Hindawi
Scientific Programming
Volume 2022, Article ID 9658735, 12 pages
https://doi.org/10.1155/2022/9658735

Research Article

Research on Evaluation Model of Music Education Informatization System Based on Machine Learning

Daliang Wang and Xiaowen Guo

Music Academy, Henan Polytechnic, Zhengzhou 450046, China

Correspondence should be addressed to Daliang Wang; daliangwang@hnzj.edu.cn

Received 6 November 2021; Revised 11 December 2021; Accepted 27 December 2021; Published 24 February 2022

Academic Editor: Hangjun Che

Music education informatization system can promote music teaching; in addition, due to the characteristics of music disciplines such as the audiovisual nature of music, the influence of informatization on music teaching is self-evident. With the rapid development of the human ability to obtain information, machine learning algorithms have been widely used in various fields of scientific research and engineering, involving chemical production statistical process control, archeology text recognition, social and criminal investigation field fingerprint and image recognition, and genomic information research in the field of biomedicine. In order to correctly evaluate the music education information system based on machine learning, through the comparison of four models, it is concluded that the construction of the GBDT model is optimal.

1. Introduction

In the information environment, new information technology has brought about the rapid development of basic education informatization, but sometimes, it also shows the lack of a good operating mechanism and the imbalance of obvious structural differences. For the development of informatization, for example, expanded audiovisual channels and forms of music discipline [1], mathematical evaluation model [2], fuzzy evaluation model of industrial park informatization [3], corporate information based on graying weight cluster analysis evaluation model for the level of integration [4], primary education information system [5], evaluation index system for preschool education in colleges and universities [6], fuzzy comprehensive method [7], evaluation model for in-service slopes of in-service roads [8], management information system [9], comprehensive performance evaluation model [10], and community information system [11].

Machine learning, located at the intersection of computer science and statistics, is one of the fastest-growing technical fields today, and it is also the core of artificial intelligence and data science. Examples of applications are as follows: predicting song types [12], the effectiveness of detecting credit card fraud [13], CASD system for skin lesion analysis [14], predicting the two-stage protein-protein interaction, providing a set designed to enhance guidelines for future anomaly detection research [15], expert system [16], MXNet [17], CNN and DBN [18], potential pitfalls of brain imaging data [19], covariate shift algorithm [20], data-intensive-based machine learning methods [21], land cover types in agricultural environments, aquaculture robot technology research project evaluation model [22], and real-time online analysis and evaluation of CPS reliability.

2. Introduction to the Theoretical Basis

2.1. Introduction to Music Education. Music education is one of the important ways for schools to implement aesthetic education. It has the effects of cultivating aesthetics, harmonious body and mind, sound personality, and purification of the soul. It has an irreplaceable role in promoting the overall development of students.

2.2. Music Education Classification. This article divides the methods and contents of music education into four categories: traditional music education, special music education, multicultural music education, and modern music education as shown in Figure 1.

2.2.1. Traditional Music Education. The breadth and depth of China's excellent traditional culture are created and continued after thousands of years of history. It is the root and soul of our Chinese nation, and it is also the foundation for China to stand firm in the world's culture. There are many types of traditional Chinese music. It is an indispensable part of our traditional culture.

Traditional music generally refers to local music with national characteristics that have been passed down from generation to generation. It mainly includes folk songs and instrumental music. Newly created music based on these artistic traditions belongs to the extension and development of traditional music.

2.2.2. Special Music Education. Special music education is an education that is of great significance to special people's attention, and it is also one of the important contents of music education. The development of special music education can benefit more special people and the education field.

2.2.3. Music Education in Multiculturalism. Music education under multiple cultures is a diversified and diversified music education based on different cultures, social structures, and nationalities. As a branch of multicultural education, multicultural music education advocates the richness and diversity of music education and strives to realize the equal right to be educated for the music of all nationalities in the world and achieve the diversity and coprosperity of music culture heterogeneity.

2.2.4. Modern Music Education. Modern music education is combined with technology to realize education informatization. Educational informatization refers to the integration of the improvement of information literacy into the educational goals and the cultivation of talents suitable for the information society, the effective application of information technology to teaching and scientific research, and the emphasis on the development and utilization of educational information resources. Teaching is the central task in the field of education, and teaching informatization is to make teaching methods scientific and technological, education dissemination informatization, and teaching methods modernized, through the use of modern information technology based on computers, multimedia, etc., to promote education reform, so as to adapt to the new requirements of the coming information society.

2.3. The Concept of Machine Learning. "Learning" is the most basic ability inherent in humans and all kinds of animals. It is a kind of intelligent behavior that humans have. Since the day when people try to express human intelligence on computers, "learning" has naturally become the main issue of research. In real life, people can summarize general laws through analysis and research on actual cases and estimate events that are difficult or impossible to observe directly. From the above, the general laws obtained through analysis can not only explain the known examples reasonably but also correctly estimate and predict future events or phenomena to meet the requirements of social development. In fact, the core issue of many scientific research studies is the process of quantitatively analyzing an object or event in life and establishing a model. The model can achieve the purpose of control, prediction, etc. The above modeling process is the machine learning process.

The learning process includes two main aspects. One is knowledge acquisition: acquiring knowledge, accumulating experience, discovering laws, etc.; second, improving ability: improving performance, adapting to the environment, etc. In the learning process, the two are closely related and serve as the core and result of learning, respectively.

As shown in Figure 2, the model contains the four most basic links in the learning process. The learning link and the execution link represent two processes. The environment and the knowledge base are the collections of information in the way of knowledge expression. The learning link similar to the learning algorithm processes the information provided by the outside world to improve the content in the knowledge base. A large amount of information expressed in some form is stored in the knowledge base. The execution link uses the information in the knowledge base to achieve a certain goal, and then the execution result is fed back to the learning link. The research of machine learning is divided into two aspects. One is that the machine automatically captures useful information to make it smarter; the other is that it summarizes the laws of human thinking and the mystery of learning, which makes people's learning efficiency increase.

2.3.1. Machine Learning Meaning. The significance of machine learning is very important. It can continue learning, avoid a lot of repeated learning, and make knowledge accumulation reach a new height. At the same time, machine learning contributes to the dissemination of knowledge.

2.3.2. Application Areas of Machine Learning. Machine learning is a very important research field in computer science and artificial intelligence. In recent years, machine learning has not only played its role in many fields of computer science but has also become an important technical support for some interdisciplinary subjects. It has absorbed the research results of cognitive science, probability and statistics, artificial intelligence, and other disciplines. In particular, in the pharmaceutical industry, data mining, robotics, bioinformatics, industrial process control, and other fields have achieved certain results. At present, the research fields of machine learning are mainly based on the following three aspects:

(1) Task-oriented research: analyze and research a set of predetermined tasks to improve their execution performance

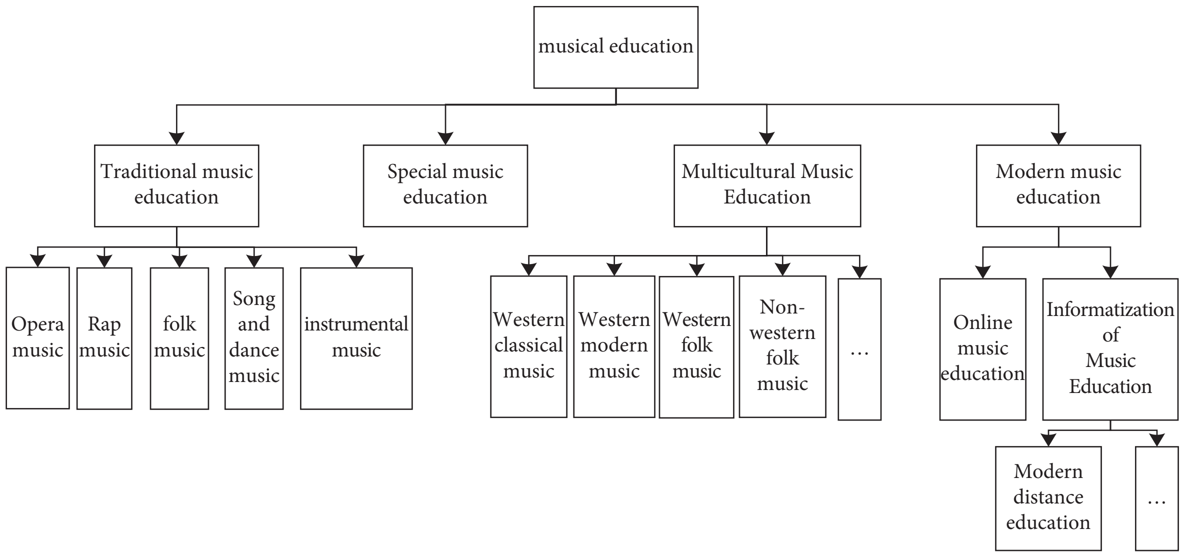

Figure 1: Classification of music education.

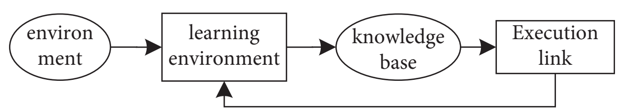

Figure 2: Machine learning model.

(2) Cognitive model: study the process of human thinking and learning, and use computers to simulate

(3) Theoretical analysis: theoretically study feasible learning methods

Machine learning is another research field of artificial intelligence after expert systems, and it is also one of the important research directions of artificial intelligence and neural computing. At present, the learning ability of artificial intelligence and computer systems is relatively poor, so they cannot meet the requirements of modern production and technology. Therefore, research and discussion on machine learning will surely lead to the rapid development of artificial intelligence and the entire science and technology.

Data mining is an important aspect of machine learning. The technology originated from the massive data caused by database technology and people's interest in the use of these data. The data is stored in a data management system, and then machine learning methods are used to analyze and mine the potential information in the massive data, which leads to the emergence of data mining. In general, data mining is to extract unknown, human-interested, and potentially useful information from a large database.

3. Overview of Machine Learning Algorithms

3.1. Introduction to Machine Learning Algorithms. Learning certain laws or features by analyzing a large number of sample data, so as to summarize, identify, and predict unknown results or unobservable data, is called a machine learning algorithm.

Machine learning algorithms include logistic regression, random forest, support vector machine, and GBDT. Among them, the logistic regression model is widely used in classification learning because of its simple implementation, fast speed, and easy update. The random forest model that is popular in machine learning mainly uses average decision trees to reduce the risk of overfitting and make the model relatively stable. Support vector machine has become one of the most commonly used classifiers with the best effect due to its excellent generalization ability. The GBDT model is known for its high prediction accuracy and flexibility to handle various types of data.

3.1.1. What Is an Algorithm? There can be many definitions of algorithms. Here, the concept of the human brain "algorithm" is proposed, and an attempt can be made to incorporate the human brain "algorithm" into the philosophical algorithm concept. The "algorithm" of the human brain has all the attributes of the full cognition of the human brain, achieving an abstract meaning of the full cognition of the human brain.

"Algorithm" has now become a term specifically referring to the logical methods used in computer programs. From an intuitive and practical perspective, algorithms are an integral part of computer science. An algorithm is a set of domains with a partial ordering of transition mapping to determine the state and a "definable recursive body" for the mapping value to determine the result. In the past, an algorithm was an entity that satisfies the assumptions of continuous time, abstract state, and limited search. The classic definition of "algorithm" is usually embodied in the mathematical context that formed early computer science. Common elements are deterministic rules or steps, calculations about a given input to produce an output, etc.

With the development of computers, algorithms have gradually become huge networks, and their original definition has obviously changed. An algorithm is a kind of effective, limited, abstract, compulsory command and completes a specific purpose compound control structure in accordance with regulations.

3.1.2. Algorithm Related Concepts

Algorithms and Computers. A computer is a machine that executes symbolic algorithms. However, not all carriers that execute algorithms are machines. For example, abacuses are not "machines" because they cannot calculate on their own. However, having certain attributes of a computer is not necessarily a computer in the true sense, because a computer is not only multifunctional but also universal and can execute all symbolic algorithms. In the age of computer networks, almost everything is controlled by algorithms. As a result, algorithms and computers are complementary and inseparable.

Algorithms and Artificial Intelligence. People with different research backgrounds often hold different views on the meaning of artificial intelligence. The artificial intelligence simulation method of the algorithm is used to make the intelligent behavior of the computer similar to the intelligent behavior of human beings. The higher the similarity, the more intelligent. This kind of research can achieve the purpose of explaining the human cognitive mechanism through the realization of algorithms so that people can have a deeper understanding of themselves and the nature of cognition. Although people have different opinions on the meaning of artificial intelligence, everyone still has the most basic consensus: to further understand human intelligence by constructing intelligent systems.

Algorithms and Cognitive Computing. Relying on the strong development of artificial intelligence, a unique way of thinking through computer principles to reflect on human's own cognitive mechanism has formed. However, regardless of the interpretation of cognitive methods, the view of computationalism is based on computers and recognizes the basic premise that cognition is computable. A series of analogies under this premise achieve the purpose of trying to unravel the mystery of cognition.

3.2. Overview of Machine Learning Algorithms

3.2.1. Logistic Regression. Basic Principle. It is a two-classification model in machine learning. Because the logistic regression algorithm is simple in principle and efficient in classification, it is very widely used in practical applications. Its idea comes from linear regression in statistics. The difference from linear regression is that its dependent variable is discrete. The basic principle is as follows: the output of the dependent variable y in the binary classification problem can only be 0 or 1, which is not a continuous value within a certain range. In order to convert the continuous output into a binary problem, the continuous value is converted into a discrete binary value through a nonlinear function.

Algorithm introduction: $X = \langle x_1, x_2, \ldots, x_n \rangle$ represents the n-dimensional feature vector, n-dimensional column vector $\theta = \langle \theta_1, \theta_2, \ldots, \theta_n \rangle$ represents the corresponding weight or coefficient, b represents the intercept, and the linear relationship is expressed as

$$F(\theta, X, b) = \theta^T X + b. \tag{1}$$

The value range of F is R. When dealing with the simplest binary classification problem, it is necessary to map the value of F to (0,1). Commonly used S-shaped Korean (sigmoid function) to express: its expression is shown in the publicity:

$$g(z) = \frac{1}{1 + e^{-z}}. \tag{2}$$

The value of z is $(-\infty\ ,\ +\infty)$, the value of g *is* $(0, 1)$, and the odd function image is as shown above: the logistic regression function can be obtained from Figure 3 and has the following characteristics: when $z \gg 0, y = 1$; when $z \ll 0, y = 0$. In terms of classification, the probability that y maps the result to a value between 0 and 1 represents the probability of belonging to the correct particle.

Regularization. Realize regularization by modifying the cost function formula of the algorithm, the expression of the hypothesis function, and updating the iterative method, so as to punish the training samples by the degree model to avoid overfitting of the model and improve the generalization ability of the model. During its operation, the regularization coefficient can be set to $L1$ or $L2$, where $L2$ is the default value. The cost function of the logistic regression model is

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(t)}\log\left(h_\theta\left(x^{(t)}\right)\right) + \left(1 - y^{(t)}\right)\log\left(1 - h_\theta\left(x^{(t)}\right)\right)\right]. \tag{3}$$

Among them, m represents the number of samples for magic training, y represents the y value in the training samples of the hospital, $h_\theta(x)$ represents the y value predicted by the parameters θ and x, and (t) represents the t-th sample. The logistic regression minimization cost function under binary classification $L2$ regularization and the logistic regression cost function under $L1$ regularization are as follows:

$$\begin{gathered} \min \frac{1}{2}\omega^T \omega + C\sum_{i=1}^{n}\log\left(\exp\left(-y_i\left(x_i^T \omega + c\right)\right) + 1\right), \\ \min_{\omega f} \omega_1 + C\sum_{i=1}^{n}\log\left(\exp\left(-y_i\left(x_i^T \omega + c\right)\right) + 1\right). \end{gathered} \tag{4}$$

3.2.2. Random Forest. Basic Principle. The random forest can explain the effect of several independent variables $X_1, X_2, \ldots, X_k$ on dependent variable Y. Among them, the dependent variable Y has n observations, and it will choose the bootstrap resampling method to randomly select n observations from the original data. Some observations are selected multiple times, and some are not selected. It will randomly select some variables from the k independent variables to determine the node of the classification tree. In this way, the classification tree constructed each time may be different. In general, random forests randomly generate hundreds to thousands of classification trees and then choose the tree with the highest degree of repetition as the final result.

Algorithm Introduction. From the original training data set, K new self-service sample sets are randomly selected using the bootstrap method with replacement, K classification and regession trees are constructed from this, and the samples that are not drawn each time constitute K out of the bag data; there are n features, randomly selected m_{try} features ($m_{\text{try}} \leq n$) at each node of each tree; by calculating the amount of information contained in each feature, select one of the m_{try} features with the most classification ability to split the node. Each tree grows to the utmost extent without any tailoring; the generated multiple trees are formed into a random forest, and the new data is classified with it, and the classification result is determined by the number of votes of the tree classifier.

3.2.3. SVM. Basic Principle. By solving the quadratic programming problem, a classifier based on structural risk minimization is searched for the optimal hyperplane that divides the data into two categories. It can also be simply described as follows: finding an optimal classification hyperplane that meets the classification requirements, so that it can maximize the blank area on both sides while ensuring the classification accuracy so that the support vector machine can achieve linearly separable data. *Optimal Classification.* Among them, the optimal classification plane should meet the following conditions:

$$\begin{aligned}&\max_{w,b} \frac{2}{w},\\&s.t\, y_i\left[(W \cdot x_i) + b\right] - 1 \geq 0, i = 1, 2, 3, \ldots, n.\end{aligned} \tag{5}$$

Among them, $2/w$ represents the size of the classification interval of the classifier. In order to make the classification plane robust, it is necessary to find the largest classification interval to reach the optimal hyperplane. In addition, in the constraint condition, $y_i[(W \cdot x_i) + b]$ means the distance from the sample point to the classification plane, and $y_i[(W \cdot x_i) + b] \geq 1$.

Algorithm Introduction. By introducing the kernel function, the core of the support vector machine, an algorithm that is nonlinear with respect to the original space can be realized in a high-dimensional space. The kernel function of a support vector machine is the inner product of a certain high-dimensional space, which plays a vital role in a support vector machine. Choosing different kernel functions will produce different support vector machine algorithms. There are three kinds of kernel functions: the first kind is a polynomial kernel function of order q, as follows:

$$K(x_i, y_i) = \left(x_i \cdot x_j + 1\right)^q. \tag{6}$$

The second kind is the radial basis function kernel function, as follows:

$$K(x_i, y_i) = \exp\left[-\frac{\left|x_i - x_j\right|^2}{\sigma^2}\right]. \tag{7}$$

The third kind is the neural network kernel function, as follows:

$$K(x_i, y_i) = \tanh\left[c_1\left(x_i \cdot x_j\right)\right] + c_2. \tag{8}$$

The standard support vector machine is using the structural risk principle: the loss function is the error selected in the optimization objective δ_i, where δ_i is the slack variable that allows misclassification. For classic support vector machines, the optimization problem is

$$\begin{aligned}&\min J(w, \delta) = -\frac{1}{2} w.w + c\sum_{i=1}^{l} \delta_i,\\&s.t \quad y_i\left[(\Psi \cdot (x_i) \cdot w) + b\right] \geq 1 - \delta_i, i = 1, 2, 3, \ldots, l.\end{aligned} \tag{9}$$

Through the Lagrangian method, the standard support vector machine optimization problem can be transformed into the following quadratic programming:

$$\begin{aligned}&\max W(a) = -\frac{1}{2}\sum_{i,j=1}^{l} a_i y_i y_j K(x_i, y_i) a_j + \sum_{i=1}^{l} a_i,\\&s.t \sum_{i=1}^{l} a_i y_i = 0, \quad 0 \leq a_i \leq c,\ i = 1, 2, 3, \ldots, l.\end{aligned} \tag{10}$$

3.2.4. Gradient Boosted Decision Trees. Basic Principle. An iterative decision tree algorithm is composed of multiple decision trees. There are usually hundreds of trees, and each tree is small in size. In model prediction, for an input sample instance, an initial value is first assigned, then each decision tree will be traversed, each tree will adjust and correct the predicted value, and finally, the results of each decision tree will be accumulated the final prediction result.

Algorithm Introduction. Including regression algorithm and classification algorithm, classification algorithm also includes two-classification and multiclassification models. The loss function that defines the two classifications is

$$L(y, f(x)) = \log(1 + \exp(-y f(x))). \tag{11}$$

Among them $y_i \int y = \{-1, +1\}$; the negative gradient error at this time is

$$\begin{aligned}r_{mi} &= -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)}\\&= \frac{y_i}{1 + \exp\left(y_{i,f(x_i)}\right)}, \ i = 1, 2, \ldots, N.\end{aligned} \tag{12}$$

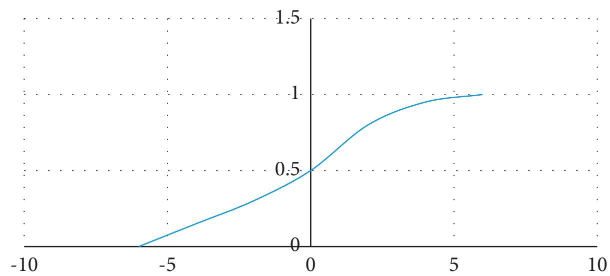

Figure 3: Logistic regression function image.

For the generated decision tree, the best residual fitting value of each leaf node is

$$c_{mi} = \arg\min_{c} \sum_{x_i \in R_{mi}} \log\left(1 + \exp\left(-y_i\left(f_{m-1}(x_i) + c\right)\right)\right). \quad (13)$$

Because the above formula is difficult to optimize, we generally use approximate values instead:

$$c_{mi} = \frac{\sum_{x_i \in R_{mi}} r_{mi}}{\sum_{x_i \in R_{mi}} \left|r_{mi}\right|\left(1 - \left|r_{mi}\right|\right)}. \quad (14)$$

Regularization is mainly used to solve the overfitting phenomenon of the model. For the GBDT algorithm, the method to deal with the regularization problem can be summarized as follows.

One is to use a method similar to Adaboost to set the step size (learning rate) parameter for the regularization setting. For the first few iterations of the weak learner, there are

$$f_x(x) = f_{k-1}(x) + h_k(x). \quad (15)$$

If the regularization term is added, there are

$$f_x(x) = f_{k-1}(x) + \nu h_k(x). \quad (16)$$

The value range is $0 \le \nu \le 1$. When it is smaller, it means that the number of iterations is more; on the contrary, the number of iterations is less. Therefore, in the process of building a model, in order to achieve a better fitting effect, the step size and the maximum number of iterations are usually adjusted together. The second is to set the regularization by changing the subsampling ratio. In general, concerning the subsampling ratio (0,1], when the value is 1, it means full sampling; that is, all samples are used; when the value is less than 1, the corresponding samples are drawn according to the ratio to fit the model to avoid overfitting. The third is to use a weak learner for regularization and pruning.

3.2.5. GP. GP (Kriging model) algorithm is a modeling technique based on the assumption that the input variables obey the joint Gaussian distribution. It can be used for both regression problems and classification problems. It can not only accurately establish complex nonlinear industrial process models but also quantify the uncertainty in predictions.

The GP algorithm based on Bayesian theory obtains its posterior distribution by studying the prior distribution of the training set parameters. Assume a set of random variables $\{Y(x) | x \in X\}$ indexed by the input space X, where p is the number of input variables $X = (x_1, x_2, \ldots, x_p)$. The Gaussian process is defined by its mean function $\mu(x) = E[Y(x)]$ and covariance function:

$$C\left(x, x^T\right) = E\left[\left(Y(x) - \mu(x)\right)\left(Y\left(x^T\right) - \mu\left(x^T\right)\right)\right]. \quad (17)$$

That is, once the mean function and covariance function are determined, the Gaussian model is determined. In general, the military assumes that the Gaussian process is zero-mean, ie $\mu(x) = 0$[39,40].

According to the training set obeys the joint Gaussian distribution, the Gaussian distribution of the test sample can be obtained from the mean function $\hat{y}(x)$ and the covariance function $\sigma^2_{\hat{y}}(x)$. The specific form is as follows:

$$\begin{aligned} \hat{y}(x) &= k^T K^{-1} t, \\ \sigma^2_{\hat{y}}(x) &= C(x, x) - k^T(x) K^{-1} k(x). \end{aligned} \quad (18)$$

In the above formula, $k(x) = (C(x, x(1)), \ldots, C(x, x(n)))^T$, K is the covariance function, in terms of the training set $K_{ij} = C(x(i), x(j)), t = (t(1), \ldots, t(n))$, and n is the number of training samples.

For the application of the GP algorithm, there are a variety of covariance functions to choose from. Generally, the covariance function has a specific requirement; that is, for all sample point sets $(x(1), \ldots, x(n))$, the covariance matrix generated by the covariance function must be nonnegative; for example,

$$\begin{aligned} C\left(x^{(i)}, x^{(j)}\right) &= \nu_0 \exp\left\{-\frac{1}{2}\sum_{i=1}^{p} w_l\left(x_l^{(i)} - x_l^{(j)}\right)^2\right\} \\ &\quad + a_0 + a_1 \sum_{l=1}^{p} x_l^{(i)} x_l^{(j)} + \nu_1 \delta(i, j). \end{aligned} \quad (19)$$

In the above formula, $\theta = (a_0, a_1, \nu_0, w_1, \ldots, w_p, \nu_1)^T$ is the hyperparameter that defines the covariance function, and T is the symbol for rank conversion. θ is obtained by finding the maximum likelihood function, and the expression is as follows:

$$L = \log p(y | \theta, X) = -\frac{1}{2}\log|C| - \frac{1}{2} y^T C^{-1} y - \frac{N}{2}\log(2\pi). \quad (20)$$

C is the covariance function.

Use gradient method to solve it; for example, conjugate gradient method. Usually, the estimation of the maximum likelihood function is a convex function solving problem, so there may be several local optimal solutions. Through multiple random selections of initial values for calculation, the shortcomings of misunderstanding of the local optimal solution as the global optimal solution are overcome.

4. Brief Introduction to the Design of Music Education Information System

4.1. Feasibility Analysis. The music education information system can collect, organize, and archive the relevant materials of the music subject. The feasibility analysis mainly analyzes the three levels of technical level, management operation level, and economic behavior.

The construction of the system framework is mainly divided into the selection of hardware equipment and the construction of the software framework, by adopting Tomcat as a service container, B/S development framework, MySQL database, and other mature technologies. In the development process, Intelli J IDEA is used as the website development tool, Microsoft's MySQL Workbench is used as the database development tool, and Visual Studio Code is used as the front-end page development tool. We use Java technology to develop the system.

The operation level of the music education information system mainly takes into account the operability of the staff. Convenience, flexibility, and simple operation are the criteria for evaluating the operability of the system. Different classifications of operating system personnel mean that operators in different positions can perform different tasks. Operators operate through dialog boxes, buttons, etc., through the comprehensive factors of the system to ensure the practicability, functionality, and convenience of the system.

4.2. Demand Analysis of Music Education Information System Design

4.2.1. System Design Principles. The design principles of the actual music education information system are completeness, systematicness, and reliability.

Completeness means that the music education information system is fully functional and complete and can fulfill the requirements of music education information. The music education information system must have music data collection, music education information detection, data processing, query, and editing in the system. There should be display, analysis, and execution results. Make it a safe and reliable operation.

Systematicity is mainly reflected in the coordination between the exterior and the various functional modules of the system to form an organic whole.

Reliability refers to the ability to accurately perceive the status changes of the various components of the system and to make accurate judgments and predictions on module queries through detection so that the system can operate stably and reliably.

4.2.2. System Function Requirements. This system focuses on the realization of the design of the integrated platform, the core framework of the integrated platform music education informatization, which plays a role in the data exchange and data integration of each platform. The establishment of this system can clearly understand the deficiencies of music education and realize the further development of the music education industry. The responsibilities of the system are divided into system administrators, user administrators, and data administrators.

5. Model Construction and Effect Analysis

The following uses a confusion matrix to judge the superiority of the model to show the prediction effect of each model.

5.1. Model Building

5.1.1. Logistic Regression. Logistic regression has a regularization item by default, and the value of the penalty parameter can be selected: $L1$ represents $L1$ regularization, $L2$ represents $L2$ regularization, and different regularization items can be selected according to different purposes. In the actual model construction process, if the main purpose is to solve overfitting, the penalty parameter generally selects $L2$ regularization. In special cases, the problem of overfitting may not be solved after $L2$ regularization is used; that is, $L1$ regularization is considered. In addition, $L1$ regularization can also solve the situation where there are more model features, and achieve sparse model coefficients by zeroing the unimportant feature coefficients. Based on this article, in order to be able to compare that model better, we select $L2$ and $L1$, respectively, for regularization settings and get the confusion matrix of the logistic regression as shown in Table 1.

Through the confusion matrix of each model, the correct rate, accuracy rate, recall rate, $F1$ value, and AUC calculated by each model can be obtained, so that the quality of each model can be judged. The specific calculation results are shown in Table 2 and Figure 4.

Through the intuitive presentation of tables and graphs, in the logistic regression model, the effect of logistic regression $L2$ is better than that of logistic regression $L1$.

5.1.2. Random Forest. Random forest requires additional attention: the number of trees' parameters and the number of subdata sets generated by sampling the original data set with replacement, that is, the number of decision trees. If the number of trees is too small, it is easy to underfit, and if the number of trees is too large, the model cannot be improved significantly. Therefore, the number of trees needs to be selected with a moderate value, the default value is 100; the maximum tree depth parameter is the maximum depth of the decision tree. If it is equal to none, it means that the

Table 1: Logistic regression model detection data set confusion matrix.

Model	Real category forecast category	Forecast excellent	Good prediction
Logistic regression $L2$	Forecast excellent	3902	1539
	Good prediction	4023	22890
Logistic regression $L1$	Forecast excellent	5291	1204
	Good prediction	3029	21378

Table 2: Logistic regression model test data set results.

	Correct rate	Accuracy	Recall rate	$F1$ value	AUG
Logistic regression $L2$	0.7900	0.7280	0.4375	0.5628	0.8728
Logistic regression $L1$	0.8427	0.8628	0.5639	0.8762	0.9827

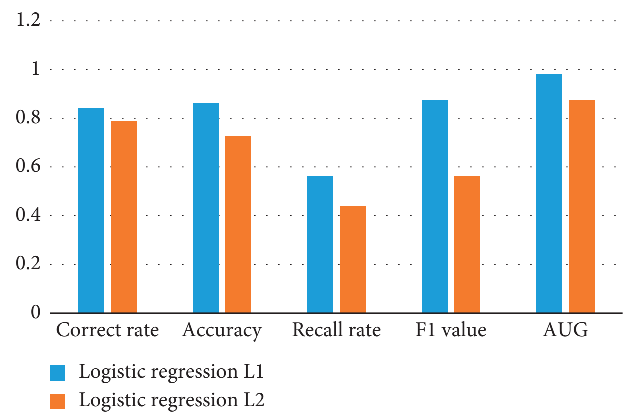

Figure 4: Logistic regression model test data set results.

decision tree will not limit the depth of the subtree when constructing the optimal model. If the model has a large sample size and features, it is recommended to limit the maximum depth; if the sample size is small or the features are few, the maximum depth is not limited; the minimum sample number parameter of the leaf node, that is, the minimum number of samples contained in the leaf node. If the number of leaf node samples is less than this parameter, then the leaf node and sibling leaf nodes are pruned, leaving only the parent node of the leaf node. The default is 50. Since the model fitting effect is not very sensitive to its parameters, parameter tuning is generally to adjust the number of decision trees. This paper keeps the maximum depth of the tree and the minimum sample of the leaf nodes unchanged and sets the number of trees to 90, 100, and 110, respectively, to obtain the confusion matrix of the random forest model, as shown in Table 3:

Judge the quality of each model. The specific calculation results are shown in Table 4 and Figure 5:

Through the visual presentation of tables and graphs, in the random forest model, random forest 3 is better than random forest 1 and random forest 2.

5.1.3. Support Vector Machines. This article uses support vector machines to build predictive models. C, which punishes misclassified training samples in the model, is the penalty factor, which is a manually set parameter, and its default value is 1. Generally, the larger the C, the more accurate the model obtained through training. In this paper, the penalty factor is set to 1, 5, and 20 to obtain the confusion matrix of different models, as shown in Table 5.

Judge the quality of each model. The specific calculation results are shown in Table 6 and Figure 6.

5.1.4. GBDT. This paper selects the GBDT model that comes with the platform in the data factory to configure the parameters. The number of iterations is n_estimators. Generally, if the number of iterations is too small, it is easy to underfit; if the number of iterations is too large, it is easy to overfit. Generally, we choose a moderate value, and the default is 100. The learning rate (learning_rate) is the weight reduction coefficient ν of each weak learner, also called the step size. From the previous chapter, the iterative formula of the strong learner is $f_x(x) = f_{k-1}(x) + \nu h_k(x)$. The value range is $0 \le \nu \le 1$. The closer the ν value is to 0, the weaker learner needs more iterations; conversely, the closer the ν value is to 1, the weaker learner needs fewer iterations. Based on the opposite relationship between the two, when adjusting the parameters of the GBDT algorithm, the weight reduction coefficient and the number of iterations are adjusted together. The parameters can be adjusted from a smaller ν, and the default is 1. The maximum depth of the tree is max_depth; the default value that can be omitted is 3. In the case of a few data features or data, the value may not be adjusted. If the amount of data is large or there are many

Table 3: Confusion matrix of random forest model detection data set.

Model	Real or forecast category	Forecast excellent	Good prediction
Random forest 1	Forecast excellent	582	290
	Good prediction	8379	24579
Random forest 2	Forecast excellent	2763	762
	Good prediction	7638	27622
Random forest 3	Forecast excellent	3729	1204
	Good prediction	8732	21378

Table 4: Random forest model test data set results.

	Correct rate	Accuracy	Recall rate	$F1$ value	AUG
Random forest 1	0.7520	0.7280	0.04375	0.0982	0.8728
Random forest 2	0.8725	0.7023	0.1892	0.2876	0.7992
Random forest 3	0.6728	0.6782	0.2108	0.3598	0.9735

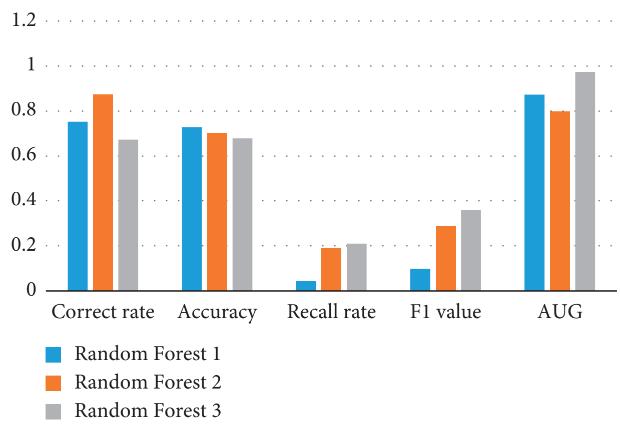

Figure 5: Random forest model test data set results.

Table 5: Confusion matrix of linear SVM model detection data set.

Model	Real or forecast category	Forecast excellent	Good prediction
SVM 1	Forecast excellent	2210	1290
	Good prediction	5289	23579
SVM 2	Forecast excellent	2763	1762
	Good prediction	6500	22622
SVM 3	Forecast excellent	4729	2204
	Good prediction	6732	18378

Table 6: Support vector machine model test data set results.

Model	Correct rate	Accuracy	Recall rate	$F1$ value	AUG
SVM 1	0.7520	0.7280	0.04375	0.0982	0.8728
SVM 2	0.8725	0.7023	0.1892	0.2876	0.7992
SVM 3	0.6728	0.6782	0.2108	0.3598	0.9738

data features, the value should be selected. The general value range is [10, 100] , and the specific value can be determined according to the distribution of the data. The other last parameter is the minimum number of samples of leaf nodes (min_samples_leaf), which is mainly used to limit the minimum number of samples of leaf nodes. If the number of samples is very large, we set a large parameter; if the parameter is set to be less than the number of samples, it will be pruned together with the sibling nodes. This paper selects different parameter combinations based on the GBDT model, as shown in Table 7 and Figure 7.

Through the intuitive presentation of tables and graphs, GBDT3 is better than GBDT1, GBDT2, and GBDT4 in the GBDT model.

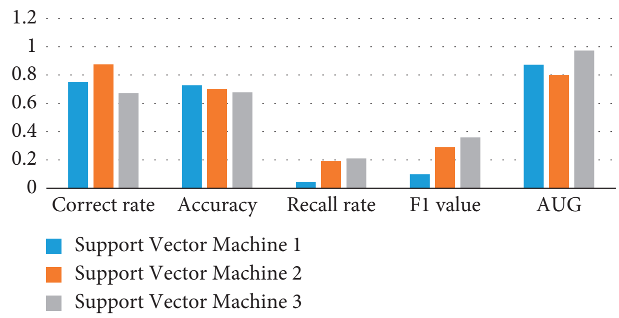

Figure 6: Support vector machine model test data set results.

Table 7: Confusion matrix of GBDT model detection data set.

Model	Real or forecast category	Forecast excellent	Good prediction
GBDT1	Forecast excellent	4472	1790
	Good prediction	4289	23479
GBDT2	Forecast excellent	6763	1362
	Good prediction	2500	22722
GBDT3	Forecast excellent	6753	1187
	Good prediction	2009	23090
GBDT4	Forecast excellent	4729	1099
	Good prediction	6732	21988

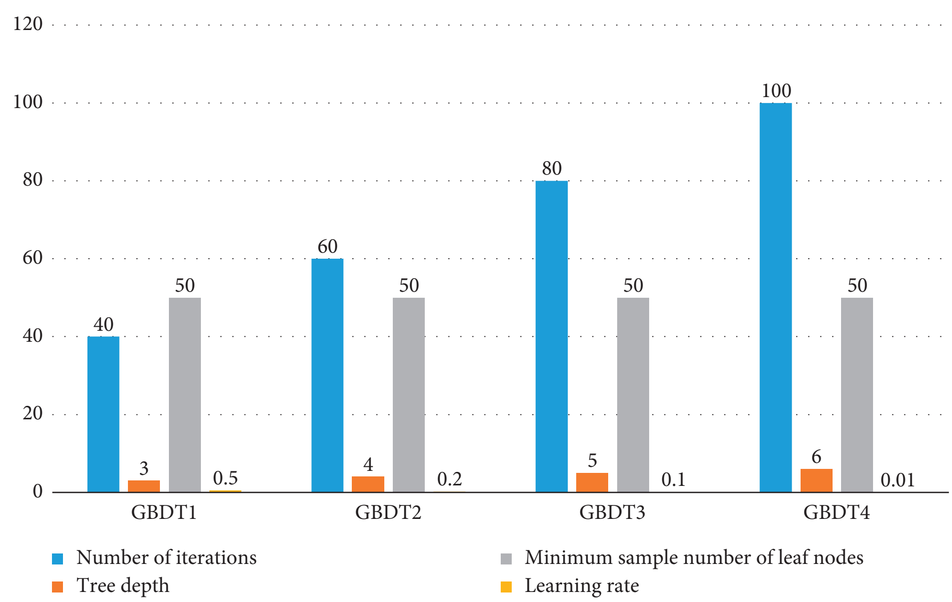

Figure 7: GBDT parameter settings.

Through different settings of model parameters, confusion matrices of different models are obtained, as shown in Table 7.

Judge the quality of each model. The specific calculation results are shown in Table 8 and Figure 8.

To sum up, through the independent comparison of the four models, logistic regression $L2$, random forest 3, support vector machine 2, and GBDT3 are the better models for each test data. By comparing the four models, it can be concluded that the model effect of GBDT3 is the best.

5.2. Effectiveness Analysis. Through the comparative analysis, the effects of the above four models are shown in Figure 9.

The model effect of GBDT3 is the best.

Table 8: GBDT model test data set results.

	Correct rate	Accuracy	Recall rate	*F*1 value	AUG
GBDT1	0.8520	0.7180	0.5375	0.5982	0.8428
GBDT2	0.8910	0.8305	0.7527	0.7798	0.9460
GBDT3	0.8925	0.8423	0.7892	0.7876	0.9492
GBDT4	0.8728	0.7882	0.5108	0.5598	0.8734

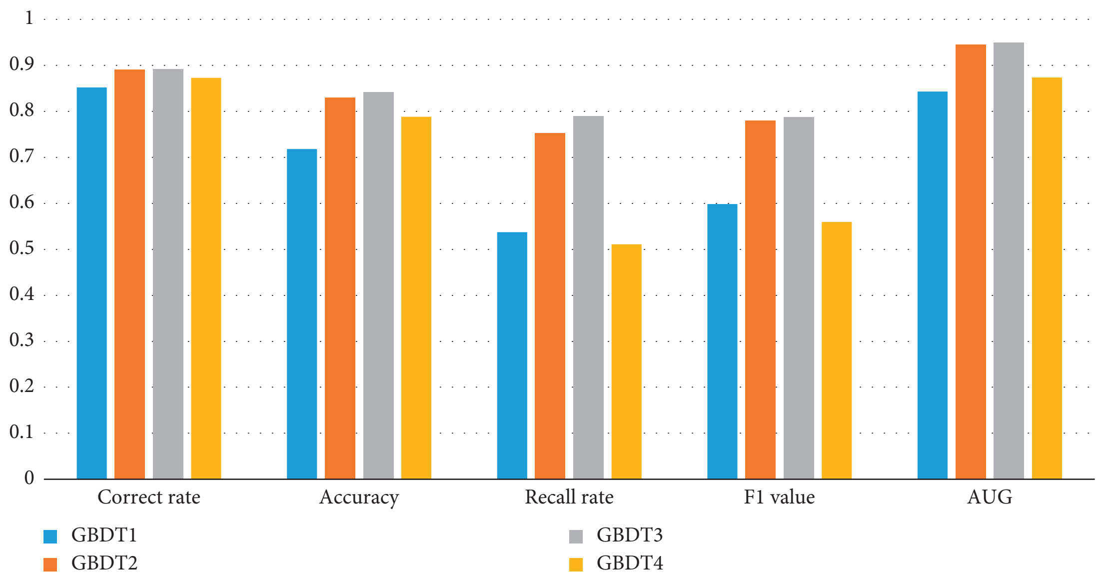

Figure 8: GBDT model test data set results.

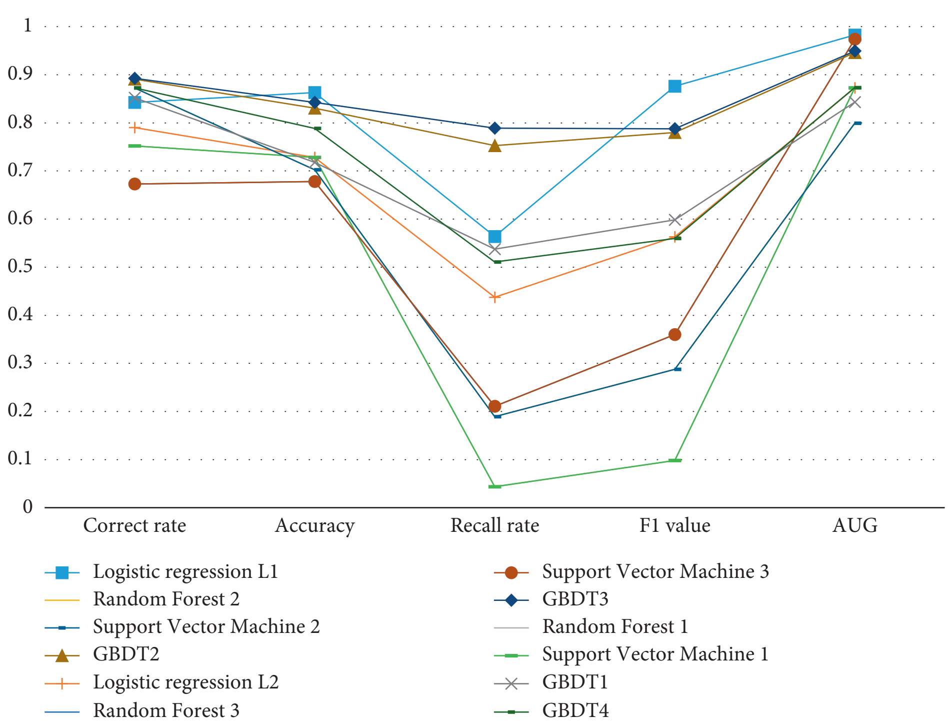

Figure 9: Model test data set results.

6. Conclusion

Through the introduction of related concepts of music education and machine algorithms, a logistic regression model, a random forest model, a support vector machine model, and a GBDT model were constructed for the music education information system based on machine learning. There is a systematic evaluation. Through the comparison of the four models, it is concluded that the construction of the GBDT model is optimal. The model used in this article still has limitations, and it is hoped that better algorithms can be found in the development of machine learning algorithms in the future.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

References

[1] S. Lalit, "Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies," *Computing Reviews*, vol. 57, no. 4, pp. 237-238, 2016.

[2] X. Danyan and X. Guangwen, "Applications based on the principal component analysis of informatization evaluation model in university informatization evaluation," in *International Conference on Information Technology and Management Science(ICITMS 2012) Proceedings*, B. Xu, Ed., Springer, Berlin, Germany, pp. 391–398, 2013.

[3] D. L. Yang and W. M. Chen, "An evaluation model of enterprise informatization level based on grey whitenization weight cluster analysis," *Advanced Materials Research*, vol. 189-193, pp. 823–829, 2011.

[4] J. Du, W. Ning, and S. N. University, "Study on the construction of informatization evaluation model of elementary education from the viewpoint of ecology," *China Educational Technology*, vol. 7, 2014.

[5] S. Hu and J. Zhang, "Cost-benefit analysis and evaluation of education informatization in colleges and universities," *Modern Distance Education Research*, vol. 188, no. 12, pp. 6278–6286, 2012.

[6] L. Zhou, F. Shan, P. Liu, and Y. Wang, "The establishment of the enterprise informatization evaluation model based on supply chain," *Journal of Xi'an Polytechnic University*, vol. 29, no. 6, pp. 772–779, 2015.

[7] G. Chen and S. Li, "Research on location fusion of spatial geological disaster based on fuzzy SVM," *Computer Communications*, vol. 153, pp. 538–544, 2020.

[8] Z. Weimei, "Selection and planning of human resource management informatization system," in *Proceedings of the 2011 National Academic and Teaching Seminar on Electronic Information Specialty in Higher Vocational Education*, pp. 250–252, Wuxi, China, September 2011.

[9] B. Huang, C. Zhu, J. Huang, and F. Li, "Performance evaluation of informatization construction on teaching administrative system in colleges from user's perspective," *China Educational Technology*, vol. 380, no. 9, pp. 69–74, 2018.

[10] W. Wang, "Community informatization evaluation: model construction and analysis," *Information Studies: Theory and Application*, 2011.

[11] A. Rahman and S. Tasnim, "Application of machine learning techniques in aquaculture," *International Journal of Computer Trends and Technology*, vol. 10, no. 4, 2014.

[12] T. Guo and G. Y. Li, "Neural Data Mining for Credit Card Fraud detection," in *Proceedings of the International Conference on Machine Learning & Cybernetics*, pp. 3630–3634, IEEE, Kunming, China, July 2008.

[13] G. Chen and S. Li, "Network on chip for enterprise information management and integration in intelligent physical systems," *Enterprise Information Systems*, vol. 15, no. 7, pp. 935–950, 2021.

[14] J. Zubek, M. Tatjewski, A. Boniecki, M. Mnich, S. Basu, and D. Plewczynski, "Multi-level machine learning prediction of protein-protein interactions inSaccharomyces cerevisiae," *PeerJ*, vol. 3, no. 1, Article ID e1041, 2015.

[15] P. M. Tag and J. E. Peak, "Machine learning of maritime fog forecast rules," *Journal of Applied Meteorology*, vol. 35, no. 5, pp. 714–724, 2010.

[16] A. Matoga, R. Chaves, P. Tomas, and N. Roma, "A Flexible Shared Library Profiler for Early Estimation of Performance Gains in Heterogeneous systems," in *Proceedings of the International Conference on High Performance Computing and Simulation*, pp. 461–470, IEEE, Helsinki finland, July 2013.

[17] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning—a new frontier in artificial intelligence research (research frontier)," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, 2010.

[18] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *NeuroImage*, vol. 56, no. 2, pp. 387–399, 2011.

[19] N. Adams, "Dataset shift in machine learning," *Journal of the Royal Statistical Society: Series A*, vol. 173, no. 1, p. 274, 2010.

[20] M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[21] D. C. Duro, S. E. Franklin, M. G. Dubé, and G. Dubé, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sensing of Environment*, vol. 118, no. 6, pp. 259–272, 2012.

[22] H. Wang, "Research on real-time reliability evaluation of CPS system based on machine learning," *Computer Communications*, vol. 157, no. 1, 2020.