Predicting the Preparation Level of Future Mathematics Teachers Using Machine Learning Models and Wolfram Alpha Data

Rima Abdualiyeva [1], Karlygash Shetiyeva [2,*], Aigul Aldabergenova [2], Gulmaral Mashripkhanova [2], and Assem Yerkinova [3]

[1] Department of Informatics and Digitalization of Education, Zhetysu University named after I. Zhansugurov, Taldykorgan, Kazakhstan
[2] Department of Information Technologies and Artificial Intelligence, Zhetysu University named after I. Zhansugurov, Taldykorgan, Kazakhstan
[3] Alem.AI Foundation, Astana, Kazakhstan
Email: rimaergaliyevna09@gmail.com (R.A.); zhgu.project@gmail.com (K.S.); a.aldabergenova@zu.edu.kz (A.A.); zmashripkhanova@gmail.com (G.M.); assemyerkinova@gmail.com (A.Y.)
*Corresponding author
Manuscript received March 17, 2026; revised April 15, 2026; accepted May 15, 2026; published September 21, 2026

***Abstract*—The study examines the use of a machine learning framework that predicts the preparation level of prospective mathematics teachers through the integration of academic, behavioral, and analytic measures based on mathematical tasks and Wolfram Alpha-supported analysis. The study adopted a quantitative predictive research design using a purposive sample of 72 students in their last year of studying mathematics teaching within a single institution (195 task-level observations). The dependent variable, preparation level, was conceptualized at the individual level by utilizing a preparation rubric of the institution and categorized into three tertile groups: low, medium, and high. Given the hierarchical structure of the data, statistical analyses were performed using linear mixed-effects models with Student_ID as a random effect. In order to avoid leakage of information among repeated tasks from the same candidate, student-level features were extracted from the task-level observations in the predictive modeling phase. Various machine learning models were tested using stratified five-fold cross-validation with hyperparameter optimization. The findings indicated statistically significant variations between the preparation levels in terms of error rate, symbolic steps of solutions, test scores, completion times, and progress rates. The Random Forest model yielded the highest performance, with accuracy of 0.889 ± 0.064 and macro-F1 score of 0.888 ± 0.064. These results indicate that integrating features based on academic and behavioral characteristics, along with analytical features, yields an efficient approach to modeling teacher preparation level. However, it should be noted that further research and validation are required on larger datasets from different institutions.**

***Keywords*—prediction, teacher preparation, mathematics education, machine learning, Wolfram Alpha, symbolic analysis, preparation level**

I. Introduction

Teacher training in mathematics is vital in helping aspiring educators develop their skills in analysis, problem solving, and communication of concepts to students. Educational systems have increasingly focused on the use of data to inform instructional decision-making [1]. It has therefore become important to have objective and quantifiable measures available to assess the level of preparation of future educators. The most common method of evaluating a student's preparedness for the teaching profession has traditionally been through the student's course grade or final examination score. These measures do not adequately measure the development of mathematical skills and analytic competence required to teach mathematics because both of these evaluations reflect only a measure of outcomes instead of reflecting the reasoning processes and/or development dynamic that occurred in learning mathematics.

Recent advances in machine learning and educational data analytics have created new possibilities for modelling the performance and learning trajectories of students as a follow-up to an initial assessment. By merging different sources of data regarding students (e.g., grades, learning system logs, behaviour), predictive modelling can provide measurable insights into a student's preparedness and progression [2, 3]. However, most current modelling techniques in mathematics teacher education rely on surface-level indicators of academic performance and do not adequately consider all the components essential to the analytical processes involved in problem-solving, which forms the basis of a teacher's professional mathematical competence [4].

The traditional methods of assessing students include final grade scores or exam results. While this information is very valuable, it does not offer comprehensive feedback about the ways in which candidates deal with mathematical problems. In other words, there is no data about the way in which students tackle mathematical tasks, their problem-solving speed, their development in this respect, or symbolic manipulation used during solving. This aspect becomes especially relevant for the case of math teachers, because their professional readiness involves both answering questions correctly and providing solutions in a structured manner.

Mathematical competence is not limited to being able to produce correct answers [5]. It includes the ability to analyse the behaviour of mathematical objects, apply symbolic transformations, explain solution steps, and adapt problem-solving strategies when task conditions change. These abilities are especially important for future mathematics teachers because they must not only solve mathematical problems themselves but also explain mathematical reasoning to learners. Therefore, assessment of teacher preparation should include both outcome-based indicators and process-oriented measures of mathematical reasoning.

In this case, computational mathematics tools like Wolfram Alpha can serve as a means for expanding the dataset with analytical indicators of student performance. Not only do they allow performing symbolic calculation and visualization, but they also assist in defining the degree of function complexity, number of steps involved in solving a mathematical problem, and types of errors made when coming up with an answer. Process-based indicators like these rarely find their way to academic transcripts, yet might prove useful when it comes to determining how analytical reasoning is developing in prospective mathematics teachers. At the same time, the current investigation did not imply any attempts at using Wolfram Alpha as a substitute for student reasoning; instead, it served as an analytical instrument which made it possible to extract process-based indicators from students' solutions to mathematical problems. Thus, the contribution of this computational tool to our investigation is connected with turning mathematical task properties into analytical variables during the feature engineering process.

Though machine learning is increasingly being used in educational prediction research, there are still three areas of research concerning the education of mathematics teachers that need to be better addressed. First, there is a tendency among educational predictors to focus on academic or behavioral measures and not consider symbolic processes involved in mathematical operations. Second, computational mathematics software tends to be used primarily as an instructional tool rather than a source of data. Finally, methods for validating predictive models in small educational datasets, particularly datasets involving repeated observations of the same task nested within individual subjects, are lacking. This paper seeks to fill these gaps in research.

This study aims to investigate the ability of a multidimensional feature set in predicting the preparation status of mathematics teachers in the future, as well as the factors that distinguish different preparation statuses. This study makes three significant contributions. First, it uses analytical task features, along with academic and behavioral variables. Second, it employs machine learning models validated via student-level cross-validation to prevent data leakage due to repetition. Third, it combines predictive models with a statistical analysis based on linear mixed-effects modeling.

II. Literature Review

Rapid growth of machine learning approaches has occurred in the area of educational assessment, especially predictive performance and learning analytics [6]. Initially, empirical evidence published in studies indicated that classical supervised learning (e.g., Logistic Regression and Support Vector Machines), using a limited number of academic indicators (i.e., grades, test scores, attendance), yielded performance prediction accuracies between 65%–75% [7]. While demonstrating the potential of using data-driven approaches for assessment, this research also pointed out limitations in relying solely on outcome indicators for assessment purposes.

To rectify these issues, behavioural indicators were included in later studies to improve predictive performance as additional numerical features. Specifically, behavioural information captured from Learning Management Systems (LMS) (e.g., Moodle, Blackboard, Canvas) used as predictive features (e.g., time spent completing tasks, frequency of clicking through course content, submission of assignments) contributed to improved prediction accuracy (8%–15%) for prediction models using only grades [8]. Therefore, behavioural analytics has emerged as a key component of using educational data mining techniques today.

Mathematics education presents unique challenges to the predictive modelling task, due to the hierarchical nature of mathematical knowledge and the abstract concepts involved. Studies comparing linear classification algorithms with non-linear alternatives reveal that tree-based classifiers have significant advantages when predicting student achievement in mathematics, with tree-based methods producing better results (up to a 12% improvement in macro F1-score) than linear classifiers [9, 10]. This information has led to a rapid increase in the number of uses of ensemble learning methods such as gradient boosting and Extreme Gradient Boosting (XGBoost) on numerous datasets in the education field [11].

Additional research, in addition to increasing predictive ability, has focused on understanding the problem-solving process rather than concentrating solely on results. For example, process-oriented research on how learners progress through problems allows researchers to determine the degree to which learners can see how all the pieces come together, articulate their reasoning for problem-solving steps, and exhibit higher retention and transfer of information in the long term [12]. Recently in mathematics, it has been shown that accuracy of students' conceptual understanding is linked to the steps taken (e.g., symbolic reasoning sequences) to arrive at their answer and the complexity of the transformation process involved with arriving at a solution [13].

Behavioural variables linked to learners' dynamics are another example of how preparatory skill levels can be evaluated. For example, empirical studies indicate that a student's average time spent completing tasks, progress rate, performance stability over multiple testing experiences, and use of self-regulated learning skills correlate highly with each other correlation coefficients ranging from 0.42 to 0.67 [14]. Thus, applying these variables in teacher education programs is also important, as having consistent engagement with learning and demonstrating adaptive behaviours is key to developing professional readiness [5].

However, even with advancements in the data collection and analysis procedures for educational assessment, the majority of the current predictive accuracy and modelling is based on using features derived through manual processes from LMS log files or assessment systems, while the analytical process structure of mathematics problems has yet to be examined adequately. Computer Algebra Systems (CAS) and computational mathematics software would allow researchers to analyse and explore both the symbolic expression and transformation complexity associated with solution development. However, CAS and computational mathematics software have been primarily used in instructional support settings instead of using data-driven methodologies to assess students [15, 16]. Examples of prior studies using mathematical software (e.g., Mathematica and Maple) have mainly focused on providing support to teach

mathematics, rather than data-driven methodologies for evaluating students [17].

Wolfram Alpha is an open-ended computational knowledge engine that allows users to do step-by-step symbolic computations, visualize functions, automatically analyse complexity and has been used in many academic environments to support student's needs in verifying problems, analysing derivatives and integrals, providing breadth of functions available for analysis by students in undergraduate mathematics courses [18, 19]. Studies indicate that using Wolfram Alpha as an analytical tool increases the efficiency of problem-solving and reduces conceptual errors; however, much of its potential as a source of structured data that could serve as input for machine learning is still not well-explored [20].

Another major issue in the application of machine learning techniques in educational environments concerns the interpretability and trustworthiness of predictive models derived from complex educational data. Although ensemble-based machine learning algorithms often achieve high predictive accuracy, their decision-making processes are frequently difficult to interpret, which raises concerns about transparency and accountability when such systems are used to support educational decision-making. The so-called "black-box" nature of many advanced models can limit stakeholders' trust and hinder their adoption in high-stakes contexts such as student assessment and academic progression [21].

Recent research therefore emphasizes the importance of integrating Explainable Artificial Intelligence (XAI) techniques, interpretable feature representations, and transparent evaluation procedures. In particular, learning analytics studies recommend using interpretable model explanations together with robust evaluation metrics such as precision, recall, and macro-averaged F1-score in order to ensure fairness, reliability, and responsible use of predictive models in education [22].

In summary, the literature review shows that machine learning could be useful for educational prediction, but it also exhibits a number of methodological and theoretical shortcomings. Investigations focused exclusively on academic variables generate interpretable, but sometimes insufficient models of readiness. On the other hand, the application of behavioral analysis enhances predictions by taking into account learning processes; nevertheless, these characteristics offer only partial knowledge about the mathematical nature of problem-solving processes. The use of Wolfram Alpha and other similar computational platforms has proven its potential to contribute to educational instruction, although they have not yet been extensively applied as structured sources of analytical predictors. Hence, the innovative element of the current investigation is the combination of academic, behavioral, and symbolic analytical predictors within a single model of predicting readiness for teaching mathematics.

III. Materials and Methods

A. Data Sources and Research Design

This study adopted a predictive research design that is quantitative in nature. The data set comprised 195 observations on tasks conducted by 72 final year undergraduate students who were enrolled in a mathematics teacher education program. Purposive sampling of a single institution was employed as the participants were nearing the end of their teacher education and had completed the mathematics and pedagogical components of the course. While the sample size was small, it covered the entire population of available final year students in the chosen program. Therefore, the data set was deemed fit for conducting a predictive study through cross-validation at the student level.

The dataset employed in this empirical research study incorporated three different sets of indicators. These indicators were academic indicators, behavioral indicators, and analytical indicators. The academic indicators included the average mathematics grade, standardized test scores, and the error rate. The behavioral indicators included the time taken to complete the task and the progress rate. The analytical indicators included the complexity of functions and the number of symbolic solution steps. This information was based on mathematical tasks analysis using Wolfram Alpha.

The nesting was due to the fact that multiple observations pertained to the same candidate. Two methods were adopted to account for the nesting problem. One method involved the use of a random intercept of Student_ID in the linear mixed effects models. The second method involved the aggregation of the task level observations into student-level features before training the machine learning algorithms.

B. Visual-analytical Representation of Mathematical Tasks

In addition to symbolic calculations, Wolfram Alpha allows for the visual-analytical exploration of mathematical functions, allowing for the development of additional structural characteristics. Fig. 1 illustrates that Wolfram Alpha includes visualisation of two-dimensional and three-dimensional functions along with contour mapping and graphical analysis based on parameters.

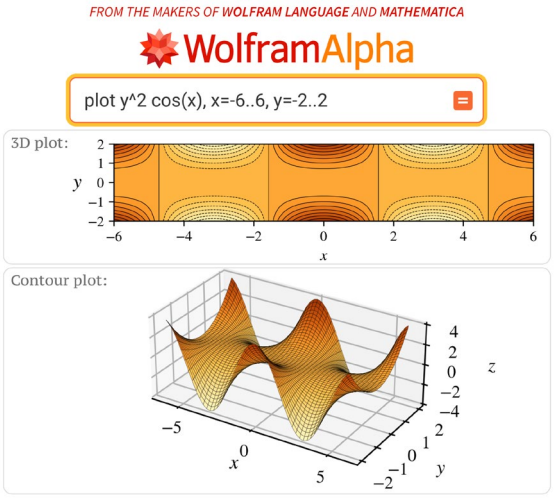


Fig. 1. Wolfram Alpha visualization of a mathematical function for analytical feature extraction, including a 2D contour plot and a 3D surface plot.

Visual and symbolic representations were employed in the characterization of mathematical problems based on their structural and analytical aspects. Within the context of this research work, the analysis conducted using Wolfram Alpha

played a crucial role in transforming the mathematical properties of solutions into quantifiable indicators of tasks, which included function complexity and symbolic solving procedures. These indicators served as features within the process of feature engineering rather than being considered as independent measures of teacher readiness.

C. Feature Set Description

The feature set was divided into three sets; academic, behavioral, and analytical. The academic set comprised the performance of candidates on the course and tasks. The behavioral set involved the effectiveness and progress made in the task completion process. The analytical set involved the structure of mathematical tasks analyzed using Wolfram Alpha and processes of symbolic reasoning. The complete feature set is presented in Table 1.

Table 1. Feature set used in the predictive framework

Feature name	Category	Description	Source
Avg_Math_Grade	Academic	Average grade in core mathematics courses	University records
Test_Score	Academic	Result of standardized mathematics assessment	Testing system
Error_Rate	Academic	Percentage of incorrect solutions in mathematical tasks	Learning analytics
Task_Completion_Time	Behavioural	Average time required to solve mathematical problems	LMS logs
Progress_Rate	Behavioural	Dynamics of academic performance over the study period	LMS analytics
Func_Complexity	Analytical	Complexity level of mathematical functions	Wolfram Alpha-supported task analysis
Symbolic_Steps	Analytical	Number of symbolic transformation steps in the solution	Wolfram Alpha-supported task analysis

Numeric data were converted into a consistent numeric format and inspected for missing values. In the statistical modelling, task-level data observations were used along with a random grouping variable Student_ID. On the other hand, in machine learning, the observations at the task level were converted into feature engineering data at the student level. The purpose of the conversion was to prevent overlapping attempts by the same candidate in training and validation samples.

D. Target Variable and Problem Formulation

The target variable in the study was the preparation level of the future math teachers. The preparation level was considered at the student level using the institutional preparation rubric. This rubric consisted of indicators that reflected mathematical preparation, analytical reasoning, performance on courses, and pedagogical preparedness. It is important to note that this classification was carried out independently from any machine learning models and did not involve optimization according to the predicted values.

In order to create an evenly balanced three-way classification task, the 72 students were grouped into tertiles depending on their institution's level of preparation. Thus, the lower tertile was labeled as low preparation, while the middle tertile was considered medium preparation, and the upper tertile was high preparation. It led to creating an equal number of student-level groups with 24 students in each group. Then, the student-level label was attached to all observations at the task-level for that particular student. Therefore, the distribution of classes at the student-level was balanced, but the task-level class distribution was not equal due to the varying number of task observations per student.

E. Machine Learning Models, Evaluation and Statistical Analysis

The problem of predicting was posed as a multi-class classification task with three classes of preparation levels-low, medium, and high. For the machine learning analysis, the task was run at the student level in order not to have data leakage from multiple runs of the same task. First, task-level data was rolled up to student-level engineered data, and then the modeling was done using this student-level dataset.

The machine learning algorithms considered for evaluation are listed below: Logistic regression with Mutual Information feature selection, Support Vector Machine with linear kernel and SelectKBest feature selection, *k*-Nearest Neighbours with SelectKBest feature selection, Random Forest, Extra Trees, Gradient Boosting, and XGBoost. Additionally, a stratified dummy classifier is used as a benchmark. The numerical features are standardized where needed, mainly in linear, distance, and support vector algorithms.

Performance of models was evaluated by stratified five-fold cross-validation in students. Tuning of hyperparameters was conducted inside folds via grid search. Metrics of choice include accuracy, precision, recall, and macro-average of F1 score. The latter is considered the most important in our case due to the classification task being multiclass. Mean and standard deviation values inside folds were used for evaluation. For model tuning, the hyperparameter grid included model-specific parameters. In case of tree-based models, number of estimators, tree depth, minimum samples per leaf, and sampling scheme for features were chosen. For the best-performing Random Forest, configurations include 200 estimators, square-root feature sampling, maximum depth of 3 or unrestricted depth, and minimum leaf sizes of 1 or 2.

For support vector and logistic regression approaches, parameter values for regularization and feature selection were optimized. For feature selection, the process was restricted to the training fold to avoid data leakage from the validation fold. In consequence, all of the final scores reported are derived from cross-validation results with optimization done internally to the model. There was no independent external dataset to test our model predictions. Model assessment thus relied on stratified student-level cross-validation, with external validation through larger, multi-institutional data sets as a focus for further work.

In order to determine whether there were significant differences between the preparation groups in terms of academic performance, behavioral outcomes, and analytics, linear mixed-effect models were fitted in JASP (Jeffreys' Amazing Statistics Program). The model specification included Preparation Level as a fixed factor along with

Student_ID as a random intercept because repeated task-level observations were nested within individual candidates. Terms in the models were evaluated using type III sums of squares and Satterthwaite degrees of freedom. The estimated marginal means were calculated for the pairwise comparisons using Holm correction.

F. Methodological Framework Overview

Methodology used for the development of the proposed approach can be found in Fig. 2. This figure provides an overview of the entire methodology from data acquisition and feature creation to statistical analysis and machine learning-based prediction. Furthermore, this figure illustrates how academic, behavioral, and Wolfram Alpha-enabled analysis measures were incorporated for preparation level estimation.

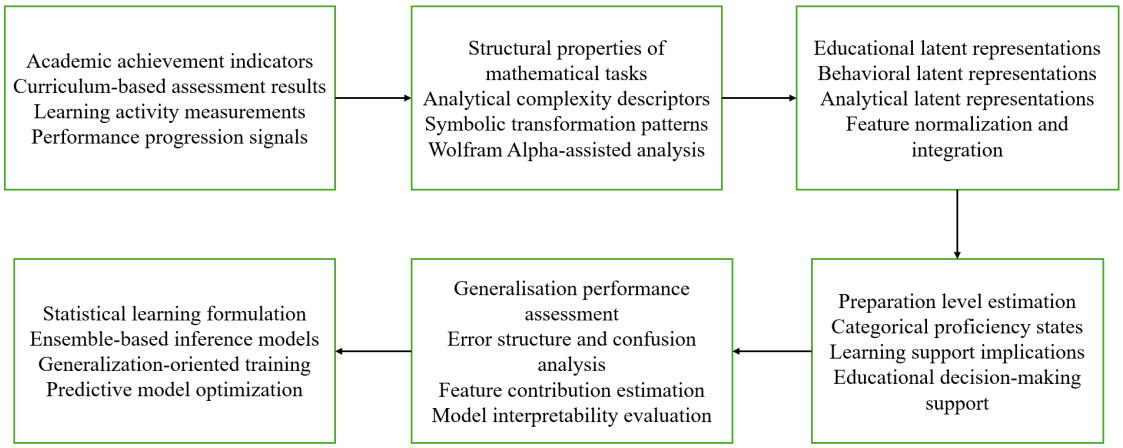


Fig. 2. Conceptual framework for estimating the preparation level of future mathematics teachers.

Fig. 2 illustrates the methodology utilized to quantify the preparation level of prospective mathematics teachers. This approach integrates indicators related to academics, behavior, and analysis into one feature space. Statistical methods were applied for analyzing differences in preparation levels, whereas machine learning algorithms were employed for prediction purposes, with the latter being used to predict the institutional preparation category.

G. Ethical Considerations and Bias Mitigation

Anonymous data was provided for analysis purposes. Student_ID was solely used as a grouping criterion for statistics and aggregation but did not have any personal data associated with it. This prediction system was supposed to provide recommendations for diagnostic monitoring and instructional decisions, not as an alternative to teacher decision-making or automated high-stakes decisions about students.

Algorithmic bias concerns were mitigated with the use of balanced preparation groups, with macro-F1 used as the primary metric and comparison with a dummy model. Yet, since the dataset is based on one institutional context, there is a chance that some biases could be present. As such, it is recommended to view the predictions as informative indicators that need to be checked by educators.

IV. Results and Discussion

The experiment was evaluated on the basis of 195 observations of task performance collected from 72 final year mathematics teaching students. On the level of student, the preparation levels were designed to be evenly distributed, with 24 students for each category. However, on the level of task, there is a slight difference in the distribution since different students contributed unequal number of task observations (65 observations for low preparation level, 62 observations for medium preparation level, and 68 observations for high preparation level).

In order to substantiate the differences in group means using statistical evidence, the linear mixed-effects model was estimated using the Student_ID variable as the random intercept. It can be seen that there are statistically significant differences among preparation levels on the variables of error rate, symbolic solution steps, test score, task completion time, and progress rate (Table 2).

Table 2. Linear mixed-effects model results for preparation-level differences

Indicator	df	F	*p*	Main direction
Error_Rate	2, 65.85	5.497	0.006	High lower than Low and Medium
Symbolic_Steps	2, 70.06	6.575	0.002	High higher than Low
Test_Score	2, 69.20	37.59	<0.001	High > Medium > Low
Task_Completion_Time	2, 70.92	24.19	<0.001	High < Medium < Low
Progress_Rate	2, 192	126.30	<0.001	High > Medium > Low

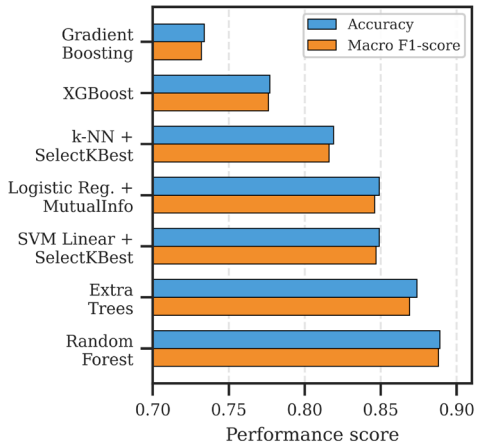


Fig. 3. Student-level cross-validated performance of machine learning models for preparation level prediction.

Fig. 3 illustrates the student-based analysis of the evaluated machine learning algorithms using the measures of accuracy and macro-F1. The comparative analysis is performed using the full set of academic, behavioral, and analytical features, which uses a five-fold cross-validation method. The figure highlights the performance of each algorithm in terms of predicting the preparation level.

From the visual analysis, it is clear that the Random Forest classifier provided the highest accuracy, followed by the Extra Trees classifier. The linear support vector machine classifier with feature selection as well as the Logistic Regression classifier with Mutual Information feature selection performed comparably to the other two classifiers. Both XGBoost and Gradient Boosting performed better than the baseline but could not beat the best-performing tree ensembles on student-level validation. These results confirm findings from past literature in educational analytics where ensemble-based approaches have been found to perform effectively on student performance data [23]. A detailed numerical summary of the cross-validated model performance is presented in Table 3.

For all models, the accuracy, precision, recall, and macro-F1 measures were averaged together with their respective standard deviations. The use of macro-F1 measure as the comparison metric was justified due to the multi-class nature of the task and equal importance of all preparation classes. All tested machine learning algorithms substantially outperformed the baseline dummy model.

The most effective algorithm used for prediction is the Random Forest algorithm. This algorithm had an accuracy score of 0.889 ± 0.064 and a macro-F1 score of 0.888 ± 0.064. These scores were substantially greater than those of the dummy algorithm, where the accuracy score was 0.235 ± 0.058 and the macro-F1 score was 0.222 ± 0.059.

Table 3. Cross-validated machine learning performance at the student level

Model	Accuracy	Precision	Recall	Macro-F1
Random Forest	0.889 ± 0.064	0.897 ± 0.069	0.893 ± 0.060	0.888 ± 0.064
Extra Trees	0.874 ± 0.060	0.899 ± 0.045	0.880 ± 0.056	0.869 ± 0.069
SVM Linear + SelectKBest	0.849 ± 0.148	0.868 ± 0.129	0.853 ± 0.145	0.847 ± 0.148
Logistic Regression + MutualInfo	0.849 ± 0.148	0.854 ± 0.146	0.853 ± 0.145	0.846 ± 0.150
k-NN + SelectKBest	0.819 ± 0.082	0.853 ± 0.086	0.823 ± 0.070	0.816 ± 0.084
XGBoost	0.777 ± 0.113	0.797 ± 0.114	0.787 ± 0.110	0.776 ± 0.114
Gradient Boosting	0.734 ± 0.136	0.767 ± 0.128	0.740 ± 0.138	0.732 ± 0.134
Dummy Baseline	0.235 ± 0.058	0.219 ± 0.061	0.237 ± 0.061	0.222 ± 0.059

To supplement the findings of predictive analysis, Fig. 4 offers an illustrative representation of the key analytical and educational indicators used in the study. Panel (a) is an illustration of the relationship between function complexity and preparation level. Panel (b) demonstrates the number of symbolic steps made for solving the equation and a clear tendency for more sophisticated solutions for the equations at advanced preparation levels. Panel (c) shows the connection between function complexity and error rate. The latter panel should be analyzed as a mere descriptive one because there was no statistically significant interaction between the two variables in question, according to a mixed-effects interaction model. Panel (d) depicts students' score distributions that rose in proportion with their preparation level.

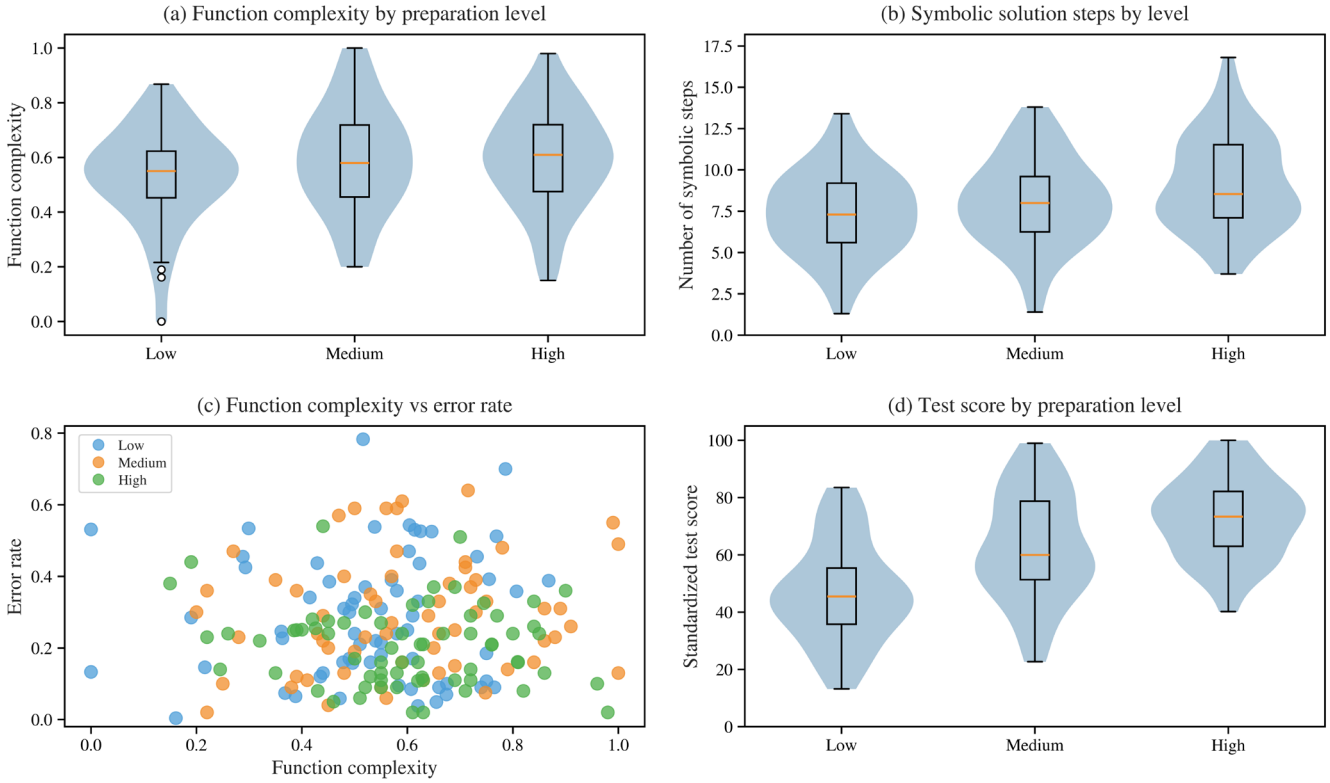


Fig. 4. Multi-panel visualization of function complexity, symbolic solution steps, error rate, and standardized test score across low, medium, and high preparation levels.

The visualization presented in Fig. 4 corroborates the findings from the mixed-effects analysis. Candidates who were well prepared tended to exhibit more symbolic problem-solving procedures and higher test scores, while those less prepared exhibited higher error frequencies. The data suggests that preparation status can be assessed on multiple dimensions, including mathematical precision, symbolic thought, academic success, behavioral efficiency, and improvement over time. Panel (c) must therefore be regarded as an exploratory charting exercise and not as one indicating the existence of any significant interaction effect. No significant interaction effect exists between the factors of function complexity and preparation level in the mixed-effects interaction analysis. This shows that task complexity is not adequate enough to explain the occurrence of errors. Preparation level must therefore be regarded in terms of its contributions from different perspectives.

This result is compatible with multi-dimensional approaches to preparing teachers. Mathematics teachers-to-be need not only correct results in mathematics but also well-structured reasoning about symbols, consistency in task implementation, effective problem-solving skills, and learning progression. This approach expands the traditional form of assessment by incorporating outcome-oriented criteria with process-oriented measures gained from mathematical tasks' analysis. In this context, Wolfram Alpha-assisted analysis was crucial since it helped to translate qualitative characteristics of mathematical tasks into quantifiable features.

Other methodological techniques might also be applied to tackle the same issue. Since the preparation level has a certain hierarchy, ordinal classification models would be suitable for analysis. In order to explicitly account for uncertainty in education data sets, Bayesian hierarchical models can be employed. Learning analytics models can take into account the development of the preparation process instead of treating the student's readiness as a static one. Additionally, in subsequent research, explainable AI methods such as SHAP values can be applied in order to understand the impact of different parameters on predictions.

V. Conclusion

This study proposed and evaluated a multidimensional framework for predicting the preparation level of future mathematics teachers by integrating academic, behavioural, and Wolfram Alpha-supported analytical indicators. The findings show that preparation level can be modelled more effectively when traditional academic measures are complemented by process-oriented indicators such as symbolic solution steps, error rate, task completion time, and progress rate.

The statistical analysis demonstrated significant differences across preparation levels for error rate, symbolic solution steps, test score, task completion time, and progress rate. The machine learning analysis further showed that the combined feature set supported accurate student-level classification, with Random Forest achieving the strongest cross-validated performance. These findings support the broader view that teacher preparation should be understood as a multidimensional construct that includes mathematical accuracy, symbolic reasoning, learning efficiency, and developmental progress.

Beyond the specific context of mathematics teacher education, the results contribute to broader discussions on data-informed assessment and learning analytics. The study shows that educational prediction should not rely only on final grades or examination scores, because such indicators may overlook important aspects of reasoning, task engagement, and learning development. By combining outcome-based and process-oriented evidence, the proposed framework aligns with a wider shift from purely summative assessment toward diagnostic and formative approaches that can support early identification of learning needs.

For teacher educators, the proposed framework may serve as a diagnostic support tool for identifying candidates who may require additional instructional guidance. Educators who wish to apply this approach should begin with low-stakes diagnostic use, ensure that input data are reliable and anonymized, and combine model outputs with professional judgement and qualitative evidence. However, the model should not be used as an automated high-stakes assessment instrument. Its outputs should be interpreted together with teacher judgement, institutional assessment practices, and qualitative evidence of pedagogical readiness.

Some limitations arise from this study. Firstly, it had only one source of its sample that consisted of 72 candidates, which limits the external validity of the results. Despite the fact that student-level cross-validation and mixed-effects modeling were used in order to minimize the methodological bias in this case, future studies should conduct external validations on larger and more diverse datasets. It is important for future studies to work with longitudinal data, ordinal models, and explainable artificial intelligence methods as well.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

R.A. developed the initial conceptualization of the study, designed the predictive machine learning framework, conducted the primary research, and wrote the original draft of the paper; K.S. secured the research funding from the Ministry of Science and Higher Education, managed the overall project administration, supervised the team, and handled corresponding author responsibilities; A.A. was responsible for software implementation, performed the data preprocessing, executed the machine learning models, and created the visual-analytical tools; G.M. provided essential resources, conducted the Wolfram Alpha-supported task analysis, and curated the academic dataset; A.Y. formulated the statistical methodology, performed the data analysis using linear mixed-effects models, validated the predictive outcomes, and co-wrote the results and discussion sections; all authors had approved the final version.

Funding

This research has been funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP25796283

"Development of a theoretical basis for the professional training of mathematics teachers using WolframAlpha neural network platform").

REFERENCES

[1] I. Issah, O. Appiah, P. Appiahene, and F. Inusah, "A systematic review of the literature on machine learning application of determining the attributes influencing academic performance," *Decis. Anal. J.*, vol. 7, 100204, 2023. doi: 10.1016/j.dajour.2023.100204

[2] N. Y. Ling, T. Tin, T. C. Keat, U. F. Khattak, and M. A. Almaiah, "Educational big data analytics: Machine learning based academic performance predictive modelling," *Pak. J. Life Soc. Sci.*, vol. 22, no. 2, pp. 7442–7477, 2024. doi: 10.57239/PJLSS-2024-22.2.00562

[3] J. Pan, Z. Zhao, and D. Han, "Academic performance prediction using machine learning approaches: A survey," *IEEE Trans. Learn. Technol.*, vol. 18, pp. 351–368, 2025. doi: 10.1109/TLT.2025.3554174

[4] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Appl. Sci.*, vol. 11, no. 1, 237, 2020. doi: 10.3390/app11010237

[5] R. Baez, H. Sanchez, and D. Pllana, "Computational thinking in high school mathematics," *Int. J. Educ. Math. Sci. Technol.*, vol. 14, no. 1, pp. 256–280, 2026. doi: 10.46328/ijemst.5250

[6] K. T. Chong, N. Ibrahim, S. H. Huspi, W. M. N. W. Kadir, and M. A. Isa, "A systematic review of machine learning techniques for predicting student engagement in higher education online learning," *J. Inf. Technol. Educ. Res.*, vol. 24, 5, 2025. doi: 10.28945/5456

[7] A. Haruna, I. D. Alhassan, L. Yakubu, M. U. Ahmad, and S. I. Gurjiya, "Exploring the potential of artificial intelligence tools in teaching and learning: Mathematics education considerations," *Fed. Univ. Gusau Fac. Educ. J.*, vol. 5, no. 6, pp. 16–21, 2026. doi: 10.64348/zije.2025215

[8] W. B. Valverde, D. G. A. Condo, and K. T. Gallegos, "The educational context 2.0 of artificial intelligence as a systemic activity: Case of mathematical analysis," *Educ. Soc.*, vol. 23, no. 3, pp. 169–181, 2025. doi: 10.5281/zenodo.18035553

[9] G. M. Agung, R. A. Zuama, and E. S. Budi, "Analysis of student academic performance using random forest and support vector machines," *Comput. Sci. (CO-SCIENCE)*, vol. 6, no. 1, pp. 57–65, 2026. doi: 10.31294/co-science.v6i1.10123

[10] M. Radek, A. Pietraszek, A. Kozień, K. Radek, and J. Pietraszek, "Matching computational tools to user competence levels in education of engineering data processing," *Mater. Res. Proc.*, vol. 34, pp. 453–459, 2023. doi: 10.21741/9781644902691-52

[11] R. Yavich, "Improving learning outcomes in advanced mathematics for underprepared university students through AI-driven educational tools," *Afr. Educ. Res. J.*, vol. 13, no. 2, pp. 224–239, 2025. doi: 10.5281/zenodo.15519390

[12] M. Nirmala and Y. D. Ravichandran, "Artificial intelligence in school-level mathematics education: A comprehensive review," *Int. J. Res. Sci. Innov.*, vol. 12, no. 7, pp. 1907–1913, 2025. doi: 10.51244/IJRSI.2025.120700187

[13] J. C. Valverde, "Wolfram|Alpha & ChatGPT in the design and analysis of mathematical models," in *Proc. 19th International Technology, Education and Development Conference*, Valencia, Spain, 2025, pp. 7003–7010. doi: 10.21125/inted.2025.1806

[14] M. Milićević, B. Marinović, and L. Jeftić, "Machine learning methods as auxiliary tool for effective mathematics teaching," *Comput. Appl. Eng. Educ.*, vol. 32, no. 6, e22787, 2024. doi: 10.1002/cae.22787

[15] S. Abramovich, "Wolfram Alpha and Maple as educational technology for future teachers of mathematics," *Int. J. Educ. Technol. Learn.*, vol. 17, no. 2, pp. 28–42, 2024. doi: 10.55217/101.v17i2.845

[16] R. C. Torres-Peña, D. Peña-González, E. Chacuto-López, E. A. Ariza, and D. Vergara, "Updating calculus teaching with AI: A classroom experience," *Educ. Sci.*, vol. 14, no. 9, 1019, 2024. doi: 10.3390/educsci14091019

[17] G. Choukade, "Artificial Intelligence (AI): A modern concept for mathematics education," *Educ. Quest—Int. J. Educ. Appl. Soc. Sci.*, vol. 16, no. 1, pp. 53–58, 2025. doi: 10.30954/2230-7311.1.2025.7

[18] M. G. Campuzano and T. Crisanto, "Learning analytic geometry with the aid of Wolfram Alpha," *Int. J. Innov. Sci. Res. Technol.*, vol. 7, no. 1, pp. 722–727, 2022. doi: 10.5281/zenodo.6045067

[19] R. Perezzan, S. Montalvo-Quirós, R. Rama-Ballesteros, and D. Herráez-Aguilar, "A proposal to integrate artificial intelligence tools in chemistry teaching in higher education," in *Proc. 18th International Technology, Education and Development Conference*, Valencia, Spain, 2024, pp. 3435–3441. doi: 10.21125/inted.2024.0910

[20] L. Chaudhary and S. Sharma, "Artificial intelligence for enhanced rainfall predictions: Leveraging sequential models and advanced callbacks for climate data analysis," *ESS Open Archive (Preprints)*, 2023. doi: 10.22541/essoar.170317284.40444299/v1

[21] H. Khosravi, S. B. Shum, G. Chen *et al*., "Explainable artificial intelligence in education," *Comput. Educ.: Artif. Intell.*, vol. 3, 100074, 2022. doi: 10.1016/j.caeai.2022.100074

[22] M. Afzaal, J. Nouri, A. Zia, P. Papapetrou, U. Fors, Y. Wu, X. Li, and R. Weegar, "Explainable AI for data-driven feedback and intelligent action recommendations to support students self-regulation," *Front. Artif. Intell.*, vol. 4, 723447, 2021. doi: 10.3389/frai.2021.723447

[23] S. Khan, T. Mazhar, T. Shahzad, M. A. Khan, W. Waheed, A. Waheed, and H. Hamam, "Predictive analytics in education: Enhancing student achievement through machine learning," *Soc. Sci. Humanit. Open*, vol. 12, 101824, 2025. doi: 10.1016/j.ssaho.2025.101824

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).