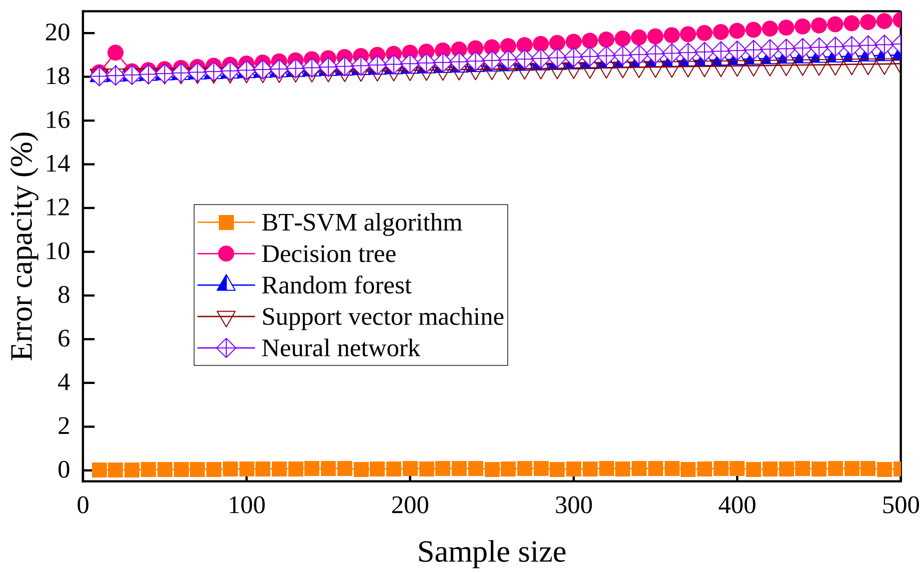

Fig. 6. Error rate under different evaluation samples

5.2.2. Fairness test. All along, due to the large gap between urban and rural education development, different education investment, and irrational layout of curriculum, it has caused problems such as the waste of education resources or the impact on education quality. The legal fairness research on the education evaluation model constructed based on machine learning algorithm, selecting the school area, the number of teachers and the number of teaching classes as the capacity influencing factors, and calculating the capacity of the school. By comparing the difference in educational resources between elementary school and the quality of education, the efficiency assessment index set at 1 is used to analyze the efficiency of educational resource utilization in elementary school in each district and between districts and districts. Figure 7 shows the capacity fairness indices of the three resources, and the efficiency indices of the number of teachers, school area, and the number of teaching classes are between 0.7-1, 0.5-1.6, and 0.1-2. Most of the schools have a very reasonable allocation of educational resources, and the schools have a good efficiency and fairness in their utilization, which verifies that the evaluation model constructed is able to maintain the compliance of the educational evaluation to a certain extent.

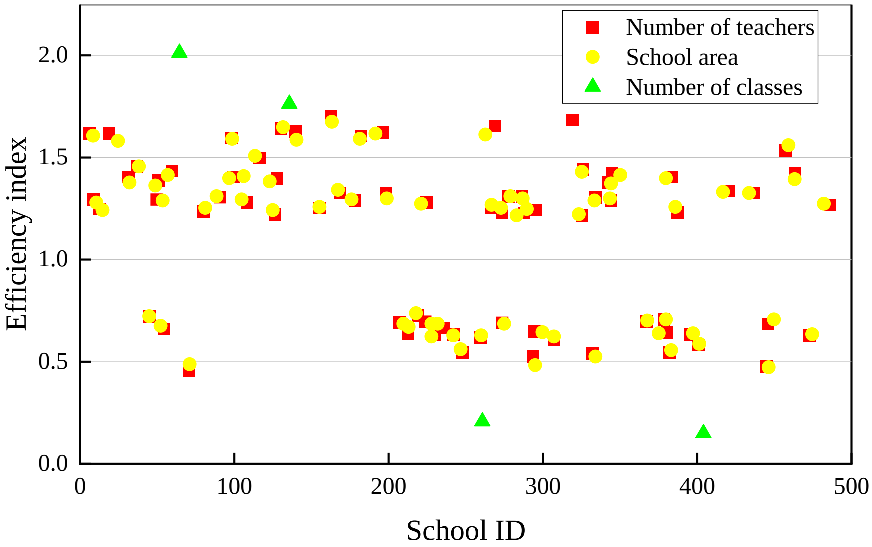

Fig. 7. Capacity equity index for three resources

5.2.3. Compliance Score. The legal compliance of the BT-SVM educational evaluation model is assessed based on the constructed compliance metrics, and the results of the measurements on each of the secondary metric dimensions are shown in Table 3. In the privacy policy information dimension, the constructed educational evaluation model focuses on data privacy and security, with only one non-compliant clause, while the neural network exists with 10 clauses, failing to meet the compliance standard. In the dimension of collection and use of educational information, the constructed educational evaluation model exists up to 2 non-compliance clauses, which is the least compared to the other four algorithms, indicating that the algorithm in this paper focuses on unhidden vulnerabilities and follows the principle of legal compliance. In the dimension of educational information storage rules, the BT-SVM algorithm has up to 3 non-compliant articles, and 7 articles less compared with neural network, in the educational evaluation model accumulates a huge amount of user data, and it should pay attention to the standardization of the development of relevant privacy policies. In the dimension of educational information sharing, transmission, and public disclosure rules, BT-SVM algorithm still performs the best, and there are at most 2 non-compliance clauses, compared with 3 non-compliance clauses in the decision tree, which is more capable of ensuring the safety of user information.

Table 3. Compliance Testing on Secondary Indicator Dimensions

Algorithm	Non compliant clauses/clauses							
	A1		A2		A3		A4	
	B1	B2	B3	B4	B5	B6	B7	B8
BT-SVM	0	1	1	2	1	3	2	1
Decision tree	5	6	7	9	10	9	9	10

Random forest	9	6	10	6	12	6	12	10
Support vector machine	9	6	6	9	6	9	10	6
Neural network	10	10	12	6	10	10	10	12

5.3. *Advantages of machine learning algorithms in educational evaluation*

5.3.1. Model Iteration. By selecting and decoding historical binary messages in the packet parsing module for historical packets in the evaluation model, the system state information in the network data stream is obtained completely, and attribute labels are added to the evaluation classification in combination with the protocol format. Although the educational evaluation data contains a lot of information that can reflect the characteristics of different events. However, the data itself is not structured. The attribute set is established through the basic features and is able to complement and categorize the relevant information on the basis of the known protocol format to form complete indicator information. Model iteration is a prerequisite for data analysis, if we want to structure the historical evaluation data and verify the effectiveness of the educational evaluation model based on machine learning algorithms, we need to iterate the model, Figure 8 shows the iteration process under 33 training cycles, and after 100 iterations, we meet the requirements for evaluation accuracy.

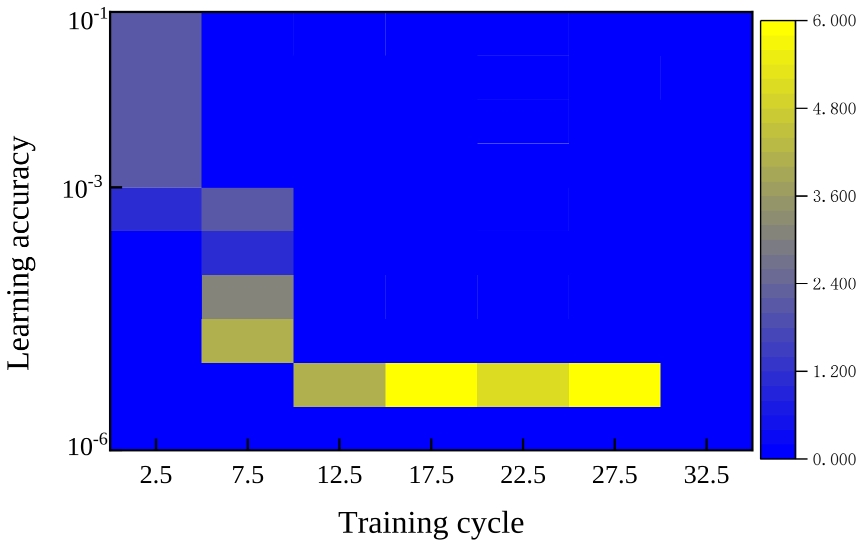

Fig. 8. Iterative process of educational evaluation model

5.3.2. Performance comparison. In addition, data on the evaluation of teaching effectiveness of freshman and sophomore students in a university for the past five years were collected. The dataset is normalized before the model starts training. In order to better understand the advantages of machine learning in educational evaluation, the BT-SVM multi-classification algorithm proposed in the paper is compared with the decision tree, random forest, support vector machine, and neural network algorithms in terms of model performance, and the performance comparison results of different methods are obtained as shown in Figure 9. The boxes in the figure indicate the accuracy intervals of each algorithm, which are BT-SVM multiclassification algorithm, decision tree, random forest, support vector machine, and neural network from left to right. The horizontal line in the middle of the box indicates the median accuracy, with 50% of the intervals less than this value. The upper and lower horizontal lines indicate the maximum and minimum values of the algorithm's extension, respectively, and the rectangle in the middle of the box is the 50% range in the middle of the accuracy results. In the comparison process, attention needs to be paid to its legal compliance to ensure that the use of the algorithm meets the requirements of the relevant laws and regulations, as well as the assessment results of the algorithm are fair, objective and accurate, to avoid any adverse effects on students. The small squares in the figure indicate the accuracy of different algorithms under 5, 10, 15, 20, 25, and 30 features. In this paper, the prediction accuracy of the BT-SVM multi-classification algorithm in machine learning is 96% on average, which is able to effectively capture the complex patterns and features in the educational evaluation data, thus providing more accurate prediction results. The accuracy of several other commonly used algorithms averages no more than 82%, and the BT-SVM multiclassification algorithm ensures the privacy protection of student data and the impartiality of the algorithm.

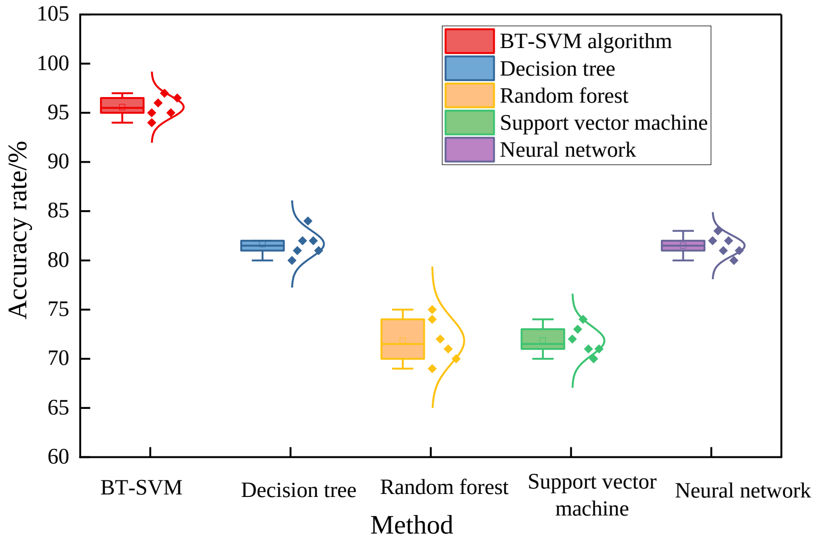

Fig. 9. Performance comparison results

6. Discussion

As education develops and progresses, it is affected by the complex environment in many ways and at many levels, and its values are embodied in many different ways, so the purposes and standards of education are correspondingly many and varied. In the future, while focusing on improving the operational efficiency and legal compliance of educational evaluation, it is necessary to explore the complexity of the issues involved in educational evaluation. Future research on educational evaluation models will focus on the following aspects:

1) Focusing on the plurality of evaluation objects, many teachers, many subjects, many schools have their own nature and characteristics, and different evaluation objects should have different applications for the same evaluation index system.

2) Focus on the subjectivity of the evaluation subject, the evaluation subject's knowledge structure and way of thinking there are differences, and subject to a certain worldview, methodology and values, so the evaluation of the same evaluation object will produce different evaluation results, to be considered in the model construction of the subjectivity.

3) The evaluation subject has constraints between the evaluation subject and the evaluation object and the decision maker, on the one hand, the evaluation subject must understand the relevant situation of the evaluation object, and even make some kind of contact with the evaluation object. On the other hand, the evaluation subject must also understand the purpose and intention of the decision maker, and the value and role of the evaluated object, but cannot be swayed by the decision maker, which requires a balanced relationship between the evaluation subject and the decision maker.

7. Conclusion

With the development of information technology and big data, the use of machine learning to evaluate the quality of higher education has become a research hotspot in academia and education. The purpose of this paper is to explore the evaluation of education quality based on machine learning and to study its legal compliance. The selected public education dataset includes 500 data, 30 index evaluation contents and 100000 evaluation results, and the error rates of BT-SVM algorithm in machine learning are all below 0.1%. In the compliance test, the BT-SVM algorithm has up to 3 non-compliant clauses in the secondary indicator system, which is the least compared to the other four algorithms. After 100 iterations, the prediction accuracy is 96%. The four algorithms of Decision Tree, Random Forest, Support Vector Machine, and Neural Network all have an error rate of 18% or more, and the evaluation model cannot play a good role in data privacy protection. The school area, the number of teachers, and the number of teaching classes are selected as the capacity influencing factors, and after being processed by the evaluation model, the efficiency indexes are between 0.7-1, 0.5-1.6, and 0.1-2, and most of the schools have a very reasonable allocation of educational resources, and the schools have a good efficiency of utilization and fairness, which verifies that the constructed evaluation model is able to maintain the compliance of the evaluation model to a certain extent. Although the good performance of the machine learning algorithm is proved, due to the limited time of the study, more targeted evaluation models, as well as legal compliance evaluation indicators, will be constructed in future studies.

Funding

Higher Education Research project of the "14th Five-Year Plan" of Guangdong Higher Education Association: "Research and Practice on the Reform of Higher Education Evaluation Mechanism in the New Era--Based on the Survey of University Learning Situation"(22GQM46);Quality Engineering and Teaching Research Reform Project of Guangdong University of Finance and Economics:" Tracking Research and Practice of Higher Education Evaluation in the New Era Under the Background of New Liberal Arts Construction".

References

[1] Sun, Z. , Zhang, S. , & Liu, M. . (2022). Education teaching evaluation method aided by adaptive genetic programming and robust scheduling. Mathematical Problems in Engineering.

[2] Hou, W. . (2022). Analysis of key indicators in english teaching evaluation based on big data model. Scientific Programming.

[3] Weng, J. , Hu, Y. , & Huang, H. . (2024). Lpsa: online educational video evaluation considering cognitive and teaching style. Computer applications in engineering education.

[4] Wang, M. . (2023). A teaching evaluation method of english online course based on conscious normal cloud model. International journal of continuing engineering education and life-long learning.

[5] Li, D. . (2024). An interactive teaching evaluation system for preschool education in universities based on machine learning algorithm. Computers in Human Behavior.

[6] Lu, L. , Tan, Y. , Klaic, M. , Galea, M. , & Zhao, E. . (2020). Evaluating rehabilitation progress using motion features identified by machine learning. IEEE Transactions on Biomedical Engineering.

[7] Tian, Y. . (2022). Teaching effect evaluation system of ideological and political teaching based on supervised learning. Journal of Interconnection Networks.

[8] He, H. , Sun, M. , Li, X. , & Mensah, I. A. . (2022). A novel crude oil price trend prediction method: machine learning classification algorithm based on multi-modal data features. Energy, 244.

[9] Gu, J. , Wang, Y. , Hu, J. , Zhang, K. , Shi, L. , & Deng, K. . (2024). Real-time prediction of fuel consumption and emissions based on deep autoencoding support vector regression for cylinder pressure-based feedback control of marine diesel engines. Energy.

[10] Kawaguchi, O. . (2024). Machine-learning and thresholding algorithms to automatically predict fishing effort of small-scale trawl fishery. Fisheries Science.

[11] Cresci, S. , Petrocchi, M. , Spognardi, A. , & Tognazzi, S. . (2022). Adversarial machine learning for protecting against online manipulation. IEEE internet computing.

[12] Li, C. , Leite, W. , & Xing, W. . (2022). Building socially responsible conversational agents using big data to support online learning: a case with algebra nation. British Journal of Educational Technology

[13] Wang, D. , & Guo, X. . (2022). Research on evaluation model of music education informatization system based on machine learning. Scientific programming.

[14] Sacco, M. A. , Ruiz, J. J. , Pulido, M. , & Tandeo, P. . (2022). Evaluation of machine learning techniques for forecast uncertainty quantification. Quarterly Journal of the Royal Meteorological Society.

[15] Li, J. , & Wang, D. . (2023). Comparison of pdem and mcs: accuracy and efficiency. Probabilistic engineering mechanics.

[16] Belkis, E. , Volkan, Y. , & Huseyin, E. . (2024). A novel microgrid islanding classification algorithm based on combining hybrid feature extraction approach with deep resnet model. Electrical engineering

[17] Zhou, X. , & Qiu, D. . (2023). Succinct quantum classification algorithm based on quantum circuit model. Chinese Journal of Physics.

[18] Ma, P. , Dong, C. , Lin, R. , Ma, S. , Jia, T. , & Chen, X. , et al. (2022). A classification algorithm of an ssvep brain-computer interface based on cca fusion wavelet coefficients. Journal of Neuroscience Methods.

[19] Yu - Eop Kang, Shon, S. , & Yee, K. . (2022). Local non - intrusive reduced order modeling based on soft clustering and classification algorithm. International Journal for Numerical Methods in Engineering

[20] Zhu, K. , Cao, W. , Ran, C. , & Gu, B. . (2023). (109488)A novel automatic crack classification algorithm of 3-d woven composites based on deep-learning u-net model. Engineering Fracture Mechanics.

[21] Han, C. . (2023). An evaluation method of english online learning behavior based on feature mining. International Journal of Continuing Engineering Education and Life-Long Learning.

[22] Schwartz, A. , King, B. A. , Mink, R. , Turner, T. , Abramson, E. , & Blankenburg, R. , et al. (2023). The appd longitudinal educational assessment research network's first decade. Pediatrics.

[23] Nizamani, S. , Nizamani, S. , Basir, N. , Laghari, G. , Khoumbati, K. , & Nizamani, S. . (2022). Heuristic evaluation versus guideline reviews: a tale of comparing two domain usability expert's evaluation methods. IEEE Transactions on Professional Communication.

[24] Clerbaux, B. , Molla, M. C. , Petitjean, P. A. , Xu, Y. , & Yang, Y. . (2021). Study of using machine learning for level 1 trigger decision in juno experiment. IEEE Transactions on Nuclear Science.

[25] Chen, C. , & Xue, X. . (2022). A novel hybrid intelligent model for the prediction of creep coefficients based on random forest and support vector machine. Ocean Engineering.

[26] Park, H. K. , Kim, Y. , Jung, J. , Lee, H. H. , Park, J. M. , & Ameyama, K. , et al. (2023). Efficient design of harmonic structure using an integrated hetero-deformation induced hardening model and machine learning algorithm. Acta materialia.

[27] Yang, H. . (2022). The evaluation method of distance learning engagement based on the multi-level linear model. International Journal of Continuing Engineering Education and Life-Long Learning.

[28] Zhou, Z. , Deng, W. , Wang, Y. , & Zhu, Z. . (2022). Classification of clothing images based on a parallel convolutional neural network and random vector functional link optimized by the grasshopper optimization algorithm:. Textile Research Journal

[29] An, Y. , & Xue, H. . (2022). Indefinite twin support vector machine with dc functions programming. Pattern

Recognition.

[30] Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. Materials today: proceedings, 80, 3782-3785.

[31] Welten, S. , Holt, A. , Hofmann, J. , Schelter, L. , Klopries, E. M. , & Wintgens, T. , et al. (2022). Synthetic rainfall data generator development through decentralised model training. Journal of Hydrology.