

Research Article

Application of Random Forest Algorithm in Physical Education

Qingxiang Xu¹ and Jiesen Yin² 

¹Jiangnan University, Wuxi, Jiangsu Province 214144, China

²Wuxi Institute of Technology, Wuxi, Jiangsu Province 214121, China

Correspondence should be addressed to Jiesen Yin; yinjs@wxit.edu.cn

Received 19 July 2021; Revised 21 August 2021; Accepted 29 August 2021; Published 10 September 2021

Academic Editor: Mian Ahmad Jan

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

Learning has been a significant emerging field for several decades since it is a great determinant of the world's civilization and evolution, having a significant impact on both individuals and communities. In general, improving the existing learning activities has a great influence on the global literacy rates. The assessment technique is one of the most important activities in education since it is the major method for evaluating students during their studies. In the new era of higher education, it is clearly stipulated that the administration of higher education should develop an intelligent diversified teaching evaluation model which can assist the performance of students' physical education activities and grades and pay attention to the development of students' personalities and potential. Keeping the importance of an intelligent model for physical education, this paper uses factor analysis and an improved random forest algorithm to reduce the dimensions of students' multidisciplinary achievements in physical education into a few typical factors which help to improve the performance of the students. According to the scores of students at each factor level, the proposed system can more comprehensively evaluate the students' achievements. In the empirical teaching research of students' grade evaluation, the improved iterative random forest algorithm is used for the first time. The automatic evaluation of students' grades is achieved based on the students' grades in various disciplines and the number of factors indicating the students' performance. In a series of experiments the performance of the proposed improved random forest algorithm was compared with the other machine learning models. The experimental results show that the performance of the proposed model was better than the other machine learning models by attaining the accuracy of 88.55%, precision of 88.21%, recall of 95.86%, and *f1*-score of 0.9187. The implementation of the proposed system is anticipated to be very helpful for the physical education system.

1. Introduction

Education is an important and a fascinating field which grows over the years and has a significant impact on everyone's life. Numerous techniques and methods were proposed to develop high quality experiences that benefit the entire educational sector, starting with learning and advancing toward E-learning. The route to improvement often needs extra hands to bring alternative perspectives and modifications, which can be done with the help of the crowd. Student/learner assessment is one of the most important practices in the educational field. Educational assessment is a technique used to evaluate a learner's degree of knowledge and enhancing his or her learning during the course. This process has a significant impact on students' motives,

development, and learning practices and has been called "one of the most powerful forces impacting education" by Crooks [1]. Similarly, Harlen et al. [2] asserted that "assessment feedback plays a significant role in predicting future learning," as it has a direct impact on students' performance and persistent effort in future projects.

In the earlier approaches of students and teachers performance evaluation systems, in most of the cases, people have used the total score or average score to evaluate the students in the college teacher evaluations throughout the last few decades. This is a very simple and easy method to implement and is still the fundamental evaluation method in numerous colleges and universities in different countries. However, in today's era of advocating students' personalized and diversified development, especially in profession

selection, it is very important to fully understand students' ability, characteristics, and comprehensive indicators in all aspects, and many people pay attention to and study them in the recent years [3–5].

In the higher education evaluation methods, it has been clearly pointed out that a reasonable and scientific evaluation system should be established for college students, including evaluation concepts, evaluation contents, evaluation form, and evaluation system. The evaluation process needs to be focused on both the student's mathematical learning as well as the learning process. The student's learning should not be the only point of attention, but the changes of their emotional attitude in the activities should be kept in consideration [6]. Among the key goals of curriculum reforms in China one is to establish an assessment system with multiple objectives and methods. The primary goal of evaluation is to fully understand the process and results of students' learning and encourage students to learn and give feedbacks to the teacher, in order to improve the teachers' teaching strategies [7–10]. Therefore, how to investigate an effective teaching assessment method in order to better serve teaching is a problem that needs more investigation. At the same time, the grade evaluation of students' performance is the most important content in teaching evaluation from beginning till now.

Machine learning (ML) is an important field of artificial intelligence (AI) which can be applied in various fields such as healthcare, industries, agriculture, media, and education, etc., and it helps to improve all of its connected tasks by giving real-time responses that save time and eliminate the need for manual intervention of users in an effective manner [11, 12]. Various methodologies are being used in the education sector to attain their goals, including various ML supervised learning algorithms and natural language processing (NLP) techniques. Therefore, in order to overcome the problems of the traditional systems, this study uses classical multivariate statistical method factor analysis and an intelligent ML iterative random forest algorithm. Further, the establishment of a comprehensive assessment model of physical education teaching is of great theoretical and practical significance for the real-time grasp of students' learning situation and the improvement of teaching methods. When we have a reasonable "evaluation model of students comprehensive performance," we can more intuitively understand which factors play a leading role in students' performance, and then we can make a comprehensive and objective evaluation of students' physical education teaching performance. Through the understanding of students' learning situations, the teachers can teach more targeted stuff and content and can help every student as much as possible. At present, the research of data mining and ML in the field of education mainly focuses on the exploration of learning environments, network-based teaching systems, improving students' performance, and other fields. In the earlier approaches the application of information mining for student evaluation is almost negligible which indeed is a serious problem. This paper studies the potential of data mining and ML in the measurement of teachers' performance perceived by the students. Four

commonly used ML classifiers, namely, decision tree (DT), support vector machine (SVM), Naïve Bayes (NB), and random forest (RF), are selected to model the dataset of students' online evaluation course information. Further, the performance of various classification techniques is compared [13–16]. The main contribution of this research study is given below:

- (1) This study proposes a novel approach of factor analysis and random forest algorithm to reduce the dimensions of students' multidisciplinary achievements in physical education into a few typical factors.
- (2) Several experiments have been performed for various ML classification algorithms in order to check the performance and stability of the proposed system.
- (3) Performance of all the utilized ML models has been evaluated in terms of numerous performance measurement metrics such as accuracy, precision, recall, and *f1*-score.
- (4) This study recommends which ML classifier is more feasible in order to develop a high level intelligent system for the student physical education.
- (5) From the experimental results it is obvious that the performance of the proposed system is much better than the traditional student physical education systems.

The remaining paper is structured in the following order. Section 2 represents the related work. Section 3 shows the material and methods. Section 4 illustrates the experimental results and analysis. Section 5 concludes the proposed work.

2. Related Work

The research field is one of the most important parts of education and has a great influence on the education of a country. The paucity of resources needed to conduct professional research can be avoided by assembling a group of researchers who can pool their resources and do joint study. Numerous platforms have been established to assist students in sharing their ideas and contributing to a big research group comprised of many researchers, each of whom performs a different character [17]. Enhancing the research field could usher education into a new era by bringing together all of the knowledge from many fields of expertise, raising the educational instruments to a new level. Further, it can also assist in giving the ideal learning experience by investigating all the flaws in various areas and can offer valuable suggestions to improve them. The mining contacts among students in online education are another serious issue of education. One of the proposed systems addresses this problem by providing a peer-to-peer debate system that lets the students from various regions share their knowledge depending on their own areas and countries [18]. Another framework was discovered to improve online course interaction by allowing the crowdsourcing technique to provide timely response on online student submissions [19]. All the proposed systems have the potential to increase the

learning process and provide opportunities to those students who have limited learning opportunities and are depending only on the online courses for continuing their education. As a result, concentrating on student engagement is critical to the learning process' efficiency. As a result, because the offered solutions are entirely reliant on nonexpert personnel, the process must be closely monitored to avoid supplying incorrect information or providing inaccurate feedback.

In the learning process, establishing a personalized learning goal is indeed a critical task. To accomplish this purpose, Whitehill and Seltzer [20] proposed a platform for online learning videos where crowd-sourced videos are developed and filmed by common people rather than professional professors. Faisal et al. [21] proposed a framework that assists the learners in selecting the appropriate learning plan based on a recommendation system, where the desired learning activity is different depending on the learner's attributes. Alghamdi et al. [22] proposed a system in which they attempted to produce exam questions of high quality. To produce large-scale tests in a professional context, this platform can incorporate as many teachers as feasible in the process of creating and evaluating question items. One of the approaches presented by De Alfaro and Shavlovsky [23] allows students to submit assignments and review and grade them collectively. This technique provides students with an overall crowd-grade, which assures the quality of their homework as well as their reviewer effort. The students can answer the test questions on another system, which uses a specific algorithm to assess the complexity and validity of the questions generated. [24]. In a similar manner, Pirttinen et al. [25] presented an embedded tool for online courses that allows students to generate assignments while also reviewing and evaluating each other's work. Farasat et al. [26] proposed the concept of crowd-learning, in which students might learn more deeply by generating their own educational materials. They presented an online platform as a solution for it that can be utilized for in-class practice or online classes. Among all of the previous ways, the evaluation process, which is the primary means of evaluating any student, can become extremely delicate. Total reliance on the audience for grading may result in grades that are unfair or erroneous.

ML can be implemented into any field to improve all of its related activities by giving real-time replies that save time and eliminate the need for manual intervention. To reach their aims in the education sector, many systems are leveraging different ML models such as supervised or natural language processing (NLP). For their preferred schemes, many approaches used supervised ML models. A system was presented to autonomously examine the effectiveness of community question answers, which uses a classification technique to evaluate the quality of every answer by examining a set of various criteria [27]. Classification algorithms are mostly supervised learning algorithms that learn from the patterns of the input data and then produce output based on the learned knowledge. The proposed method begins with feature extraction and the collection of historical data about the community member, which will aid in the development of the categorization model. Finally, the

trained models will be used to assess the quality of all new answers. Li [28] suggested a method that used a supervised model and followed a regression technique, with the output variable having a continuous and real value. By using the model to identify the key variables affecting the student's learning performance, the author hoped to gain a better understanding and analysis of the reasons behind each student's results. As a result, teachers will be able to track the learning effect and change their teaching technique to meet the demands of their students using this method. The supervised model is completely reliant on the teaching practice and previous experiences. The offered collection of data is the user's previous experience with a computer language. If the improper collection of data is used to train, the results may be inaccurate, affecting the suggested approach's intended use. As a result, the problem for each new educational research is to choose the correct data with the right model because it is critical to its success. In the studied literature, one of the platforms that employs NLP is a web-based application that visualizes the output from a ML-based model trained to guess MCQs, with the primary purpose of identifying and manipulating questions within a well-organized and high quality bank of MCQs [29]. Furthermore, NLP is utilized in the exam evaluation process to compare the similarity of students' answers to the perfect answer when the ideal solutions are accessible. The proposed method will compute the student's recommended score as a consequence of this comparison [30]. This paper investigated the potential of ML in the measurement of teacher's performance perceived by the students. Four ML models, i.e., NB, SVM, DT, and improved RF, are used on the students' online evaluation course information dataset. The improved RF showed sublime performance in terms of various performance measures.

3. Material and Methods

This section of our study represents the methods followed and the material used for the conduction of the research study.

3.1. Data Set Collection and Characterization. The data collection process is an important step for building an intelligent system. In this study, the recent 3-year student physical education data is selected which comprises 3216 randomly selected instances in which 2000 were positive samples while the remaining 1216 were negative samples of data. Each data instance has 28 attributes, which come from students' online scores, involving physical education teaching preparation, physical education teaching performance, teaching methods implementation, case organization, curriculum implementation, teachers' attitude, curriculum construction, and so on. The evaluation indexes are shown in Table 1.

3.2. The Classification Models Used in This Study. One of the most popular applications of data mining and ML is the classification. The main task of classification is to assign a

TABLE 1: Description of different evaluation indexes.

Attributes	Description	Value range
V1	How often do you take this course?	1~5
V2	The teacher is well prepared before class?	1~5
V3	Does the teacher come to class on time?	1~5
V4	Do teachers often encourage students to ask questions? Participate in the discussion?	1~5
V5	The teacher has a good grasp of the subject?	1~5
V6	Is the teacher approachable and easy to communicate?	1~5
V7	Is the teacher effective in organizing the materials in the demonstration, lectures, and discussions?	1~5
V8	Homework helps me?	1~5

class label among possible categories for a sample represented by a set of feature vectors and is accomplished by a classification model. The model constructs a learning algorithm on a training set, in which the class label of each instance in the training set is known before training. At the end of the learning phase, the test set is used to evaluate the performance of the classification model. Decision tree algorithm is one of the classical ML classification algorithms, which classifies and induces data through a top-down and clear-cut process. The purpose of the decision tree algorithm is to recursively divide the observation results into mutually exclusive subgroups until there is no difference in the given statistics. Information gain, gain ratio, and Gini index are the most commonly used statistics for finding classification attributes of different nodes of the tree. Generally, iterative dichotomy (ID3) uses information gain, C4.5 and C5.0 (the successor of ID3) use gain ratio, and classification and regression tree (CART) uses Gini index. Support vector machine (SVM) tries to find a hyperplane to separate classes, minimize the classification error, and maximize the edge. SVM is a good classification and regression technology proposed by Vapnik at Bell Laboratories. SVM has four kernel types including linear, rbf, sigmoid, and polynomial. The use of kernel type depends on the nature of the problem. Among the kernel types linear and rbf are the most commonly used kernel types of SVM. Naïve Bayes (NB) classifier classifies samples by calculating the probability that an object belongs to a certain category. The theoretical basis of classification is Bayesian theorem. According to the Bayesian formula, the posterior probability is calculated according to the prior probability of an object, and the class with the largest posterior probability is selected as the class of the object. In other words, Bayesian classifier is the optimization in a sense of minimum error rate. Random forest (RF) is a kind of ensemble learning classification algorithms, which integrate the classification effect of multiple decision trees. It consists of multiple base classifiers, each of which is a decision tree (DT). Each DT is used as a separate classifier to learn and predict independently. Finally, these predictions are integrated to get the total prediction which is better than a single classifier. Figure 1 shows the basic diagram of the utilized ML classification models training and testing process.

3.3. The Proposed Random Forest Classification Algorithm. Random forest (RF) algorithm is an integrated learning method proposed by Leo Breiman and Adele Cutler, which

means that it is composed of many small submodels, and the output of each small submodel is combined to give the final output. RF algorithm is a typical ML algorithm, which is usually used for classification, regression, or other learning tasks. The RF algorithm is based on bagging algorithm to group data from the original dataset. After training for each group, the corresponding decision tree model is obtained. Finally, all the decision data results of the subsmall models are combined and analyzed to get the final RF model. The final prediction result of the RF algorithm is based on the voting algorithm, and the classification with the largest number of votes is the final output of the RF algorithm [31–35].

By using multiple classifiers for voting classification, RF algorithm can effectively reduce the error of a single classifier and improve the classification accuracy [36–42]. Practical experience shows that, compared with artificial neural network (ANN), regression tree, SVM, and other algorithms, RF algorithm has higher stability and robustness, and the corresponding classification accuracy is also in the leading level. RF algorithm is efficient for large-scale data processing and can adapt to high-dimensional data application scenarios. At the same time, it can still maintain high classification in missing data scenarios. The style and working process of RF algorithm are shown in Figure 2.

Compared with other classification algorithms, RF algorithm has better classification performance. It can process large-scale data, support large-scale variable parameters, and intuitively evaluate the importance of variable features. More and more algorithm competitions and practices have proved that the RF algorithm has a high classification performance and has better robustness and stability while maintaining high efficiency [43–46].

3.4. Applications of Different ML Algorithms in the Prediction of Students Course Performance

3.4.1. Description of Student Achievement Dataset. There are many factors that affect the physical education teaching. Among them, there are some uncontrollable factors and controllable factors, which directly or indirectly affect students' performance. This study attempts to integrate the improved RF algorithm into the prediction data of physical education teaching, through the improved RF algorithm to more accurately predict the students' physical performance in order to focus on the factors that affect students'

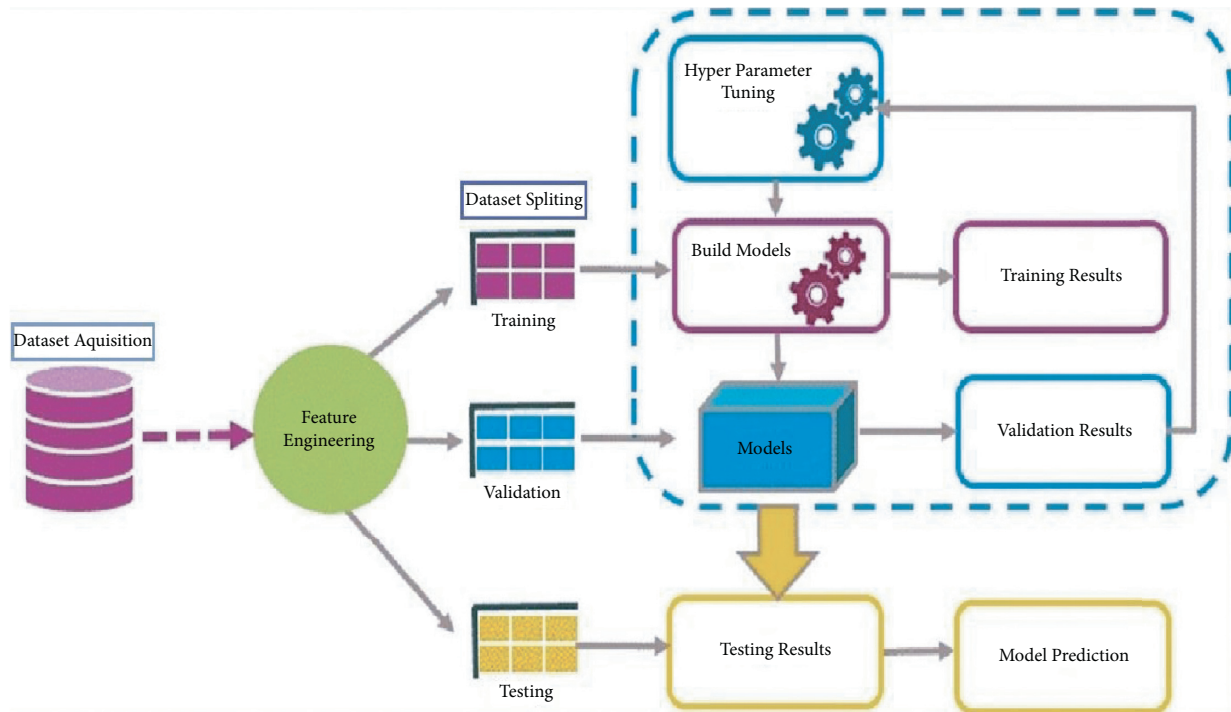


FIGURE 1: Basic diagram of the utilized ML classification algorithms.

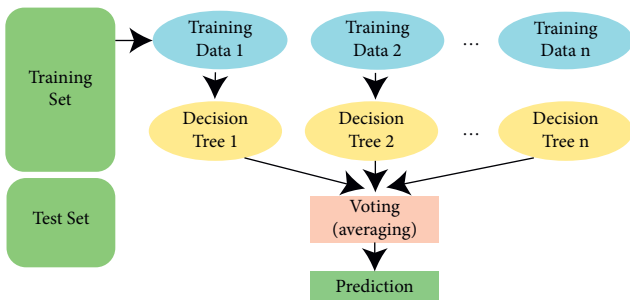


FIGURE 2: Working process of the RF algorithm.

performance and to focus on the process of curriculum reform. We firmly believe that the continuous cycle of this prediction practice improvement way can promote the school teaching reform, promote the level of school physical education to be more scientific and efficient, at the same time make the students' performance more excellent, and enhance the competitiveness of students' job opportunities after graduation. The description of student achievement dataset is shown in Table 2.

In this paper, the dataset selected for the experimental work is the student achievement dataset. The dataset of students' achievement and characteristics was collected from the course of college students' public physical education. The dataset includes 9 characteristics/attributes of students. The nine characteristics are divided into three categories, i.e., students' statistical characteristics, educational background characteristics, and students' behavior characteristics. The classification of students' behavior characteristics includes students' activity in the class, i.e., students' absence, the number of times students visit teaching resources after the

class, the number of times the students participated in the course discussion, and students' satisfaction from the course. In addition, the course also collects students' course scores, which are divided into two categories: positive (score between 60 and 100) and negative (score below 60).

3.4.2. Normalization and Characterization of Dataset Features. The dataset is collected by the author from daily physical education work, which involves multidimensional original data collection. The work involves a long time range, and the workload is relatively large. Then, the data preprocessing for the original data is carried out, including data cleaning, data discretization, removal of missing values, data filtering, and so on. Due to the differences and diversity of the expression forms of the collected data, it is necessary to preprocess the data: for unreasonable data or illegal data, the records should be cleaned out. For a wide range of characteristic data, "students' curriculum activity," discretization, and normalization are needed.

Feature normalization is mainly to scale and normalize the feature value, so that the feature value is reduced to a specific range, such as $[-1.0, 1.0]$ or $[0, 1.0]$. The work of feature normalization helps to reduce the excessive difference of different feature value ranges and the dependence of the algorithm on feature measurement units. Feature normalization mainly uses linear transformation, log transformation, Tan transformation, and other methods for data standardization, so as to transform the data into a small common space.

For the data of students' comprehensive scores, the characteristic variables need to be planned, while the other characteristic variables are calculated with the original

TABLE 2: Characteristics of student achievement dataset.

Serial number	Meaning	Features	Value description
1	Teaching objectives	Mubiao	{0: OK; 1: Bad}
2	Teaching method	Fangfa	{0: OK; 1: Bad}
3	Exercise intensity	Yundong	{0: OK; 1: Bad}
4	Organizing teaching	Zuzhi	{0: Bad; 1: General; 2: OK}
5	Demonstration capability	Shifan	{0: OK; 1: Bad}
6	Responsibility	Zeren	{0: OK; 1: Bad}
7	The atmosphere of physical education	Qifen	{0: OK; 1: Bad}
8	Venue equipment	Qicai	{0: No; 1: Yes}
9	Teaching effectiveness	Xiaoguo	{0: OK; 1: Bad}

values. For the characteristics of students' absence times, the feature programming method mainly uses the maximum and minimum normalization method and linear transformation of memory. Next, we choose the most representative student sports teaching activity as an example to illustrate the complex characteristics of standardization. For the characteristics of students' active degree of physical education teaching, it is mainly to collect the number of users' hands raised to answer the questions, the number of exchanges in physical education teaching, and the degree of personal physical education teaching concentration for comprehensive evaluation. In the process of implementation, the discrete data is normalized to [0, 1] based on the log function standardization method, and each factor is assigned the corresponding weight coefficient, so as to get the data of students' physical education activity. The expression of students' active degree in physical education teaching is given as follows:

$$\begin{aligned} \text{KTScore} = & W_{js} * \frac{\log_{10}(JSx)}{\log_{10}(\max JS)} + W_{jl} * \frac{\log_{10}(JLx)}{\log_{10}(\max JL)} \\ & + W_{ZZ} * \text{ZZFactor}, \end{aligned} \quad (1)$$

where W_{js} , W_{jl} , and W_{ZZ} represent the weight coefficients of the three factors, which are set accordingly to 0.5, 0.2, and 0.3 by default, and ZZFactor indicates that the teachers evaluate the students' concentration subjectively, and its value is also between 0 and 1.

3.5. Performance Evaluation Metrics. There are many performance measures to evaluate the performance of the classification models according to the correctness of classification decision. Suppose that, in a binary class task, class variable values can be assumed to be positive (P) or negative (N). The actual positive cases (P) correctly classified by the model as positive cases are named as true positive (TP) cases, and when the actual positive cases are wrongly classified by the model as negative cases they are named as false negative (FN) cases. In a similar way, the actual negative cases (N) correctly marked as negative cases by the model are regarded as true negative (TN) cases, while the actual negative cases wrongly marked as positive cases by the model are regarded as false positive (FP) cases. These terms are given in the confusion matrix as shown in Table 3.

TABLE 3: Confusion matrix.

	Predicted (-)	Predicted (+)
Actual (-)	TN	FP
Actual (+)	FN	TP

Formulas (2)–(5), respectively, give the calculation results of performance metrics, such as accuracy, precision, recall, and $F1$ -score.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} * 100\%, \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} * 100\%, \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} * 100\%, \quad (4)$$

$$F1 \text{ score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \quad (5)$$

4. Experimental Results and Analysis

This section of the paper demonstrates the experimental results attained via different ML classification models. The performance of each ML model is then analyzed with the help of various performance metrics. Figure 3 shows the platform infrastructure of the experimental environment. The platform consists of both the hardware and software components and is carried out via various software configurations on the basis of hardware environment. PostgreSQL is used as a data storage platform to store the data sets and provides data sources for the R language programs. R is used as a language for the implementation of all the ML classification models and R-Studio IDE environment is used to write the relevant algorithm code to realize the improved RF algorithm.

The experimental environment of this paper basically consists of the two parts, i.e., software and hardware. The software part is mainly Linux 64-bit operating system, the database is PostgreSQL relational database, the programming language used is the R language, and the IDE environment is R-Studio. In addition, there are some other software, whose configuration and version description are shown in Table 4. The hardware part includes a computer

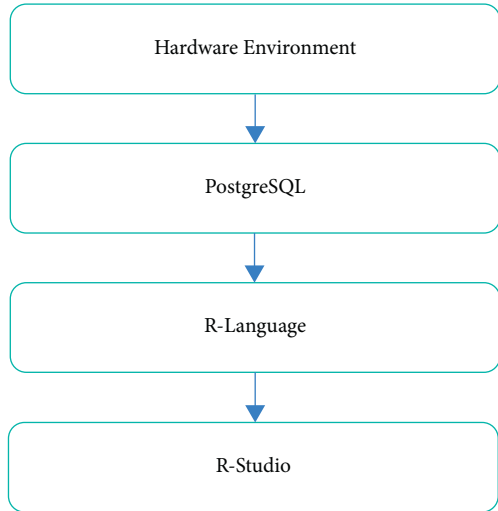


FIGURE 3: Experimental platform infrastructure.

TABLE 4: Software configuration and environment platform.

Function	Software version
Database	PostgreSQL 9.6.12
IDE environment for R development	R-Studio 1.1363
R language	R3.1
Database backup tool	Pg_dump
Database import tool	Pg_restore
Data synchronization tool	Sync
Code management tools	Git
Document authoring tools	VIM, Office

system having the specification: Intel Core i7, 3.4GHz processor, 16 GB RAM, and 256 GB SSD.

4.1. Experimental Results of All the Investigated ML Classification Models. This subsection represents the experimental results attained via the investigated ML classification models using the students’ achievement dataset. Table 5 shows all the experimental results attained by the utilized ML classification models.

Table 5 shows that the proposed iterative RF algorithm performed really well by attaining the accuracy of 88.55%, precision of 88.21%, recall of 95.86%, and f_1 -score of 0.9187. The second best results were observed for the LR model. LR attained the accuracy of 87.99%, precision of 89.19%, recall of 93.90%, and f_1 -score of 0.9148. The lowest performance was observed for the GRNN model. GRNN achieved the accuracy of 84.35%, precision of 86.40%, recall of 90.75%, and f_1 -score of 0.8852.

Figure 4 shows the accuracy, precision, and recall results while Figure 5 illustrates the f_1 -score results of all the investigated ML classifiers used in this study.

From Figures 4 and 5, we can see that the RF algorithm has advantage over the other algorithms in terms of the mentioned performance measures using the student characteristics dataset.

TABLE 5: Experimental results of all the investigated ML classification models.

Measures	LR ($C = 1$)	NB	SVM (kernel = “rbf”)	GRNN	RF
TP	231	214	223	216	232
FN	15	32	13	22	10
FP	28	24	42	34	31
TN	84	88	80	86	85
Accuracy	87.99	84.36	84.64	84.35	88.55
Precision	89.19	89.91	84.15	86.40	88.21
Recall	93.90	86.99	94.49	90.75	95.86
F_1 score	0.9148	0.8842	0.8902	0.8852	0.9187

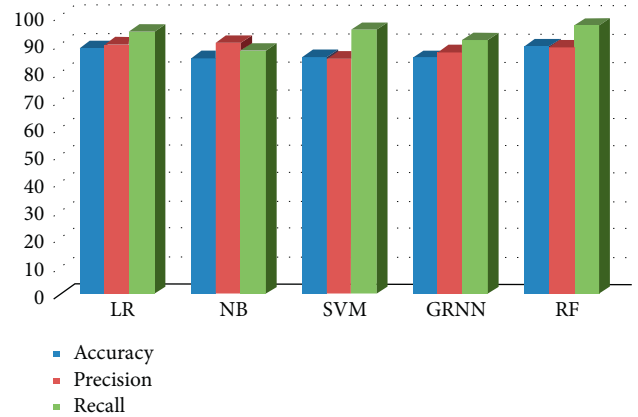


FIGURE 4: Performance (accuracy, precision, and recall) of all the investigated ML models.

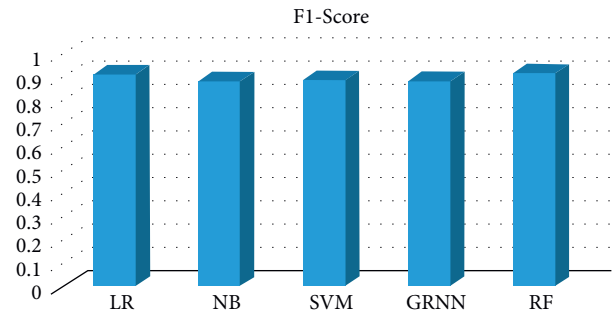


FIGURE 5: F_1 -score of all the investigated ML classification models.

4.2. Algorithm Evaluation and OOB Simulations. In the evaluation process of improved RF algorithm (IRFC), the same parameter configuration was used, and the common parameter settings are maxgen: 20 and $T_{max} = 10$. In the process of algorithm performance evaluation, the average method of 10-fold cross validation was used to calculate the value of each index. For the dataset of students’ characteristics, our main idea is to construct a dataset based on the behavior records of historical students and subject scores (three-year data), select the scores of the first two years as the training set, train the parameters and characteristics based on the improved RF algorithm, and predict the scores of the last year as the test set and compare them with the real score

classification. The scale of the student achievement dataset (three years) was 2002.

Reducing the OOB error is one of the goals of the improved RF algorithm proposed in this paper. Figure 6 shows the OOB error comparison chart of the improved RF algorithm in UCI dataset and student comprehensive score dataset.

The abscissa of the improved RF algorithm is $k = 1, m, \sqrt{M} \log 2(M) + 1$, and the ordinate of the improved RF algorithm is the OOB error value. It can be seen from the figure that the OOB error value of the IRFC is obviously lower than that of the traditional fixed parameter values. Through the verification of the above two datasets, it also shows that the fixed k value of the traditional RF algorithm is not the optimal scheme, which has a great impact on the performance of the algorithm, and its tuning can significantly improve the performance of the algorithm. The improved RF algorithm also provides a reference method to optimize the value of k parameter. Figure 7 shows the relationship between OOB error and the number of decision trees.

In order to verify the influence of the number of decision trees on OOB error, we calculate the corresponding relationship between the number of decision trees and OOB error in the calculation process of the improved RF algorithm, as shown in Figure 6. As can be seen from the figure, OOB error decreases rapidly between 0 and 300 decision trees and reaches stability between 300 and 600 decision trees. From the previous results, the improved RF algorithm achieves the best performance on 339 decision trees.

The improved RF algorithm in the previous sections is developed in R-Studio IDE environment based on R language. It can complete the training in an acceptable processing time when processing the current small-scale datasets, but it will show a very slow model training process when the data scale is relatively large. In addition, since we will carry out many years of data training and prediction of large-scale courses in the future, we must also consider the efficiency of algorithms in large-scale data.

In order to improve the efficiency of the improved RF algorithm, this paper plans to use the task parallelization mode to improve the classification speed of the algorithm. In this study, the parallel processing mode of SparkR is used to process the algorithm, and the data set of students' performance is processed in the cluster environment. The improved RF algorithm of parallel transformation is used to predict the final performance classification of students' physical education teaching. The cluster environment consists of four servers, one of which is the master (driver) node, and the other three are the slave (worker) nodes. The basic configuration of each server is CentOS 6.8, spark version is 2.2, memory is 16 GB, and CPU model is Intel Core i7 and 3.4 GHz processor. In order to implement the improved RF algorithm in parallel based on SparkR, it is necessary to adjust the algorithm properly, so that it can be executed in parallel. The parallel strategy of improved RF algorithm includes the parallel strategy of RF algorithm and the parallel strategy of simulated annealing algorithm.

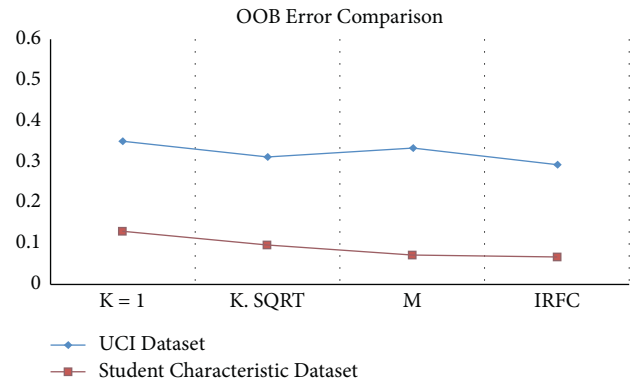


FIGURE 6: Comparison effect of OOB error.

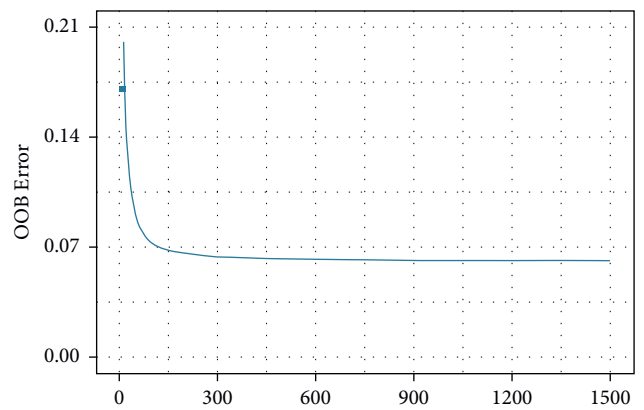


FIGURE 7: Relationship between OOB error and the number of decision trees.

Parallel strategy of RF algorithm: the decision tree construction in the execution process of RF algorithm is evenly distributed to each node in the cluster, so that the construction of decision tree can be executed in different cluster nodes, so as to realize the execution parallelization of decision tree. On each node, the growth of different decision trees can also be executed concurrently.

This experiment is mainly for the improved RF algorithm and the traditional RF algorithm in the student physical education teaching evaluation characteristic dataset test. The two algorithms were implemented on the spark cluster with three computing nodes, and the performance of the two algorithms was compared by using the time measurement method. The specific experimental results are shown in Table 6.

From Table 6 we can see that, after parallel processing, the improved RF algorithm has been shortened from 1486 seconds to 230 seconds, and the efficiency has been improved about 6 times. Due to the addition of simulated annealing algorithm for parameter optimization, the running time of the improved RF algorithm is longer than that of the traditional RF algorithm. In the single machine running process, the running time of the improved RF algorithm is almost three times longer than that of the traditional RF algorithm, but after data parallelization, the improved RF algorithm is only two times longer than the

TABLE 6: Running time comparison of the two algorithms.

	Parallelization run time (s)	Single machine running time (s)
Improved RF algorithm	230	1486
Traditional RF algorithm	132	528

traditional RF algorithm, which is mainly due to the parallel optimization of RF algorithm and simulated annealing algorithm.

5. Conclusion

The field of education is extremely important and has a significant influence on many civilizations. The evaluation of physical education teaching quality is one of the most important methods to improve the physical education teaching system. Keeping the significance of the physical education system in consideration various ML algorithms are used in this study to generate a high quality experienced and intelligent system that will improve the whole physical educational sector. The improved RF algorithm proposed in this study is simulated on the student achievement dataset. Through a number of experiments, it is confirmed that the simulated annealing algorithm, feature selection process, weight optimization, and other processes can provide assistance for the effectiveness of the algorithm. In addition, it can also identify the characteristic factors that have a significant impact on students' course performance and is helpful for the teaching curriculum reform. The improved RF algorithm is used in the evaluation of college physical education classroom and teaching quality, and through the mining and analysis of the survey data, it tries to find the factors that affect the quality of college physical education classroom teaching and provide scientific suggestions for future physical education classroom teaching reform. The future work of this study is to use more optimization and ML techniques to improve the accuracy of the student achievements and improve the physical education system.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] T. J. Crooks, "The impact of classroom evaluation practices on students," *Review of Educational Research*, vol. 58, no. 4, pp. 438–481, 1988.
- [2] W. Harlen, R. D. Crick, P. Broadfoot et al., *A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning*, Evidence for Policy and Practice, London, UK, 2002.
- [3] G. M. Snodgrass, A. B. Rosay, and A. R. Gover, "Modeling the referral decision in sexual assault cases: an application of random forests," *American Journal of Criminal Justice*, vol. 39, no. 2, pp. 267–291, 2014.
- [4] J. I. Woo, J. G. Yang, Y.-H. Lee, and U. G. Kang, "Healthcare decision support system for administration of chronic diseases," *Healthcare Informatics Research*, vol. 20, no. 3, p. 173, 2014.
- [5] J.-G. Yoon, N. Zarayeneh, and S. C. Suh, "Interrelationship between the general characteristics of Korean stroke patients and the variables of the sexual functions: random forest and boosting algorithm," *Journal of Physical Therapy Science*, vol. 29, no. 4, pp. 613–617, 2017.
- [6] D. Kim, O. Hilliges, and S. Izadi, *Three-Dimensional User Interaction*, Infotechopen, London, UK, 2012.
- [7] Z. Zhang, Y. Sun, and T. Xiaoqing, *Predictive Value of Random Forest Algorithms for Diabetic Risk in People Underwent Physical Examination*, Chinese General Practice, Shanghai, China, 2019.
- [8] B. C. Blair and A. Hartman, *Predicting Local Violence*, Social Science Electronic Publishing, Rochester, NY, USA, 2014.
- [9] Y. Wang, C. Song, and S. T. Xia, "Unifying decision trees split criteria using tsallis entropy," *Computer Science*, vol. 21, 2015.
- [10] K. Devendran, S. K. Thangarasu, P. Keerthika, R. Manjula Devi, and B. K. Ponnarasee, "Effective prediction on music therapy using hybrid SVM-ANN approach," *ITM Web of Conferences*, vol. 37, no. 1, p. 01014, 2021.
- [11] Y. Muhammad, M. Tahir, M. Hayat, and K. T. Chong, "Early and accurate detection and diagnosis of heart disease using intelligent computational model," *Scientific Reports*, vol. 10, pp. 19747–19817, 2020.
- [12] Y. Muhammad, M. D. Alshehri, W. M. Alenazy, T. Vinh Hoang, and R. Alturki, "Identification of pneumonia disease applying an intelligent computational framework based on deep learning and machine learning techniques," *Mobile Information Systems*, vol. 2021, Article ID 9989237, 20 pages, 2021.
- [13] S. G. Trost, "Physical activity classification in youth using raw accelerometer data from the hip," *Measurement in Physical Education and Exercise Science*, vol. 24, 2020.
- [14] J. Sun, S. Cong, H. Mao, H. Wu, and N. Yang, "Quantitative detection of mixed pesticide residue of lettuce leaves based on hyper spectral technique," *Journal of Food Process Engineering*, vol. 41, no. 2, 2018.
- [15] L. Iughetti, P. Bruzzi, and B. Predieri, "Evaluation and management of hyperlipidemia in children and adolescents," *Current Opinion in Pediatrics*, vol. 22, no. 4, pp. 485–493, 2010.
- [16] N. K. Shrestha and S. Shukla, "Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment," *Bioresource Technology*, vol. 128, no. 2, pp. 351–358, 2013.
- [17] S. N. S. Rajan Vaish, G. Kovacs, A. Veit et al., *Crowd Research*, Open and Scalable University Laboratories, HangZhou, China, 2017.
- [18] Y. Kotturi, C. E. Kulkarni, M. S. Bernstein, and S. Klemmer, "Structure and messaging techniques for online peer learning systems that increase stickiness," in *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pp. 31–38, Vancouver, Canada, March 2015.
- [19] M. Krause, M. Mogalle, H. Pohl, and J. J. Williams, "A playful game changer: fostering student retention in online education with social gamification," in *Proceedings of the Second (2015)*

- ACM conference on Learning@ Scale, pp. 95–102, Vancouver, Canada, March 2015.
- [20] J. Whitehill and M. Seltzer, “A crowdsourcing approach to collecting tutorial videos--toward personalized learning-at-scale,” in *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, pp. 157–160, Cambridge, MA, USA, April 2017.
- [21] M. H. Faisal, A. W. AlAmeeri, and A. A. Alsumait, “An adaptive e-learning framework: crowdsourcing approach,” in *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, pp. 1–5, Brussels, Belgium, December 2015.
- [22] E. A. Alghamdi, N. R. Aljohani, A. N. Alsaleh, W. Bedewi, and M. Basher, “CrowdyQ: a virtual crowdsourcing platform for question items development in higher education,” in *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, pp. 1–4, Brussels, Belgium, December 2015.
- [23] L. De Alfaro and M. Shavlovsky, “CrowdGrader: a tool for crowdsourcing the evaluation of homework assignments,” in *Proceedings of the 45th ACM Technical Symposium on Computer science education*, pp. 415–420, Atlanta, GA, USA, March 2014.
- [24] A. Taniguchi and S. Inoue, “A method for automatic assessment of user-generated tests and its evaluation,” in *Proceedings of the Adjunct 2015 ACM international Joint cConference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pp. 225–228, Osaka, Japan, September 2015.
- [25] N. Pirttinen, V. Kangas, I. Nikkarinen, H. Nygren, J. Leinonen, and A. Hellas, “Crowdsourcing programming assignments with CrowdSorcerer,” in *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, pp. 326–331, Larnaca, Cyprus, July 2018.
- [26] A. Farasat, A. Nikolaev, S. Miller, and R. Gopalsamy, “Crowdlearning: towards collaborative problem-posing at scale,” in *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, pp. 221–224, Cambridge, MA, USA, April 2017.
- [27] L. T. Le, C. Shah, and E. Choi, “Evaluating the quality of educational answers in community question-answering,” in *Proceedings of the 2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*, pp. 129–138, Austin, TX, USA, June 2016.
- [28] Y. Li, “An application of EDM: design of a new online system for correcting exam paper,” in *Proceedings of the 2018 13th International Conference on Computer Science & Education (ICCSE)*, pp. 1–6, Colombo, Sri Lanka, August 2018.
- [29] P. Watson, T. Ma, R. Tejwani, M. Chang, J. Ahn, and S. Sundararajan, “Human-level multiple choice question guessing without domain knowledge: machine-learning of framing effects,” in *Proceedings of the Companion Proceedings of the the Web Conference 2018*, pp. 299–303, Lyon, France, April 2018.
- [30] A. Kashi, S. Shastri, A. R. Deshpande, J. Doreswamy, and G. Srinivasa, “A score recommendation system towards automating assessment in professional courses,” in *Proceedings of the 2016 IEEE Eighth International Conference on Technology for Education (T4E)*, pp. 140–143, Mumbai, India, December 2016.
- [31] A. Rajwa-Kuligiewicz, R. J. Bialik, and P. M. Rowiński, “Dissolved oxygen and water temperature dynamics in lowland rivers over various timescales,” *Journal of Hydrology and Hydromechanics*, vol. 63, no. 4, pp. 353–363, 2015.
- [32] R. L. Dart and S. M. Ausbrooks, *Earthquakes in Arkansas and Vicinity 1699–2010*, Dongzer University, Bikaner, India, 2011.
- [33] I. Period, “MAC medium access control zhibin Wu lecture overview introduction random access aloha slotted aloha CSMA CSMA/CD,” *Journal of Transstellar Letter*, vol. 33, no. 2, pp. 21–25, 2018.
- [34] J. A. Blumenthal, F. J. Keefe, M. A. Babyak et al., “Caregiver-assisted coping skills training for patients with COPD: background, design, and methodological issues for the INSPIRE-II study,” *Clinical Trials*, vol. 6, no. 2, pp. 172–184, 2009.
- [35] W. Wu, M. Zhang, Y. Ozeren, and D. Wren, “Analysis of vegetation effect on waves using a vertical 2D RANS model,” *Journal of Coastal Research*, vol. 287, no. 2, pp. 383–397, 2013.
- [36] F. Villaescusa-Navarro, P. Bull, and M. Viel, “Weighing NCNH,” *The Astrophysical Journal*, vol. 814, no. 2, p. 146, 2015.
- [37] A. Feizollah, B. Anuar, R. Salleh, and F. Amalina, “Comparative evaluation of ensemble learning and supervised learning in android malwares using network-based analysis,” in *Proceedings of the International conference on communication & computer engineering*, Zhengzhou, China, July 2015.
- [38] E. Acorn, N. Dipsis, T. Pincus, and K. Stathis, *Sit-to-Stand Movement Recognition Using Kinect*, Springer International Publishing, New York, NY, USA, 2015.
- [39] F. Mualla, S. Schöll, B. Sommerfeldt, and J. Hornegger, “Using the monogenic signal for cell-background classification in bright-field microscope images,” in *Bildverarbeitung für die Medizin 2013* Springer, New York, NY, USA, 2013.
- [40] J. Riback, A. G. Hatzigeorgiou, and M. Reczko, “Computational classification of microRNAs in next-generation sequencing data,” *Theoretical Chemistry Accounts*, vol. 125, no. 3–6, pp. 637–642, 2010.
- [41] Y. Di, C. Wang, H. Wu, and X. Yu, “Comparative assessment of data sets of protein interaction hot spots used in the computational method,” in *Proceedings of the International Conference on Intelligent Computing*, Springer International Publishing, Bhubaneswar, India, November 2014.
- [42] S. Gondi, S. Y. Li, M. S. Kanwal et al., “Portable assistive device for combating autism spectrum disorders,” US20140379352 [P], 2016.
- [43] S. J. Webb, T. Hanser, B. Howlin, P. Krause, and J. D. Vessey, “Feature combination networks for the interpretation of statistical machine learning models: application to Ames mutagenicity,” *Journal of Cheminformatics*, vol. 6, pp. 1–8, 2014.
- [44] C. R. Clark, M. J. Ommerborn, K. Moran et al., “Predicting self-rated health across the life course: health equity insights from machine learning models,” *Journal of General Internal Medicine*, vol. 36, pp. 1–8, 2021.
- [45] S.-W. Chou, G.-R. Shiu, H.-C. Chang, and W.-P. Peng, “Wavelet-based method for time-domain noise analysis and reduction in a frequency-scan ion trap mass spectrometer,” *Journal of the American Society for Mass Spectrometry*, vol. 23, no. 11, pp. 1855–1864, 2012.
- [46] M. Farhadian, S. Torkaman, and F. Mojarad, “Random forest algorithm to identify factors associated with sports-related dental injuries in 6 to 13-year-old athlete children in Hamadan, Iran-2018 -a cross-sectional study,” *BMC Sports Science Medicine and Rehabilitation*, vol. 12, no. 1, 2020.