scientific reports

OPEN

Supervised machine learning models for predicting student mathematics performance in Somaliland primary examinations 2023

Mukhtar Abdi Hassan[1], Abdisalam Hassan Muse[1], Saralees Nadarajah[2✉] & Yahye Hassan Muse[1]

The declining mathematics performance among primary school students in Somaliland, as evidenced by the increasing failure rate from 51.9% in 2020 to 65.58% in 2023, has prompted the need to investigate the influencing factors and potential predictive models. This study leverages data from the 2022/2023 Somaliland National Examinations to identify and analyze these factors. This study aimed to compare the effectiveness of various supervised machine learning models in predicting mathematics performance among primary school students and identify the factors contributing to performance disparities. Data were drawn from the 2022/2023 Somaliland National Examination database, which covered 20,950 students. Six supervised machine learning models— logistic regression, decision tree, random forest, Naïve Bayes, support vector machine (SVM), and K-Nearest Neighbors (KNN)—were applied to predict student performance. Performance metrics, such as accuracy, sensitivity, specificity, F1-score, and AUC, were used to evaluate the models. Significant regional and demographic differences were observed between the groups. Regions such as Awdal and Maroodi Jeeh showed high failure rates, whereas the Sheekh and Sanaag regions demonstrated higher success rates than the others. Males (67.17%) failed more frequently than females (63.52%), and urban schools (67.64%) performed worse than rural schools (45.21%). The Naïve Bayes model achieved the highest accuracy of 98.6%, followed by the KNN model at 80.3%. Other models, such as Random Forest and Logistic Regression, demonstrated moderate success, whereas SVM performed the least effectively. The findings indicate that regional, sex, and school-type disparities significantly influence mathematics performance. The Naive Bayes model was the most effective in predicting performance, and these insights can be used for targeted interventions to improve the educational outcomes.

Keywords Mathematical performance, Machine learning, Logistic regression, Prediction

Mathematics is one of the most important subjects taught in schools today because it plays an essential role in shaping how individuals deal with various aspects of their private, social, and civic lives[1]. Mathematics is a universal language of science that helps people communicate and describe different situations in their daily lives[2]. According to[3], mathematics is essential for admission to engineering, technology, social science, and music education, as well as for understanding other subjects, such as science and social studies. It plays a crucial role in economic growth and development and serves all areas of science and technology. The underachievement of primary school students in mathematics is a critical issue that has attracted the attention of educators and governments worldwide.

Recent assessments of mathematical ability[4–6] It has been shown that East Asian countries, including China, Japan, South Korea, and Singapore, consistently lead the world in terms of mathematical performance. Studies have consistently found an East Asian advantage in mathematical ability across all age groups, starting in preschool[7]. The issue of underachievement in mathematics among learners is a global concern that has prompted

[1]Faculty of Science and Humanities, School of Postgraduate Studies and Research (SPGSR) , Amoud University, Borama 25263, Somalia. [2]Department of Mathematics, University of Manchester, Manchester M13 9PL, UK. ✉email: mbbsssn2@manchester.ac.uk

developing countries to participate in initiatives aimed at bringing about positive changes in their communities[8], including South Africa.

Assessments such as the Southern and Eastern Consortium for Monitoring Education Quality (SACMEQ) and the Trends in International Mathematics and Science Study (TIMSS) have confirmed the state of mathematics performance in South Africa. As reported by TIMSS, between 68% and 90% of African boys and girls in grade eight failed to meet the low international benchmark in mathematics in 2012[5]. Furthermore, no significant progress was observed in TIMSS 2007 or TIMSS 2011. Although the mathematics performance of Grade 9 South African learners improved in the 2015 TIMSS compared to the 2011 results, South Africa remains a low-performing country in mathematics[9].

In Somaliland, the poor performance of primary school students in mathematics is a cause for concern and requires further investigation. Academic workload, assessment techniques, teacher preparation, and curriculum design contribute to primary school students' low performance in mathematics.

Primary education is crucial to a nation's development, providing the highest average public return on investment and serving as the foundation for subsequent educational attainment and economic progress[10]. Primary education spans eight years, from Grade 1 to Grade 8, and provides essential, comprehensive education to prepare students for secondary education and training. While the official age range for primary education is 6–13 years, provisions are made for both younger and older pupils. Complementary to primary education, Alternative Basic Education (ABE) is considered part of the primary education subsector, emphasizing its role in widening access to basic education[10].

According to[10], the educational landscape is characterized by a network of 1,385 functional primary schools, with a significant majority (73.4%) being government-owned institutions. Additionally, 26.6% of primary schools are privately owned, reflecting a diverse educational landscape. This distribution underscores the collaborative efforts of public and private entities in providing accessible and functional primary education across the region.

According to the data in Table 1, there was a clear decline in the mathematics performance of primary school pupils over the three years from 2020 to 2023. The average score fell from 51.9 in 2020 to 42.3 in 2023, with 65.58% of the students failing mathematics, indicating a significant decline in primary school performance in this subject. This trend raises concerns about the likely causes of this decline and highlights the need for further research and targeted interventions to address the observed decline in mathematical skills.

Research on primary school mathematics performance in Somaliland and East Africa is limited, particularly in studies focusing on primary school pupils. The existing literature focuses mainly on secondary education, leaving a knowledge gap in understanding the factors that influence performance. This is crucial because primary education is the foundation of a student's academic journey. Targeted research initiatives can provide insights into these dynamics and support the development of informed policies and interventions to improve educational outcomes in Somaliland.

Although student performance prediction has been widely studied in various educational contexts, research on primary-level mathematics achievement in Somaliland remains extremely limited. Existing studies primarily focus on secondary education or broader academic outcomes, leaving a substantial knowledge gap in understanding the early predictors of mathematics success among primary school learners. This study offers a novel contribution by applying a comprehensive suite of supervised machine learning models—including Naïve Bayes, Random Forest, Logistic Regression, KNN, SVM, and Decision Tree—to a large, nationally representative dataset of over 20,000 primary school students. By integrating demographic, linguistic, and academic variables, this work provides the first machine–learning–based examination of primary-level mathematics performance in Somaliland. The study uniquely highlights the strong predictive power of language proficiency and regional educational disparities, offering evidence-driven insights to inform targeted interventions, resource allocation, and policy reforms within the Somaliland education sector. This represents a significant advancement in data-driven educational research in low-resource contexts, where predictive analytics remains underdeveloped.

Despite its strengths, this study has several limitations that should be considered when interpreting the findings. First, the analysis relies solely on examination records and does not include socioeconomic, pedagogical, or psychological variables that could offer a more holistic understanding of students' learning environments.

Exam Results			
Secondary		**Primary**	
Subject	Mean score	Subject	Mean score
Islamic	62.70	Islamic	62.10
Somali	61.10	Somali	71.40
Arabic	75.10	Arabic	65.00
English	63.10	English	52.70
Maths	**54.30**	**Maths**	**42.30**
Biology	81.50	Science	58.00
Chemistry	63.00	Social	57.30
Physics	67.00		
Geography	63.30		
History	84.50		

Table 1. Somaliland National examination result 2023.

Factors such as household income, parental education, teacher qualifications, classroom practices, and access to learning materials were not captured in the dataset but may significantly influence mathematics performance. Second, the cross-sectional design limits the ability to draw causal inferences, as the data reflect only a single examination year rather than longitudinal learning trajectories. Third, although six supervised machine learning models were tested, more advanced techniques—such as gradient boosting, deep neural networks, or hybrid ensemble models—were not explored and may offer additional predictive gains. The imbalance in the pass/fail outcome, while addressed through stratified sampling and imbalance-aware metrics, may still have influenced some models' performance, particularly SVM and Decision Tree.

Review of the related literature

Mathematics plays a crucial role in a wide range of academic disciplines studied by primary school students. This provides a strong foundation for future research. This study aimed to assess factors influencing mathematics performance by examining demographic characteristics, including gender, school type (urban and rural), school location (districts and regions), and school ownership (public and private).

[11] used TIMSS exam data to examine gender differences in mathematics, particularly liking, confidence, and value. The results show similar performance for boys and girls, but cross-country sociocultural, political, and educational equality does not always predict positive effects on gender equality. Some countries have shown larger gender differences in mathematics than others.

[12] examined the influence of gender on mathematics performance among primary school students in the Keiyo South District in Kenya. This included 300 students and mathematics teachers. The results showed that while most students believed that learning mathematics was worthwhile, male students had more positive attitudes. This study suggests that gender differences in mathematics achievement may contribute to these discrepancies.

According to[13], gender differences in mathematical achievement and the concept of mathematical greatness may have been eliminated because girls are expected to have similar abilities. The study found no gender differences in academic achievement but acknowledged the limitations of extending the findings to all school districts in the state.

Differences in economic and social development between rural and urban areas in Ghana have resulted in an unequal distribution of educational resources and varying levels of academic achievement among students in different regions[14].

Educational achievement is influenced by individual, cognitive, social, and environmental factors. However, rural students often perform worse than their urban counterparts across the country, regardless of educational backwardness, and vice versa[15].

Private school teachers have higher levels of commitment to teaching and lower absenteeism, despite lower levels of formal education and income. They also had better test scores and lower pupil cost. Private schools offer better quality teaching in mathematics and literacy and are more likely to be younger and from the same area as public schools. However, public schools still excel in some areas, such as learning outcomes and infrastructure. In conclusion, private schools have strengths in teacher engagement and cost-effectiveness, which may lead to better educational outcomes if spending is in line with that of public schools[16–21].

Machine learning (ML) has emerged as a transformative tool in education, particularly for predicting student performance and enhancing personalized learning. Studies have shown that ML applications not only characterize educational behavior but also identify key factors influencing academic success, providing mechanisms for forecasting outcomes based on complex data interactions[22,23]. The shift toward digitalized education underscores how supervised algorithms, such as Support Vector Machines, Random Forests, and Decision Trees, contribute to both performance prediction and adaptive learning experiences[24]. Empirical evidence highlights their effectiveness, with models such as Decision Trees achieving high accuracy rates in predicting academic achievement[25] and reviews confirming the growing integration of ML into curriculum design and instructional strategies[24,26].

In the specific context of mathematical performance, ML models demonstrate strong potential for analyzing diverse factors that influence outcomes, including student demographics, perceptions, and instructional practices. Research shows that algorithms such as gradient boosting, neural networks, and k-nearest neighbors support predictive insights into mathematics achievement while also enabling educators to adapt their pedagogical approaches[27,28]. Comparative analyses of supervised learning models further reveal that algorithmic effectiveness varies by context and dataset characteristics, requiring rigorous validation and careful model selection[29,30].

Language proficiency plays a crucial role in students' mathematics performance, serving as a foundational element that shapes the cognitive processes underlying mathematical reasoning and problem-solving. Research by[31]. Indicates that language skills significantly correlate with mathematics achievement, highlighting the dependency on language for understanding mathematical concepts, particularly those articulated through word problems[32]. support this perspective by demonstrating that early language skills are instrumental in facilitating informal math skills, which positively impact formal math competencies. Moreover[33], Emphasizes the necessity of comprehension skills concerning mathematical instructions, indicating that without adequate language understanding, children may struggle to grasp essential mathematical concepts. Thus, proficient language command not only enhances verbal mathematical communication but also supports cognitive tasks, such as interpreting and solving mathematical problems, underscoring the intricate connection between language and mathematical performance.

Methodology
Source of the data
The data for this study were obtained from the Somaliland National Examination Database, which is managed and certified by the Somaliland National Examination and Certification Board (SLNECB) under the supervision of the Ministry of Education and Science, Somaliland. The dataset includes information on various variables, such as place of residence (districts and regions), school location (urban or rural), students' gender, school type (public or private), and school accommodation (boarding or day). The dataset included 20,950 students, including 9,111 girls and 11,839 boys, who participated in the 2023 national primary school exams.

Data preprocessing and handling
To ensure full reproducibility, the dataset underwent a rigorous preprocessing workflow before model development. Examination records were complete, as the Somaliland National Examination Database includes only students who sat for the examination; therefore, no students absent from the exam appeared in the dataset. Categorical variables such as region, district, residence, and school type were encoded using one-hot encoding, while binary variables, including gender and school ownership, were retained as dichotomous. For continuous variables related to subject performance (Somali, Arabic, English, and Science), each was converted into a binary "pass/fail" indicator based on national grading criteria: scores ≥ 50 were coded as "pass," and scores < 50 as "fail." There were no other continuous predictors in the dataset. The dependent variable (mathematics pass/fail) presented a natural imbalance of 64/36; therefore, no resampling techniques were used. Instead, we applied stratified train–test splitting to preserve the class distribution and evaluated models using imbalance-sensitive metrics such as F1-score, balanced accuracy, and AUC.

Hyperparameter optimization was conducted via a grid search embedded within 10-fold stratified cross-validation to enhance robustness and reduce overfitting. For Decision Trees, parameters such as maximum depth, minimum samples per leaf, and splitting criteria were tuned. At the same time, Random Forest models were optimized using the number of estimators, the maximum number of features, and the maximum depth. SVM models were evaluated using both linear and RBF kernels, with penalty parameter (C) and kernel coefficient (γ) tuned. KNN optimization included testing a range of k values (3–15) and alternative distance-weighting schemes. Naïve Bayes and Logistic Regression required minimal parameter adjustments and were implemented using standard configurations. All preprocessing, variable transformations, and hyperparameter-tuning procedures were executed with fixed random seeds to guarantee methodological transparency and full reproducibility.

Study area
This study focuses on assessing primary school students' mathematics achievement in Somaliland using datasets corresponding to the 2022/2023 academic year. It is concerned with students' mathematics performance, particularly in the context of the Somaliland National Primary School Examination (SNPE). To ensure methodological rigor and the acquisition of a representative dataset for subsequent analysis and evaluation, the scope of the data was extended to include all primary education institutions in the 14 regions and 23 constituencies of Somaliland, as shown in Fig. 1. This comprehensive approach ensured a sophisticated examination of the educational panorama across diverse administrative and geographical strata in Somaliland. The methodology builds upon previous studies, such as[34], which utilized similar approaches to predict student dropout rates in Somaliland, and[35], which examined the determinants of student academic performance using multi-level logistic regression to estimate unobserved effects at both the student and school levels. Furthermore[36], machine learning-based analysis of academic performance determinants, using insights from the 2021/2022 National Secondary School Exams, underscores the value of integrating advanced analytical techniques to uncover critical educational patterns, which this study also endeavors to achieve.
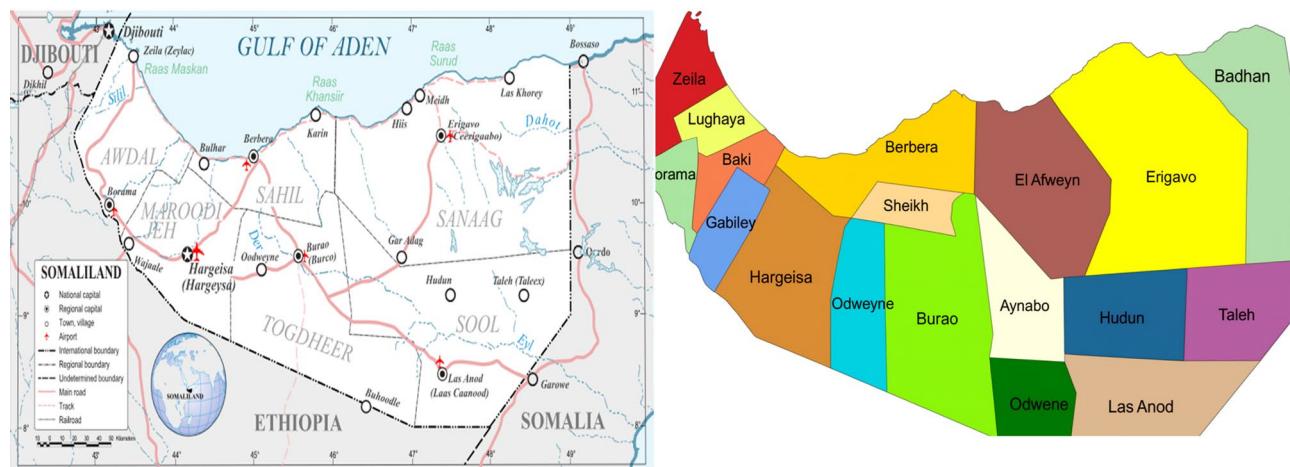

Fig. 1. Somaliland map (regions and districts).

Variables of study

Dependent variable

The dependent variable in this study was students' mathematics performance in primary school examinations, categorized as pass (coded as 1) or fail (coded as 0). The dichotomous classification signified a binary outcome, with one category designated as the event (coded as 1) and the other as the reference level (coded as 0). In the context of this research, success in mathematics is defined as students scoring 50 and above, which is termed a pass (coded 1), whereas those scoring below 50 are classified as failures (coded zero). Therefore, the variable 'Student Math Score' takes the value of 1 for scores equal to or above 50, denoting a pass, and 0 for scores below 50, denoting a failure in the primary examinations.

Independent variables

Various factors are expected to be associated with students' mathematics performance in primary school examinations, as shown in Table 2. In this study, the following independent variables were examined as potential influences on students' mathematics performance, particularly in the context of primary mathematics examinations:

Specification of supervised machine learning models

Logistic regression

Logistic regression (LR) is a widely used classification algorithm that predicts the probability of binary outcomes by modeling the relationship between independent variables and the dependent variable using logistic functions. It is a popular choice among data analysts and statisticians because of its simplicity and effectiveness[36]. The formula for the LR is

$$P\left(Y = 1|X\right) = \frac{1}{\left(1 + \exp\left(-z\right)\right)},$$

where $P(Y = 1|X)$ represents the probability of the positive class for a given instance, X represents the input variables, and z is the linear combination of the input variables and their respective coefficients.

Decision tree

A decision tree (DT) is a type of non-parametric supervised machine learning algorithm that constructs a tree-like model of decisions and their possible outcomes. It uses a data-splitting process based on different attribute values to form branches and leaf nodes, and the decision-making process is guided by a sequence of if-else conditions. The Decision Tree algorithm relies on various metrics, such as information gain, Gini index, and entropy, to identify the optimal attribute for partitioning the data[35].

Random forest

Random forests (RFs) are ensemble learning methods that combine multiple decision trees to improve prediction accuracy and reduce overfitting. A collection of Decision Trees was generated using random subsets of the training data and features. The final prediction was made by pooling the predictions of each tree. There is no single formula for random forests because they involve the integration of Decision Trees[37].

Naïve Bayes

Naive Bayes (NB) is a probabilistic classifier that uses Bayes' theorem and assumes independence between features. It determines the probability of each class based on a set of input features and selects the class with the highest probability[35]. More precisely, the following formula for NB is derived from Bayes' theorem:

$$P\left(C|X\right) = \frac{\left(P\left(X|C\right)\ P\left(C\right)\right)}{P\left(X\right)},$$

where $P\left(C|X\right)$ represents the probability of the class C Given the input features X, $P\left(X|C\right)$ is the probability of features X given class C, $P\left(C\right)$ is the prior probability of class C, and $P\left(X\right)$ is the prior probability of features X.

Support vector machine

Support vector machines (SVMs) are widely employed for both classification and regression tasks because they identify an optimal hyperplane that effectively separates data points from different classes. The SVM formula involves determining the decision boundary by solving a quadratic optimization problem[38].

K nearest neighbors

The K-nearest neighbors (KNN) algorithm is a simple yet highly efficient method for categorizing an instance based on its nearest neighbors in the feature space. The KNN algorithm can make accurate predictions by determining the majority class of k-nearest neighbors. The KNN algorithm involves calculating the distances between the target instance and all other instances in the training set to identify the K nearest neighbors[39].

Independent variables	Explanation	Category coding
	Awdal	0
	Badhan	1
	Buuhoodle	2
	Daadmadheedh	3
Regions	Gabiley	4
	Hawd	5
	Maroodi Jeeh	6
	Sahil	7
	Salal	8
	Sanaag	9
	Saraar	10
	Sool	11
	Togdher	12
	Haysimo	13
	Ainaba	0
	Badhan	1
	Baki	2
	Baligubadle	3
	Berbera	4
	Borama	5
	Burao	6
	Buuhoodle	7
Districts	Dhahar	8
	El Afwein	9
	Erigavo	10
	Gabiley	11
	Garadag	12
	Hargeisa	13
	Hudun	14
	Las'anod	15
	Lasqoray	16
	Lughaya	17
	Odweine	18
	Salahley	19
	Sheikh	20
	Taleh	21
	Zeila	22
Residence	Urban	0
	Rural	1
School accommodation	Boarding	0
	Day	1
Gender	Male	0
	Female	1
School type	Private	0
	Public	1
Science	Pass	0
	Fail	1
Somali	Pass	0
	Fail	1
English	Pass	0
	Fail	1
Arabic	Pass	0
	Fail	1

Table 2. Variables used in the study and their codes.

These six supervised machine learning models offer a diverse set of techniques for predicting mathematics performance using the 2022/2023 Somaliland National Primary Examination Results. Each model has its own strengths, assumptions, and formulas, enabling a comprehensive assessment of its predictive ability.

Model adequacy measures

The use of model adequacy measures provides essential information about a model's performance, precision, and reliability. Our analysis included several measures, such as accuracy, sensitivity, specificity, F1 score, precision, and recall. Each of these measures plays a vital role in assessing the predictive ability of the models and their effectiveness in correctly categorizing the instances.

Accuracy

Accuracy is a measure of the proportion of correctly predicted instances relative to the total number of cases. This value was calculated using the following equation:

$$Accuracy \ = \ \frac{(TP \ + \ TN)}{(TP \ + \ TN \ + \ FP \ + \ FN)},$$

Where TP (True Positives) represents the number of correctly predicted positive instances, TN (True Negatives) represents the number of correctly predicted negative instances, FP (False Positives) represents the number of incorrectly predicted positive instances, and FN (False Negatives) represents the number of incorrectly predicted negative instances.

A higher accuracy score indicates a more dependable model.

Sensitivity (recall)

Sensitivity, often referred to as recall or true positive rate, measures the proportion of correctly predicted positive instances (i.e., identifying students who performed well in mathematics) among all actual positive instances. The formula used to calculate this value is as follows

$$Sensitivity \ = \ \frac{TP}{(TP \ + \ FN)},$$

Where TP and FN are as described previously.

Specificity

Specificity is a metric that assesses the proportion of correctly predicted negative instances (i.e., identifying students who did not perform well in mathematics) to all actual negative instances. The following formula was used for the calculation:

$$Specificity \ = \ \frac{TN}{(TN \ + \ FP)},$$

Where TN and FP are the same as those described above.

F1 score

The F1 score is a combined measure of precision and recall. It provides a balanced assessment of the model's performance by considering both false positives and false negatives. This was calculated using the following formula.

$$F1 \ Score \ = \ \frac{2 \ * \ (Precision \ * \ Recall)}{(Precision \ + \ Recall)},$$

Precision is the proportion of correctly predicted positive instances relative to the total number of predicted positive cases, and recall, also known as sensitivity, measures the accuracy of predicting positive instances by calculating the proportion of correctly predicted positive instances among all actual positive cases.

A higher F1 score indicates better precision and recall, which reflects a more accurate and reliable model.

Precision

Precision is a metric that assesses a model's ability to identify positive instances among all accurately predicted positive instances. This was determined using the following equation:

$$Precision \ = \ \frac{TP}{(TP \ + \ FP)},$$

Where TP and FP are as described above.

A higher precision score indicates fewer false positives and a more dependable model.

Area under the curve

The area under the curve (AUC) is a commonly used measure of model performance that assesses a supervised machine learning model's overall effectiveness across a range of classification thresholds. This is particularly useful for evaluating the predictive ability of models and comparing their performance.

In the context of our study, which assessed the ability to predict mathematical performance using data from the 2022/2023 Somaliland National Primary Examination Results, the AUC served as a valuable tool for evaluating models' ability to discriminate between positive and negative outcomes.

A higher AUC score indicates that the model has a superior ability to correctly classify instances, as it assigns higher probabilities to positive than to negative cases.

Using these model adequacy measures, we gained valuable insights into the performance and effectiveness of supervised machine learning models for predicting mathematical performance using the 2022/2023 Somaliland National Primary Examination Results.

Figure 2 Illustrates the methodology employed in this study. This study used the Somaliland National Examination Result 2022/2023 dataset, which was split into training and test sets at 80% and 20%, respectively. A suite of machine learning models, including Logistic Regression, Decision Tree, Random Forest, Naive Bayes, SVM, and KNN, were developed and trained on the training data. Subsequently, the performance of each model was evaluated based on the test data using a comprehensive set of accuracy metrics, including accuracy, Sensitivity, F1-score, Precision, specificity, and AUC. This rigorous methodology enabled the identification of the most suitable model for predicting student performance in the Somaliland National Examination.

Model validation strategy

To evaluate robustness, we implemented 10-fold stratified cross-validation in addition to the conventional 80/20 train–test split method. The dataset was divided into 10 folds, preserving the pass/fail ratio within each fold. Each model was trained on nine folds and tested on the remaining fold, with each fold serving as the test set in turn. The performance metrics were averaged across folds, and their standard deviations were reported. This approach mitigated the risk of overfitting and enabled us to assess the model's stability.

Hyperparameter tuning and implementation

The hyperparameters were tuned using a grid search within the cross-validation framework. For the Decision Tree (DT) and Random Forest (RF), the maximum depth, number of estimators, and minimum samples per split were optimized. For Support Vector Machines (SVM), both linear and radial basis function kernels were tested by tuning the penalty parameter (C) and kernel coefficient (γ). For K-Nearest Neighbors (KNN), the number of
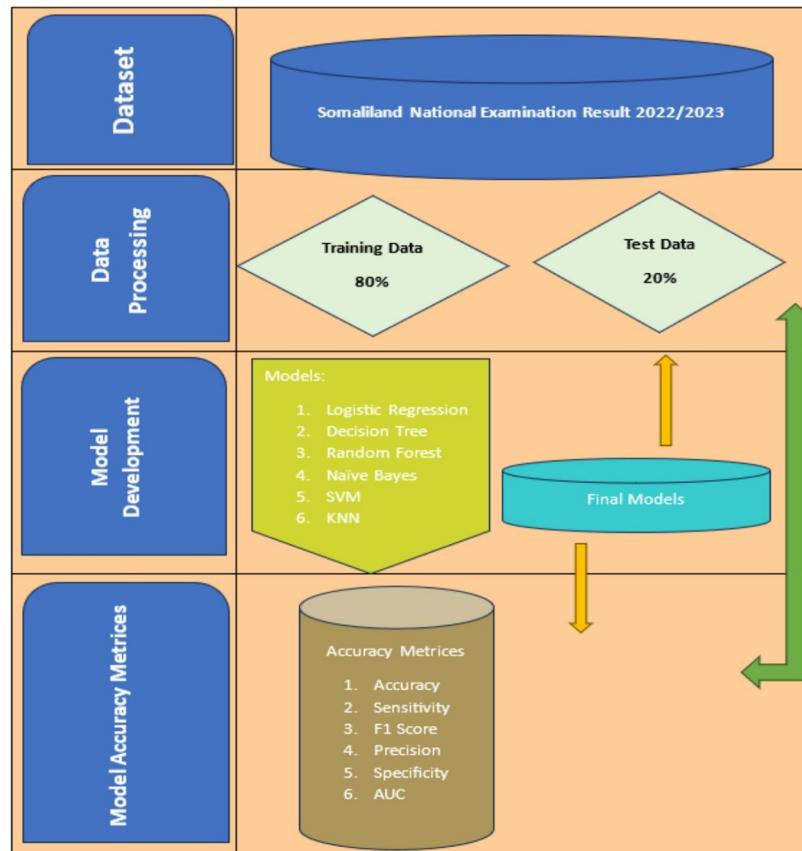

Fig. 2. Diagram of methodology.

neighbors (k) was tuned to 3–15, with alternative distance weighting schemes tested. Naïve Bayes and Logistic Regression required minimal tuning and were implemented with standard parameterization.

Results
Descriptive statistics
Table 3 Presents a comprehensive analysis of mathematics performance across regions, districts, school types, school types, school accommodations, and demographic characteristics. The regions of Awdal and Maroodijeex had the highest failure rates of 78.88% and 81.49%, respectively. This indicates the need for targeted educational reforms and additional support for students. In contrast, the Sanaag and Buhodleh regions had significantly higher pass percentages of 83.48% and 80.00%, respectively. This suggests that specific educational practices or resources in these regions may be responsible for their success. Laasqoray and Sheekh districts have achieved exceptional pass rates of 100% and 98.95% respectively. In contrast, the Hargeisa and Borama districts performed poorly, with failure rates of 81.55% and 83.38%, respectively.

Figures 3 and 4 further patterns emerged when analyzing factors such as school type, place of residence, and gender differences. Urban schools exhibited a significantly higher failure rate (67.64%) than rural schools (45.21%). This finding suggests that students in urban areas face specific challenges. The failure rate in private schools was 68.08%, slightly higher than the 62.21% reported in public institutions. Gender disparities were evident, with males exhibiting a higher failure rate (67.17%) than that of females (63.52%). This indicates the presence of different degrees of support and external influences. The patterns shown in Table 1 call for targeted interventions to address the underlying factors that contribute to gaps in mathematics performance, with the aim of creating a more equitable and inclusive educational environment.

Feature importance
The feature importance Fig. 4 features importance for all models analysis across the six supervised machine learning models revealed that *English proficiency consistently emerged as the most influential predictor of mathematics performance*, highlighting the central role of language comprehension in academic achievement. Models such as *Naïve Bayes and Random Forests achieved strong performance in related subjects*, *such as Science and Arabic*, *demonstrating cross-subject generalization*. In contrast, *support vector machines and logistic regression* highlighted demographic and contextual variables such as region, residence, and school type, indicating that educational disparities are also significant drivers of performance differences. These variations demonstrate that different algorithms capture distinct aspects of the problem, with probabilistic models accentuating subject strengths and regression-based approaches reflecting systemic inequities.

Tree-based methods further illustrate the complexity of variable interactions. The *Decision Tree* emphasized Somali proficiency and regional context, whereas the *Random Forest* distributed importance more evenly, reducing instability through ensemble averaging. Meanwhile, the *KNN model* placed disproportionate weight on English, suggesting a potential over-reliance on a single predictor despite achieving moderate accuracy. Overall, the comparative analysis underscores that mathematics outcomes are shaped by both linguistic and contextual factors, with each model offering complementary insights. This supports the need for integrated interventions that simultaneously address *language barriers*, *regional disparities*, *and school-level inequalities* to improve mathematics performance in Somaliland.

Logistic regression
An Logistic regression study conducted on the 2022/2023 Somaliland National.

Examination revealed several notable findings regarding the elements that influence mathematics performance. Table 4 indicates that several factors, including geographical region, district, gender, school type, place of residence, school facilities, and performance in other courses, significantly impacted mathematics performance.

First, the odds ratios indicate that specific regions have significantly higher or poorer mathematics performance than the reference region, Awdal. The Sanaag region had the highest odds ratio (OR:18.878), indicating a substantially higher likelihood of achieving higher mathematics performance. Similarly, Buhodleh, Badhan, Sahil, Salal, Saraar, and Daad-Madheedh had higher levels of mathematics achievement. Marodijeex, on the other hand, was slightly less proficient than Awdal, with an odds ratio of 0.849.

Furthermore, the district variable indicates that mathematics proficiency is associated differently across districts. Sheekh had a significantly higher odds ratio (OR: 30.122), suggesting a stronger correlation with superior math proficiency than Badhan. However, Hargaysa, Boorama, Lug-haya, Gabiley, Baligubadle, and Burco had significantly lower mathematics achievement than Badhan.

However, when sex was considered, the odds ratio (OR: 0.851) indicated that males had a slightly lower probability of achieving higher arithmetic performance than females. Conversely, the variable "school type" suggests that students attending public schools (OR: 1.295) were more likely to have higher arithmetic performance than those attending private schools.

The residence variable indicated that children from urban areas (OR: 0.395) were significantly less likely to achieve higher mathematics proficiency than those from rural areas.

In addition, the variable for school accommodation indicates that children who do not live at school show a modest decrease in arithmetic performance (OR: 0.945).

Furthermore, academic achievement in other subjects significantly impacted mathematics performance. There was a high correlation between success in Somali, Arabic, English, and Science, and improved mathematical skills. Somali (OR: 3.076), Arabic (OR: 4.926), English (OR: 7.114), and science (OR: 7.018) were associated with a significant increase in the likelihood of improved mathematics performance.

Characteristics	Frequency	Maths performance (%)	
		Fail	Pass
Region			
Awdal	2,259	78.88	21.12
Badhan	345	24.06	75.94
Buhodleh	50	20.00	80.00
Daad-madheedh	124	31.45	68.55
Gabiley	1,410	63.26	36.74
Hawd	45	82.22	17.78
Maroodijeex	10,501	81.49	18.51
Sahil	965	26.01	73.99
Salal	257	26.46	73.54
Sanaag	1,138	16.52	83.48
Togdheer	3,603	49.04	50.96
Saraar	253	25.69	74.31
District			
Badhan	342	24.27	75.73
Baki	148	20.95	79.05
Baligubadle	45	82.22	17.78
Berbera	680	36.47	63.53
Borama	1,955	83.38	16.62
Buhodle	50	20.00	80.00
Burco	3,603	49.04	50.96
Caynabo	253	25.69	74.31
Ceelafweyn	127	18.90	81.10
Ceerigaabo	937	15.37	84.63
Gabiley	1,410	63.26	36.74
Garadag	74	27.03	72.97
Hargeysa	10,458	81.55	18.45
Laasqoray	3	00.00	100.0
Lug-haya	156	77.56	22.44
Oodweyne	124	31.45	68.55
Salaxley	43	67.44	32.56
Saylac	257	26.46	73.54
Sheekh	285	1.05	98.95
Residence			
Rural	1,922	45.21	54.79
Urban	12,870	67.64	32.36
School type			
Private	12,032	68.08	31.92
Public	8,918	62.21	37.79
Gender			
Female	9,111	63.52	36.48
Male	11,839	67.17	32.83
School accommodation			
Boarding	830	64.34	35.66
Day	20,120	65.63	34.37
Science			
Fail	8,056	88.16	11.84
Pass	12,894	51.47	48.53
Somali language			
Fail	1,574 (4%)	84.56	15.44
Pass	19,376 (96%)	64.04	35.96
English Language			
Fail	9,641	86.62	13.38
Pass	11,309	47.64	52.36
Continued			

		Maths performance (%)	
Arabic language			
Fail	4,270	88.01	11.99
Pass	16,690	59.84	40.16

**Table 3**. Percentage of students' mathematics performance by characteristics.
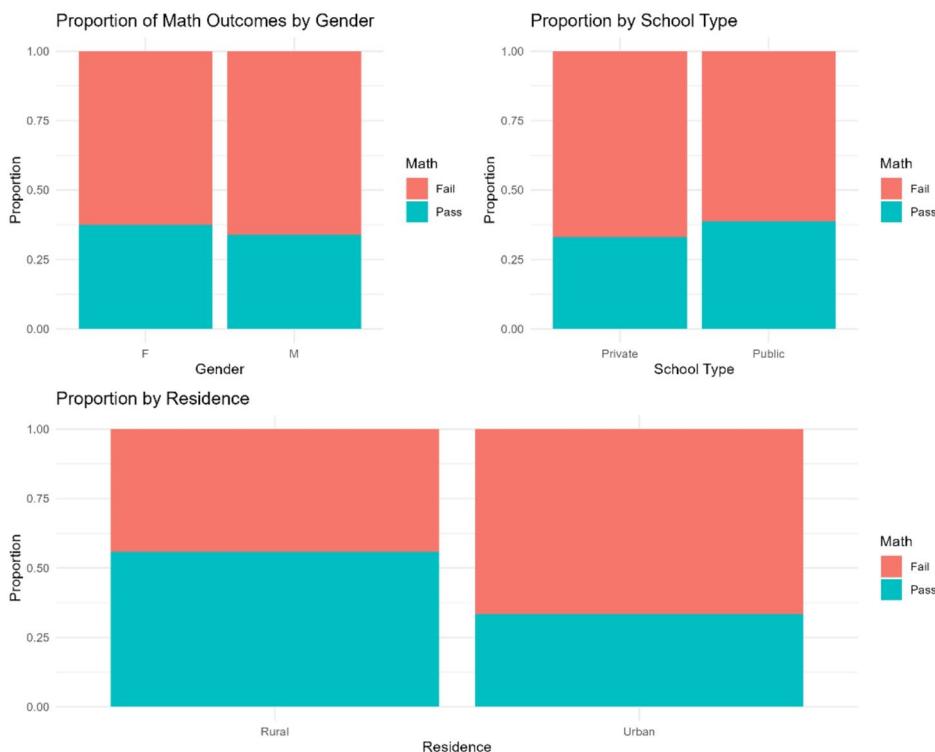
Fig. 3. Count plot for categorical variables.

In summary, the Logistic regression study identified several factors that influenced mathematical performance in the 2022/2023 Somaliland National Examination. Several factors, including geographical region, district, gender, type of school, place of residence, school facilities, and success in other subjects, had a significant impact. The odds ratios provide valuable insights into the magnitude of the relationships between these parameters and mathematics performance, facilitating the identification of areas for improvement and targeted interventions to improve mathematics education in Somaliland.

Predicting students' mathematics performance

Table 5; Figs. 5, and 6 provide valuable insights into the comparative effectiveness of different machine learning models, including logistic regression. Naïve Bayes, Random Forest, Decision Tree, Support Vector Machine, and K-nearest neighbor.

The NB model emerged as a clear standout, with the highest overall accuracy of 98.6%. This indicates that it is the most successful method for correctly classifying instances. Furthermore, NB demonstrated exceptional sensitivity (98.5%) and specificity (99.0%), demonstrating its ability to accurately identify both the "Fail" and "Pass" classes. The model's positive and negative predictive values (99.4% and 97.3%, respectively) further reinforce its reliability in predicting positive and negative outcomes. In particular, the AUC of the NB model (0.999) was exceptionally close to a perfect score of 1, underlining its excellent performance in discriminating between the two classes.

In contrast, the Random Forest, Logistic regression, and Decision Tree models exhibited relatively comparable performances, with accuracies ranging from 78.0% to 79.4%. These models showed balanced sensitivity and specificity, indicating their ability to discriminate between the "Fail" and "Pass" classes. However, their negative predictive values were relatively lower (66.0%, 64.3%, and 64.9%, respectively), suggesting that they are better at predicting the "Fail" class than the "Pass" class. The AUC values for the Random Forest and Logistic Regression models were both 0.856, indicating good discrimination between the two classes, whereas the Decision Tree model's AUC was slightly lower at 0.806.
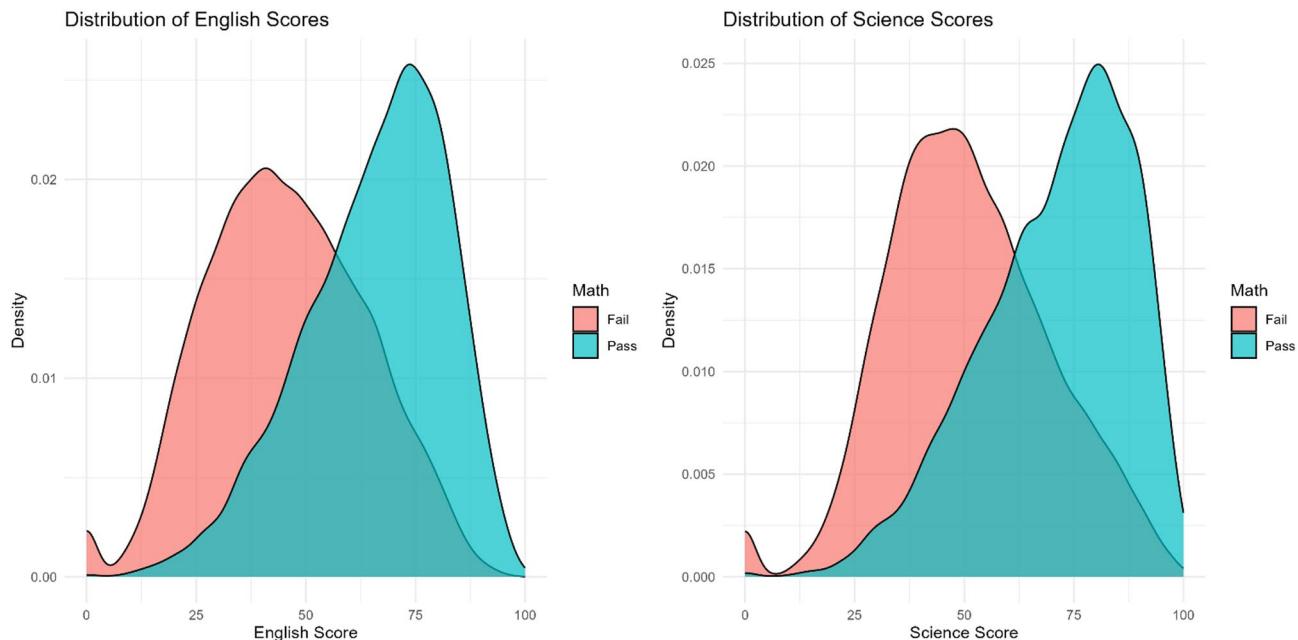
Fig. 4. Boxplot for continuous variables.

The Support Vector Machine model, with an accuracy of 70.3%, was the weakest of the six models. Its sensitivity and specificity of 69.4% and 76.0%, respectively, suggest an unbalanced performance in identifying the two classes. Although the model's positive predictive value was very high at 95.2%, its negative predictive value was extremely low at 26.4%, indicating a much stronger ability to predict the "Fail" class than the "Pass" class. The SVM model's AUC (0.745) was the lowest among the models, further confirming its relatively weak performance at discriminating between the two classes.

Finally, the K-nearest neighbor model occupied middle ground with an accuracy, sensitivity, and specificity of 80.3%, balanced sensitivity (77.5%), and specificity (91.3%). Its positive predictive value was very high at 97.2%, but its negative predictive value was relatively low at 50.5%, suggesting a stronger ability to predict the Real-Fail class. The K-nearest neighbor model had an AUC of 0.868, the third-highest among the models, indicating good performance at discriminating between the two classes.

The table provides an analysis of the different machine learning models employed to predict mathematical failure rates. It includes metrics for predicted failure, observed failure, and model accuracy. The observed mathematical failure rate was 64.4%. The models' predicted failure rates ranged from 70.71% to 70.98%, suggesting a tendency to overestimate failure rates.

In conclusion, a comprehensive analysis of the performance metrics highlighted the Naïve Bayes model as a clear standout, with exceptional accuracy, sensitivity, specificity, and AUC. The Random Forest, Logistic Regression, and Decision Tree also demonstrated strong performance, whereas the Support Vector Machine model was the weakest among the six models. These findings can inform the selection of the most appropriate model for a given classification task based on the desired balance between accuracy, class-specific performance, and overall discriminative ability.

Model comparisons

Figure 5, the comparative performance results clearly differentiate the supervised machine learning models evaluated in this study. Naïve Bayes excelled, achieving nearly perfect accuracy, sensitivity, F1-score, and AUC, indicating strong alignment with predictors such as language proficiency and science outcomes. KNN and Random Forest also performed well, showing balanced results and effectively modeling non-linear relationships. Logistic Regression and Decision Tree showed moderate performance, with some predictive ability but limited capacity to capture the full complexity of student performance. Conversely, the Support Vector Machine performed worst across all metrics, highlighting challenges in handling data imbalance and multiple categorical predictors in the Somaliland primary examination dataset.

The ROC curves Fig. 6 further support these results by demonstrating each model's ability to differentiate between outcomes. Naïve Bayes achieved an almost-perfect ROC curve with an AUC of 99.8%, confirming its superior ability to distinguish students who passed from those who failed the mathematics exam. Random Forest, Logistic Regression, and K-nearest neighbor also yielded smooth ROC curves, each with an AUC above 85%, indicating strong and reliable classification performance. Meanwhile, the Decision Tree had moderate predictive power, and the SVM, with an AUC of only 65.3%, demonstrated limited discrimination. Overall, the results validate Naïve Bayes as the most effective for predicting mathematics performance and highlight the importance of choosing algorithms suited to educational datasets in Somaliland (Fig. 7).

Math performance	Odds Ratio	Std. Err.	z	P>|z|	95% Conf.	
Region (Ref: Awdal)						
Badhan	11.793	1.605	18.13	**0.000**	9.032	15.398
Buhodleh	14.943	5.339	7.57	**0.000**	7.419	30.101
Daad-Madheedh	8.142	1.630	10.48	**0.000**	5.500	12.054
Gabiley	2.169	0.164	10.25	**0.000**	1.871	2.516
Hawd	0.808	0.318	-0.54	**0.587**	0.374	1.746
Maroodijeex	0.849	0.049	-2.86	**0.004**	0.758	0.950
Sahil	10.627	0.953	26.35	**0.000**	8.914	12.669
Salal	10.383	1.563	15.55	**0.000**	7.731	13.946
Sanaag	18.878	1.794	30.92	**0.000**	15.670	22.743
Togdheer	3.882	0.238	22.09	**0.000**	3.442	4.378
saraar	10.805	1.652	15.57	**0.000**	8.008	14.579
District (Ref: Badhan)						
Baki	1.209	0.288	0.8	**0.424**	0.758	1.929
Baligubadle	0.069	0.028	-6.51	**0.000**	0.031	0.155
Berbera	0.558	0.083	-3.91	**0.000**	0.417	0.748
Boorama	0.064	0.009	-19.65	**0.000**	0.049	0.084
Buhodle	1.282	0.481	0.66	**0.508**	0.614	2.675
Burco	0.333	0.043	-8.43	**0.000**	0.258	0.430
Caynabo	0.927	0.177	-0.4	**0.691**	0.637	1.349
Ceelafwayn	1.375	0.357	1.23	**0.219**	0.827	2.287
Ceerigaabo	1.765	0.274	3.66	**0.000**	1.302	2.393
Gabiley	0.186	0.026	-12.21	**0.000**	0.142	0.244
Garadag	0.865	0.251	-0.5	**0.618**	0.490	1.529
Hargaysa	0.073	0.009	-20.4	**0.000**	0.056	0.093
Lug-haya	0.093	0.021	-10.36	**0.000**	0.059	0.145
Oodweyne	0.698	0.161	-1.55	**0.120**	0.444	1.098
Salaxlay	0.155	0.054	-5.35	**0.000**	0.078	0.307
Saylac	0.891	0.169	-0.61	**0.541**	0.614	1.291
Sheekh	30.122	17.891	5.73	**0.000**	9.404	96.485
Gender (Ref: Female)						
Male	0.851	0.025	-5.51	**0.000**	0.804	0.901
School type (Ref: Private)						
Public	1.295	0.038	8.83	**0.000**	1.223	1.372
Residence (Ref: Rural)						
Urban	0.395	0.019	-19.21	**0.000**	0.359	0.434
Accommodation (Ref: Boarding)						
Days	0.945	0.070	-0.77	**0.442**	0.817	1.092
Somali (Ref: Fail)						
Pass	3.076	0.219	15.75	**0.000**	2.675	3.538
Arabic (Ref: Fail)						
Pass	4.926	0.245	32.09	**0.000**	4.469	5.430
English (Ref: Fail)						
Pass	7.114	0.251	55.51	**0.000**	6.638	7.624
Science (Ref: Fail)						
Pass	7.018	0.272	50.32	**0.000**	6.505	7.572

Table 4. LR analysis of students' mathematics performance.

Cross-validation and robustness checks

Cross-validation results showed that the Naïve Bayes model achieved a mean accuracy of 95.8% (SD ± 0.7), compared with 98.6% in the holdout test set. This indicates that while Naïve Bayes remained the strongest performer overall, the single-split estimate slightly overestimated the performance. Random Forest and Logistic Regression showed mean accuracies of 79.0% and 78.5%, respectively, with lower fold-to-fold variability, indicating stable but less extreme performance.

These results highlight that although Naïve Bayes is promising, its performance should be interpreted cautiously. Correlated predictors, such as Science and English grades, may partly explain the strong results,

	LR		NB		RF		DT		SVM		KNN	
Predicted	Fail	Pass	Fail	Pass	Fail	Pass	Fail	Pass	Fail	Pass	Fail	Pass
Fail	2349	322	2656	15	2322	349	2285	386	2545	126	2598	73
Pass	541	978	40	1479	516	1003	532	987	1118	401	751	768
%												
Accuracy	0.794		0.986		0.793		0.780		0.703		0.803	
Sensitivity	0.812		0.985		0.818		0.811		0.694		0.775	
Specificity	0.752		0.990		0.741		0.718		0.760		0.913	
Pos pred value	0.879		0.994		0.869		0.855		0.952		0.972	
Neg pred value	0.643		0.973		0.660		0.649		0.264		0.505	
Prevalence	0.689		0.643		0.677		0.672		0.874		0.799	
Detection rate	0.560		0.633		0.554		0.545		0.607		0.620	
Detection prevalence	0.637		0.637		0.637		0.637		0.637		0.637	
Balanced accuracy	0.782		0.987		0.780		0.765		0.727		0.844	
AUC	0.856		0.999		0.856		0.806		0.745		0.868	

Table 5. Model accuracy and comparison.
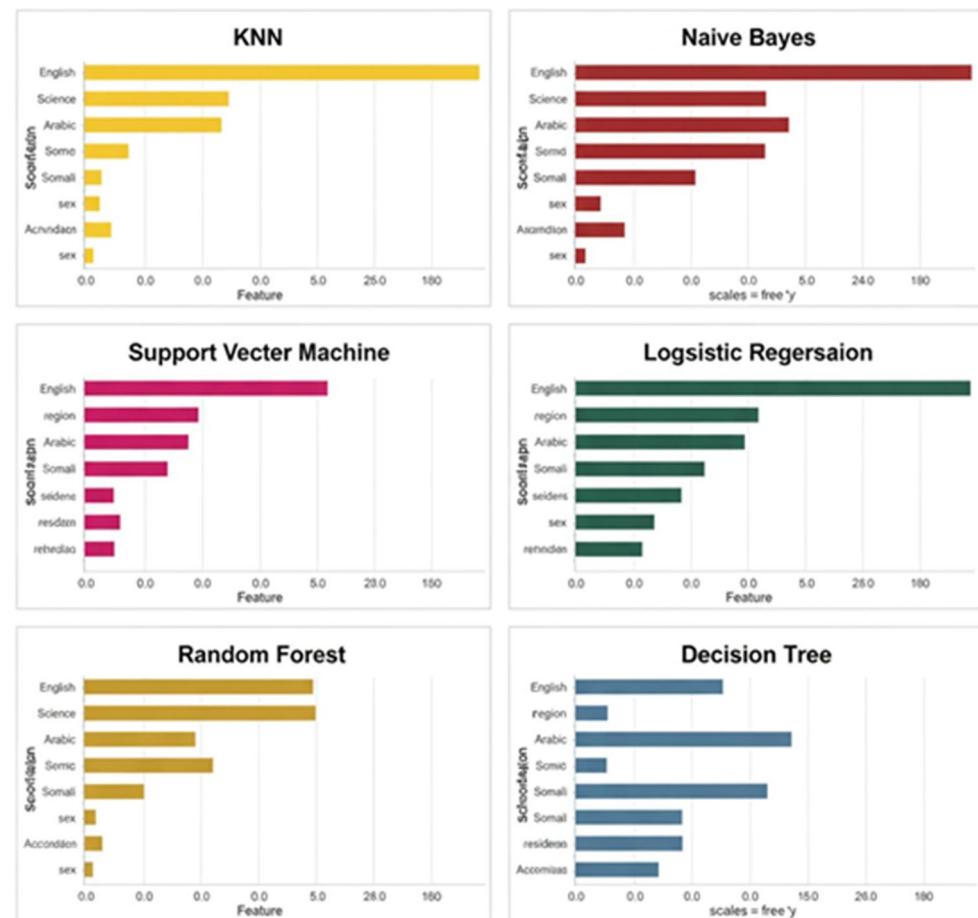
Fig. 5. Features importance for all models.
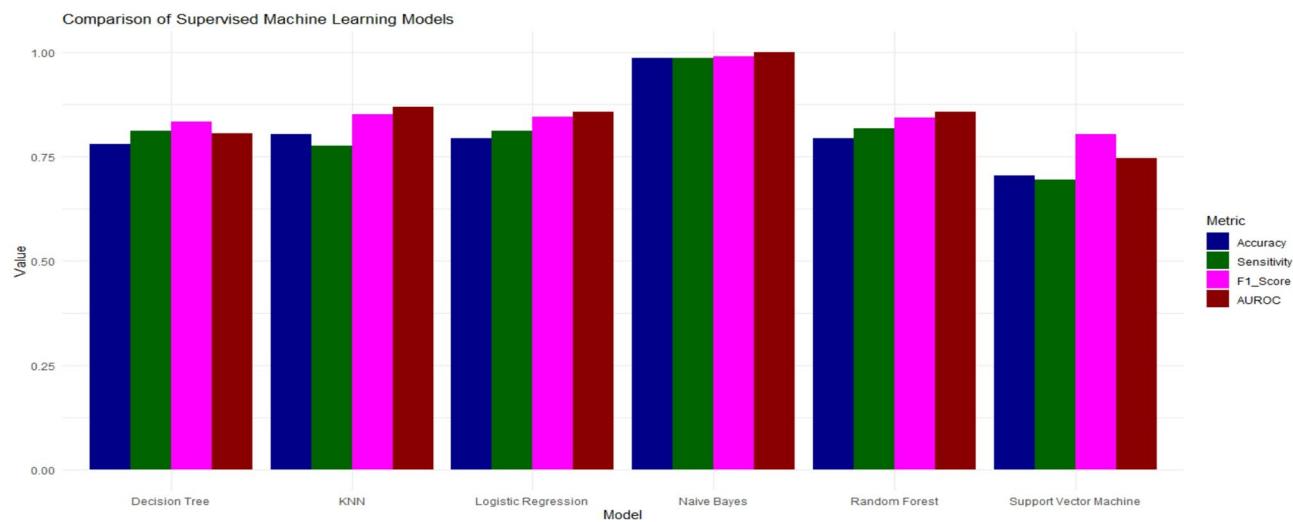
Fig. 6. Accuracy for all six model.
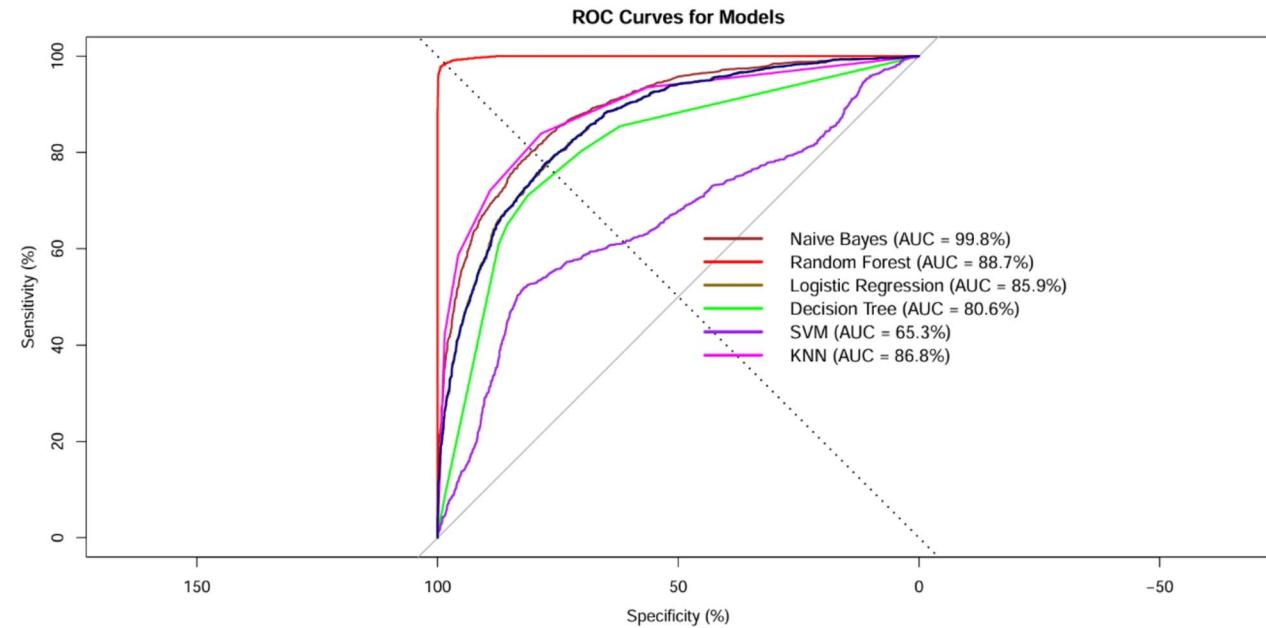
Fig. 7. ROC curve for the six ML models.

raising the possibility of inflated predictive accuracy. Nevertheless, cross-validation confirmed that all models generalized reasonably well across folds.

Discussion

This study provides valuable insights into the factors influencing primary school students' mathematics performance in Somaliland, analyzed through a comprehensive comparison of six supervised machine-learning models. The findings underscore critical disparities across regions, school types, gender, and residence, reflecting broader systemic and contextual influences on mathematics achievement.

The outstanding performance of the naïve Bayes model, with an accuracy of 98.6%, highlights its potential as a reliable tool for predicting student outcomes. Its high sensitivity and specificity demonstrate its ability to effectively differentiate between students at risk of failing and those likely to succeed academically. This finding aligns with the existing literature advocating the use of probabilistic models in educational settings. For instance[40], demonstrated the effectiveness of machine learning in predicting educational performance, emphasizing the utility of these models in shaping educational strategies and interventions. Similarly[41], highlighted the role of

machine learning frameworks, particularly Naïve Bayes, in identifying at-risk students, reinforcing our findings regarding their predictive capabilities.

The comparatively moderate performance of other models, such as Random Forest and Logistic Regression, suggests their potential utility in contexts in which balanced class performance is prioritized. This aligns with prior studies, such as those by[42,43], which explored the adaptability of various machine learning techniques in diverse educational scenarios.

The higher failure rates observed among urban students compared to rural students, and among males compared to females, merit further investigation. These disparities could be due to variations in resource availability, learning environments, and social expectations, aligning with previous findings[44]. , who highlighted the importance of contextual factors in influencing students' academic success. Regional analyses showed notable differences in performance, with areas like Sanaag and Buhodleh achieving higher success rates than others. These findings are consistent with those of[45], highlighting the significance of region-specific educational methods in improving performance.

The feature importance analysis reveals a strong correlation between language proficiency and mathematics performance, underscoring the interconnectedness of academic subjects. Proficiency in English, Somali, and Arabic is a key predictor of success, highlighting the importance of integrated educational strategies that strengthen both language and math skills. This aligns with earlier research[46]. , which underscored the intertwined relationship between linguistic and mathematical abilities in educational settings. The identified English proficiency was the single most influential predictor of mathematics performance, surpassing even Science scores. This suggests that the barrier to mathematical achievement in Somaliland may not be purely numerical, but linguistic. If students struggle to comprehend the syntax of English word problems or instructions, their mathematical reasoning cannot be accurately assessed. This implies that 'poor math performance' may reflect 'low English reading comprehension,' necessitating a curriculum that integrates Content and Language Integrated Learning (CLIL) approaches.

Interestingly, the observation that public school students outperform their private school counterparts questions the common belief of private education's superiority. This indicates that public schools in Somaliland might have strengths in curriculum delivery and teacher involvement, which could enhance mathematics achievement. The finding that public schools outperformed private schools challenges the conventional perception of private education superiority in developing contexts. This disparity may stem from a lack of regulatory standardization among private institutions in Somaliland, where teacher qualifications and adherence to curriculum may vary significantly compared to the government-monitored public sector. Additionally, private schools may be prioritizing rote learning or other subjects, whereas public schools may focus more strictly on the national curriculum requirements for the examination.

These findings carry important implications for educational policy and practice in Somaliland. Targeted actions like teacher training, directing resources to areas with poor performance, and modifying the curriculum to overcome language barriers could help reduce the disparities identified. Additionally, using predictive models such as Naïve Bayes can assist educators and policymakers in detecting vulnerable students early and taking proactive support steps[47].

This study is grounded in the theory of educational inequality, which emphasizes how structural and contextual disparities—such as region, gender, and school type—shape students' learning outcomes. By applying this lens, the analysis interprets performance gaps in mathematics as reflections of unequal access to resources, teacher quality, and supportive environments. In addition, cognitive frameworks linking language proficiency to mathematical achievement provide an essential conceptual bridge, highlighting how mastery of instructional language enhances the comprehension of mathematical concepts. Together, these perspectives offer a robust foundation for understanding the observed disparities and situating them within the broader debates on equity and learning.

The Naïve Bayes model's superior accuracy of 98.6%, outperforming more complex algorithms like SVM at 70.3%, is due to the dataset's characteristics. This dataset mainly consists of categorical variables such as Region, District, Gender, and School Type. Naïve Bayes excels with high-dimensional categorical data and is less affected by the 'curse of dimensionality' compared to distance-based methods like KNN or SVM. Additionally, SVM faced difficulties in identifying an optimal hyperplane, possibly because of the non-linear overlap between passing and failing students in the feature space. In contrast, Naïve Bayes effectively utilized conditional probabilities of key predictors—particularly language proficiency—to achieve higher classification accuracy.

Conclusions

The results of the 2022/2023 Somaliland National Examination revealed a decline in primary school students' mathematics performance, highlighting global concerns about poor performance in mathematics, particularly in East Asia and Africa. This study used data from the Somaliland National Examination Database, comprising over 20,000 students, to assess factors influencing mathematics performance, including demographic characteristics such as gender, school type, school accommodation, place of residence, school ownership, and other subjects. The analysis revealed performance disparities across regions, districts, school types, and genders, indicating the need for targeted interventions to address them.

Machine learning models, including Logistic regression, Naïve Bayes, Random Forest, Decision Tree, and Support Vector Machine, were used to predict mathematical outcomes. NB had the highest accuracy, highlighting its reliability in predicting outcomes. Sensitivity and specificity analyses demonstrated the models' ability to detect positive and negative cases accurately. Naïve Bayes and K-nearest neighbor exhibited superior discriminative powers, as reflected by their higher AUC values.

In conclusion, this study provides valuable insights into the factors influencing primary school students' mathematics performance in Somaliland, enabling researchers to understand better and predict students'

academic outcomes and paving the way for targeted interventions to improve mathematics education in the region.

With data from over 20,000 students across 14 regions and 23 constituencies, the study's findings can be reasonably generalized to Somaliland's primary-school population. The large, diverse sample and robust machine learning framework also make this approach adaptable to similar educational settings in sub-Saharan Africa.

Recommendations

This study highlights the need for a multifaceted approach to address disparities in mathematics performance among primary school students in Somaliland. Policymakers should prioritize targeted regional interventions by allocating resources to underperforming areas, such as Maroodi Jeeh and Awdal, while studying and replicating successful practices from high-performing regions, such as Sanaag and Buhodleh. Furthermore, urban schools require enhanced support, particularly in teacher training and access to learning resources, to mitigate the challenges that contribute to higher failure rates in urban settings. Addressing gender disparities is crucial, and strategies must include fostering female participation through mentorship and role models while investigating and addressing the unique challenges faced by male students.

Recognizing the strong correlation between language proficiency and mathematics performance, curriculum developers should integrate language enhancement into mathematics education to improve both areas simultaneously. Additionally, public and private schools should collaborate to exchange best practices and share innovative teaching methodologies that can strengthen educational delivery across sectors. The use of predictive models, such as Naïve Bayes, should be expanded to identify at-risk students and proactively guide interventions to ensure that resources are allocated effectively.

Teacher training remains a cornerstone of improving educational outcomes, and investment should focus on equipping educators with advanced teaching techniques, technology, and classroom management skills. Future research should explore broader determinants of mathematics performance, including socioeconomic status, parental involvement, and extracurricular activities, to provide a more comprehensive understanding of the factors that influence it. Finally, continuous monitoring and evaluation of the implemented strategies are essential for measuring their effectiveness, refining approaches, and adapting to evolving educational needs. Engaging stakeholders, including parents and local communities, in these initiatives fosters a supportive environment that promotes equitable and sustainable improvements in mathematics education.

The high predictive accuracy of the Naïve Bayes model suggests it can be deployed as an Early Warning System (EWS). rather than waiting for exam failure, the Ministry of Education can input student demographic and preliminary language scores into the model at the start of Grade 8. This would allow for the identification of 'at-risk' students' months before the final exam, enabling targeted remedial math camps specifically for those predicted to fail.

Limitations and future studies

The study on the determinants of mathematics performance among primary school students in Somaliland suggests that future research should focus on implementing changes to the mathematics curriculum to align with Vision 2030 requirements. It is also recommended that essential mathematical skills be identified and prioritized in the future to improve the technical and vocational environment. Future studies should examine the effectiveness of mathematics education and integrate recommendations into teacher training programs to improve teaching practices and student outcomes. Research on pedagogical innovations in primary mathematics can contribute to more engaging and practical education. Investigating the impact of teachers' mathematical knowledge on students' achievement and the development of confidence in mathematical concepts can guide policy initiatives to improve mathematics education.

However, the study is limited by its reliance solely on examination records, without socioeconomic or pedagogical variables that may influence performance. The cross-sectional design restricts causal inference, and only six algorithms were tested, excluding newer methods, such as deep learning. Thus, findings should be applied cautiously outside Somaliland's context.

Future research should aim to incorporate richer datasets that include socioeconomic indicators, teacher-level variables, and school-resource profiles to provide deeper explanatory power. Longitudinal studies following students across multiple grades would allow researchers to analyze learning progressions and identify long-term predictors of mathematics achievement. Exploring contemporary machine learning approaches, including XGBoost, LightGBM, deep learning, and explainable AI frameworks, could yield more nuanced insights and support more robust prediction models. Additionally, qualitative studies could contextualize the quantitative findings by examining how language proficiency, regional disparities, and school environments shape students' learning experiences. Finally, integrating predictive analytics into school-level decision-making systems could support early identification of at-risk learners, enabling timely interventions that enhance mathematics outcomes across Somaliland.

Data availability

The datasets analyzed in this study are available from the corresponding author upon reasonable request.

References

1. Ekwueme, C. *Mathematics Teaching and Learning in Schools.* (Radiant Venture Nig Ltd., 2013).

2. Edoho, E. & Esuong, U. Proficiency in mathematics language and word problem performance. *Int. J. Adult Contin. Dev. Educ.* **3**(1), 1 (2016).

3. Sarma, M. & Ahmed, M. A study on the difficulty of teaching and learning mathematics in under graduate level with special reference to Guwahati City. *Int. J. Soft Comput. Eng. IJSCE* **3**(1), 1 (2013).

4. Askew, M., Hodgen, J., Hossain, S. & Bretscher, N. *Values and Variables: Mathematics Education in High-Performing Countries* (Nuffield Foundation, 2010).

5. Mullis, I. V., Martin, M. O. & Sainsbury, M. PIRLS 2016 reading framework. *PIRLS*. 11–29 (2016).

6. Sturman, L. What is there to learn from international surveys of mathematical achievement? *Oxf Handb. Numer. Cogn*. 430–444 (2015).

7. Geary, D. C., Nicholas, A., Li, Y. & Sun, J. Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *J. Educ. Psychol.* **109** (5), 680 (2017).

8. Sinyosi, L. B. *Factors Affecting Grade 12 Learners' Performance in Mathematics at Nzhelele East Circuit* (Vhembe District in Limpopo, 2015).

9. Reddy, V. & Hannan, S. *TIMSS in South Africa: Making Global Research Locally Meaningful* (Human Sciences Research Council, 2018).

10. Ministry of Education and Science. *Annual Education Statistics Yearbook 2021 /2022*. https://moe.govsomaliland.org/article/annual-education-statistics-yearbook-20212022. Accessed 29 Nov 2023 (2023).

11. Ghasemi, E. & Burley, H. Gender, affect, and math: A cross-national meta-analysis of Trends in International Mathematics and Science Study 2015 outcomes. *Large-Scale Assess. Educ.* **7**(1), 1. https://doi.org/10.1186/s40536-019-0078-1 (2019).

12. Kiptum, J. K., Rono, P. K., Too, J. K., Bii, B. K. & Too, J. Effects of students gender on mathematics performance in primary schools in Keiyo South District, Kenya. *Language* **47** (2013).

13. Ding, C. S., Song, K. & Richardson, L. I. Do Mathematical gender differences continue? A longitudinal study of gender difference and excellence in mathematics performance in the U.S. *Educ. Stud.* **40**(3), 3 (2006). https://doi.org/10.1080/00131940701301952

14. Opoku-Asare, N. A. A. & Siaw, A. O. Rural–urban disparity in students' academic performance in visual arts education: Evidence from six senior high schools in Kumasi, Ghana. *Sage Open* **5**(4), 4 (2015).

15. Islam, R. & Khan, Z. N. Impact of locality and type of school management on academic achievement of school students at national level: An analysis. *Asian J. Res. Soc. Sci. Hum.* **7**(11), 11 (2017). https://doi.org/10.5958/2249-7315.2017.00536.6

16. Zanolli, M. B. et al. Differences in clerkship development between public and private Brazilian medical schools: An overview. *BMC Med. Educ.* https://doi.org/10.1186/s12909-020-02193-3 (2020).

17. Brunello, G. & Rocco, L. Educational standards in private and public schools. *Econ. J.* https://doi.org/10.1111/j.1468-0297.2008.02194.x (2008).

18. Adhikari, D. B. & Aryal, G. N. Factors determining performance of institutional schools in Chitwan, Nepal. *Econ. J. Dev. Issues*. https://doi.org/10.3126/ejdi.v25i1-2.25079 (2019).

19. Nurkolis, M. & Yuliejantiningsih, Y. Factors that are strongly associated with student learning outcomes and their policy implications. *Kne Soc. Sci.* https://doi.org/10.18502/kss.v7i19.12445 (2022).

20. Muralidharan, K. & Sundararaman, V. The aggregate effect of school choice: Evidence from a two-stage experiment in India. (2013).

21. Aslam, M. The relative effectiveness of government and private schools in Pakistan: Are girls worse off? *Educ. Econ.* https://doi.org/10.1080/09645290903142635 (2009).

22. Zhao, L., Ren, J., Li, Z. & Hong, Z. Quantitative analysis and prediction of academic performance of students using machine learning. *Sustainability* **15** (16), 12531. https://doi.org/10.3390/su151612531 (2023).

23. Chen, L., Chen, P. & Lin, Z. Artificial intelligence in education: A review. *Ieee Access.* **8**, 75264–75278. https://doi.org/10.1109/access.2020.2988510 (2020).

24. Deng, Y. A systematic review of application of machine learning in curriculum design among higher education. *J. Emerg. Comput. Technol.* **4**(1), 15–24. https://doi.org/10.57202/ject.1475566 (2024).

25. Ahmed, A. & Tolera, D. Multi-category prediction of students' academic performance using machine learning: For students joining higher educational institutions in Ethiopia. https://doi.org/10.21203/rs.3.rs-3342736/v1 (2023).

26. Alhabeeb, S., Alrusayni, N., Almutiri, R., Alhumud, S. & Al-Hagery, M. A. Blockchain and machine learning in education: A literature review. *Iaes Int. J. Artif. Intell. Ij-Ai*. **13** (1), 581. https://doi.org/10.11591/ijai.v13.i1.pp581-596 (2024).

27. Yang, B. An empirical study on the application of machine learning for higher education and social service. *J. Glob Inf. Manag.* **30** (7), 1–16. https://doi.org/10.4018/jgim.296723 (2022).

28. Ibarra-Vázquez, G., Soledad, M. & Fernández, M. B. Forecasting gender in open education competencies: A machine learning approach. *IEEE Trans. Learn. Technol* **17**, 1224–1235. https://doi.org/10.1109/tlt.2023.3336541 (2024).

29. Animashaun, E. S., Familoni, B. T. & Onyebuchi, N. C. Advanced machine learning techniques for personalising technology education. *Comput. Sci. It Res. J.* **5** (6), 1300–1313. https://doi.org/10.51594/csitrj.v5i6.1198 (2024).

30. Munir, H., Vogel, B. & Jacobsson, A. Artificial intelligence and machine learning approaches in digital education: A systematic revision. *Information* **13** (4), 203. https://doi.org/10.3390/info13040203 (2022).

31. Rittle-Johnson, B., Fyfe, E. R., Hofer, K. G. & Farran, D. C. Early math trajectories: Low-income children's mathematics knowledge from ages 4 to 11. *Child. Dev.* **88** (5), 1727–1742. https://doi.org/10.1111/cdev.12662 (2016).

32. Zhang, J. et al. The role of early language abilities on math skills among Chinese children. *Plos One.* **12** (7), e0181074. https://doi.org/10.1371/journal.pone.0181074 (2017).

33. Hardy, J. K. & Hemmeter, M. L. Systematic modeling and prompting to teach math skills to preschoolers with disabilities. *Top. Early Child. Spec. Educ.* **43** (2), 103–115. https://doi.org/10.1177/02711214211012772 (2021).

34. Hassan, M. A., Muse, A. H. & Nadarajah, S. Predicting student dropout rates using supervised machine learning: Insights from the 2022 national education accessibility survey in Somaliland. *Appl. Sci.* **14** (17), 7593 (2024).

35. Jama Ali, M., Chesneau, C. & Muse, A. H. Examining the determinants of student academic performance in Somaliland: Estimating unobserved effects at student and school levels using multi-level logistic regression. *Cogent Educ.* **11** (1), 2436293 (2024).

36. Ali, M. J., Hassan Muse, A. & Chesneau, C. Machine learning-based analysis of academic performance determinants in Somaliland: Insights from the 2021/2022 National Secondary School Exams. In *Springer Proceedings*. 24 (2024).

37. Naing, L., Winn, T. & Rusli, B. Practical issues in calculating the sample size for prevalence studies. *Arch. Orofac. Sci.* **1**, 9–14 (2006).

38. Tyralis, H., Papacharalampous, G. & Langousis, A. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water* **11** (5), 910 (2019).

39. Joachims, T. Making large-scale Svm learning. *Pract. Adv. Kernel Methods-Support Vector Learn.* (1999).

40. Nhu, V. H. et al. Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve Bayes tree, artificial neural network, and support vector machine algorithms. *Int. J. Environ. Res. Public. Health*. **17** (8), 2749 (2020).

41. Hossain, M. A., Ahammad, I., Ahmed, M. K. & Ahmed, M. I. Prediction of the computer science department's educational performance through machine learning model by analyzing students' academic statements. *Artif. Intell. Evol.* 70–87 (2023).

42. Evangelista, E. A hybrid machine learning framework for predicting students' performance in virtual learning environment. *Int. J. Emerg. Technol. Learn. Ijet.* **16**, 255–272. https://doi.org/10.3991/ijet.v16i24.26151 (2021).

43. Soyoye, T. O. Predicting academic performance of university students using machine learning: A case study in the UK. 431–434. https://doi.org/10.1109/wi-iat59888.2023.00071 (2023).
44. Alsariera, Y. A. et al. Assessment and evaluation of different machine learning algorithms for predicting student performance. *Comput. Intell. Neurosci.* **2022**, 1–11. https://doi.org/10.1155/2022/4151487 (2022).
45. Kaddoura, S., Popescu, D. & Hemanth, D. J. A systematic review on machine learning models for online learning and examination systems. *Peerj Comput. Sci.* **8**, e986. https://doi.org/10.7717/peerj-cs.986 (2022).
46. Dias, S. B., Hadjileontiadou, S. J., Diniz, J. A. & Hadjileontiadis, L. J. DeepLMS: A deep learning predictive model for supporting online learning in the Covid-19 era. *Sci. Rep.* **10** (1). https://doi.org/10.1038/s41598-020-76740-9 (2020).
47. Aziz, S. F. A. Students' performance evaluation using machine learning algorithms. *Coll. Basic. Educ. Res. J.* **16** (3), 977–986. https://doi.org/10.33899/berj.2020.166006 (2020).

Author contributions

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to S.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.