Journal of Computer Science and Digital Technology, 2026, 2(2), 1-29
https://doi.org/10.71204/n0hwaz15

A Machine Learning–Driven Analysis of Learning Behavior and Learning Atmosphere in Predicting Junior High School English Achievement

Manman Li [1,*]

[1] Foreign Languages Department, Taiyuan Normal University, Jinzhong 030619, China

*** Correspondence:**

Manman Li

lmm051411@163.com

Received: 2 July 2026 / Accepted: 26 August 2026 / Published online: 27 March 2026

Abstract

Accurately predicting students' academic achievement is challenging because learning outcomes are shaped by complex interactions between individual behaviors and contextual learning environments. This study develops a machine learning framework to examine the joint effects of learning behavior and learning atmosphere on English academic performance among 256 junior high school students from three public schools in China. Survey data and standardized English scores were collected across 17 behavioral and environmental indicators. Four machine learning models—Multilayer Perceptron (MLP), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR)—were evaluated using RMSE, MAE, MAPE, and R^2. SVR achieved the best performance (RMSE = 1.9481, MAE = 0.7462, R^2 = 0.7315), outperforming the other models. SHAP analysis identified learning motivation, sleep duration, resource availability, peer influence, and parental education as the most influential predictors, while attendance, household income, and teacher quality showed relatively limited effects. These findings underscore the importance of intrinsic motivation and learning context in predicting English achievement and provide empirical support for learner-centered instruction and data-informed educational interventions under China's "double reduction" policy.

Keywords: Machine Learning; Educational Data Mining; Performance Prediction; Feature Importance Analysis; Supervised Learning; Ensemble Methods; SHAP Analysis

1. Introduction

Against the backdrop of the ongoing implementation of the "double reduction" policy and the deepening integration of educational informatization, junior high school English education is undergoing a fundamental transformation from knowledge transmission to competence-oriented development. This transformation has also prompted a shift in students' learning approaches from passive knowledge reception to active learning and knowledge construction. Within this context, increasing attention has been paid to the role of non-intellectual factors, such as learning motivation, learning behavior, and learning atmosphere, in shaping students' academic achievement. This issue is particularly important at the junior high school level, where students are undergoing critical stages of cognitive, emotional, and behavioral development. Differences in students' learning engagement, behavioral patterns, and perceptions of the learning environment may substantially affect the effectiveness of English learning and contribute to disparities in academic achievement.

Learning motivation provides an important theoretical perspective for understanding students' engagement in English learning. Previous research has suggested that motivation not only directly influences students' willingness to participate in learning activities but may also affect academic performance indirectly through behavioral choices, task persistence, self-regulation, and learning strategies. However, motivation does not operate in isolation. Students' actual learning behaviors represent the behavioral manifestation of their motivational states, reflecting the extent to which students actively participate in learning activities, allocate learning time and effort, complete learning tasks, and adopt effective learning strategies. Meanwhile, the learning atmosphere constitutes an important external condition for learning, encompassing factors such as teacher support, peer interaction, classroom climate, learning resources, and perceived learning support. A supportive learning atmosphere can reinforce students' learning motivation, encourage positive learning behaviors, and provide favorable conditions for sustained academic engagement. Conversely, an unsupportive or unfavorable learning environment may weaken students' engagement and hinder the development of effective learning behaviors.

In actual junior high school English classrooms, however, problems such as insufficient learning initiative, superficial engagement, ineffective learning behaviors, and a lack of supportive learning environments remain common. These problems may reduce learning efficiency and further exacerbate differences in students' English achievement. Therefore, understanding how students' learning behaviors and learning atmosphere are associated with English academic performance is not only theoretically relevant but also practically important for developing more effective and individualized English teaching strategies.

Although existing studies have examined the relationships between learning motivation, learning behavior, learning atmosphere, and academic achievement, several limitations remain. First, a considerable proportion of existing research has focused on university students or adult learners, while empirical evidence concerning junior high school students remains comparatively limited. Second, previous studies have frequently examined learning behavior or learning atmosphere as isolated factors, with relatively less attention being given to their potential synergistic relationships with English academic achievement. Third, limitations in variable measurement and analytical modeling may make it difficult to identify the specific indicators that are most closely associated with students' English performance and to clarify the relationships between behavioral and environmental factors. Consequently, a systematic investigation that integrates learning behavior and learning atmosphere within a unified analytical framework is needed to better understand the determinants of junior high school students' English achievement.

To address these gaps, this study first conducted a systematic literature review to identify the core indicators associated with junior high school students' learning behavior and learning atmosphere and to establish a theoretically grounded measurement framework. Based on the indicators identified from the literature, an empirical questionnaire survey will then be conducted to collect data on junior high school students' learning behaviors, perceptions of the learning atmosphere, and English academic achievement. The collected data will be analyzed using correlation and regression analyses to examine the associations between learning behavior, learning atmosphere, and English scores. Through this process, the study seeks to move beyond a simple description of students' learning characteristics and identify the specific behavioral and environmental factors that are most strongly associated with English achievement.

Furthermore, considering the increasing application of educational informatization and data-driven approaches to educational research, this study extends the conventional statistical analysis by incorporating predictive modeling and feature-importance analysis. A machine learning-based English achievement prediction framework will be developed using learning behavior and learning atmosphere indicators as predictive features. Compared with traditional correlation and regression analyses, machine learning methods can provide an additional perspective for identifying nonlinear relationships and determining the relative predictive importance of different factors. More importantly, feature-importance analysis can help identify potentially actionable intervention points, thereby providing empirical evidence for teachers and schools to design targeted instructional interventions, optimize classroom environments, and develop more individualized English learning support strategies.

Accordingly, this study aims to address the following two research questions:

RQ1: What learning behavior and learning atmosphere indicators are associated with junior high school students' English academic achievement?

RQ2: What are the relationships between junior high school students' learning behavior and learning atmosphere indicators and their English academic achievement?

By integrating literature-based indicator identification, empirical questionnaire analysis, statistical modeling, and machine learning-based prediction, this study aims to provide a more comprehensive understanding of the behavioral and environmental factors associated with junior high school students' English achievement. The findings are expected to contribute to the theoretical understanding of how learning behavior and learning atmosphere are related to language learning outcomes, while also providing practical evidence for constructing a more supportive, engaging, data-informed, and individualized junior high school English learning ecosystem.

2. Related Work

2.1. Learning Motivation

The word "motivation" comes from the Latin word "Movere", which means the internal driving force that drives individual actions. In the field of learning, motivation is regarded as an important psychological factor that stimulates, maintains and guides learning behavior. Although scholars have different definitions of learning motivation, it is generally believed that it is the key force that motivates individuals to actively engage in learning activities and persist in achieving goals.

Gardner et al. (2006) believes that motivation is the core factor that drives individual actions, motivates them to engage in tasks and achieve success. Lambert et al. (1972) conducted a systematic study on foreign language learning motivation from the 1950s to the 1980s and proposed a dichotomy between integrative motivation and instrumental motivation. The former stems from learners' interest in the culture and people of the target language country, while the latter emphasizes the function of language as a tool to achieve goals such as academic qualifications, employment or social status. Dörnyei et al. (1994) reflected on Gardner's dichotomy and believed that the causes of foreign language learning motivation are more complex. He proposed a three-dimensional motivation model, which takes the language itself, the learner individual and the learning environment as the three key dimensions for analyzing motivation, and pointed out that learning motivation is not static, but changes with the learning situation. Subsequently, Dörnyei et al. (1994) proposed the concept of "Directed Motivational Currents" (DMCs) based on empirical research, revealing the continuous motivation mechanism of learners in the process of long-term and high-intensity learning, and promoted the dynamic development of motivation theory research. Williams and Burden (1997) expanded Dörnyei's theory from the perspective of social psychology, believing that learning motivation is not only related to learners' cognition and interest in learning activities, but also includes their willingness to make efforts to achieve learning goals. This view highlights the interactive relationship between learning motivation and social environmental factors, and further enriches the understanding of the essence of second language learning motivation.

2.2. Learning Behavior

Learning behavior refers to the learning process and learning activities, which are influenced by comprehensive factors such as personal learning goals, learning attitudes, learning methods, and learning habits. In recent years, learning behavior has received widespread attention from the academic community as an important topic in educational research. Related studies have systematically explored the characteristics and influencing mechanisms of learning behavior from multiple dimensions such as teaching strategies, psychological factors, data analysis, behavioral modeling, and technology applications. In terms of teaching strategies, Xu et al. (2021) emphasized the student-centered teaching concept through blended learning research, combined flipped classrooms with SPOC, and explored the practical path of personalized learning analysis, providing useful inspiration for understanding individual differences in learning behavior. The goal-oriented active learning (GOAL) system proposed by Li et al. (2021) verified the effectiveness of goal-oriented strategies in promoting learning behavior by improving students' reading participation, autonomous learning behavior, and learning motivation. In the study of learning behavior in special situations, Holzer et al. (2021) focused on the relationship between psychological needs satisfaction and learning behavior of adolescents during the COVID-19 epidemic, emphasized the important influence of perceived relevance on positive emotions and intrinsic learning motivation, and revealed the key role of psychological characteristics in the influencing mechanism of learning behavior. Ghani et al. (2021) analyzed effective learning behavior elements such as critical thinking, problem solving and cooperation ability from the perspective of problem-based learning (PBL), and provided a theoretical basis for the optimization of learning behavior in occupational medical education.

With the rise of data-driven educational research, learning behavior analysis has gradually become a research hotspot. Akhuseyinoglu et al. (2021) used learning analysis technology to identify student behavior patterns and associate them with learning performance, providing data support for the design of personalized learning paths. Wang et al. (2023) analyzed the emotional state in online learning through learning behavior data, emphasizing the important role of emotional factors in learning behavior. Li et al. (2023) further explored the dynamic interaction between learning behavior and learning environment, pointed out the significant impact of environmental factors on students' learning participation, and provided a theoretical basis for optimizing the learning environment. In the research of behavioral models and algorithms, Nasiriany et al. (2022) proposed a behavioral primitive framework based on enhanced reinforcement learning, which realized complex task operations through predefined behavioral modules, demonstrating the application potential of behavioral primitives in learning behavior modeling. Wei et al. (2022) proposed a multi-behavior comparative learning framework, which improved the behavioral understanding ability in the recommendation system by transferring knowledge of different types of behaviors, highlighting the deep mining of diverse behavioral characteristics. In addition, the research on learning behavior under technology empowerment has also shown new development trends. Chen et al. (2024) studied the impact of AI-assisted gamification learning on scientific learning outcomes, intrinsic motivation and cognitive load, and emphasized the application potential of intelligent technology in promoting learning behavior.

The above research on learning behavior covers multiple fields such as teaching strategies, psychological mechanisms, data analysis, behavioral modeling and technology application, forming a multi-dimensional and multi-level theoretical system and practical path, which provides a solid theoretical foundation and methodological support for in-depth understanding and effective promotion of learning behavior. However, despite the fruitful results of related research, there are still problems such as relatively concentrated research objects and the need to expand research perspectives. To further understand the current hot topics and shortcomings in learning behavior research, the author conducted an advanced search on the topic of "learning behavior" through China National Knowledge Infrastructure (CNKI) on August 4, 2025. The search results show that current journal research focuses primarily on areas such as "online learning," "autonomous learning," "empirical research," "learning motivation," and "influencing factors." Existing empirical research has primarily focused on contexts such as "online learning," "network learning," and "higher education," while systematic empirical analysis of junior high school students' learning behavior remains relatively scarce. Based on this, this article, based on the perspective of learning motivation, conducts an empirical study on the relationship between junior high school students' learning behavior, learning atmosphere, and English achievement, attempting to address the research gap in this area at the basic education level. Regarding the specific dimensions of learning behavior, this article analyzes the relationship between "study hours" "extracurricular activities," "sleep time," "physical activities," "learning motivation," and "number of tutoring frequency" and English achievement. This research also provides a more practical perspective for a deeper understanding of the characteristics and influencing mechanisms of junior high school students' learning behavior.

2.3. Learning Atmosphere

The definition of learning climate is multifaceted, covering multiple levels such as psychological, social, cultural and physical. According to relevant literature, learning climate refers to the place where learning takes place and the various elements it contains, which together affect learners' experience and learning outcomes (Müller et al.,2021). Specifically, the learning environment includes not only physical spaces such as classrooms, schools or workplaces, but also virtual spaces, emphasizing the diversity and inclusiveness of spaces (Berliana et al.,2024). In addition, learning climate also involves psychological and social factors of the learning place. For example, classroom climate emphasizes the social, emotional and physical environmental characteristics, emphasizing that the learning space should make students feel safe, supported and motivated (Virtanen et al.,2022). A positive learning climate should make learners feel comfortable, encouraged and motivated to learn. At a broader level, learning climate also includes cultural factors and psychological safety, which are important for promoting effective learning. For example, a psychologically safe learning environment is considered to be key to supporting learners to express their opinions and try new things. Therefore, the definition of learning climate is not just the physical attributes of the space, but also emphasizes its psychological and social support role, aiming to create an environment conducive to learning and development.

Regarding the study of learning atmosphere, scholars have explored its importance and influence mechanism in education from multiple perspectives. Ma'ruf et al. (2021) pointed out that creating an effective learning atmosphere is inseparable from the continuous professional development of teachers. Teachers should actively build an open, inclusive and dynamic learning environment by adopting a variety of teaching strategies. The study emphasized that the professional quality and teaching ability of teachers directly determine the quality of the learning atmosphere. Therefore, teachers need to continuously improve their teaching level to achieve more efficient teaching results. In the field of college English teaching, Xu et al. (2021) found that the rational application of information and communication technology (ICT) can effectively optimize the learning experience and learning atmosphere based on the hybrid teaching model. Through observation and questionnaire surveys, the study confirmed that the hybrid teaching model combining the Discover, Learn, Practice, Collaborate and Assess (DLPCA) model can significantly improve students' learning enthusiasm and cooperative spirit, thereby building a more interactive and attractive learning environment. In the context of special education, Tarigan et al. (2022) successfully stimulated children's learning interest and motivation by designing children's activity manuals. The study shows that incorporating interesting and interactive teaching materials can help create a relaxed and enjoyable learning atmosphere and promote the cognitive and skill development of young children.

At the same time, digital learning systems have also shown significant advantages in creating an academic atmosphere. Tsani et al. (2023) used Islamic junior high schools as the research object to evaluate the impact of digital learning systems on the academic atmosphere. The results showed that the system can effectively stimulate students' enthusiasm for learning, cultivate their critical and innovative thinking, and thus form a positive learning culture. In a special teaching environment, Aras et al. (2022) explored the lecture teaching method during the "new normal" period and pointed out that in the context of the epidemic, reasonable teaching preparation and clear teaching objectives are crucial to maintaining teaching continuity and a stable learning atmosphere. The study emphasized that teaching strategies that flexibly respond to environmental changes are key factors in ensuring the healthy development of the learning atmosphere.

In addition, Zheng et al. (2024) studied the relationship between learning atmosphere, negative emotions and willingness to communicate, and found that a positive learning atmosphere can effectively reduce students' negative emotions, enhance their willingness to communicate and cooperate, and thus promote deeper learning interactions. Aisyah et al. (2025) further pointed out that digital technology, as an interactive learning medium, can not only improve the learning atmosphere in primary schools, but also enhance the fun and participation of learning, which is in line with the development trend of digital transformation of contemporary education. In summary, the optimization of learning atmosphere is the result of the combined effect of multiple factors, including the improvement of teachers' professional ability, the effective application of information technology, the innovation of teaching strategies, and the fun and interactivity of teaching resources. Different studies have revealed the important paths to create a positive learning atmosphere from different perspectives, providing important theoretical support and practical inspiration for the improvement of educational practice and future research.

3. Research Methods

3.1. Research Design

This study employed a quantitative research design, utilizing a structured questionnaire survey combined with actual student English achievement data. The study aimed to explore the relationship between junior high school students' learning behaviors, learning atmosphere, and English academic achievement. This approach helps to fully understand the combined influence of internal behavioral and external environmental factors on English learning outcomes.

3.2. Research Subjects and Samples

The study subjects were students from grades 7 to 9 from three public junior high schools in Linfen, China. Stratified random sampling was used to ensure representativeness of grade and gender. A total of 272 questionnaires were distributed, of which 256 were valid, yielding a validity rate of 94.15.

3.3. Instruments and variable measurement

English scores were derived from standardized test scores at the end of the semester. Cronbach's coefficients for each subscale were all above 0.80, indicating high reliability. KMO sampling adequacy test values were all greater than 0.85, and Bartlett's test of sphericity was significant ($p < 0.001$), indicating that the data were suitable for factor analysis (Table 1).

https://doi.org/10.71204/n0hwaz15

In this study, indicators of learning behavior and learning atmosphere were considered independent variables, and English scores were the dependent variable to explore the relationships between multiple independent and dependent variables. Descriptive statistical analysis, correlation analysis, and predictive analysis were performed on the collected learning behavior and learning atmosphere data. Data analysis was conducted using SPSS 27.0 software and included four main aspects: First, descriptive statistics were used to calculate the mean, standard deviation, and distribution characteristics of each variable to understand the basic characteristics of the sample; second, Pearson correlation analysis was used to test the correlations between variables; third, machine learning and other models were used to analyze the predictive effects of various dimensions of learning behavior and learning atmosphere on English scores.

Table 1. Learning Behavior Indicators and Learning Environment Indicators

Indicator Name	Description
Learning Behavior Indicators	
Learning Time	Average daily time spent on extracurricular English learning.
Extracurricular Activities	Frequency of participation in English-related clubs or
Sleep Time	Average nightly sleep hours.
Physical Activity	Frequency of physical exercise and sports.
Learning Motivation	Intrinsic and extrinsic motivation for learning English.
Tutoring Frequency	Frequency of participation in extracurricular tutoring or
Learning Environment Indicators	
Attendance Rate	Regularity of school attendance.
Parental Involvement	Degree of parental support for learning.
Resource Availability	Provision of learning materials and tools.
Internet Access	Convenience of internet learning conditions.
Household Income	Level of family economic support.
Parental Education Level	Educational attainment of parents.
Teacher Quality	Teachers' professional and teaching capabilities.
School Type	Differences in school nature and resources.
Peer Influence	Role of learning interactions among classmates.
Distance	Commuting distance between home and school.
Learning Barriers	Obstacles encountered during the learning process.

4. Data Analysis and Results

4.1. General Description of Related Work

This study conducted descriptive statistical analysis of questionnaire data on learning behavior and learning atmosphere variables. The results showed that the overall mean for English scores was 77.26, which was above average, but the standard deviation and variance were large, indicating significant inter-student variability and a relatively flat distribution, with significant individual differences. Regarding learning behavior, the mean for study hours was 19.39, the average attendance rate was 79.35.

4.2. Overall Representation of Learning Behavior

Learning behavior constitutes a key indicator of students' engagement and motivational commitment in the learning process. Drawing on questionnaire responses, this study examines student learning behavior across six dimensions: study hours outdoor activities, sleep hours, physical activities, learning motivation, and number of tutoring frequency. These six dimensions also serve as critical factors in assessing and influencing students' English scores. The corresponding visualizations are presented in Figure 1, which displays: a) Study Hours, b) Outdoor activities, c) Sleep Hours, d) Physical Activities, e) Learning motivation, and f) Tutoring frequency.

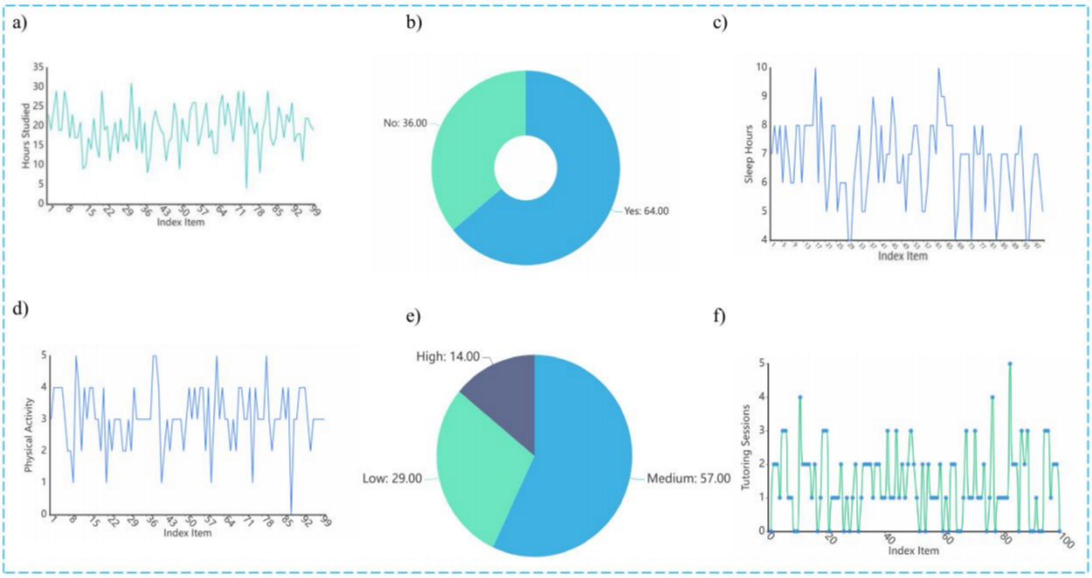

Figure 1. Overall representation of learning behavior.

The following is a detailed explanation of three key indicators that influence English scores: a) Study Hours, e) Learning Motivation, and f) Number of Tutoring frequency. a) Study Hours Analysis: The horizontal axis represents student number (approximately 100 students), and the vertical axis represents study hours. This chart shows that most students study between 15 and 25 hours, which is relatively balanced overall. However, a small number of students study less than 10 hours or more than 30 hours, indicating either insufficient or excessive study commitment. There are significant differences between students, and extreme values may reflect specific study habits or stressors. Targeted attention should be given to low- and high-motivated students. e) Learning Motivation Analysis: The majority of students' learning motivation is at a medium level (57.00), followed by a low level (29.00), with the highest level accounting for the smallest proportion (14.00). Overall, student motivation is concentrated at a medium level, with a lack of highly motivated students. Teaching should focus on further stimulating students' enthusiasm for learning. f) Tutoring Frequency Analysis: Figure 2 shows a line chart of tutoring frequency. The horizontal axis represents different students or learning stages (index items), and the vertical axis represents the number of tutoring frequency. Overall, the graph shows significant fluctuations, reflecting a wide range of tutoring frequency. Most students' tutoring frequency ranges from 0 to 5, with relatively few extreme highs. There is also significant variation among students, with some students receiving almost no tutoring records, while others receive more frequently. Frequent fluctuations in the curve indicate that tutoring behavior is scattered and lacks a consistent pattern.

4.3. Overall Representation of the Learning Atmosphere

The influence of the learning environment on students' English scores should not be underestimated. In this study, eleven indicators related to the learning environment were collected, namely peer influence, household income, internet access, teacher quality, resource availability, attendance rate, learning barriers, parental involvement, parental education level, home–school distance, and school type. The corresponding visualizations are presented in Figure 2, which illustrates: a) peer influence, b) household income, c) internet access, d) teacher quality, e) resource availability, f) attendance, g) learning barriers, h) parental involvement, i) parental education level, j) home–school distance, and k) school type.

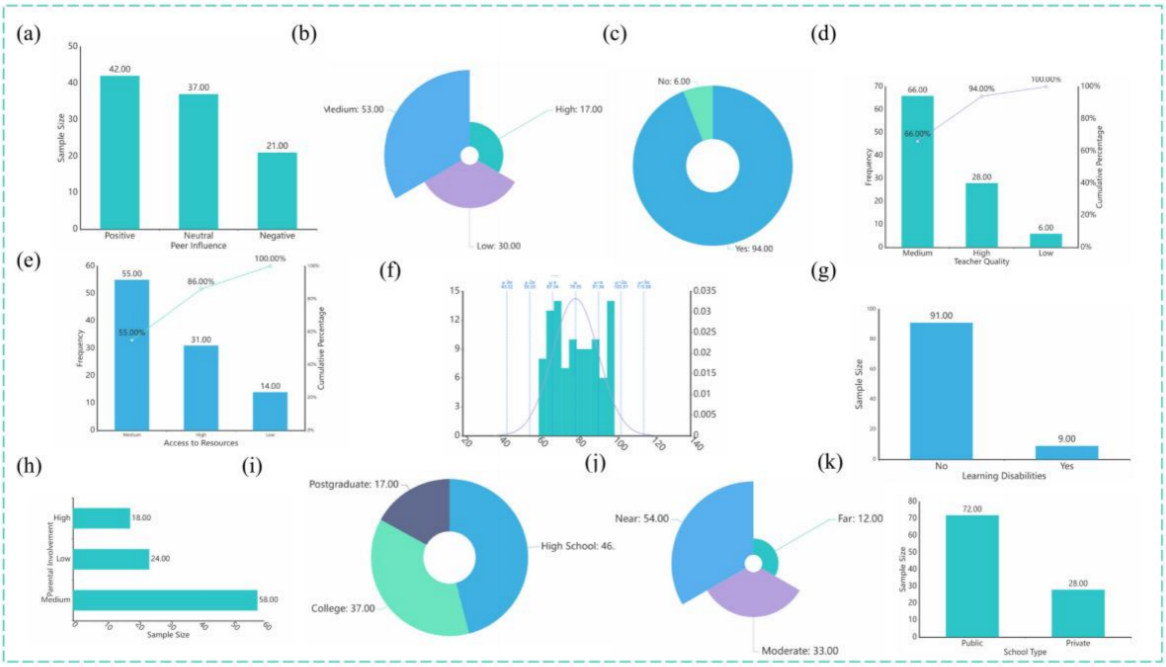

Figure 2. The overall performance of the learning atmosphere.

https://doi.org/10.71204/n0hwaz15

The following presents a statistical analysis of five key learning climate indicators: teacher quality, resource availability, attendance, parental involvement, and parental education.

(d) Teacher Quality: The distribution of teacher quality levels indicates that teachers of moderate quality constitute the largest proportion. (e) Resource Availability: A cumulative percentage line graph illustrates the distribution of learning resources. (f) Attendance Rate: The distribution approximates a normal curve with a mean of 79.35, reflecting generally good classroom attendance. Most students' attendance rates fall within one standard deviation of the mean (67.34 to 91.36), indicating relatively stable and consistent attendance behavior. This high level of regularity fosters sustained student engagement in class activities—crucial for English learning—and helps create a positive learning atmosphere. Furthermore, the proportion of extreme cases is very low (tail probability < 0.035), underscoring the stability of attendance rates and suggesting that attendance serves primarily as a source of consistent, positive support rather than a major driver of fluctuations in English proficiency. (h) Parental Involvement: The distribution of parental involvement is shown. (i) Parental Education Level: The distribution of parental education levels indicates that a high school education is the most common.

4.4. Correlation Analysis Between Learning Behavior and English Scores

To understand the correlations and educational significance of learning behavior indicators, this study focused on behavioral aspects that influence English scores. Subsequently, correlation analysis was conducted using Spss 27.0 to explore the relationship between these behavioral indicators and English scores. Using English scores as the dependent variable and behavioral indicators as the independent variables, this study aimed to determine correlation coefficients, using this method to identify significance and assess significant relationships between variables. The results of this analysis are shown in Figure 3, which demonstrates the correlation between learning behavior and English scores.

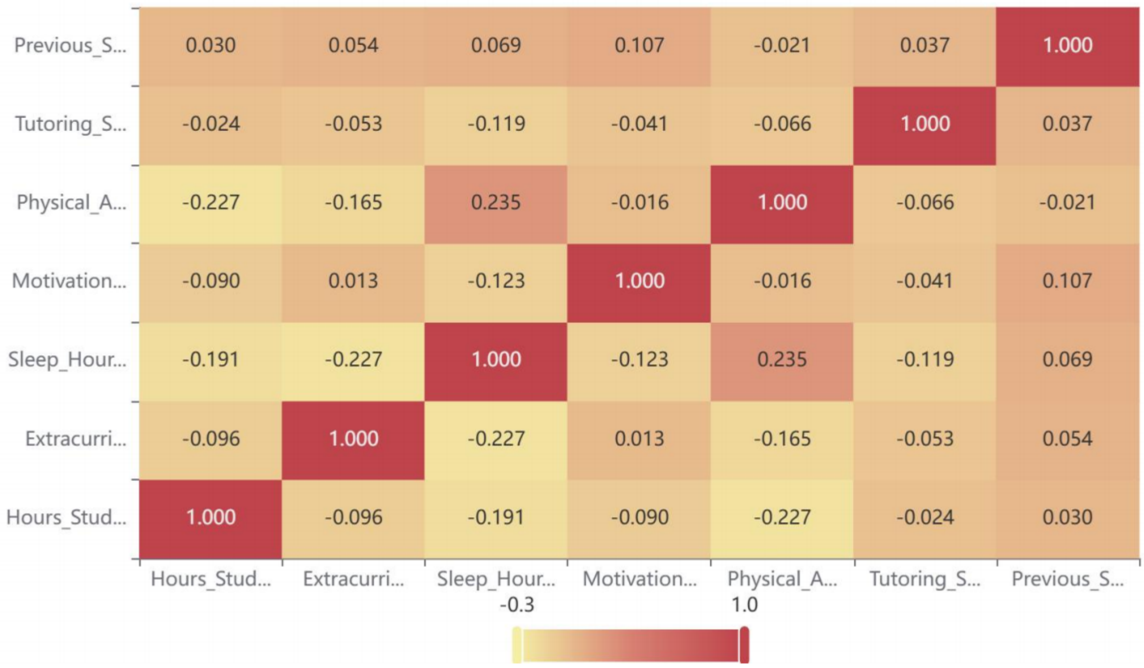

Figure 3. Correlation coefficient heat map between learning behavior and English scores.

Journal of Computer Science and Digital Technology, 2026, 2(2), 1-29
https://doi.org/10.71204/n0hwaz15

Overall, the correlations between existing English scores and the six learning behavior indicators were weak, but certain directional characteristics were observed. Among them, learning motivation (0.107) was positively correlated with academic performance, indicating that stronger motivation is associated with better academic performance, highlighting the critical role of motivation in learning success. sleep hours (0.069) was also positively correlated with academic performance. While the correlation was not strong, it suggests that adequate sleep contributes to cognitive recovery and learning outcomes. Extracurricular activities (0.054) also showed a slight positive correlation with academic performance, indicating that appropriate extracurricular activities can promote well-rounded development and indirectly support academic performance. Study hours (0.030) and tutoring frequency (0.037) showed very limited correlations with academic performance, but both were positive, suggesting that they may provide some support but are not decisive factors. In contrast, physical activity (-0.021) showed a slight negative correlation with academic performance, suggesting that excessive physical activity may squeeze study hours but the impact was minimal and negligible.

4.5. Correlation Analysis between Learning Atmosphere and English Achievements

Using English scores as the dependent variable and learning atmosphere indicators as the independent variables, we used the same method to investigate and analyze the correlation between learning atmosphere and English scores. This method was used to identify significance and assess significant relationships between variables. The results of this analysis are shown in Figure 4, which demonstrates the correlation between learning atmosphere and English scores.

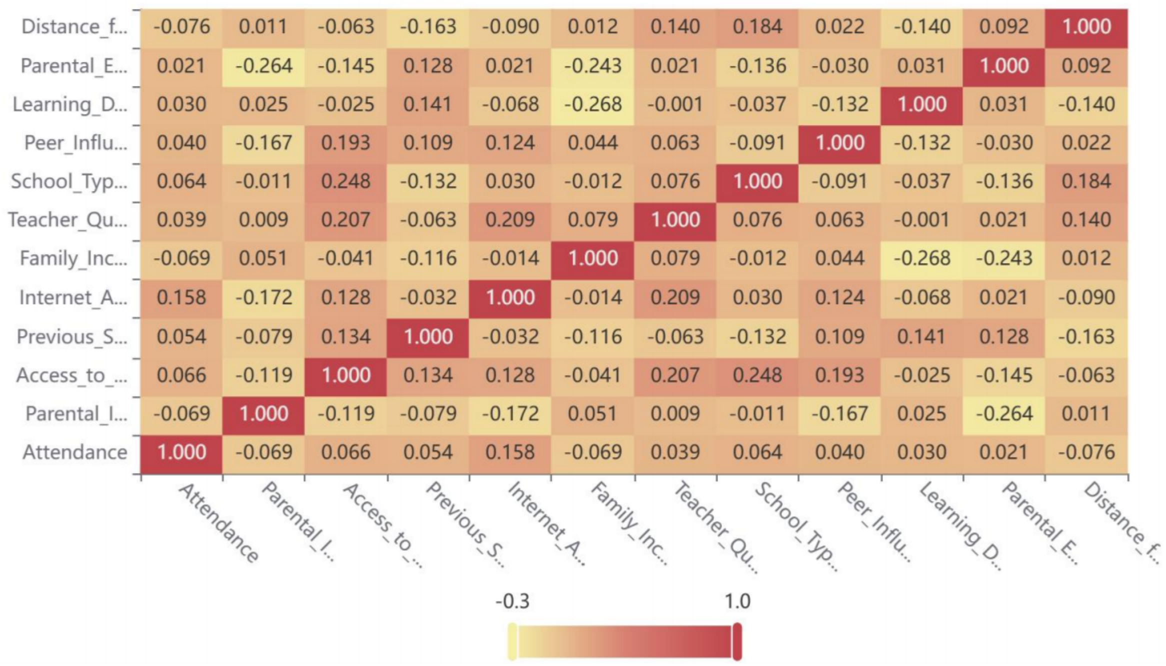

Figure 4. Correlation coefficient heat map between learning atmosphere and English scores.

Correlation analysis results show significant positive correlations between previous English scores and the four learning climate indicators. First, resource availability (0.134) is positively correlated with academic performance, indicating that sufficient learning materials, technical equipment, library, and online support provide students with more learning opportunities and convenience, thereby improving academic performance. Second, peer influence (0.109) demonstrates that positive learning habits and academic attitudes have a positive impact on peers, and a positive learning atmosphere contributes to overall academic improvement. Furthermore, the positive correlation with learning barriers (0.141) needs to be interpreted in light of the variable definition: if a larger value indicates "more learning barriers," this result may reflect a problem with the coding direction; if a larger value indicates "greater ability to cope with learning barriers," it indicates that students possess greater adaptability and problem-solving skills in learning, leading to better academic performance. Finally, parental education level (0.128) is also positively correlated with academic performance, indicating that parents with higher levels of education tend to create a more favorable learning environment and provide more effective academic support for their children, thereby promoting academic development. Indicators with weak or near-zero correlations with existing English scores include attendance rate, internet access, family income, teacher quality, school type, distance between home and school, and parental involvement. Specifically, attendance rate (0.054) has almost no direct correlation with academic performance, suggesting that "attending school" does not necessarily equate to "effective learning," and that classroom quality and student autonomy may be more critical. Internet access (-0.032) has a near-zero or even slightly negative correlation, possibly because the internet can provide both learning resources and entertainment distractions. household income (-0.116) has a slight negative correlation with academic performance, suggesting that economically advantaged families do not necessarily guarantee academic excellence for their children; they may even be distracted by extracurricular resources, which can affect their grades. Similarly, teacher quality (-0.063) did not directly translate into academic advantage in the sample, possibly due to differences in assessment methods or schools. The slight negative correlation between school type (-0.132) and academic performance may be due to the more competitive environment in some "key schools," resulting in greater disparity in academic performance. The negative correlation with home-school distance (-0.163) suggests that commuting costs consume learning time and energy, thus affecting academic performance. Finally, the weak negative correlation with parental involvement (-0.079) may indicate that greater parental involvement tends to occur among students with relatively low academic performance, thus showing a statistically negative correlation.

5. Predictive Analysis of Various Indicators and English Scores

Correlation analysis has identified the core factors influencing junior high school English scores (such as learning motivation and resource availability), but the practical question of how to predict performance based on these factors remains unresolved. To further translate these findings into actionable educational intervention tools and validate the comprehensive predictive power of these core factors, this section constructs an English scores prediction model based on the collected data from 256 valid samples. The detailed process is as follows.

In this study, the data was first preprocessed, and the "English score" was treated as the prediction target variable, leaving the original score as the regression prediction target. At the same time, the 17 indicators of learning behavior and learning atmosphere were standardized (such as using Z-score standardization) to eliminate the impact of dimensional differences on the model effect. The data set was divided into a training set (205 samples) and a test set (51 samples) in an 8:2 ratio to ensure scientific verification of the model's generalization ability. In terms of model selection and construction, three typical models were selected for comparison: multi-layer perceptron model, random forest model, XGBoost regression model and support vector machine regression SVR model.

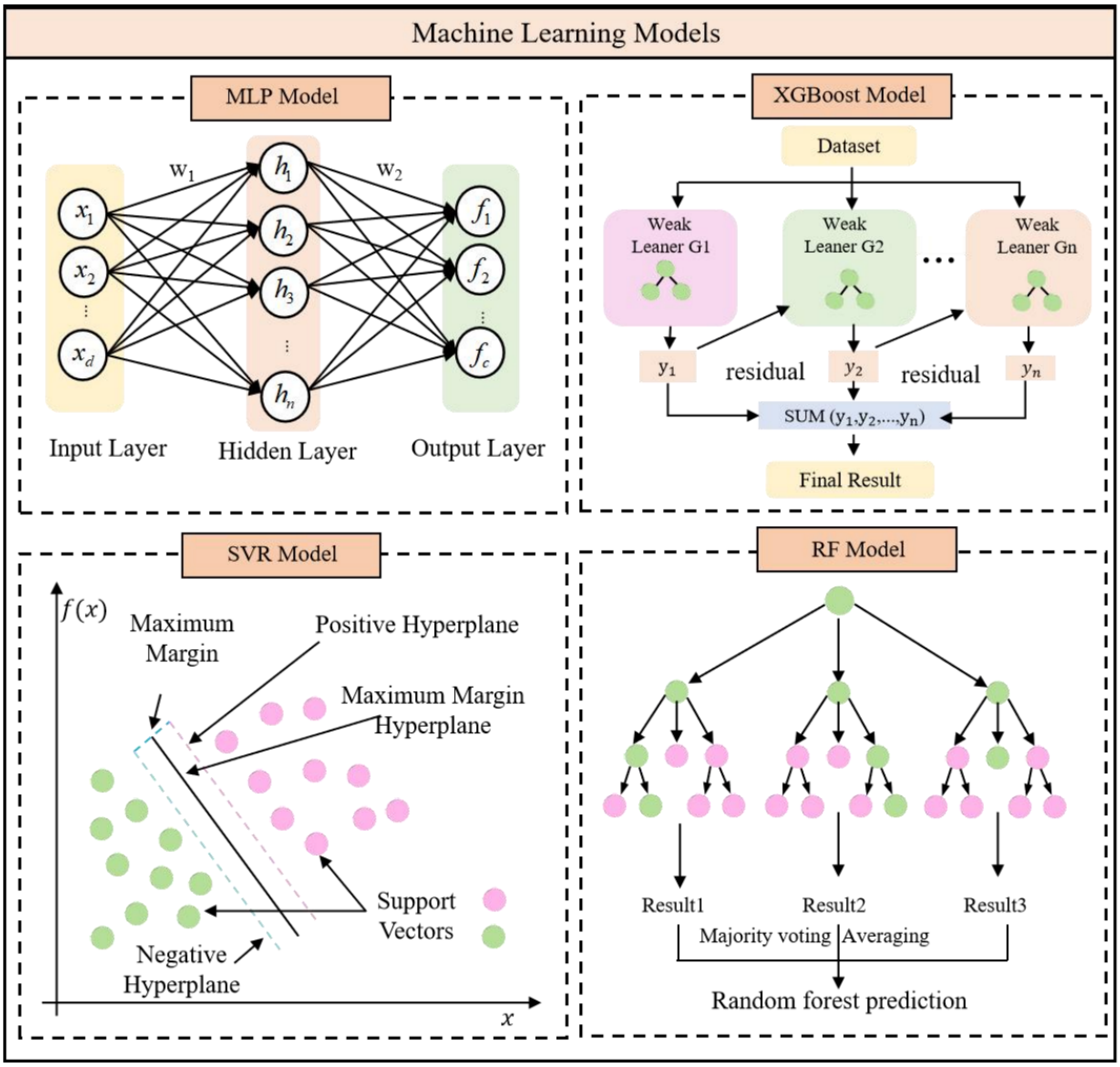

Figure 5. Introduction to Machine Learning Models.

5.1. Introduction to each prediction model

Multilayer Perceptron (MLP) is a fully connected feed-forward network that alternates linear transformations with element-wise nonlinear activations. (Wang et al., 2026) universal approximation theory guarantees that a single hidden layer of sufficient width can represent any Borel-measurable function, while stochastic back-propagation efficiently minimises the empirical risk even in high-dimensional parameter spaces.

XGBoost is a gradient-boosting framework that additively constructs an ensemble of regularised regression trees by minimising a second-order Taylor approximation of the loss, coupled with sparse-aware split finding and implicit missing-value handling, yielding state-of-the-art accuracy on tabular data with minimal manual feature engineering (Shokri et al., 2019).

Support Vector Regression (SVR) maps inputs into a high dimensional feature space induced by a kernel function, then fits an ε insensitive tube that penalises deviations only outside the margin. the resulting convex quadratic programme depends on a sparse subset of support vectors, thereby controlling model capacity through the principle of large margin geometry.

Random Forest (RF) generates an ensemble of axis-parallel decision trees using bootstrap sampling and random feature subsampling (Gaur et al., 2022). the average of de-correlated trees reduces variance without increasing bias, while out-of-bag data provide an internal unbiased estimate of generalisation error, making the method robust to outliers and computationally scalable.

5.1.1. Multilayer Perceptron Model

The Multilayer Perceptron (MLP) is a class of feedforward artificial neural networks that consists of at least three layers: an input layer, one or more hidden layers, and an output layer (Yuan et al., 2025). Each neuron in one layer connects with a certain weight to every neuron in the following layer, enabling the model to learn complex non-linear relationships between input features and output variables.

In this study, the MLP model was implemented to capture the potential non-linear interactions among the 17 learning behavior and learning atmosphere indicators. The model was configured with two hidden layers containing 128 and 64 neurons, respectively, utilizing the ReLU activation function to introduce non-linearity. The Adam optimizer with a learning rate of 0.001 was employed to minimize the mean squared error loss function. Dropout regularization (rate = 0.3) was applied to prevent overfitting. The model was trained for 200 epochs with early stopping (patience = 20) to ensure optimal convergence.

5.1.2. XGBoost Model

XGBoost (Extreme Gradient Boosting) is an advanced implementation of gradient boosting machines that constructs an ensemble of weak prediction models, typically decision trees, in a sequential manner. Each subsequent tree is built to correct the errors made by the previous ensemble, thereby progressively improving prediction accuracy (Keshtegar et al., 2020). The objective function of XGBoost combines a loss function and a regularization term:

$$L(\varphi)=\Sigma_i l(y_i,\hat{y}_i)+\Sigma_k(f_k)$$

where l is a differentiable convex loss function measuring the difference between the prediction $\hat{y}_i$ and the target y_i, and (f_k) penalizes the complexity of the model to control overfitting.

The XGBoost model was selected for its robust performance on structured tabular data and its ability to handle feature interactions automatically (Yuan et al., 2025). Key hyperparameters were optimized through grid search: maximum tree depth = 6, learning rate (eta) = 0.05, number of estimators = 200, minimum child weight = 3, and subsample ratio = 0.8. The model also provides feature importance scores based on gain, frequency, and coverage metrics, offering valuable insights into which learning behavior and atmosphere indicators contribute most significantly to English score prediction.

5.1.3. Support Vector Machine Regression

Support Vector Regression (SVR) extends the principles of Support Vector Machines to regression problems by finding a function that deviates from the actual targets by a value no greater than ε for each training point, while remaining as flat as possible.

In this study, the SVR model employed a Radial Basis Function (RBF) kernel to capture non-linear patterns in the relationship between learning indicators and English scores. The kernel coefficient (gamma) was set to `scale' (1/(n_features × X.var())), the regularization parameter C was set to 1.0, and the epsilon parameter defining the margin of tolerance was set to 0.1. The model was trained using the epsilon-insensitive loss function, which provides robust predictions by ignoring errors within the epsilon tube.

5.1.4. Random Forest Model

Random Forest (RF) is an ensemble learning method that constructs multiple decision trees during training and outputs the mean prediction of individual trees for regression tasks. The algorithm introduces randomness in two ways: bootstrap sampling of training data for each tree (bagging) and random selection of feature subsets at each split node. This dual randomization mechanism reduces variance and mitigates overfitting.

The Random Forest model was implemented to leverage its robustness to outliers and its capacity to handle high-dimensional data without explicit feature selection. The model was configured with 100 estimators (trees), a maximum tree depth of 15 to control model complexity, and a minimum of 5 samples required to split an internal node. The model evaluated all 17 features at each split (max_features = `sqrt') to maintain diversity among trees. Additionally, Random Forest provides built-in feature importance measures based on mean decrease in impurity (Gini importance), facilitating interpretation of which learning behavior and atmosphere variables exert the greatest influence on English academic performance.

These four models represent different machine learning paradigms: MLP captures complex non-linear patterns through neural architectures; XGBoost excels in handling structured data with gradient boosting; SVR provides robust predictions through margin-based optimization; and Random Forest offers stability through ensemble averaging. By comparing these diverse approaches, this study ensures comprehensive evaluation of the predictive relationships between learning indicators and English scores, with model selection based on performance metrics including RMSE, MAE, MAPE, and R^2.

5.2. Evaluation Metrics

Predictive accuracy on the independent test set was quantified by Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R^2). Feature importance was extracted from tree-based models via Gini impurity reduction.

$$RMSE=(1)/(n)\Sigma_i=1^n(y_i-\hat{y}_i)^2$$

$$MAE=(1)/(n)\Sigma_i=1^n|y_i-\hat{y}_i|$$

$$R^2=1-\Sigma_i=1^n(y_i-\hat{y}_i)^2\Sigma_i=1^n(y_i-\bar{y})^2$$

5.3. Parameter Settings for Each Model

5.3.1. Random Forest Model

Random Forest is an ensemble learning method that builds multiple decision trees and aggregates their predictions to enhance model generalization and accuracy. As shown in Table 2. In this study, 100 trees were used (n_estimators=100), achieving a good balance between performance and computational efficiency. Too few trees may cause underfitting, while too many trees can increase computational cost with diminishing returns. The maximum tree depth was set to 20 (max_depth=20), which limits the complexity of each tree to prevent overfitting. Deeper trees capture complex feature interactions but may also learn noise from the training data. This setting ensures a balance between model capacity and generalization. To further control tree growth, the minimum samples required to split a node (min_samples_split=5) and to form a leaf node (min_samples_leaf=2) were specified. These parameters act as regularization terms to avoid creating nodes containing only a few samples, reducing overfitting risk. Parallel computation (n_jobs=-1) was enabled to utilize all CPU cores and significantly reduce training time.

Table 2. Parameter settings of the Random Forest model

Parameter	Value	Description
n_estimators	100	Number of decision trees in the forest
max_depth	20	Maximum depth of each tree
min_samples_split	5	Minimum number of samples required to split an internal node
min_samples_leaf	2	Minimum number of samples required at a leaf node
random_state	42	Random seed to ensure reproducibility
n_jobs	-1	Number of CPU cores for parallel computation

5.3.2. Support Vector Machine Model

Support Vector Machine (SVM) is a powerful supervised learning algorithm, particularly effective for nonlinear regression problems. As shown in Table 3. The Radial Basis Function (RBF) kernel (kernel='rbf') was used to map input features into a high-dimensional space, capturing nonlinear relationships among features. The RBF kernel is known for its strong generalization ability, making it ideal for complex data distributions. The regularization parameter was set to 100 (C=100), controlling the trade-off between model complexity and training error. A larger C value emphasizes minimizing training errors, which works well when the dataset has low noise. The kernel coefficient was set to automatic scaling (gamma='scale'), ensuring adaptive adjustment according to the number and variance of features. The -tube parameter was set to 0.1 (epsilon=0.1), defining a margin of tolerance within which prediction errors are not penalized. This setting provides robustness to small prediction deviations. Considering the heavy computational cost of kernel-based matrix operations, a large cache size (cache_size=1000) was used to accelerate training. All features were standardized before model fitting, as SVMs are highly sensitive to feature scaling.

Table 3. Parameter settings of the Support Vector Machine model

Parameter	Value	Description
kernel	rbf	Type of kernel function (Radial Basis Function)
C	100	Regularization parameter
gamma	scale	Kernel coefficient (automatically scaled)
epsilon	0.1	-tube parameter
cache_size	1000	Kernel cache size (MB)

5.3.3. Multi-Layer Perceptron Model

The Multi-Layer Perceptron (MLP) is a feedforward neural network that learns complex data patterns through multiple nonlinear transformations. As shown in Table 4. This study designed a three-layer hidden structure (hidden_layer_sizes=(128, 64, 32)) with decreasing neuron counts across layers. The first layer with 128 neurons captures high-dimensional features, the second layer (64 neurons) compresses and abstracts them, and the third layer (32 neurons) extracts key representations. This pyramid-shaped structure helps the model progressively learn hierarchical representations and improves generalization. The activation function used was ReLU (activation='relu'), which accelerates convergence and mitigates the vanishing gradient problem. The Adam optimizer (solver='adam') was adopted due to its efficiency and adaptive learning rate mechanism. The initial learning rate was set to 0.001 (learning_rate_init=0.001) with an adaptive strategy (learning_rate='adaptive'), allowing dynamic adjustment when the loss plateaus.

Table 4. Parameter settings of the Multi-Layer Perceptron model

Parameter	Value	Description
hidden_layer_sizes	(128, 64, 32)	Structure of hidden layers
activation	relu	Activation function (ReLU)
solver	adam	Optimization algorithm (Adam optimizer)
alpha	0.001	L2 regularization parameter
batch_size	32	Batch size for mini-batch gradient descent
learning_rate	adaptive	Learning rate strategy (adaptive)
learning_rate_init	0.001	Initial learning rate
max_iter	500	Maximum number of iterations
early_stopping	True	Whether to enable early stopping
validation_fraction	0.1	Proportion of validation set
n_iter_no_change	20	Patience parameter for early stopping

5.3.4. XGBoost Model

XGBoost (Extreme Gradient Boosting) is an efficient gradient boosting algorithm that sequentially builds decision trees to optimize the loss function. As shown in Table 5. In this study, 200 trees were used (n_estimators=200), providing a balance between model complexity and performance. The tree depth was set to 6 (max_depth=6), a typical configuration in XGBoost that favors shallow trees to prevent overfitting by leveraging an additive learning process. The learning rate was set to 0.1 (learning_rate=0.1), controlling the contribution of each tree to the final prediction. A smaller learning rate requires more trees but enhances model stability and generalization. Random sampling was applied at both sample and feature levels (subsample=0.8, colsample_bytree=0.8), functioning as a dropout-like regularization mechanism to reduce overfitting. To further regularize the model, the minimum child weight was set to 3 (min_child_weight=3), while both L1 (reg_alpha=0.1) and L2 (reg_lambda=1) regularization terms were added to prevent overcomplex trees. The parameter = 0 (gamma=0) indicates that no additional loss reduction threshold was required for node splitting (Figure 7). During training, an evaluation set was used to monitor performance and detect overfitting early. Parallel computation (n_jobs=-1) was enabled to fully utilize multicore processors, greatly improving training efficiency. These configurations ensure that XGBoost performs robustly when modeling complex tabular data such as student performance prediction.

Table 5. Parameter settings of the XGBoost model

Parameter	Value	Description
n_estimators	200	Number of boosting trees
max_depth	6	Maximum depth of each tree
learning_rate	0.1	Step size shrinkage (learning rate)
subsample	0.8	Subsample ratio of training instances
colsample_bytree	0.8	Subsample ratio of features per tree
min_child_weight	3	Minimum sum of instance weights per leaf
gamma	0	Minimum loss reduction required to make a split
reg_alpha	0.1	L1 regularization parameter
reg_lambda	1	L2 regularization parameter
random_state	42	Random seed for reproducibility
n_jobs	-1	Number of CPU cores used for parallel computation

https://doi.org/10.71204/n0hwaz15

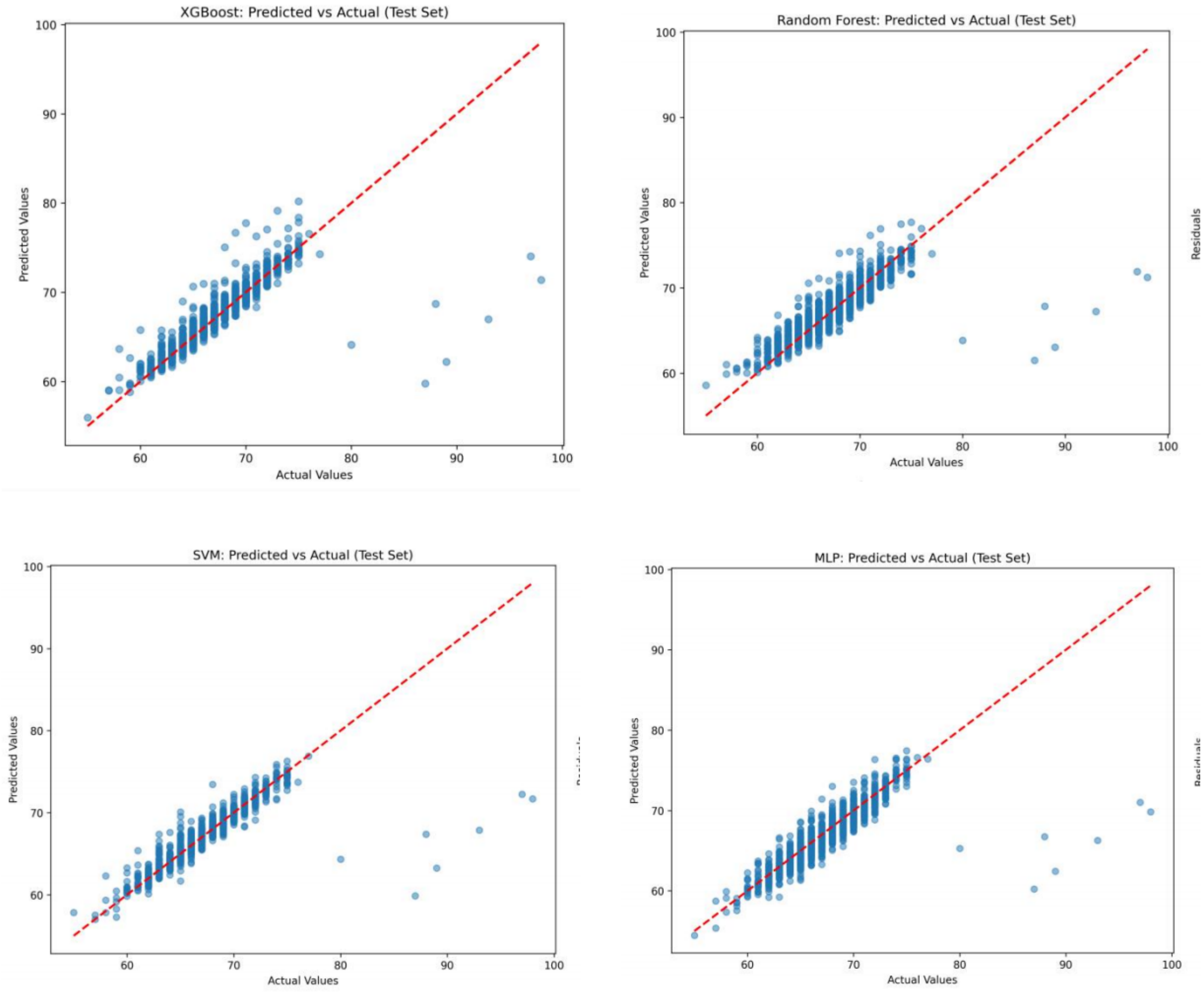

Figure 7. Comparison of the true value and predicted value of each model(Multi-Layer Perceptron Model; Random Forest Model; Support Vector Machine Model; XGBoost Model)

5.4. Experimental Results Analysis

Table 6 presents the comparative performance of four machine learning models—Multilayer Perceptron (MLP), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Support Vector Regression (SVR)—in predicting students' English achievement. The evaluation metrics include the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2).

Table 6. Performance comparison of different models

Model	RMSE	MAE	R^2
MLP	2.1663	1.0786	0.6680
RF	2.1594	1.0996	0.6701
XGBoost	2.0285	0.8212	0.7089
SVR	1.9481	0.7462	0.7315

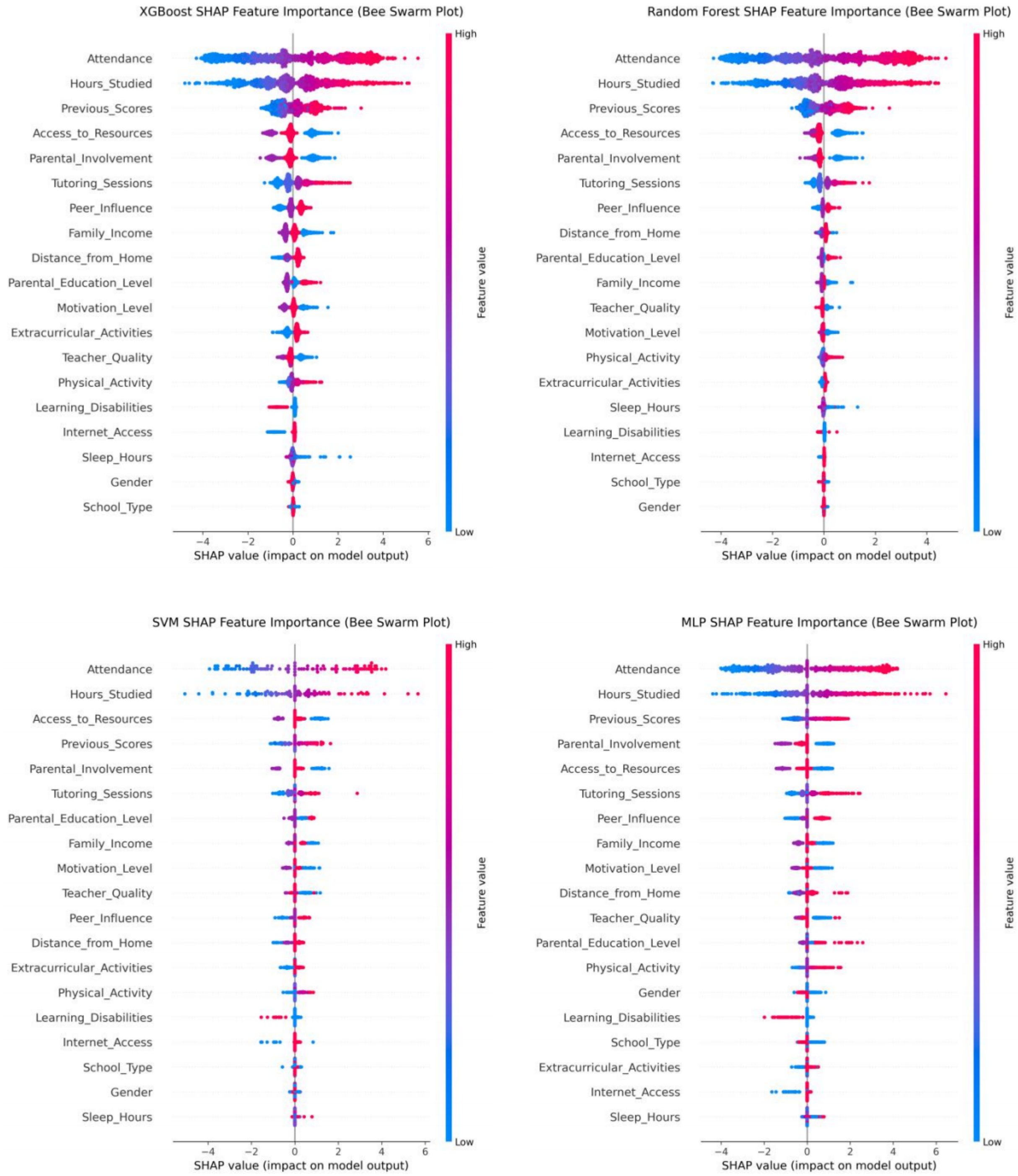

Figure 8. Shap feature maps of each model(XGBoost Model; Random Forest Model; Support Vector Machine Model; Multi-Layer Perceptron Model)

Among the models tested, the SVR model achieved the highest predictive accuracy, with the lowest RMSE (1.9481) and MAE (0.7462), and the highest R^2 (0.7315). This indicates that SVR provides a superior balance between bias and variance, effectively capturing nonlinear relationships among learning behavior, learning atmosphere, and English scores. The XGBoost model also demonstrated strong performance (R^2 = 0.7089), reflecting its robustness in handling structured educational data and its ability to automatically model complex feature interactions.

The Random Forest and MLP models exhibited comparable but slightly lower accuracy (R^2 0.67), suggesting that while both models possess acceptable generalization capability, their sensitivity to small-sample fluctuations may lead to minor instability in prediction. weaker adaptability.

Overall, the comparative analysis demonstrates that kernel-based regression (SVR) outperforms tree-based and neural architectures in the given dataset, underscoring the nonlinear and multidimensional nature of the factors influencing students' English scores.

Figures 7 and 8 visualize the prediction results and error distributions for the four models. As illustrated, both the SVR and XGBoost models exhibit prediction curves that closely follow the actual English scores, indicating high consistency between predicted and observed values. The prediction errors are minimal and symmetrically distributed around zero, confirming the models' stability and robustness.

In contrast, the MLP and Random Forest models show slightly larger deviations, particularly for students with extreme or atypical learning patterns. This may be attributed to overfitting in the neural model or the limited interpretive capacity of tree ensembles in handling high-dimensional psychological and behavioral indicators.

Figure 8 further confirms that the residuals of the SVR model are narrowly concentrated, suggesting its strong ability to generalize across diverse samples. The XGBoost model also maintains low residual dispersion, while MLP and RF display a broader spread of residuals, implying comparatively.

6. Discussion and Outlook

6.1. Discussion of the Results

This study focused on the impact of junior high school students' English learning behaviors and learning atmosphere on their academic performance, revealing a complex and nuanced relationship between multiple factors. From the perspective of learning behavior, learning motivation is significantly positively correlated with English scores, consistent with the motivation-engagement-performance model proposed by Pintrich and De Groot (1990).

Regarding the learning atmosphere, resource availability demonstrates strong predictive power, suggesting that the richness and ease of access to learning resources are crucial for English learning. However, the current distribution of resources shows a significant trend toward centralization, with core resource types covering the vast majority of usage scenarios. This lack of diversity may limit students' ability to broaden their learning horizons. Peer influence also significantly and positively predicts English scores, demonstrating that the learning culture and cooperative atmosphere within a group play an important role in socializing learning. This finding provides empirical support for building learning communities. The lack of a significant correlation between attendance and academic performance challenges the traditional educational assumption that simply attending school ensures learning success, demonstrating that physical presence does not equate to cognitive engagement. The true key to classroom effectiveness lies in teaching quality, student engagement, and the depth of teacher-student interaction. Among family environmental factors, parental education level is positively correlated with English achievement, indicating that highly educated parents tend to provide more professional language tutoring and learning strategy support. However, household income shows a weak negative correlation with parental involvement, possibly reflecting "over-intervention" or a "resource substitution effect" in high-income families.

6.2. Educational Significance and Practical Implications

The findings of this study have important practical implications for junior high school English teaching and family education. First, schools and teachers should prioritize cultivating students' intrinsic learning motivation by designing challenging and realistic learning tasks and employing teaching strategies such as project-based learning, contextualized instruction, and gamification to enhance students' learning experience and sense of self-efficacy. Second, educational administrators should optimize resource allocation and establish diverse and open learning resource platforms to reduce inequality in access. This includes introducing graded reading materials, English videos, and online interactive platforms to meet the needs of students at different levels and with different interests.

In classroom instruction, teachers should move beyond the traditional knowledge-inculcation model and focus on building interactive, inquiry-based, and generative classrooms. Scientifically designing problem scenarios to guide students in deep thinking and communication will not only enhance learning engagement but also effectively stimulate their interest and autonomy in learning. Furthermore, teachers should strengthen immediate feedback in the classroom and provide differentiated guidance and support tailored to students' diverse performance to meet individual learning needs. Furthermore, activities such as group learning, peer support, and presentations and exchanges can not only enhance students' motivation but also cultivate their critical thinking and collaborative skills, thereby creating a positive, supportive, and diverse classroom learning atmosphere.

At the family education level, parents' roles should not be limited to monitoring grades, but should instead shift to supporting learning and providing emotional support. Establishing a scientific approach to education means parents should pay closer attention to their children's thinking patterns, habit formation, and psychological state during the learning process, rather than simply focusing on grades. Providing children with a quiet learning environment, a reasonable schedule, and consistent emotional care can help them develop a positive learning attitude and a healthy psychological state. Especially in the context of the "double reduction" policy, parents should avoid blindly "piling up" learning resources by increasing tutoring and purchasing supplementary materials. Instead, they should focus on the quality of their time with their children, adopting a respectful, communicative, and guiding approach to help their children develop strong learning motivation and self-regulation skills.

At the policy support level, national and local education departments should further refine and implement the "double reduction" policy, ensuring its practical implementation in classrooms and families. Specifically, schools should be strengthened as the primary venue for education, encouraging teachers to engage in innovative methods and educational research in the classroom, exploring teaching models based on learning science, and improving classroom effectiveness and personalized student learning experiences. At the same time, education administrations need to strengthen supervision and guidance on the equitable allocation of educational resources, ensuring balanced development of educational opportunities and conditions across urban and rural areas, regions, and schools. Furthermore, long-term mechanisms for collaborative education between families, schools, and communities should be established to foster complementary relationships among families, schools, and communities, thereby creating an educational ecosystem conducive to students' well-rounded development.

In short, learning English in junior high school is a multi-factor, systematic project that requires the collaborative efforts of students, teachers, families, and schools to build a positive learning ecosystem, thereby achieving a true transition from "knowledge infusion" to "literacy development." This study provides empirical evidence and practical paths for this transition, offering a scientific basis for comprehensively improving the quality of English education in junior high schools.

6.3. Limitations and Future Directions

This study's sample size and regional representativeness are limited. Therefore, caution is warranted when generalizing the findings to regions with varying levels of economic development, educational resource allocation, and cultural backgrounds. Future research could expand the sample size and conduct cross-regional and multi-center comparative analyses to enhance the model's external validity and generalizability. Furthermore, variable measurement primarily relied on self-report questionnaires, which, despite their demonstrated reliability and validity, cannot fully eliminate the influence of social desirability bias and common method bias. Furthermore, English scores was measured solely through final standardized test scores, failing to fully capture the multidimensional aspects of language proficiency, such as oral expression, writing, and comprehensive application. Future research could incorporate multi-source data (such as teacher evaluations, classroom observations, student learning diaries, and behavioral data from digital learning platforms) for comprehensive measurement to enhance the comprehensiveness and objectivity of the data.

7. Conclusion

This study, using junior high school students as the research subjects, systematically explored the mechanisms by which learning behaviors and learning atmosphere influence English academic achievement through empirical investigation. The study found that factors such as learning motivation, sleep hours, resource availability, peer influence, and parental education level were significantly positively correlated with English achievement; however, factors such as attendance rate, family income, and teacher quality did not show significant positive predictive effects. These results indicate that intrinsic motivation, resource diversity, peer interaction, and family support are key factors in improving students' English achievement, while time investment, financial resources, or attendance alone do not directly contribute to academic achievement. Based on these findings, this study proposes the following conclusions: First, junior high school

English teaching should move beyond the traditional "time-performance" approach and shift to a synergistic "motivation-strategy-environment" model, prioritizing the stimulation and maintenance of student motivation. Second, educational resource allocation should prioritize diversity and equity, avoiding over-concentration of resources and ensuring that every student receives appropriate learning support. Third, families, schools, and society should work together to build a positive learning ecosystem, promoting students' holistic growth through a positive peer culture, effective family support, and high-quality teacher-student interaction. This study not only enriches the empirical evidence for learning motivation and educational ecology theory in junior high schools, but also provides a scientific reference for improving English teaching, family education guidance, and optimizing educational resources under the "double reduction" policy. Future research can build on this foundation by further expanding research design, broadening the sample size, deepening variable measurement, and exploring multi-level interaction mechanisms to promote the continuous improvement of basic education quality.

Author Contributions:

Manman Li drafted the manuscript, implemented the coding, and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding:

This research received no external funding.

Data Availability Statement:

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of Interest:

The authors declare no conflict of interest.

Ethical and Informed Consent for Data Used:

The data used in this paper are all publicly available and have no ethical violations.

References

Aisyah, S., Ramadani, A. F., Wulandari, A. E., & Astutik, C. (2025). Pemanfaatan teknologi digital sebagai media pembelajaran interaktif untuk siswa sekolah dasar. Jurnal Sadewa: Publikasi Ilmu Pendidikan, Pembelajaran dan Ilmu Sosial, 3(1), 388–401. https://doi.org/10.61132/sadewa.v3i1.1565

Akhuseyinoglu, K., & Brusilovsky, P. (2021). Data-driven modeling of learners' individual differences for predicting engagement and success in online learning. In Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (pp. 201–212). Association for Computing Machinery. https://doi.org/10.1145/3450613.3456834

Aras, I., & Arhas, S. H. (2022). Lecture methods during the new normal era. Jurnal Office, 8(1), 11–18. https://doi.org/10.26858/jo.v8i1.32079

Berliana, R. P., & Habiby, W. N. (2024). Teacher perceptions and implementation of a democratic classroom atmosphere in elementary schools. Jurnal Cakrawala Pendas, 10(2), 318–335. https://doi.org/10.31949/jcp.v10i2.8915

Chen, C. H., & Chang, C. L. (2024). Effectiveness of AI-assisted game-based learning on science learning outcomes, intrinsic motivation, cognitive load, and learning behavior. Education and Information Technologies, 29(14), 18621–18642. https://doi.org/10.1007/s10639-024-12553-x

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. The Modern Language Journal, 78(3), 273–284. https://doi.org/10.1111/j.1540-4781.1994.tb02042.x

Gardner, R. C. (2006). The socio-educational model of second language acquisition: A research paradigm. EUROSLA Yearbook, 6, 237–260. https://doi.org/10.1075/eurosla.6.14gar

Gardner, R. C., & Lambert, W. E. (1972). Attitudes and motivation in second-language learning. Newbury House Publishers.

Gaur, H., Khidhir, B., & Manchiryal, R. K. (2022). Solution of structural mechanic's problems by machine learning. International Journal of Hydromechatronics, 5(1), 22–43. https://doi.org/10.1504/IJHM.2022.122459

Ghani, A. S. A., Rahim, A. F. A., Yusoff, M. S. B., & Hadie, S. N. H. (2021). Effective learning behavior in problem-based learning: A scoping review. Medical Science Educator, 31(3), 1199–1211. https://doi.org/10.1007/s40670-021-01292-0

Holzer, J., Korlat, S., Haider, C., Mayerhofer, M., Pelikan, E., Schober, B., Spiel, C., & Schultes, M.-T. (2021). Adolescent well-being and learning in times of COVID-19—A multi-country study of basic psychological need satisfaction, learning behavior, and the mediating roles of positive emotion and intrinsic motivation. PLOS ONE, 16(5), e0251352. https://doi.org/10.1371/journal.pone.0251352

Hurskainen, H., & Tynjälä, P. (2023). The role of positive atmosphere on learning generic skills. Generic Skills in Higher Education, 82.

Keshtegar, B., & Nehdi, M. L. (2020). Machine learning model for dynamical response of nano-composite pipe conveying fluid under seismic loading. International Journal of Hydromechatronics, 3(1), 38–50. https://doi.org/10.1504/IJHM.2020.105499

Li, H., Majumdar, R., Chen, M. R. A., & Ogata, H. (2021). Goal-oriented active learning (GOAL) system to promote reading engagement, self-directed learning behavior, and motivation in extensive reading. Computers & Education, 171, 104239. https://doi.org/10.1016/j.compedu.2021.104239

Li, J., & Xue, E. (2023). Dynamic interaction between student learning behaviour and learning environment: Meta-analysis of student engagement and its influencing factors. Behavioral Sciences, 13(1), 59. https://doi.org/10.3390/bs13010059

Ma'ruf, M. W., & Syaifin, R. A. (2021). Strategi pengembangan profesi guru dalam mewujudkan suasana pembelajaran yang efektif. Al-Musannif, 3(1), 27–44. https://doi.org/10.56324/al-musannif.v3i1.54

Müller, C., & Mildenberger, T. (2021). Facilitating flexible learning by replacing classroom time with an online learning environment: A systematic review of blended learning in higher education. Educational Research Review, 34, 100394. https://doi.org/10.1016/j.edurev.2021.100394

Nasiriany, S., Liu, H., & Zhu, Y. (2022). Augmenting reinforcement learning with behavior primitives for diverse manipulation tasks. In 2022 IEEE International Conference on Robotics and Automation (ICRA) (pp. 7477–7484). IEEE. https://doi.org/10.1109/ICRA46639.2022.9812140

Shokri, M., & Tavakoli, K. (2019). A review on the artificial neural network approach to analysis and prediction of seismic damage in infrastructure. International Journal of Hydromechatronics, 2(4), 178–196. https://doi.org/10.1504/IJHM.2019.104386

Tarigan, G. R. T., & Putri, M. (2022). Meningkatkan semangat belajar dan kemampuan calistung anak usia dini di Panti Asuhan Mercy Clement Indonesia. ABDISOSHUM: Jurnal Pengabdian Masyarakat Bidang Sosial dan Humaniora, 1(1), 124–130. https://doi.org/10.55123/abdisoshum.v1i1.516

Tsani, I., Nisa', A. L., & Sufirmansyah, S. (2023). Evaluation of academic atmosphere formation in Islamic junior high school through digital learning system. Munaddhomah: Jurnal Manajemen Pendidikan Islam, 4(2), 352–367. https://doi.org/10.31538/munaddhomah.v4i2.436

Virtanen, A., Lauritsalo, K., Mäkinen, T., Hurskainen, H., & Tynjälä, P. (2022). The role of positive atmosphere on learning generic skills in higher education—Experiences of physical education students. Frontiers in Education, 7, 886139. https://doi.org/10.3389/feduc.2022.886139

Wang, L., Shen, L., Yi, J., Yang, X., Peng, Y., Ding, J., Tian, Y., & Yan, S. (2026). Prediction model of dynamic fracture toughness of nickel-based alloys: Combination of data-driven and multi-scale modelling. European Journal of Mechanics - A/Solids, 116, 105892. https://doi.org/10.1016/j.euromechsol.2025.105892

Wang, Y. (2023). Affective state analysis during online learning based on learning behavior data. Technology, Knowledge and Learning, 28(1), 1–16. https://doi.org/10.1007/s10758-022-09597-8

Wei, W., Huang, C., Xia, L., Xu, Y., Zhao, J., & Yin, D. (2022). Contrastive meta learning with behavior multiplicity for recommendation. In Proceedings of the 15th ACM International Conference on Web Search and Data Mining (pp. 1120–1128). Association for Computing Machinery. https://doi.org/10.1145/3488560.3498527

Williams, M., & Burden, R. L. (1997). Psychology for language teachers: A social constructivist approach. Cambridge University Press.

Xu, N. (2021). Research on blended teaching model applied in college English grammar teaching from the perspective of TDT approach. In Proceedings of the 2021 4th International Conference on Humanities Education and Social Sciences (ICHESS 2021) (pp. 1268–1274). Atlantis Press. https://doi.org/10.2991/assehr.k.211220.217

Xu, Z., Yuan, H., & Liu, Q. (2021). Student performance prediction based on blended learning. IEEE Transactions on Education, 64(1), 66–73. https://doi.org/10.1109/TE.2020.3008751

Yuan, F., Huang, X., Jiang, H., Jiang, Y., Zuo, Z., Wang, L., Wang, Y., Gu, S., & Peng, Y. (2025). An xLSTM–XGBoost ensemble model for forecasting non-stationary and highly volatile gasoline price. Computers, 14(7), 256. https://doi.org/10.3390/computers14070256

Yuan, F., Ye, C., Huang, A., Xiao, K., Qin, J., & Chen, X. (2025). Machine learning-based search strategy for water object retrieval in cultural tourism safety contexts. Journal of Global Trends in Social Science, 2(4), 31–37. https://doi.org/10.70731/r97znf68

Zheng, M., & Guo, Y. (2024). Exploring latent profiles in the relationship between learning atmosphere, negative emotions, and willingness to communicate among EFL learners. Innovation in Language Learning and Teaching, 1–20. https://doi.org/10.1080/17501229.2024.2402459