

# Personalized Learning Ability Classification Using SVM for Enhanced Education in System Modeling and Simulation Courses

Chao Liu, Shengyi Yang

School of Physics and Mechatronic Engineering, Guizhou Minzu University, Guiyang 550025, China

© Higher Education Press 2024

**Abstract** This study investigates the application of a support vector machine (SVM)-based model for classifying students' learning abilities in system modeling and simulation courses, aiming at enhancing personalized education. A small dataset, collected from a pre-course questionnaire, is augmented with integer data to improve model performance. The SVM model achieves an accuracy rate of 95.3%. This approach not only benefits courses at Guizhou Minzu University but also has potential for broader application in similar programs in other institutions. The research provides a foundation for creating personalized learning paths using AI technologies, such as AI-generated content, large language models, and knowledge graphs, offering insights for innovative educational practices.

**Keywords** SVM-based classification, personalized education, learning ability classification, system modeling

## 1 Introduction

In modern higher engineering education, tailoring learning experiences and materials to meet each student's specific needs is essential for improving teaching effectiveness and enhancing students' capabilities. This approach, which aligns with the principles of Education 4.0, emphasizes personalized and student-centered learning by leveraging advanced technologies to create customized educational pathways. Personalized and self-paced learning, a key component of Education 4.0, focuses on adapting teaching methods and materials to individual learning needs, abilities, and

preferences, ultimately providing a more meaningful and effective learning experience.

Innovative technologies such as AI-generated content (AIGC), large language models (LLMs), and knowledge graphs have simplified the creation of personalized learning materials. Nevertheless, evaluating students' genuine learning abilities remains a significant challenge. This difficulty arises from the complexity and diversity of both learning behaviors and questionnaire data. As a result, an effective classification of student ability has become increasingly important, requiring sophisticated models that can analyze and interpret large and varied datasets to support the goals of Education 4.0.

This study aims at tackling the urgent challenge associated with assessing learning abilities and offering an effective solution to categorize students' learning capacities within higher engineering education to enable personalized learning interventions. Focusing on the systems modeling and simulation course at a university in southwest China, Guizhou Minzu University, this research leveraged support vector machine (SVM) technology to develop a classification model for students' learning abilities. This model formed the foundation for creating personalized learning paths tailored to individual students.

First, the research group devised a self-assessment questionnaire that focused on learning abilities associated with the course, utilizing questionnaire outcomes as data source for the learning ability classification model. The experienced educational team meticulously preprocessed the questionnaire data to create a pivotal dataset for the student learning ability model.

Second, the research group used a decision tree to define the correlation between the questionnaire data and classification outcomes. The method selectively chose data with correlations surpassing the average for model input. Considering the constraints of limited

Received August 30, 2024; revised November 26, 2024; accepted December 10, 2024

Chao Liu (✉)

E-mail: liuchaoaero@sina.com

data samples, especially for the advanced and novice student categories, we introduced a specialized data augmentation mechanism tailored to small-sample learning data. These efforts culminated in the creation of an SVM model for classifying students' learning abilities. The experimental findings confirmed the feasibility and high precision of the adopted model design approach.

## 2 Literature Review

The rapid development of education digitalization has made educational data mining technology increasingly crucial in improving the quality of higher education. Aiming at discovering valuable information pertinent to the educational domain, such technologies use advanced data analysis methods to thoroughly explore, analyze, and mine data.

Educational data mining technology, specifically, is dedicated to providing personalized learning support for each student by analyzing multidimensional data, such as students' learning behaviors, grades, and interests. This personalized support helps educators design more specific teaching methods that better meet students' needs. Various studies have used decision tree methods to predict students' academic performance and define the influential factors such as financial status, learning motivation, and gender on academic outcomes. Those studies have provided educators with insights into how student characteristics correlate with their academic achievements (Kolo & Adepoju, 2015). A survey of research conducted between 2010 and 2020 focuses on intelligent technologies used for predicting students' performance (Namoun & Alshantqi, 2020). It presents a series of models used to predict students' learning and explores the influential factors of academic outcomes. This survey offers a historical perspective for researchers in educational data mining and serves as a valuable reference for future research (Namoun & Alshantqi, 2020). Kumar et al. (2023) conducted a systematic review of machine learning technologies used in education and models to predict students' performance, exploring key factors that influence students' learning outcomes. This systematic review contributes to a comprehensive understanding of current applications of machine learning technologies in education and reveals the problems being addressed urgently in the future research.

Furthermore, educational data mining technology can detect potential problems in students' learning processes early on and intervene before problems escalate. Proactive intervention based on educational data mining results can provide additional support, contributing to the improvement of students'

academic performance. Such interventions are based on various machine learning algorithms, including deep neural networks (DNNs), decision trees, random forests, gradient boosting, logistic regression, support vector classifiers, and  $K$ -nearest neighbors. These algorithms are employed to develop predictive models for assessing students' future academic performance. Methods like DNNs are used to construct a predictive model based on students' grades in early courses of their first academic year. Nabil et al.'s (2021) model achieved an accuracy of 89% in predicting students' performance in a data structure course and identified students at risk of failure early in the semester. In another study, a gradient boosting decision tree algorithm predicts students' performance in final exams, successfully identifying students who require special attention and offering the necessary assistance (Ahmed et al., 2021). A comparison of these models to other machine learning algorithms such as SVMs, logistic regression, Naive Bayes models, and gradient boosting trees demonstrates their higher accuracy. Qazdar et al. (2019) proposed a machine learning algorithm-based framework that analyzed and tested students' data collected from a school management system, which exhibited high predictive accuracy. Meanwhile, Xu et al. (2017) developed a novel machine learning approach that was used to predict students' performance in degree programs, helping to assess whether students could complete degree programs on time. Addressing the issue of small sample sizes students' learning data in higher education, Zohair and Mahmoud (2019) demonstrated the feasibility of training and modeling on small datasets, creating a predictive model with credible accuracy. Imran et al. (2019) introduced a students' performance predictive model based on a supervised learning decision tree classifier and an ensemble method to enhance classifier performance. Rastrollo-Guerrero et al. (2020) studied widely used models and methods to assess and predict students' learning performance, producing findings that could aid in the design of effective mechanisms, improve academic performance, and mitigate issues such as students' dropouts.

In the context of the early identifying of learning obstacles and providing additional supports, educational institutions can leverage data analysis. Multidimensional student datasets that incorporate attendance records, assignments, and course grades are established. Muraina et al. (2022) successfully employed techniques such as neural networks, logistic regression, and decision trees to successfully predict students' academic performance with an accuracy of 96%. Furthermore, Vergaray et al. (2022) constructed a predictive model based on students' course progress. Compared to other learning algorithms, the model demonstrated improved accuracy of predictive

performance, ultimately achieving a precision of 92.86%. This research delves into key issues related to predicting students' performance in the learning environment. [Razaque and Alajlan \(2020\)](#) analyzed and evaluated six machine learning models, including decision trees, random forests, SVMs, logistic regression, AdaBoost, and stochastic gradient descent. The study provides in-depth information on their accuracy and sensitivity in assessing students' performance. These findings contribute to the development of alternative recommendation systems for academically challenged students. The effectiveness of machine learning and deep learning models in predicting students' early performance in higher education institutions has also been investigated, with different models used to forecast students' learning outcomes ([Balcioglu & Artar, 2023](#)). The results underscore the potential of data-driven technologies in the educational decision-making process to support targeted interventions and personalized learning strategies.

In addition, the tree-based machine learning algorithms were used for the precise identification of students at risk of low grades, with targeted measures proposed to improve the quality of professional teaching ([Zhang et al., 2022](#)). The various machine learning techniques were utilized to predict students' academic performance based on real data. [Verma et al. \(2022\)](#) compared these technologies based on different assessment indicators. The results of the research could aid students in tracking their academic performance and supporting their future academic success. Meanwhile, [Li and Liu \(2021\)](#) predicted students' learning abilities using neural networks, providing supports for students in selecting courses and planning their future learning. Moreover, such methods can assist teachers and administrators in monitoring overall students' learning progress.

Data mining techniques can be used to comprehensively assess students' abilities across various domains, monitoring their progress in different subjects. The results and analyses can aid in the precise formulation of personalized learning plans and the holistic evaluation of students' overall development. [Guo et al. \(2015\)](#) extensively explored the significance of employing machine learning and data mining techniques in education. The research results indicate that leveraging these methods to enhance students' academic performance is very important. A classification model is developed by unsupervised learning algorithms for layer-wise pretraining of hidden layer features in the research, followed by the fine-tuning of parameters through supervised training. The experimental results demonstrate the excellent predictive performance of the model, presenting the applicability of this method to warning systems of higher education academic performance.

[Alsariera et al. \(2022\)](#) investigated existing machine learning methods and critical features to predict students' academic performance. The study indicates that artificial neural networks outperform other models in terms of their academic assessment accuracy, emphasizing the benefits of machine learning in recognizing and improving academic performance. Furthermore, [Vijayalakshmi and Venkatachalapathy \(2019\)](#) underscored the importance of predicting students' performance to improve their academic achievement, proposing a system that employed DNNs to predict students' academic performance. In a comparative experiment involving six algorithms, DNNs exhibits optimal performance with an accuracy of 84%. These research findings further highlight the potential value of data mining techniques in assessing students' overall capabilities and elevating their academic standards.

Introduced by [Cortes and Vapnik \(1995\)](#) in the mid-1990s, SVMs had become one of the most widely used machine learning algorithms, particularly for classification tasks. SVMs' abilities to handle high-dimensional feature spaces and provide robust generalizations make them a powerful tool in various domains, including pattern recognition, medical diagnosis, and educational modeling.

In recent years, there has been significant research in enhancing the performance of SVMs in small-sample and high-dimensional data settings, which are particularly common in educational data analysis. Several studies have explored novel kernel functions in improving classification accuracy. For instance, [Zhang et al. \(2022\)](#) proposed an adaptive kernel SVM that adjusted kernel function based on distribution of the data, improving its performance in small-sample learning problems. Their method demonstrates that SVMs can be effective even with limited data, which is a common challenge in designing and executing personalized education systems.

Furthermore, SVMs have been successfully applied in various educational contexts. For example, researchers utilize SVMs to classify students' learning abilities based on various features, such as their previous academic performance, behavioral patterns, and responses to pre-course surveys. [Wang et al. \(2022\)](#) used an SVM to classify students' learning styles and develop personalized learning paths, demonstrating that SVMs could improve the adaptability of educational systems to meet individual student's needs.

This study focuses on the system modeling and simulation course within the engineering discipline at a university in southwest China, Guizhou Minzu University. The objective is to construct and introduce a student learning capability classification model tailored to local students' learning. The study explores methods to build this model and evaluates its application effects.

The designed model of student learning capacity classification can not only accurately assess students' learning abilities at the beginning of the course, but also help instructors in guiding personalized learning paths and providing materials. Moreover, the model aids instructors in understanding the key features that enhance students' academic performance in the system modeling and simulation course, thereby optimizing the course structure and educational objectives. Therefore, the results provide valuable guidance and support for improving course teaching effectiveness and enhancing student learning experiences.

### 3 Materials and Methods

This study aims at classifying students' abilities based on their questionnaire responses using machine learning techniques. First, the process begins with the collection of students' survey data, contributing to a rich set of information regarding students' academic behaviors, performance, and engagement. These raw data are quantized into numerical features suitable for analysis.

Second, we apply a random forest for feature selection, which helps identify the most relevant attributes related to student ability classification. By ranking the importance of each attribute, we reduce the dimensionality of the dataset and focus on the most significant factors that contribute to ability assessment.

To enhance the robustness and generalizability of the model, we apply data augmentation techniques. These techniques generate additional synthetic samples from the existing dataset, increasing its size and mitigating potential class imbalance.

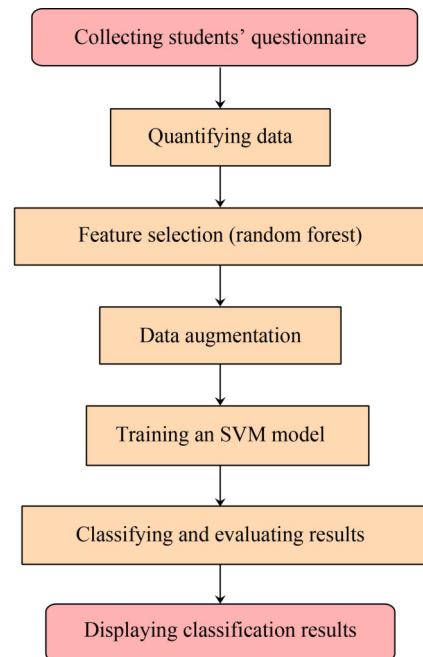
After preprocessing the data, we train an SVM model to classify students into distinct ability levels, including low, medium, and high levels. We choose an SVM model for several reasons: First, SVMs are highly effective at handling high-dimensional data, which are common in educational datasets where multiple features, such as exam scores, attendance, and participation, can influence classification. Second, SVMs are known for their abilities to handle nonlinear relationships among features by using kernel functions, such as the radial basis function (RBF), making it particularly suitable for complex and nonlinear students' data. Third, SVMs have strong generalization capabilities and are less prone to overfitting, especially when the number of samples is limited or the dataset is noisy.

An important advantage of SVMs is their simplicity and ease of implementation, which makes them an ideal choice for teachers and educators who may not have extensive experience with machine

learning or advanced technical expertise. The model's straightforward nature allows instructors to easily develop and customize a classification model tailored to students' specific characteristics. This feature is particularly valuable in educational settings, allowing teachers to build models based on the data they have already collected and providing personalized insights into students' academic performance and abilities.

Finally, the classification results are evaluated using metrics F1 score and receiver operating characteristic and area under the curve (ROC-AUC) to ensure accuracy and reliability.

The following sections detail each step of the process as shown in [Figure 1](#), outlining the specific methods used and the rationale behind the chosen techniques.



**Figure 1** Flowchart of the process for student ability classification. SVM: support vector machine.

#### 3.1 | Course Overview

The authors have regularly taught the system modeling and simulation course in Guizhou Minzu University in the southwestern region of China over the years. The research focus is undergraduates who have taken this course in the past three years. The system modeling and simulation course is a mandatory course for majors such as automation in the institutions, carrying three credits with a total duration of 48 hours.

Through the curriculum, students acquire specialized knowledge and professional skills in system modeling and simulation. Students develop the capability to employ interdisciplinary theories and

methods for establishing dynamic system models, as well as proficiently use MATLAB to facilitate tasks such as system modeling, controller design, and system analysis. Simultaneously, the course aims at enhancing students' abilities to abstract scientific problems based on complex engineering issues and interpret real-world physical phenomena through simulation results.

The curriculum incorporates collaborative group projects to elevate students' cooperative, analytical, and expressive skills. Approximately 100 students participate annually, underscoring its widespread popularity and demand among students. This course endeavors to impart a comprehensive academic foundation and practical expertise in system modeling and simulation, laying robust foundation for students' future professional endeavors in related fields.

In traditional course instruction, students use the same teaching materials and are required to complete identical learning tasks. However, due to various reasons, there is a significant disparity among students in terms of their mathematical proficiency, understanding of modeling and simulation techniques, and programming abilities. The uniformity of course tasks results in students with stronger capabilities finding the course relatively easy, making it challenging for them to acquire new knowledge. On the contrary, students with weaker skills perceive the course contents to be difficult, leading to a gradual loss of interest in continuing their studies.

To address this issue, the course team undertakes efforts to tailor learning tasks to students with varying abilities, providing personalized learning materials, and offering customized course learning paths for students with diverse learning capabilities. To acquire efficient and accurate assessments of students' learning abilities, the team uses questionnaires, pre-course performance records, and other information and data. The team, specifically, constructs a classification model for students' learning abilities, using data such as questionnaire responses and pre-course performance records. This model aims at matching the personalized needs of the institution's students, facilitating effective and precise evaluations of students' learning capacities within the course.

### 3.2 | Data Acquisition

The research data are derived from 204 undergraduates who participate in the system modeling and simulation course in 2022 and 2023. Before starting the course, these students are required to complete a specially designed survey tailored to capture their circumstances. The survey encompasses various aspects, including students' understandings and knowledge of system modeling and simulation concepts, as well as their familiarity with differential equations, methods of

solving differential equations, perceived effectiveness of pre-course programming instructions, self-assessment of programming abilities, interests in the course, and learning expectations.

All survey options are presented in text format. During the initial stages of data processing, the course team assigns values to different options based on the questions formulated in Table 1. Numeric values associated with the survey items are presented as integers. This comprehensive data collection strategy aims at gathering detailed insights into the students' perspectives, knowledge levels, and expectations, contributing to a robust analysis of their experiences and performance in the course.

**Table 1** Sample survey question 3 and options

Item	Content
Question 3	Have you ever heard of the term "system modeling" in your previous courses or extracurricular studies?
Option A	Never heard of it.
Option B	Heard of it but not sure what it is.
Option C	Heard of it and understand what the modeling is.
Option D	Very familiar. I have studied it specifically.
Option E	Extremely familiar. I have used it in competitions or projects.

#### 3.2.1 Main Feature Selection

Regarding the survey results, the course team manually excludes data related to teaching recommendations and students' background information. Teachers believe that this portion of the data is either unrelated or minimal correlation with students' learning abilities. Therefore, the data obtained comprise 10 self-assessment items, including awareness of system modeling, understanding the relationship between system modeling and differential equations, awareness of system simulation, understanding the relationship between system simulation and differential equations, importance of modeling and simulation techniques, mastery of higher mathematics knowledge, programming abilities, interest and adaptability in programming, interest in extracurricular learning, and willingness to choose courses. As shown in Table 2, the research group provides quantitative data corresponding to the features of nine students, with the level label being the ability evaluation provided by experts.

Based on the questionnaire contents, certain features play a significant role in shaping students' learning abilities, while others have a relatively minor impact on learning capabilities. To identify features of substantial relevance for the assessment of learning abilities, the data preprocessing phase uses a recency, frequency, and monetary (RFM) model. By examining the importance of each feature, it becomes possible to



Another issue arises from the distribution of students’ learning abilities. Most students are expected to have moderate learning abilities, while those with poor or strong learning abilities are relatively scarce. In terms of sample distribution, the number of students representing moderate learning abilities tends to be higher than the numbers representing the other two categories. This imbalance in sample distribution may affect the model’s performance and necessitate careful consideration and handling during the modeling process as shown in Figure 4.

During the process of converting questionnaire results into numerical values, the obtained results are typically integers. To increase the sample size and balance the quantity of result samples, consider making small numerical perturbations around the integer results as shown in Equation (1):

$$\hat{Z}_{ij} = Z_{ij} + \Delta, \tag{1}$$

where the enhanced sample data, denoted as  $\hat{Z}_{ij}$ , are derived from the original sample data, represented by  $Z_{ij}$ , and  $\Delta$  represents randomly generated perturbation parameters uniformly distributed between  $-0.4$  and  $0.4$ . Note that because the original sample data are integers, perturbing the data will not alter the results of the learning ability classification as shown in Table 3.

### 3.3 | Modeling

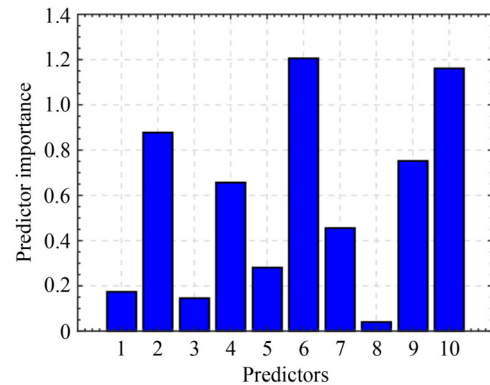
#### 3.3.1 Support Vector Machines

SVMs are powerful supervised learning algorithms widely used in classification and regression tasks. The main objective is to find an optimal hyperplane in the feature space to effectively separate samples from different classes. Due to their strong generalization capabilities and elegant handling of feature space patterns, SVMs remain one of the preferred algorithms for many machine learning problems.

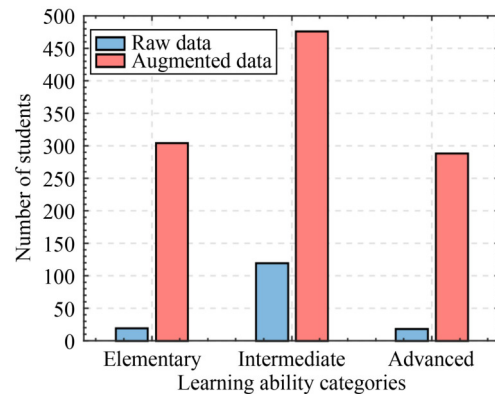
The process of modeling an SVM for multi-class classification involves two key mathematical steps. First, the training data are represented as a feature matrix  $X$  with corresponding class labels  $Y$ . For multi-class scenarios, appropriate encoding of class labels is

essential. The construction of the SVM model adopts a one-vs-one strategy, where for each pair of classes  $C_i$  and  $C_j$ , a binary SVM classifier is established using a training subset  $X_{ij}$  and the corresponding labels  $Y_{ij}$ .

Second, the SVM optimization problem is solved for each binary classifier. This involves minimizing a cost function that includes a regularization term to ensure a balance between achieving a wide margin



**Figure 3** Predictor importance measurement. 1: awareness of system modeling, 2: understanding the relationship between system modeling and differential equations, 3: awareness of system simulation, 4: understanding the relationship between system simulation and differential equations, 5: importance of modeling and simulation techniques, 6: mastery of higher mathematics knowledge, 7: programming abilities, 8: interest and adaptability in programming, 9: interest in extracurricular learning, 10: willingness to choose courses.



**Figure 4** Raw data and augmented data.

**Table 3** Schematic illustration of sample augmentation principles

Student	Feature 2	Feature 4	Feature 6	Feature 9	Feature 10	Classification	Category
1	2.00	3.00	3.00	2.00	1.00	Medium	Raw data
11	2.11	3.24	2.95	2.12	0.82	Medium	Augmented
12	1.87	3.01	3.32	2.01	1.23	Medium	Augmented

Notes. Feature 2: understanding the relationship between system modeling and differential equations, Feature 4: understanding the relationship between system simulation and differential equations, Feature 6: mastery of higher mathematics knowledge, Feature 9: interest in extracurricular learning, Feature 10: willingness to choose courses.

and minimizing misclassification. The resulting decision functions  $f_{ij}(x) = w_{ij} \cdot x - b_{ij}$  define the hyperplanes that best separate the samples of two classes.

The SVM optimization problem is given as shown in Equation (2):

$$\min_{w,b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \max(0, 1 - y_i (\mathbf{w} \cdot \mathbf{x}_i - b)), \quad (2)$$

where  $\mathbf{w}$  represents the weight vector,  $b$  is the bias term,  $C$  serves as the regularization parameter,  $\mathbf{x}_i$  denotes the feature vector of the  $i$ -th sample,  $y_i$  indicates the label of the  $i$ -th sample,  $m$  is the total number of training samples, and  $i$  acts as the index from training samples. During the testing phase, each decision function is applied to new data points, and a voting mechanism is employed to determine the final predicted class. The class with the majority of votes across all binary classifiers is assigned as the predicted class for multiclass problems.

Overall, this mathematical approach allows SVMs to elegantly handle multiclass classification by decomposing a problem into binary subproblems and finding optimal hyperplanes in the feature space.

### 3.3.2 Classification Modeling

After extracting important features for the student learning ability classification model and enhancing the dataset to increase it to 500 samples while balancing the number of samples for elementary, intermediate, and advanced learning levels, the course team uses SVM method to construct a personalized student learning ability multiclassification model.

The dataset is divided into a training set and a testing set. Specifically, 70% of the data are used to train the SVM student learning ability classification model, while the remaining 30% are used to assess the performance of the constructed model.

During the training process, the SVM, using the sequential minimal optimization algorithm, with a linear kernel is chosen and a one-vs-one multiclass coding scheme is employed.

**Table 4** Key parameters used in the SVM model training

Parameter	Value	Description
Learner	SVM	Specify SVM as the base learner for classification.
Coding	One-vs-one	Use the one-vs-one strategy for multiclass classification, creating a classifier for each pair of classes.
Kernel function	Linear	The kernel function is used for the SVM. A linear kernel is used by default, but this can be adjusted to nonlinear kernels, like RBF.
BoxConstraint	1	The regularization parameter controls the trade-off between training errors and model complexity.
Kernel scale	Auto	Automatically scales the kernel, adjusting it based on the data.

Notes. SVM: support vector machine, RBF: radial basis function.

## 4 Results

This study uses the `fitcecoc` function in MATLAB to train an SVM model for classifying students' learning abilities. The `fitcecoc` function implements the error-correcting output codes suitable for solving multiclass classification problems. An SVM is chosen as the base learner and specific parameter settings are configured to suit the characteristics of the dataset. The key parameters used in the SVM training process are summarized in Table 4.

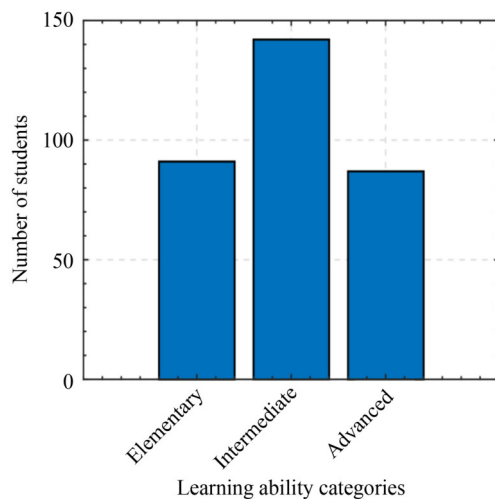
Through an in-depth analysis of the sample data, we successfully developed an SVM model for classifying students' learning abilities. This model is based on the actual learning situations of our school and is integrated into the system modeling and simulation course. The model achieves an impressive evaluation accuracy of 95.3% for the test set. Furthermore, the cross-validation results yield an average out-of-sample loss of 0.1003, providing additional confirmation of the model's excellent performance as shown in Figure 5. This indicates that the SVM model for classifying students' learning abilities, develops through cross-validation, demonstrates outstanding performance, and possesses strong generalization capabilities. The SVM model can accurately predict previously unseen data. Hence, it is a reliable tool for accurately assessing students' learning potential, offering robust support for personalized education and academic assistance in practical applications.

Regarding the 320 samples in the test set, the students' learning abilities are classified into three categories, including elementary, intermediate, and advanced categories. The classification results are shown in Figure 6.

The matrix, as shown in Figure 6, demonstrates the strong performance of the SVM-based model across the three categories, including elementary, intermediate, and advanced classes. The majority of predictions lies along the diagonal, indicating that the model achieves high accuracy in correctly classifying instances for all classes. The model correctly classifies

True class	Advanced	85		1
	Elementary		87	4
	Intermediate	8	2	133
		Advanced	Elementary	Intermediate
		Predicted class		

**Figure 5** Confusion matrix of the classification model.



**Figure 6** Classification result with trained model.

85 advanced, 87 elementary, and 133 intermediate instances, highlighting its robust performance across the board.

When examining misclassifications, the matrix shows minimal confusion among categories. For the advanced class, only one instance is misclassified as elementary, demonstrating excellent performance for this category as shown in Figure 5. The elementary class has four misclassifications as intermediate, while the intermediate class exhibits slightly higher misclassification rates, with eight instances predicted as advanced and two as elementary. These errors, while small, indicate that the intermediate class shares some overlapping characteristics with the other two categories, particularly advanced.

The confusion between intermediate and advanced classes reflects similarities in these levels, making them more challenging to distinguish. However, given the high number of correct classifications and the limited number of misclassifications, the model shows high reliability in classifying, even when dealing with potentially overlapping features.

In summary, the SVM model exhibits strong

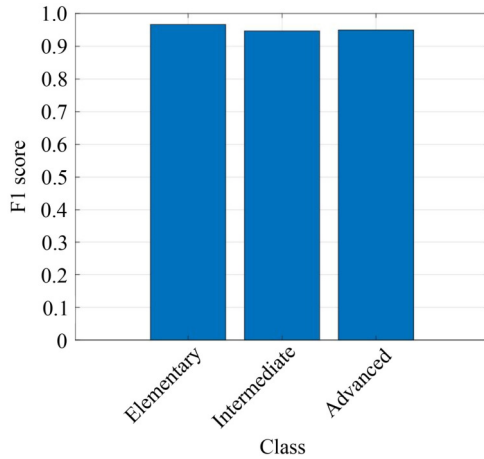
classification performance, as supported by the confusion matrix, with minimal errors that do not significantly impact its reliability. Further improvements, such as fine-tuning model parameters and augmenting data for overlapping regions, potentially mitigate these minor misclassifications. However, as it stands, the model is highly effective and well-suited for practical applications in assessing students' learning abilities.

After conducting a thorough examination of misclassified data, we discover that the majority of erroneous samples originate from students with intermediate-level abilities. To delve deeper into this phenomenon, a detailed analysis is conducted on the relevant sample data. In terms of feature selection, the focus is placed on features related to three modules, including mathematical foundations, programming skills, and learning attitudes.

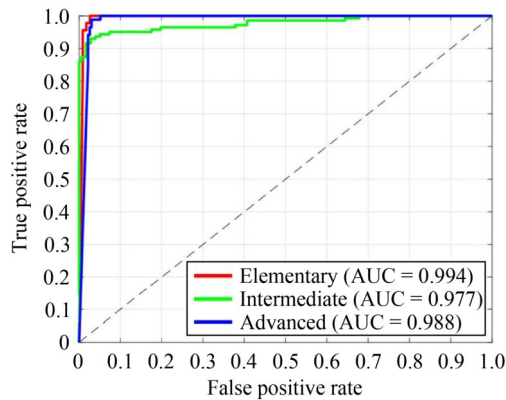
Within the misclassified samples, a clear pattern has emerged. These samples predominantly exhibit excellent programming skill levels, but weak mathematical foundations. When training the student learning ability classification model, encountering such data leads to a tendency for the model to inaccurately classify students as having advanced learning abilities, primarily due to the elevated programming-related assessments. This phenomenon may be correlated with the actual situations of students in our university. The majority of students in our university do not excel in programming technologies, and there exists a certain level of apprehension toward programming. This results in students providing lower self-assessments of their programming abilities in questionnaires. Due to the relatively low number of students with high programming skill level, this contributes to inaccuracies in evaluating these samples.

The designed model demonstrates strong performance and high reliability, as indicated by the F1 scores as shown in Figure 7 for the three classes, including elementary in 0.967, intermediate in 0.947, and advanced in 0.950. These scores, all close to 1, suggest that the model achieves an excellent balance between precision and recall across all categories. Furthermore, the relatively small variation in F1 scores across the classes reflects the model's ability to handle potential imbalances and ensure performance consistency. The results highlight the robustness of the model in accurately predicting multiple categories and its suitability for real-world applications requiring balanced and reliable classification.

The receiver operating characteristic (ROC) curve as shown in Figure 8 highlights the exceptional performance of the constructed SVM-based student ability classification model across three categories. The area under curve (AUC) values for the respective classes are 0.994, 0.977, and 0.988, all of which are remarkably



**Figure 7** F1 score for each class.



**Figure 8** ROC curves (one-vs-all). ROC: receiver operating characteristic, AUC: area under curve.

high and indicate excellent discriminatory level. These results demonstrate that the model can effectively distinguish different ability levels while maintaining a strong balance between true positive rate (TPR) and false positive rate (FPR).

The ROC curves for all classes rise steeply toward the top-left corner, showing that the model achieves a high TPR with minimal FPR across various classes. Specifically, the elementary class with 0.994 AUC exhibits near-perfect performance, ensuring highly accurate classification with almost no misclassifications. The intermediate class with 0.977 AUC and advanced class with 0.988 AUC also perform at an exceptional level, with minimal differences in performance, showcasing the model's robustness and consistency across all categories.

These findings, combined with the consistently high AUC values above 0.97, emphasize the reliability and effectiveness of the SVM model in accurately predicting students' ability levels. The balanced performance across all classes further indicates the model's ability to handle potential imbalances or overlaps in class distributions, making it a powerful and

practical tool for educational assessments and targeted interventions.

## 5 Discussion

In this study, we conduct a questionnaire survey to collect students' self-assessments of areas such as mathematical foundations, programming skills, course interests, and prerequisite knowledge. These results are quantified to obtain the initial sample data. Subsequently, we process the sample data and train the SVM multiclass learning ability classification model tailored to the characteristics of students in our university. This model objectively categorizes students' learning abilities in the system modeling and simulation course into three levels, including elementary, intermediate, and advanced levels.

However, during the analysis of the model's classification results, we discover the misclassification of students with intermediate learning abilities, requiring further in-depth investigation. Based on the current analysis of sample data, these misclassifications appear to be related to students' programming skill levels and corresponding mathematical foundations. More exploration is warranted to determine whether this student subgroup is influenced by specific characteristics.

We observe that the misclassification of this subgroup may be linked to the relatively low number of corresponding sample data points, constrained by the actual situations of students in our university. This calls for a further discussion on how to address students' apprehension toward programming and whether there are alternative methods to more accurately assess their actual programming proficiency.

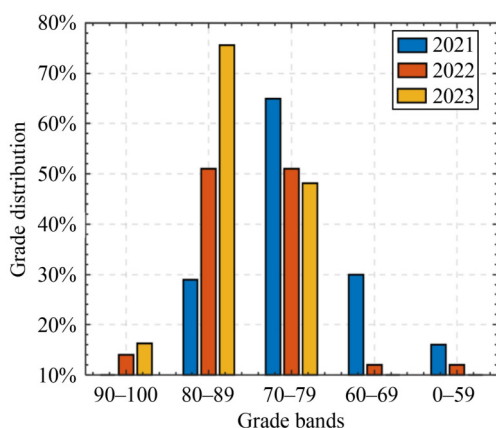
It is worth noting that the model primarily considers features related to the course, such as mathematical foundations, programming skills, and learning interests. We recommend further consideration of the rationality of the selected features and exploration of additional features related to course proficiency assessment to enhance the model's accuracy.

Finally, in the teaching process, modern tools such as AIGC, LLMs, and knowledge graphs can be employed to classify students' learning ability levels. This allows for the customization of personalized learning paths for students in different categories, including learning resources, learning tasks, and personalized teaching, to better cultivate students' learning abilities. Such methods can provide valuable insights and directions for future educational practices.

In the initial stage of teaching practice of the system modeling and simulation course, the model is

used to assess and classify students' learning abilities. Taking into account the different students' learning abilities, the course objective is divided into three levels, including advanced, intermediate, and elementary levels. For advanced students, the course focuses on fostering and assessing their comprehensive application abilities in system modeling and simulation techniques. They provide untreated and complex engineering system cases and are required to conduct in-depth analyses, simplify models, and interpret simulation results in conjunction with real engineering phenomena. For intermediate students, the emphasis is on cultivating their mastery and application abilities in system modeling and simulation techniques, requiring them to analyze some simple simulation phenomena. For elementary students, the emphasis is on cultivating their fundamental grasp of and practice abilities in simulation and modeling techniques, requiring them to model and simulate some simple systems.

The Figure 9 illustrates the course grade distribution of students enrolled in the system modeling and simulation course over the past three years. Since 2022, the course team has used a student learning capability assessment model to classify and assess students' learning abilities. Tailored learning objectives have been set for students with different learning capabilities, along with personalized learning materials. The observed shift in grade distribution indicates a significant improvement in students' performance since implementing student learning capability assessment. The decrease in the number of students in lower grade bands and the increase in those in higher grade bands suggest the effectiveness of the designed student learning capability classification model, providing a pathway for personalized instruction and learning.



**Figure 9** Students' grade distribution in the past three years.

Although the SVM-based student ability classification model demonstrates good performance in the study, there are still three limitations that need to be addressed. First, the performance of the model is highly

dependent on feature selection and data quality. In this study, we use the results of student survey to classify students' learning abilities. However, the question items in these surveys may not have fully included the complexity of students' learning abilities. For example, factors such as students' learning behaviors, cognitive abilities, and socioeconomic backgrounds may also influence learning abilities, but these factors are not included in the model. Therefore, the study still has a big room for improvement in terms of data accuracy. Second, although the SVM performs well on small-scale datasets, the training time and computational complexity of SVMs may become bottlenecks when dealing with large-scale datasets. Especially in relation to large educational datasets, the increase in the number of students and feature dimensions lead to excessively long training times, limiting the model's scalability in real-world applications. Third, SVM, as a black-box model, lacks interpretability, making it difficult for teachers to understand how the model arrives at its classification decisions. This impacts the trust and usability of the model in educational settings.

Future research should consider incorporating more features into the model, such as students' learning behavior data, cognitive test results, and psychological factors, to more comprehensively reflect students' learning abilities, thereby improving accuracy and reliability of the model. In addition, comparing SVMs with other machine learning algorithms, such as random forests, gradient boosting machines, and deep learning models, could help evaluate their performance in accuracy, scalability, and model interpretability.

To address the challenge of growing dataset sizes, future work explores parallelization techniques or uses more scalable models, such as deep learning and decision tree-based models, to improve training efficiency and handle large-scale data. Moreover, improving the interpretability of the model is an important direction for future research. Researchers should work on feature importance analysis and integrate more interpretable models, such as decision trees, with SVM to enhance the transparency of the model and to increase trust among educators.

## 6 Conclusions

This study investigates the construction of an SVM-based student learning ability classification model, emphasizing its significance in educational settings of our university's system modeling and simulation course. By applying this method, we successfully developed an efficient learning ability classification model that accurately categorized students' learning levels based on precourse questionnaire data. The

experimental results demonstrate an impressive accuracy rate of 95.3%.

Of particular note is the model's high accuracy in addressing issues related to student learning ability classification, coupled with its commendable generalization capability. The results not only provide robust support for our university's system modeling and simulation course, but also can be extended and generalized to similar courses in other institutions. This model enables course instructors to acquire more precise assessments of students' course learning abilities and achieve effective categorization.

The findings lay a groundwork for more personalized teaching and learning, advocating the use of AI technologies, such as AIGC, LLMs, and knowledge graphs, to construct personalized learning paths for students with different learning abilities. Moving forward, our research group will continue to leverage these findings to explore the possibilities of using AI technologies to build personalized learning pathways for students, contributing to further innovation and progress in educational settings.

**Acknowledgments** This work was supported by the 2021 Higher Education Teaching Reform Research and Practice Project of SEAC (Grant No. 221057), 2021 Ministry of Education Industry–University Cooperation Collaborative Education Project (Grant No. 202102646007), and 2022 Guizhou Province Gold Course Construction Project.

**Conflict of Interest** The authors declare that they have no conflict of interest related to the content of this paper.

**Ethics Statements** The authors declare that their Institutional Ethics Committee confirmed that no ethical review was required for this study. Written informed consent for participation was not required because all participants' data was anonymized before the statistical analyses were done.

**Data Availability Statements** The data supporting the findings of this study are available from the corresponding author upon reasonable request.

## References

- Ahmed, D. M., Abdulzeez, A. M., Zeebaree, D. Q., & Ahmed, F. H. Y. (2021). Predicting university's students performance based on machine learning techniques. In: *Proceedings of 2021 IEEE International Conference on Automatic Control & Intelligent Systems*. Shah Alam: IEEE, 276–281.
- Alsariera, Y. A., Baashar, Y., Alkaws, G., Mustafa, A., Alkahtani, A. A., & Ali, N. (2022). Assessment and evaluation of different machine learning algorithms for predicting student performance. *Computational Intelligence and Neuroscience*, 2022(1), 4151487.
- Balcioglu, Y. S., & Artar, M. (2023). Predicting academic performance of students with machine learning. *Information Development*.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). Predicting students performance in educational data mining. In: *Proceedings of 2015 International Symposium on Educational Technology*. Wuhan: IEEE, 125–128.
- Imran, M., Latif, S., Mehmood, D., & Shah, M. S. (2019). Student academic performance prediction using supervised learning techniques. *International Journal of Emerging Technologies in Learning*, 14(14), 92–104.
- Kolo, D. K., & Adepoju, S. A. (2015). A decision tree approach for predicting students academic performance. *International Journal of Education and Management Engineering*, 5(5), 12–19.
- Kumar, J., Vashistha, R., Kaur, K., & Singh, S. K. (2023). Machine learning techniques of predicting student's performance. In: *Proceedings of 2023 International Conference in Advances in Power, Signal, and Information Technology*. Bhubaneswar: IEEE, 693–698.
- Li, S., & Liu, T. (2021). Performance prediction for higher education students using deep learning. *Complexity*, 2021(1), 1–10.
- Muraina, I. O., Aiyegbusi, E., & Abam, S. (2022). Decision tree algorithm use in predicting students' academic performance in advanced programming course. *International Journal of Higher Education Pedagogies*, 3(4), 13–23.
- Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of students' academic performance based on courses' grades using deep neural networks. *IEEE Access*, 9, 140731–140746.
- Namoun, A., & Alshantiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, 11(1), 237.
- Qazdar, A., Er-Raha, B., Cherkaoui, C., & Mammas, D. (2019). A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in Morocco. *Education and Information Technologies*, 24, 3577–3589.
- Rastrollo-Guerrero, J. L., Gómez-Pulido, J. A., & Durán-Domínguez, A. (2020). Analyzing and predicting students' performance by means of machine learning: A review. *Applied Sciences*, 10(3), 1042.
- Razaque, A., & Alajlan, M. A. (2020). Supervised machine learning model-based approach for performance prediction of students. *Journal of Computer Science*, 16(8), 1150–1162.
- Vergaray, A. D., Guerra, C., Cervera, N., & Burgos, E. (2022). Predicting academic performance using a multiclassification model: Case study. *International Journal of Advanced Computer Science and Applications*, 13(9), 881–889.
- Verma, U., Garg, C., Bhushan, M., Samant, P., Kumar, A., & Negi, A. (2022). Prediction of students' academic performance using machine learning techniques. In: *Proceedings of 2022 International Mobile and Embedded Technology Conference*.

- Noida: IEEE, 151–156.
- Vijayalakshmi, V., & Venkatachalapathy, K. (2019). Comparison of predicting student's performance using machine learning algorithms. *International Journal of Intelligent Systems and Applications*, 11(12), 34–45.
- Wang, F., Zhang, L., Chen, X., Wang, Z., & Xu, X. (2022). A personalized self-learning system based on knowledge graph and differential evolution algorithm. *Concurrency and Computation: Practice and Experience*, 34(8), e6190.
- Xu, J., Moon, K. H., & van der Schaar, M. (2017). A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing*, 11(5), 742–753.
- Zhang, W., Wang, Y., & Wang, S. (2022). Predicting academic performance using tree-based machine learning models: A case study of bachelor students in an engineering department in China. *Education and Information Technologies*, 27(9), 13051–13066.
- Zohair, A., & Mahmoud, L. (2019). Prediction of student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, 16(27), 1–18.