Hindawi
Scientific Programming
Volume 2023, Article ID 9808730, 1 page
https://doi.org/10.1155/2023/9808730

Retraction

Retracted: The Use of Genetic Algorithm, Multikernel Learning, and Least-Squares Support Vector Machine for Evaluating Quality of Teaching

Scientific Programming

Received 19 September 2023; Accepted 19 September 2023; Published 20 September 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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

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

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

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

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

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

References

[1] Y. Yi, H. Zhang, H. Karamti et al., "The Use of Genetic Algorithm, Multikernel Learning, and Least-Squares Support Vector Machine for Evaluating Quality of Teaching," *Scientific Programming*, vol. 2022, Article ID 4588643, 11 pages, 2022.

Hindawi
Scientific Programming
Volume 2022, Article ID 4588643, 11 pages
https://doi.org/10.1155/2022/4588643

Research Article

The Use of Genetic Algorithm, Multikernel Learning, and Least-Squares Support Vector Machine for Evaluating Quality of Teaching

Yingying Yi,[1] Hao Zhang ,[1,2] Hanen Karamti,[3] Shasha Li,[1] Renmei Chen,[1] Huan Yan,[4] and Chenguang Wang [5]

[1]*Institute of Education, Guizhou Normal University, 116 Baoshan Bei Lu, Guiyang, Guizhou, China*
[2]*Guizhou Provincial Educational Governance Modernization Research Center, Guiyang 550025, China*
[3]*Department of Computer Sciences College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia*
[4]*Baiyun District Vocational and Technical School, Guiyang 550000, China*
[5]*School of Business, Lingnan University, 8 Castle Peak Road, Tuen Mun, Hong Kong*

Correspondence should be addressed to Hao Zhang; haozhang_guizhou@sina.com

Received 12 January 2022; Revised 23 January 2022; Accepted 28 January 2022; Published 21 February 2022

Academic Editor: Rahman Ali

The educational data mining (EDM) methods are increasingly diversified. In this research, a hybrid method of multikernel learning (MKL), least-squares support vector machine (LSSVM), and genetic algorithm (GA) is employed to evaluate teaching quality through nine indicators; the reliability of our proposed method is evaluated by confidence interval and prediction interval. First, English teaching quality samples occurring from three age groups at Guizhou Normal University are collected. Next, an intelligent method MK-LSSVM is proposed. Finally, the test sets are regression by the proposed model, and regression results are evaluated by confidence interval, prediction interval, and several error calculation methods; we also develop an ablation experiment for our proposed model. The experiment indicates that the MKL-LSSVM-GA outperforms other benchmark methods at three age-group levels. Additionally, at all three age-group levels, the experiment indicates that three indicators are crucial for the evaluation of teaching quality. Therefore, the proposed model in this paper can evaluate the English teaching quality effectively.

1. Introduction

With the development of English education and the orderly development of discipline construction, more and more people are learning English in China. The related academic research has received unprecedented attention. This paper is aimed to establish a set of scientific and reliable teaching quality evaluation methods and then provide a reliable basis for the school teaching quality management department to formulate corresponding measures.

The development of academic research on teaching quality evaluation is generally as follows: early educational data mining (EDM) methods are limited by data volume and data type, mainly focusing on correlation mining analysis, and supplemented by clustering and predictive analysis [1, 2]. Other research types including weighted average method, single factor evaluation method, fuzzy clustering analysis method, multiple linear regression, analytic hierarchy process, and other methods have been applied in teaching quality evaluation and achieved excellent performance. For example, teaching quality is evaluated more objectively by both qualitative and quantitative methods based on the analytic hierarchy process (AHP) [3], and a TOPSIS (technique for order preference by similarity to an ideal solution) based method is proposed for the evaluation of the physical education teaching quality [4]. Moreover, multiple linear regression was used to evaluate the teacher education information system [5]. However, these methods are still difficult to determine the problems such as weight, strong subjective factors, large randomness, and nonlinear problems, which restrict the promotion and use of these methods. In addition, fuzzy mathematics is a mathematical tool to study many problems in the real world that are not well defined or even very fuzzy, and it is already applied in the field of teaching [6].

Recently, fuzzy mathematics is widely used in artificial intelligence, and artificial intelligence analysis methods are beginning to be applied in the fields of education. Researchers have carried out research on the application of neural networks such as BP neural networks in the evaluation of teaching quality and achieved better results than traditional methods [7, 8]. Similar to the BP model, as a new technology, radial basis function networks (RBFs) are also one of the most commonly used types of neural networks. The RBFs method has achieved a range of applications in the field of the evaluation of teaching quality [9, 10], and neural network (NN) algorithms have also achieved success in other fields [11, 12]. However, the neural network algorithm is easy to be overfitting [13], and it performs not very well when the sample size is not enough; in other words, insufficient educational evaluation data sets will limit the application of the ANNs (artificial neutral network) model.

In the meantime, some machine learning (ML) models, such as hidden Markov model (HMM) and KNN, are commonly applied in the social sciences such as psychology, education, and economics [14]. As a classical machine learning model, a hybrid Markov chain is applied to evaluate the quality of teaching of universities [15]. Unlike general statistical methods, Bayesian statistics makes full use of prior information besides model information and data information; a hybrid model by Bayesian-based method is proposed to simulate the English teaching quality [16]. Unlike Bayesian linear regression, the KNN algorithm is one of the simpler machine learning methods. The idea of this method is that the sample will be divided into category, and every sample can be represented by its K nearest neighbors [17]. Other than KNN, a type of tree-based machine learning algorithms are also increasingly used in classification and regression models such as RF (random forest) and AdaBoost (adaptive Boosting). AdaBoost is an iterative algorithm, and it has already been applied to teaching evaluation [18].

At present, another frequently used ML method is SVM. Compared with the ensemble learning algorithm such as tree-based model, by using multiple weak classifiers, combining the SVM kernel with predictors is yet another promising approach both theoretically and empirically [19]. In general, variable selection is necessary for teaching quality evaluation; more variables may improve the model accuracy of predictions; and it may also have adverse effects because of the mutual dependence between variables. However, even there are many relevant indicators; SVM will still maintain a high accuracy [20]. In particular, SVM can still achieve good performance in a small sample data size, and it has been widely used in the field of education [21, 22]. Unlike SVM, LSSVM is the least-squares formulation of a standard SVM, and it proposed equality constraints in the formulation [23, 24]. Compared with the SVM, the solution process of LSSVM has high efficiency and less computational encumbrance [25]. The choice of kernel functions such as linear kernel, Gaussian kernel, polynomial kernel, and RBF kernel has a great influence on the evaluation accuracy, and the selection of kernel functions is usually based on the sample size of the feature number of the input and the training set.

In recent research, multiple kernel learning (MKL) is used in many fields. MKL avoids the risk of kernel function selection by combining different types of kernel functions and reduces the error of the model by choosing the weighted kernel function group that is most suitable for training data. MKL usually exhibits better performance than the SVM with a single kernel both in classification and regression [26, 27]. For example, the recent method shows excellent forecasting performance with the wind speed data predicted through the integration of different kernel functions [28]. A robust low-rank MKL approach is proposed, and experiments demonstrate superior performance than other state-of-the-art competitors by using six real data sets from different fields [29].

In our proposed model, MK-LSSVM is applied to evaluating the education quality, during the training processing; the evaluation performance may be affected by parameters of MK-LSSVM. That means it is critical to optimize the MK-LSSVM's parameters. Usually, DE (differential evolution) [30] and PSO (particle swarm optimization) [31] are mostly applied in recent studies. As a frequent and classic algorithm, GA (genetic algorithm) [32] has been proved to be very useful in the field of education [10]. GA mainly includes four steps: initialization, individual evaluation, population evolution, and termite inspection. Therefore, GA is employed for the English teaching quality evaluation method.

In English teaching quality evaluation processing, whether the evaluation is successful depends on the evaluation indicators. In the specific construction process of the evaluation indicators system, the following tips are followed: (1) consistency, for the whole process of teaching, full consideration of teaching methods and teaching means can improve the teaching quality evaluation better, (2) relatively independent, in the overall design of teaching quality evaluation indicators, avoiding overlapping indicators is necessary, and (3) data availability, although quantitative indicators can overcome the subjectivity of qualitative indicators in reflecting the teaching quality, not all indicators can be quantified. Therefore, the design of indicators selection, a combination of quantitative and qualitative methods should be considered. In this research, the data of 72 English teaching samples are collected for three age groups (A level, B level, and C level) from the Guizhou Normal University with a time range from 2019 to 2020. This paper constructs the teaching quality evaluation model by integrating GA, multikernel learning, and LSSVM. We find that the proposed MK-LSSVM-GA evaluation model outperforms the benchmark methods in teaching quality evaluation. And we also find the most three important indicators about the evaluation of English teaching. Our research has the following four key contributions:

(1) Enhanced support vector machine (MK-LSSVM) method is proposed: multikernel least squares support vector machine was proposed with combined kernel function to enhance the performance of the prediction model
(2) Through the prediction model, we set different parameters on the prediction model of each group of students, which can be better applied in practical problems
(3) A intelligent hybrid model, MKL-LSSVM–GA, is employed to evaluate the English teaching quality and gets the best evaluation results in each indicator among all the other algorithms
(4) This study found three important factors affecting the quality of English education

In addition, our research is expected to expand and deepen theoretical and methodological research about scientific decision-making in education, precisely allocate teaching resources, provide better service support for teaching, and provide important insights into the upgrading and accelerated development of intelligent and informative education forms. The rest of the paper is structured as follows. Section 2 describes preliminaries of the method adopted in this research. Section 3 describes the proposed methodology of the research conducted. Sections 4 and 5 are experiment design and results analysis. The conclusion of this paper is provided in Section 6.

2. Preliminaries of the Method Adopted

In this paper, multikernel learning, least-squares support vector machine, and genetic algorithm are proposed for the quality of English education.

2.1. Multikernel LSSVM. For the training sample set $T = \{(x_1, y_1), \cdots, (x_N, y_N)\}, x_i \in R^n, y_i \in R$, the decision function of a single kernel LSSVM is expressed as follows:

$$f(x) = \sum\nolimits_{i=1}^{N} a_i K(x, x_i) + b, \tag{1}$$

where b is the deviation, $K(x, x_i)$ is the kernel function that meets Mercer's condition, and $K(x, x_i)$ is defined as follows:

$$K\left(x_i, x_j\right) = \langle \varphi(x_i), \phi\left(x_j\right)\rangle, \tag{2}$$

where $\phi(\cdot)$ represents that map the input data to feature space transformation. Linear kernel, RBF kernel, and polynomial kernel are commonly used three types of kernel functions. The calculation formula is as follows:

$$\begin{cases} K_L(x, x_i) = x^T x_i \\ K_R(x, x_i) = \exp\left(-\dfrac{\left\|x - x_i\right\|^2}{2\sigma^2}\right), \\ K_P(x, x_i) = \left(\dfrac{x^T x_i}{2\delta^2 + \tau}\right)^d, \end{cases} \tag{3}$$

where the subscripts L, R, and P indicate the linear kernel, RBF kernel, and polynomial kernel, respectively. MK-LSSVM method was used to construct a new ensemble kernel by a weighted combination of the above three types of kernel functions, to make full use of the characteristics of different kernel functions and establish a prediction model with better robustness.

$$\begin{aligned} K = {} & \lambda_{11} K_{L1} + \lambda_{12} K_{L2} + \cdots + \lambda_{1M_1} K_{LM_1} + \lambda_{21} K_{R1} + \lambda_{22} K_{R2} \\ & + \cdots + \lambda_{2M_2} K_{RM_2} + \lambda_{31} K_{P1} + \lambda_{32} K_{P2} + \cdots + \lambda_{3M_3} K_{PM_3}, \end{aligned} \tag{4}$$

where $\lambda \geq 0$ is the weight of the corresponding kernel function and M_1, M_2, M_3 are for the use of the linear kernel, the number of RBF, and polynomial kernel, respectively. Write $M = M_1 + M_2 + M_3$, write $\lambda_{11}, \cdots, \lambda_{1M_1}, \lambda_{21}, \cdots, \lambda_{2M_2}, \lambda_{31}, \cdots, \lambda_{3M_3}$ for $\lambda_1, \cdots, \lambda_{M_1}, \lambda_{M_1+1}, \cdots, \lambda_{M_1+M_2}, \lambda_{M_1+M_2+1}, \cdots, \lambda_M$, and write $K_{L1}, \cdots, K_{LM_1}, K_{R1} \cdots, K_{RM_2}, K_{P1}, \cdots, K_{PM_3}$ for $K_{L1}, \cdots, K_{M_1}, K_{M_1+1}, \cdots, K_{M_1+M_2}, K_{M_1+M_2+1}, \cdots, K_M$. For the training sample T, the linear regression function in the high-dimensional feature space can be calculated as follows:

$$f(x) = \sum_{k=1}^{M} \lambda_k \omega_k^T \phi_k(x) + b, \tag{5}$$

where ω_k is the weight of ϕ_k. Based on the principle of structural risk minimization, formula (5) is transformed into a constraint optimization problem in the input data space, that is:

$$\begin{aligned} & \min_{\lambda_k} \min_{\omega_k, b_i, e_i} \frac{1}{2} \sum_{k=1}^{M} \lambda_k \omega_k^T \omega_k + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2, \\ & \text{s.t. } y_i - \sum_{k=1}^{M} \lambda_k \omega_k^T \phi_k(x_i) - b = e_i, \quad i = 1, 2, \ldots, N, \end{aligned} \tag{6}$$

where γ is the penalty factor and e_i is the modeling error. Based on the Karush–Kuhn–Tucker (KKT) condition to solve formula (6), the output $\widehat{y}$ of the MK-LSSVM model can be obtained from the following formula:

$$\widehat{y} = \sum_{i=1}^{N} a_i \left[\sum_{k=1}^{M} \lambda_k K_k(x, x_i)\right] + b. \tag{7}$$

2.2. Genetic Algorithm. Figure 1 and the following steps show the GA procedures.

Step 1. Initialization.

The initial population size is N, crossover probability p_c, the number of maximum iterations T, the number of current iterations $t = 0$, and mutation probability P_m; the initial population $P(0)$ consists of N randomly generated individuals.

Step 2. Individual evaluation.

Calculate the fitness of different individuals in the current population $P(t)$ and evaluate each individual by the fitness function.

Step 3. Population evolution. It is divided into the following four steps:

(1) Maternal selection: $M/2$ pairs of parent population $(M \geq N)$ are selected from $P(t)$.
(2) Crossover: crossover operator is the core of GA; this step applies the crossover operator to the population and gets M intermediate individuals.
(3) Mutation: M intermediate individuals are mutated with the probability P_m and get new M individuals.
(4) Descendant selection: the fitness of new M individuals in step (3) was calculated, and N individuals were selected for $P(t + 1)$ according to the fitness.

Step 4. Termite inspection. If $t = T$, output the optimal solution.

```
Begin
  t ← 0;
    initialize population P (t);
    evaluate P (t);
    While (not termination condition) do
        Begin
              t ← t + 1;
                  Select P (t) from P (t − 1);
```

Figure 1: Process used by the genetic algorithm.

3. Proposed Methodology

Our proposed evaluation model evaluates the teaching quality by nine indicators from the following four aspects:

(1) Teachers' effort in extracurricular time: such as the number of discussions about key and difficult questions (NDQ), the number of homework assignments (NHA) and, the number of homework corrections (NHC)
(2) Diversified teaching methods: including the number of times the questions are asked by the teacher in class (NQ&A) and the number of times the class discussions are organized by the teacher in class (NOD)
(3) Attendance rate: including the average attendance rate of students (AAR), the median number of absences (MNA), and the number of absences from the classroom more than three times (NAT)
(4) Student satisfaction rate: including the evaluation grade of the course by students (EGS); the rating ranges from 0 to 5.

Table 1 shows the details for our selected indicators.

The proposed method mainly includes the following stages:

3.1. Phase 1: Data Preprocessing. Hand-collected English teaching samples from the Guizhou Normal University with a time range from 2019 to 2020 and obtained the following information: (1) the contents of the English teaching course and the teacher's personal information and (2) a complete teaching record of an English teaching course, including the information of students, the date of completion of the course, and so on. This paper excluded the students who quit the course due to personal reasons and collected the corresponding indicators data. Age-group levels of the evaluation are 12 to 14, 15 to 18, and 19 to 22 years old; in this paper, it is represented by the A level, B level and, C level, respectively.

3.2. Phase 2: Model Training. MK-LSSVM algorithm is employed to train the English teaching quality evaluation model at three different age-group levels (A level, B level, and C level), and GA is employed to optimize the MK-LSSVM's parameters.

3.3. Phase 3: Applying Evaluation Model. Testing sample data set of English teaching is evaluated through the trained model.

3.4. Phase 4: Evaluation. Trained models are calculated by using different measures, and the reliability of the proposed method is analyzed by confidence interval (level 80% and level 95%) and prediction interval (level 80% and level 95%).

Figure 2 introduces the flow of the method training and English teaching quality evaluation. First, English teaching samples are extracted from the Guizhou Normal University, with a time range from 2019 to 2020. Then, English teaching samples and the corresponding nine relevant indicators are collected. Finally, the English teaching quality is outputted and evaluated.

4. Experiment Design

In this section, we describe each step of the experiment in detail.

4.1. Data Description. For this study, in every age group, a total of 24 samples (16 samples for training and 8 samples for test) of English teaching quality are collected from Guizhou Normal University during 2019 to 2020; one class is treated as one sample; and the samples for the same age group corresponds to the same course. For English teaching quality, the existing research on

Table 1: The details for selected indicators.

Aspect for indicators	Indicators	Introduction
	NDQ	The number of discussions about key and difficult questions
Teachers' effort	NHA	The number of homework assignments
	NHC	The number of homework corrections
Diversified teaching methods	NQ&A	The number of times the questions are asked by the teacher in class
	NOD	The number of times the class discussions are organized by the teacher in class
	AAR	The average attendance rate of students
Attendance rate	MNA	Median number of absences
	NAT	The number of absences from the classroom more than three times
Student satisfaction rate	EGS	Student satisfaction rate (the rating ranges from 0 to 5)

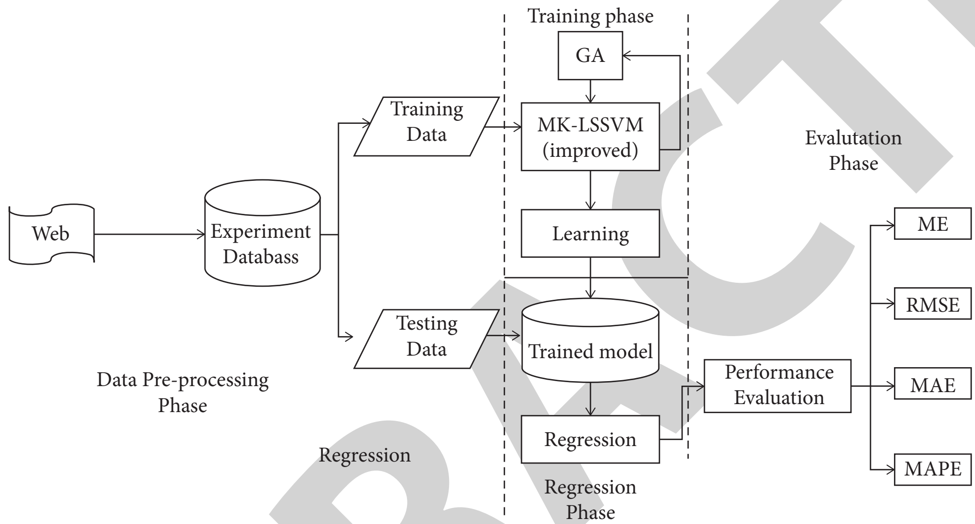

Figure 2: The flow of the method training and English teaching quality evaluation.

teaching quality usually selects expert scoring to measure. Meanwhile, the evaluation of teaching quality by expert scoring is subjective, and the training data set of the model cannot be extended to update the parameters of the model to adapt to different situations. To measure the teaching quality more objectively, we choose students' mastery of knowledge at the end of the course to measure teacher's teaching quality. Students' mastery of knowledge can be replaced by the grade-point average of the class for the following reasons: for the same age group, (1) the course has the same admission criteria; (2) the exam papers are corrected by a team of teachers from different classes, and students' papers were graded anonymously; and (3) the student structure of each class is similar, including the number of students, the ratio of boys to girls, and the student's international background.

The nine indicators used in our experiment are based on expert experience and our manual selection, and they are all important indicators that affect students' performance. They are from the teachers' effort in extracurricular time, diversified teaching methods, attendance rate, and student satisfaction rate four aspects. We added correlation numerical analysis and correlation visualizations as Figure 3 for each indicator with student performance and each indicator with each other. SP stands for student performance, and we use student performance as a proxy. In general, there is no strong correlation between different indicators, but all indicators have a strong correlation with student performance; the data sets in B level and C level are the same as A level.

4.2. MK-LSSVM Parameters. The results of MK-LSSVM model learning and prediction are highly dependent on the value of the penalty factor γ; kernel parameters σ, δ, τ, and d; and their weight coefficient $\lambda_1, \ldots, \lambda_M$; we will use an optimization algorithm (GA) to solve this problem in this paper.

4.3. Benchmark Methods. Table 2 introduces the benchmark methods (GBDT, ANN, RF, XGBoost, AdaBoost, KNN, BN, and ANN) in this paper. In Table 2, MK-LSSVM-GA is our proposed method for English teaching quality evaluation, which uses the MK-LSSVM for the evaluation model, and the MK-LSSVM's initial parameters are optimized by GA. Method 2 applied the LSSVM for English teaching quality evaluation without GA. Methods 3 to 6 are usually used tree-based models such as GBDT, AdaBoost, XGBoost, and RF. Methods 7 to 9 employ KNN, BN, and ANN, for English teaching quality evaluation.

In addition, compared with MK-LSSVM-GA, this paper applied the GA for the LSSVM model based on three single kernels; they are LSSVM (polynomial kernel) GA, LSSVM (linear kernel) GA, and LSSVM (RBF kernel) GA. Whether

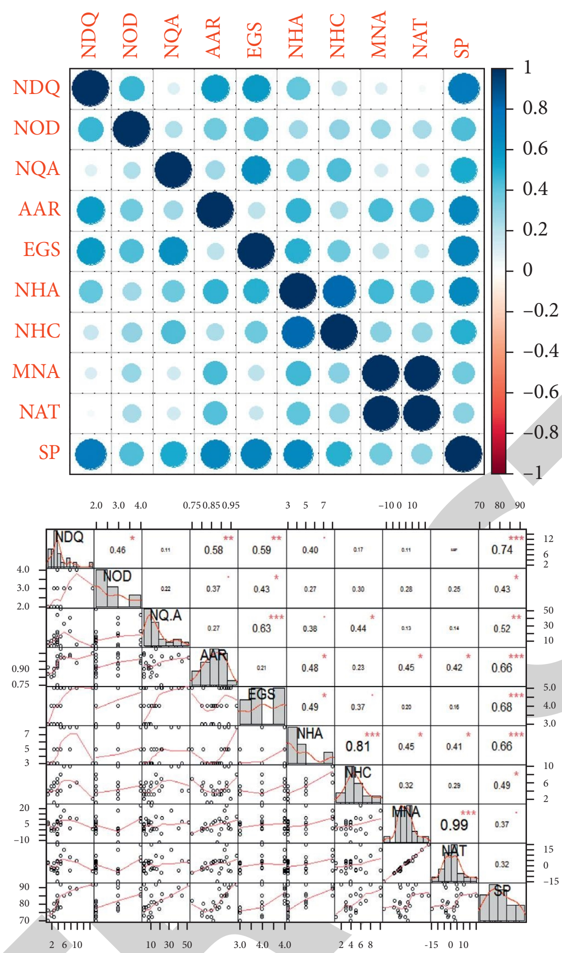

Figure 3: Correlation numerical analysis of A-level.

Table 2: A list of the experimental benchmark models.

No.	Methodology	Description
1	MK-LSSVM-GA (proposed)	The proposed method: a method of MK-LSSVM and GA. MK-LSSVM is employed for the quality of teaching regression, and GA is used for the optimization
2	LSSVM	LSSVM is proposed to evaluate teaching quality without GA and multikernel learning
3	GBDT	GBDT is proposed to evaluate teaching quality
4	RF	Random forest is proposed to evaluate teaching quality
5	XGBoost	XGBoost is proposed to evaluate teaching quality
6	AdaBoost	AdaBoost-based method is proposed to evaluate teaching quality
7	KNN	K-nearest neighbor is proposed to evaluate teaching quality
8	BN	Bayesian network is proposed to evaluate teaching quality
9	ANN	Method based on artificial neural network is proposed to evaluate teaching quality

the application of MK can enhance the evaluation performance is introduced in Tables 3–5.

4.4. Evaluation Criteria. According to the evaluation results, four indicators including ME (mean error), RMSE (root-mean-squared error), MAE (mean absolute error), and MAPE (mean absolute percentage error) are calculated and used to judge the proposed model performance. ME is the mean deviation of a distribution of accidental errors. RMSE makes an excellent general purpose error metric for numerical predictions [33], and the formula is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}. \tag{8}$$

MAE is calculated regardless of whether the score is above or below the mean value [34]; the MAE function is given by the following formula:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|. \tag{9}$$

Table 3: Reliability analysis results of English teaching quality evaluation at A level.

A level	ME	RMSE	MAE	MAPE
MK-LSSVM-GA	0.113	0.586	0.486	0.620
LSSVM (polynomial) GA	0.146	0.668	0.641	0.822
LSSVM (linear) GA	0.160	0.673	0.654	0.821
LSSVM (RBF) GA	−0.166	0.671	0.619	0.748
LSSVM	0.212	0.678	0.659	0.834

Table 4: Reliability analysis results of English teaching quality evaluation at B level.

B level	ME	RMSE	MAE	MAPE
MK-LSSVM-GA	0.144	0.495	0.473	0.589
LSSVM (polynomial) GA	−0.166	0.552	0.525	0.675
LSSVM (linear) GA	−0.165	0.674	0.566	0.691
LSSVM (RBF) GA	0.171	0.700	0.518	0.646
LSSVM	0.214	0.731	0.563	0.710

Table 5: Reliability analysis results of English teaching quality evaluation at C level.

C level	ME	RMSE	MAE	MAPE
MK-LSSVM-GA	0.111	0.512	0.461	0.588
LSSVM (Polynomial)-GA	0.190	0.584	0.536	0.678
LSSVM (linear)-GA	−0.148	0.594	0.572	0.729
LSSVM (RBF)-GA	−0.146	0.516	0.473	0.608
LSSVM	0.220	0.688	0.604	0.760

MAPE is the mean or average of the absolute percentage errors of forecasts; the error expressed as a percentage is more intuitive [35]. The formula is as follows:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right|. \tag{10}$$

The four indicators, ME, RMSE, MAE, and MAPE, are commonly used to calculate the prediction error of the model [33]. They are widely used in various fields to evaluate the performance of a model in the training set and prediction set. In our experiment, we will also use these four indicators to measure the effectiveness of the model. For further evaluation, the reliability of the proposed model is analyzed by CI (confidence interval) and PI (prediction interval). The CI is the interval for the mean of the dependent variable [36], The evaluation criteria for the proposed method applied the 95% and 80% confidence level, while the PI is the interval for the individual value of the dependent variable [37]. For the evaluation of English teaching quality, set a probability value for PI that means the actual quality value should lie within the interval with the giving probability value. The evaluation criteria for the proposed method applied over the 95% and 80% prediction interval.

5. Evaluation Results

In this section, we describe each of the evaluation results in detail.

5.1. The Results of ME and RMSE. Tables 6 and 7 show the results of ME and RMSE, respectively, for English teaching quality evaluation at three different age-group levels. Some findings could be obtained in Table 6: (1) our proposed model MK-LSSVM-GA performs best in every different age group (0.113 for A level, 0.144 for B level, and 0.111 for C level); (2) average ME results of LSSVM (0.216) is superior to that of GBDT (0.306), RF (0.265), XGBoost (0.241), AdaBoost (0.228), KNN (0.319), BN (0.332), and ANN (0.266), demonstrating that in general, the LSSVM model's evaluation accuracy outperforms other benchmark methods; and (3) the results of average ME of MK-LSSVM-GA (0.123) outperforms LSSVM (0.216) for teaching quality evaluation, indicating that the combination of multikernel learning and GA has significantly improved the model evaluation accuracy.

Moreover, from the evaluation results of RMSE in Table 7, we can also obtain some findings: (1) our proposed model MK-LSSVM-GA produces the best RMSE results in every age-group level: 0.586 for A level, 0.495 for B level, and 0.512 for C level; (2) average RMSE result of LSSVM (0.699) outperforms other benchmark methods GBDT (0.762), RF (0.906), XGBoost (0.814), AdaBoost (0.836), KNN (0.922), BN (0.912), and ANN (0.885); and (3) our proposed model MK-LSSVM-GA performs an overall 0.531 of the result in average RMSE for English teaching quality evaluation, which is significantly better than a single LSSVM and all other benchmark methods. Furthermore, regarding the TPR result, there are no benchmark methods less than 0.5 except the proposed method RMSE result (0.495) at B level. These experimental evaluation results indicate that the proposed model MK-LSSVM-GA is an efficient method for the evaluation of teaching quality in English teaching.

5.2. The Results of MAE and MAPE. Tables 8 and 9 show the evaluation of English teaching quality results of MAE at three different age-group levels. For evaluation accuracy,

Table 6: ME results of English teaching quality evaluation at three different age-group levels.

Age-group levels	MK-LSSVM-GA	LSSVM	GBDT	RF	XGBoost	AdaBoost	KNN	BN	ANN
A level	0.113	0.212	0.235	0.255	0.260	0.264	−0.300	0.316	0.292
B level	0.144	0.214	0.379	0.243	−0.218	0.209	−0.347	0.329	−0.279
C level	0.111	0.220	0.304	−0.297	0.245	0.211	0.309	0.350	0.226
Average	0.123	0.216	0.306	0.265	0.241	0.228	0.319	0.332	0.266

Table 7: RMSE results of English teaching quality evaluation at three different age-group levels.

Age-group levels	MK-LSSVM-GA	LSSVM	GBDT	RF	XGBoost	AdaBoost	KNN	BN	ANN
A level	0.586	0.678	0.752	0.972	0.698	0.902	0.946	0.976	0.874
B level	0.495	0.731	0.818	0.898	0.861	0.851	0.908	0.858	0.844
C level	0.512	0.688	0.715	0.848	0.883	0.755	0.912	0.901	0.937
Average	0.531	0.699	0.762	0.906	0.814	0.836	0.922	0.912	0.885

Table 8: MAE results of English teaching quality evaluation at three different age-group levels.

Age-group levels	MK-LSSVM-GA	LSSVM	GBDT	RF	XGBoost	AdaBoost	KNN	BN	ANN
A level	0.486	0.659	0.710	0.853	0.662	0.764	0.890	0.834	0.764
B level	0.473	0.563	0.712	0.818	0.716	0.750	0.838	0.792	0.747
C level	0.461	0.604	0.711	0.715	0.727	0.722	0.818	0.813	0.745
Average	0.474	0.609	0.711	0.795	0.701	0.745	0.849	0.813	0.752

Table 9: MAPE results of English teaching quality evaluation at three different age-group levels.

Age-group levels	MK-LSSVM-GA	LSSVM	GBDT	RF	XGBoost	AdaBoost	KNN	BN	ANN
A level	0.620	0.834	0.907	1.041	0.834	0.931	1.076	1.027	0.924
B level	0.589	0.710	0.885	1.027	0.866	0.924	1.031	1.001	0.921
C level	0.588	0.760	0.905	0.881	0.895	0.911	1.005	1.009	0.912
Average	0.599	0.768	0.899	0.983	0.865	0.922	1.037	1.013	0.919

both MAE and MAPE, the smaller the better. From Table 8, we can find that: (1) our proposed model MK-LSSVM-GA gets the best results of average MAE (0.474); (2) for the results of average MAE results from the benchmark methods, LSSVM (0.609) based methods perform better than GBDT (0.711), RF (0.795), XGBoost (0.701), AdaBoost (0.745), KNN (0.849), BN (0.813), and ANN (0.752) based methods; (3) from MAE results, the proposed MK-LSSVM-GA (0.486 for A level, 0.473 for B level, and 0.461 for C level) consistently perform better than LSSVM (0.659 for A level, 0.563 for B level, and 0.604 for C level) based methods, demonstrating that the multikernel learning and GA optimized has successfully improved the evaluation accuracy of our proposed model.

Furthermore, this study also focuses on the results of MAPE; the error expressed as a percentage is more intuitive. Following findings can be obtained From Table 9: (1) our proposed model MK-LSSVM-GA (0.599) performs better in average for English teaching quality evaluation than all other benchmark methods (0.899 for GBDT, 0.983 for RF, 0.865 for XGBoost, 0.922 for AdaBoost, 1.037 for KNN, 1.013 for BN and 0.919 for ANN); (2) same with the MAE results, MK-LSSVM-GA (0.62 for A level, 0.589 for B level, and 0.588 for C level) performs better than GBDT (0.834 for A level, 0.71 for B level, and 0.76 for C level); and (3) the MAPE results show that GA improved the evaluation accuracy same with the result of Table 8. Furthermore, except for the LSSVMs-based method, XGBoost performed well both in average MAE and MAPE and similar to the LSSVM results of the MAE and MAPE at A level (0.662 for XGBoost' MAE with 0.659 for LSSVM' MAE and the same MPAE for 0.834).

5.3. Three Age-Group Levels Comparison and Error Analysis. For English teaching, people of different age groups have different learning characteristics. Our paper establishes education quality evaluation models for three different age groups. Combined with the results of Tables 6–9, in terms of ME, the model for C level (0.111) outperforms other models (0.144 for B level and 0.113 for C level). Model for B level (0.495) significantly outperforms model for A level (0.586) and C level (0.512) of RMSE. For the indicators of MAE and MAPE, the model for C level is the best and is slightly better than the model for B level in MAPE. Our approach provides practical application options in different age groups.

5.4. Indicators Importance Analysis. The measurement of indicators importance in our teaching evaluation system will have practical significance. As a feature selection criterion, the square of the weights from the hyperplane generated by SVM is proposed by Guyon et al. [38], and it has been used by Bellotti T and Crook J to calculate the importance of indicators for credit scoring [39]. Based on this approach, this study calculated and compared the importance of input indicators in teaching quality evaluation processing. Figures 4–6 show the results of indicator's importance at

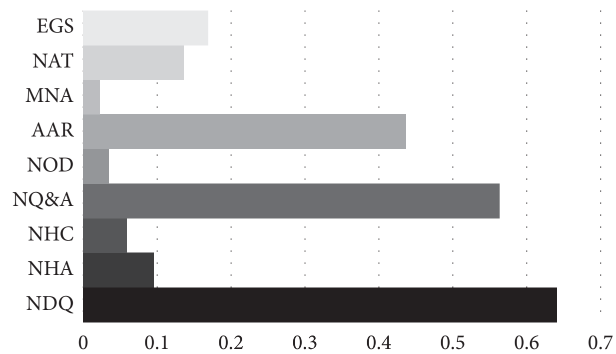

Figure 4: ge group of 12 to 14 years.

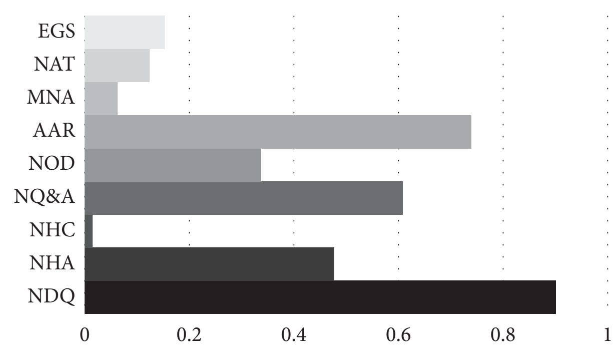

Figure 5: ge group of 15 to 18 years.

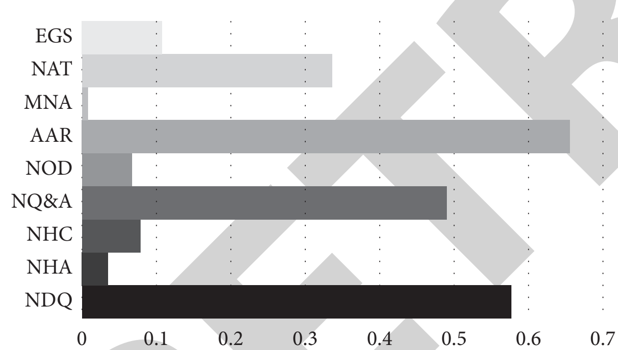

Figure 6: ge group of 19 to 22 years.

three different age-group levels. At the A-level age group, indicators that obtain the most relative importance are NDQ (0.64), NQ&A (0.56), and AAR (0.44). At the B-level age group, the top three indicators that obtain the most relative importance are NDQ (0.91), NQ&A (0.61), and AAR (0.74), while at the C-level age group, they are NDQ (0.58), NQ&A (0.49), and AAR (0.66). These results indicate that: (1) at three age-group levels, the importance of NDQ is always larger than 0.5; moreover, NDQ is the most important one for the teaching evaluation at A and B levels, indicating that the number of discussions about key and difficult questions would be an essential indicator for teaching evaluation and (2) the number of discussions about key and difficult questions (NDQ), the number of times the questions are asked by a teacher in class (NQ&A), and the average attendance rate of students (AAR) are always top three most

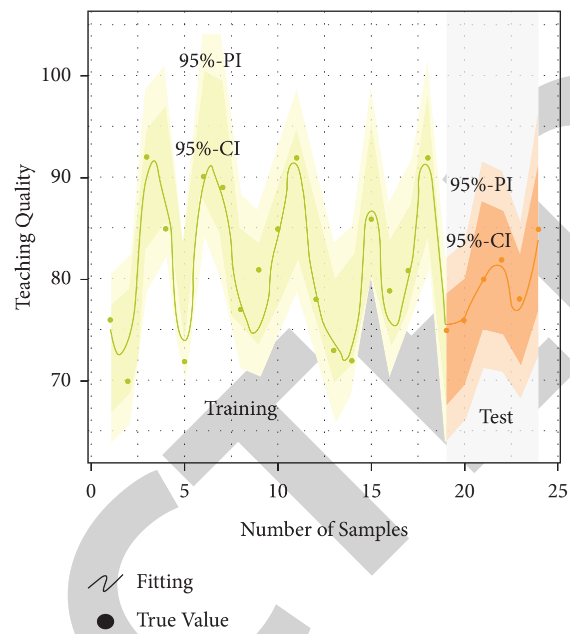

Figure 7: The results about CI and PI in A level.

important factors at three different age-group levels, indicating that they are very important for evaluation of teaching quality.

5.5. Robustness and Reliability Analysis Results. The choice of kernel function has a great influence on the accuracy of evaluation results to the proposed model; to further evaluate the efficacy of multikernel learning, we applied the GA for the LSSVM model based on three single kernels (polynomial kernel, linear kernel, and RBF kernel). Tables 3–5 report the analysis results; we obtained that at the age level of C level, RBF kernel-based method performed well both in MAE and RMSE (0.473 for RBF kernel-based method's MAE with 0.461 for the proposed method's MAE and 0.516 for RBF kernel-based method's RMSE with 0.512 for proposed method's RMSE); MK-LSSVM-GA overalls performs better than three LSSVM-GA methods in all three different age-group levels (A level, B level, and C level).

In addition, the reliability of the proposed model is analyzed by CI (confidence interval) and PI (prediction interval); our evaluation criteria for the proposed method applying the 95% level and 80% level, respectively, for confidence interval and prediction interval. The results of CI and PI are shown in Figure 7:

In Figure 7, the abscissa represents the sample; the ordinate represents the student performance; the yellow part represents the training set; the orange part with the shaded area represents the prediction; the point represents the original data observation value; and the curve represents the model fitting value. It can be seen that the model fitting value is basically around the actual value. In addition, light orange and light yellow shadows represent 95% PI, and dark yellow and dark orange areas represent 95% CI. Our result shows that the predicted values of the proposed model are all within the 95% level of CI and PI; the data sets in B and C levels are the same with A level, demonstrating that the proposed model is a robust evaluation method of English teaching quality.

6. Conclusion

It is very necessary to predict students' academic performance in advance for the management of higher education institutions and students themselves [40]. In this research, the data of 72 English teaching samples are collected for three age groups (A level, B level, and C level) from the Guizhou Normal University with a time range from 2019 to 2020. This paper constructs the teaching quality evaluation model by integrating GA, multikernel learning, and LSSVM. Teaching quality samples with nine indicators are collected from the following four aspects: (1) teachers' effort in extracurricular time, (2) diversified teaching methods, (3) attendance rate, and (4) student satisfaction rate. Evaluation results are measured by ME, RMSE, MA, and MAPE, and the reliability of our proposed model is analyzed by confidence intervals and prediction intervals at different levels. Moreover, this paper uses the benchmark experiment to test the robustness and effectiveness of GA and multikernel learning in multi-level English teaching quality samples. In the end, we find that: (1) the proposed MK-LSSVM-GA evaluation model outperforms the benchmark methods in teaching quality evaluation. (2) Due to multikernel learning's advantages, it avoids the risk of kernel function selection by combining different types of kernel functions, reduces the error of the model by choosing the weighted kernel function group that is most suitable for the training set, and enhances the evaluation processing. (3) At three different age-group levels, the predicted values of the proposed model are all within the 95% level of CI and PI, indicating that the MK-LSSVM-GA is a suitable evaluation model for English teaching quality. (4) According to the indicator importance results, top three indicators including NDQ, NQ&A, and AAR get larger importance at three age levels, demonstrating that NDQ, NQ&A, and AAR are very important for the evaluation of English teaching.

Furthermore, researchers could employ our method for other subjects such as school comprehensive strength evaluation and employee professional competence evaluation, through different indicators and any other stratified sampling method. Apart from the application of the proposed methodology in other fields and varying grouping, based on demographics characteristics, we plan to compute weights and importance of the indicators by adopting pairwise comparison of AHP and weighted sum model (WSM), as adopted by researchers in their studies [41–43].

Data Availability

The experiment data in this paper are available from the corresponding author.

Conflicts of Interest

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

Acknowledgments

This paper was supported by Princess Nourah Bint Abdulrahman University Researchers Supporting Project (number: PNURSP2022R192), Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia.

References

[1] J. Luis Cavalcanti Ramos, R. Euller Dantas e Silva, J. Carlos Sedraz Silva, R. Lins Rodrigues, and A. Sandro Gomes, "A comparative study between clustering methods in educational data mining," *IEEE Latin America Transactions*, vol. 14, no. 8, pp. 3755–3761, 2016.

[2] S. Trivedi, Z. Pardos, G. Sárközy, and N. Heffernan, "Spectral clustering in educational data mining," *Educational Data Mining*, vol. 2011, 2010.

[3] Z. Xie and Z. Su, "Evaluation of college English classroom teaching quality dependent on triangular fuzzy number," *International Journal of Electrical Engineering Education*, Article ID 002072092110020, 2021.

[4] S. Liu, "Research on the teaching quality evaluation of physical education with intuitionistic fuzzy TOPSIS method," *Journal of Intelligent & Fuzzy Systemoks*, vol. 40, no. 5, pp. 1–10, 2021.

[5] S. Li and X. Yuan, "Application of linear regression mathematical model in the evaluation of teachers' informatization quality," *Complexity*, vol. 2021, Article ID 5599655, 2021.

[6] J. Huang, "An internet of things evaluation algorithm for quality assessment of computer-based teaching," *Mobile Information Systems*, vol. 2021, Article ID 9919399, 10 pages, 2021.

[7] L. Jiang and X. Wang, "Optimization of online teaching quality evaluation model based on hierarchical PSO-BP neural network," *Complexity*, vol. 2020, Article ID 6647683, 2020.

[8] Y. G. Luo, "Improved genetic algorithm based student score prediction model," *Bulletin of Science and Technology*, vol. 28, no. 10, pp. 223–225, 2012.

[9] Y. Chen, "College English teaching quality evaluation system based on information fusion and optimized RBF neural network decision algorithm," *Journal of Sensors*, vol. 2021, no. 5, pp. 1–9, 2021.

[10] C. Lu, B. He, and R. Zhang, "Evaluation of English interpretation teaching quality based on GA optimized RBF neural network," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 3185–3192, 2021.

[11] Q. Xu, Q. Guo, C. X. Wang, S. Zhang, and B. Wen, "Network differentiation: a computational method of pathogenesis diagnosis in traditional Chinese medicine based on systems science," *Artificial Intelligence in Medicine*, vol. 118, Article ID 102134, 2021.

[12] Q. Xu, Y. Zeng, W. Tang et al., "Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network," *IEEE journal of biomedical and health informatics*, vol. 24, no. 9, pp. 2481–2489, 2020.

[13] M. Qi and G. P. Zhang, "An investigation of model selection criteria for neural network time series forecasting," *European Journal of Operational Research*, vol. 132, no. 3, pp. 666–680, 2001.

[14] S. Deng, C. Wang, M. Wang, and Z. Sun, "A gradient boosting decision tree approach for insider trading identification: an empirical model evaluation of China stock market," *Applied Soft Computing*, vol. 83, Article ID 105652, 2019.

[15] T. Yuan, "Algorithm of classroom teaching quality evaluation based on Markov chain," *Complexity*, vol. 2021, Article ID 9943865, 12 pages, 2021.

[16] W. Huang, "Simulation of English teaching quality evaluation model based on Gaussian process machine learning," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2373–2383, 2021.

[17] M. Kück and M. Freitag, "Forecasting of customer demands for production planning by local k-nearest neighbor models," *International Journal of Production Economics*, vol. 231, Article ID 107837, 2021.

[18] H. Wu, "Evaluation of AdaBoost's elastic net-type regularized multi-core learning algorithm in volleyball teaching actions," *Wireless Networks*, pp. 1–12, 2021.

[19] A. Widodo, I. Budi, and B. Widjaja, "Automatic lag selection in time series forecasting using multiple kernel learning," *International Journal of Machine Learning and Cybernetics*, vol. 7, no. 1, pp. 95–110, 2016.

[20] T. Joachims, "Text categorization with support vector machines: learning with many relevant features," in *Proceedings of the European Conference on Machine Learning*, pp. 137–142, Chemnitz, Germany, April 1998.

[21] S. Dong, "Intelligent English teaching prediction system based on SVM and heterogeneous multimodal target recognition," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 153, pp. 1–10, 2020.

[22] S. Wang and M. Mu, "Exploring online intelligent teaching method with machine learning and SVM algorithm," *Neural Computing & Applications*, vol. 6, pp. 1–14, 2021.

[23] R. Noori, A. R. Karbassi, and A. Moghaddamnia, "Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction," *Journal of Hydrology*, vol. 401, no. 3-4, pp. 177–189, 2011.

[24] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.

[25] P. K. Soni, N. Rajpal, and R. Mehta, "Semiautomatic road extraction framework based on shape features and LS-SVM from high-resolution images," *Journal of the Indian Society of Remote Sensing*, vol. 48, no. 3, pp. 513–524, 2020.

[26] M. Gnen and E. Alpaydn, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.

[27] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," *Proceedings of the Twenty-First International Conference on Machine learning*, vol. 6, 2004.

[28] S. Sun, J. Fu, A. Li, and P. Zhang, "A new compound wind speed forecasting structure combining multi-kernel LSSVM with two-stage decomposition technique," *Soft Computing*, vol. 25, no. 11, 2021.

[29] H. Jiang, C. Tao, Y. Dong, and R. Xiong, "Robust low-rank multiple kernel learning with compound regularization," *European Journal of Operational Research*, vol. 295, no. 2, pp. 634–664, 2021.

[30] R. Storn and K. Price, "Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[31] Eberhart and Y. Shi, "Particle swarm optimization: developments, applications and resources," in *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No. 01TH8546)*, vol. 1, pp. 81–86, 2001.

[32] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Machine Learning*, vol. 3, no. 2, pp. 95–99, 1988.

[33] N. Levinson, "The wiener (root mean square) error criterion in filter design and prediction," *Journal of Mathematics and Physics*, vol. 25, no. 1–4, pp. 261–278, 1946.

[34] L.-M. Wun and W. L. Pearn, "Assessing the statistical characteristics of the mean absolute error or forecasting," *International Journal of Forecasting*, vol. 7, no. 3, pp. 335–337, 1991.

[35] A. D. Myttenaere, B. Golden, B. L. Grand, and F. Rossi, "Using the mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, 2015.

[36] S. Strippoli, "Confidence interval (CI)," *Central African Journal of Medicine*, vol. 45, p. 1, 1999.

[37] D. L. Shrestha and D. P. Solomatine, "Machine learning approaches for estimation of prediction interval for the model output," *Neural Networks*, vol. 19, no. 2, pp. 225–235, 2006.

[38] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.

[39] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3302–3308, 2009.

[40] F. Aman, A. Rauf, R. Ali, F. Iqbal, and A. M. Khattak, "A predictive model for predicting students academic performance," in *Proceedings of the 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1–4, Patras, Greece, July 2019.

[41] R. Ali, M. Sajjad, F. Iqbal, M. S. H. Zada, and M. Hussain, "Reducing CO_2 emission using EDA and weighted sum model in smart parking system," *International Journal of Computer and Information Engineering*, vol. 15, no. 2, pp. 107–112, 2021.

[42] R. Ali, S. Lee, and T. C. Chung, "Accurate multi-criteria decision making methodology for recommending machine learning algorithm," *Expert Systems with Applications*, vol. 71, pp. 257–278, 2017.

[43] R. Ali and A. M. Khatak, "Multicriteria decision making using TOPSIS method: an accurate selection of deserving candidates," in *Proceedings of the Future Technologies Conference*, pp. 506–517, Vancouver, Canada, November 2021.