Predicting Students' Cognitive Profiles Using Explainable Machine Learning

Sonia Corraya[1], Fahmid Al Farid[2], M Shamim Kaiser[3],
Shamim Al Mamun[4], Jia Uddin[5]*, Hezerul Abdul Karim[6]*
Institute of Information Technology-Jahangirnagar University and Computer Science and Engineering Department,
Jagannath University, Dhaka, Bangladesh[1]
Faculty of Artificial Intelligence and Engineering, Multimedia University, Cyberjaya, Selangor, Malaysia[2]
Institute of Information Technology, Jahangirnagar University, Dhaka, Bangladesh[3,4]
Artificial Intelligence and Big Data Department, Woosong University, Daejeon, Republic of Korea[5]
Faculty of Artificial Intelligence and Engineering, Multimedia University, Cyberjaya, Selangor, Malaysia[6]

***Abstract*—Conventional educational strategies fail to comprehend and leverage the diversity of learners' cognitive strengths and overlook their innate intelligence, a fundamental driver of learning. To address this gap, this study proposes a machine learning (ML) framework to predict students' overall innate intelligence scores, independent of subject domain or exam structure, using the Learning Meta-Learning dataset, which includes data from 1,021 university students. Seven regression models, including Decision Tree, Random Forest, Extra Trees, Gradient Boosting, Extreme Gradient Boosting, LightGBM, and CatBoost, along with their ensembles have been trained and evaluated. Explainable Artificial intelligence (XAI) technique SHAP is used for important feature selection among 54 features and recursive feature elimination to further enhance model accuracy and interpretability. In comparison to the conventional method, the proposed SHAP-based ML approach is lightweight, trained with selected features, and has shown improvements in accuracy. The accuracy without XAI on CatBoost is 98.32%, whereas with XAI on CatBoost it is 98.53% using only 35 features out of 54. These findings suggest that integrating learners' cognitive profile prediction model can aid the design of personalized educational strategies, moving beyond one-size-fits-all educational strategies.**

***Keywords*—Explainable AI; SHAP feature selection; machine learning; innate intelligence prediction; cognitive profiles; student diversity**

I. Introduction

Academic success is often measured with scores in uniform exams, which measure subject-specific performance but fail to capture a learner's overall intellectual abilities. Although these scores are extensively being used as a qualifying measure for academic admissions, employment, and future education opportunities, they are a limited and impersonal assessment of talent. Multiple studies argue that standardized exams and Cumulative Grade Point Average (CGPA) do not adequately reflect genuine cognitive profiles or learning abilities [1], [2], [3], [4]. The CGPA is not a reliable indicator of innate intellect or a reliable predictor of success in life or the workplace. This is because CGPA and intelligence are two different notions: while exam scores represent achievement in particular assessment formats, intelligence is an innate element of humans and a more precise representative of talents. Traditional assessments are designed within fixed frameworks and consequently overlook intellectual diversity and may misrepresent learners' true potential [5], [6]. This misalignment highlights a fundamental flaw in how education systems comprehend and measure talents.

Intelligence is innate and the fundamental driver of learning and problem solving. Traditional education systems generally promote conformance to fixed rules over recognizing and nurturing varied intellectual strengths of learners. The belief that uniform educational strategies, whether in teaching and learning, curriculum design, or assessments, are sufficient to meet or assist all students undermines the objective of inclusive and equitable education. For example, while evaluation methods may differ between subjects, evaluations within each subject remain the same for all students. Such approaches focus on curricular content but fail to acknowledge the varied ways in which students understand, process, and apply knowledge. In order to be truly equitable and effective, educational approaches need to be learner-centered, holistic, and adaptable to the diverse intelligence scores. Nevertheless, present strategies continue to be mostly content-focused and disregard learners' intrinsic potentials [7].

With the advent of Machine Learning (ML) in educational studies, data-driven decision-making continues to gain popularity. Many ML models are aimed at predicting academic success (e.g., CGPA or performance scores) by using parameters like attendance, prior examination results, socioeconomic background, and geographic location [8], [9], [10], [11], [12]. Although these models provide valuable insights, they continue to depend on the conventional exam-centric paradigm, inheriting its limitations and disregarding deeper cognitive notions like intrinsic intelligence.

Exam performance is influenced by assessment style and primarily depends on linguistic and logical intelligence, whereas individuals possess spatial, bodily-kinesthetic, musical, interpersonal, and intrapersonal type intelligences as well [13], [14], [15], [16]. Therefore, exam scores are inadequate for measuring overall intelligence, and this narrow scope limits both scientific understanding and educational innovation. It creates a disparity between how competence is measured and what intelligence implies, which results in misguided educational decisions and inadequate learner support [17], [18], [19].

*Corresponding author.

To bridge this gap, this research aims to predict the overall innate intelligence, defined as the cumulative Multiple Intelligences (MI) score [14], [20], [21], [22]. Using the Learning Meta-Learning (LML) dataset, which includes 54 features capturing biological, psychological, behavioral, and metacognitive data, intelligence is modeled independently of test formats and study domains [23], [24]. The main contributions are as follows:

- Training and evaluation of seven regression models, namely Decision Tree, Random Forest, Extra Trees, Gradient Boosting, Extreme Gradient Boosting (XGBoost), LightGBM, and CatBoost, along with their ensembles to predict intelligence scores using the 1021 students' data from the LML dataset.
- Employment of SHAP (Shapley Additive Explanations) to identify and rank the most important features contributing to the predictions.
- Utilization of SHAP-based Recursive Feature Elimination (RFE) to improve model performance and interpretability.
- Examination of the contribution of learners' external factors, including age, gender, sleep duration, chronotype, illusion of competence, and impostor phenomenon, to the prediction of intelligence scores.
- Outlining the prospects and applicability of the proposed methodology, leading to targeted and equitable educational interventions.

Successful prediction of intelligence scores has the potential to advance personalized education. It allows educators to tailor teaching-learning techniques, determine learning outcomes, and create assessment methods that align with individuals' intrinsic intelligence, providing a more equitable alternative to one-size-fits-all approaches. This research supports the Sustainable Development Goal 4, which advocates for inclusive and equitable quality education [25]. It can also advance educational AI and learner-centered heutagogy by demonstrating that intelligence is both predictable and actionable, beyond what exam scores alone can do.

II. Literature Review

Conventional exam-based assessments have long dominated academic evaluations, admissions, and recruitment. Although standardized assessments are intended to ensure uniformity across learners, their operational simplicity comes at the cost of educational depth [1], [2], [3], [4], [17]. They often fail to reflect true talents, as fixed-format assessment performance is influenced by various external factors such as access to resources, language barriers, and test-taking strategies [18], [19], [26], [27]. Hence, more holistic, learner-centered approaches are needed to promote educational equity and inclusivity [28], [29]. Fixed-format exams, which primarily measure linguistic and mathematical abilities only, reinforce a narrow view of learner intelligence and thus privilege certain learners while marginalizing others who possess comparable overall cognitive strengths distributed across different intelligence types [5], [6].

Models such as Spearman's general intelligence factor (g-factor) and the Intelligence Quotient (IQ) score have historically molded the concept of intelligence by proposing a general cognitive ability that can be measured through standardized processes. Though these models are frequently employed in the fields of education and psychology, they are increasingly being criticized for being reductionist and biased [30], [31]. IQ tests concentrate on linguistic and logical thinking above creativity, interpersonal abilities, social skills, and practical reasoning capabilities important for the real world [32], [33], [34], [35], [36]. Furthermore, these scores show how well people perform in specific test contexts rather than their overall ability for learning or knowledge application. In summary, the target variable of this paper, intrinsic intelligence score, is a superior, holistic, and actionable measure compared to the regular intelligence models, such as IQ and the g-factor, which provide limited and biased perspectives on human potential.

The intelligence score predicted in this study includes the entire set of MI, embracing the complete spectrum of human cognitive skills, in contrast to IQ-based models [37]. This MI-based intelligence score has gained popularity among educators for its potential to provide individualized educational approaches [38], [39]. However, MI remained underutilized in educational practices and data-driven applications due to challenges with operationalizing it as a single, compatible construct [20]. Therefore, in this study the aggregated measure that encompasses all types of intelligences have been used, providing a comprehensive and scalable metric suitable for ML and educational decisions. Human intelligence types function interactively rather than in isolation; hence, treating them as separate and unrelated is both impractical and misleading [22]. The overall intelligence score provides a unified, quantitative representation of an individual's cognitive potential, more feasible for modeling and real world application than focusing only on dominant intelligences [40]. Moreover, many educational decisions, such as identifying students for enrichment or support, require comprehensive judgments rather than intelligence-type-specific analysis.

To the best of our knowledge, most ML applications in education continue to focus on predicting conventional outcomes, such as grades, test scores, course completion rates, or dropout risk, based on features like attendance, quiz scores, or demographic data. While these models can forecast academic performance, they inherit the limitations of traditional assessments [8], [9], [10], [11], [12], [41]. Additionally, most of the existing ML models lack transparency. Black-box techniques make it difficult to understand and rely on the reasoning behind output generations. Explainable Artificial Intelligence (XAI) approaches, such as SHAP, provide a solution by dictating which features are most important for predictions [30].

In summary, this study addresses three gaps in educational ML research: 1) the persistent over-reliance on traditional academic outcomes as representations for talents, 2) the limited integration of learners' innate intelligence into ML-based modeling for personalized education, and 3) the lack of interpretability in ML applications for education.

III. Proposed Methodology

The proposed methodology integrates feature selection with traditional ML models, significantly enhancing the accuracy and interpretability of learners' intelligence prediction models. As shown in Fig. 1, the study is divided into two parts to show the difference between the traditional approach and the proposed approach. The integration of RFE with SHAP for important feature selection stands out as a unique aspect. This method iteratively refines the feature set, enabling the identification of the most informative features, which are then used to train various ML models for comparative analysis. 70% of the data is utilized for training, with the remaining 30% being used to evaluate model performance. In this study, SHAP values are computed using CatBoost to generate a global feature importance ranking, which serves as the basis for a SHAP-based RFE procedure [42]. From the traditional approach part, CatBoost is found as the best performer; therefore, in the proposed approach, CatBoost is used as the SHAP explainer model for the LML dataset. Through rigorous cross-validation and hyper-parameter adjustment, this iterative approach validates the efficacy of the feature selection strategy.

After data cleaning and preprocessing of the dataset, which included the removal of null, inconsistent, and noisy values, along with label encoding to handle data types, seven ML regressors have been employed [43], [44]. Their ensemble combinations are also examined as well, as they potentially may integrate the strengths of individual models [45], [46]. Their ensembles, ensemble-2 (combinations of 2 models) are created by combining the best-performing individual model with one additional model (e.g., CatBoost + LightGBM), while ensemble-3 models include the best model along with two others (e.g., CatBoost + LightGBM + XGBoost). Within the ensembles, only those that outperformed the best individual regressor (CatBoost) are retained for further optimization. Algorithm 1 provides the pseudocode for the workflow of research methodology. Data analyses are conducted using Python 3.10.

A. Dataset and Feature Description

The utilized LML dataset consists of 1,021 responses to 54 survey items designed to measure various learner parameters. The data was acquired using a voluntary and anonymous online survey conducted at eleven universities in Bangladesh. Participants met two inclusion criteria: they were currently enrolled in a Bangladeshi higher education institution and were at least 18 years old. At the time of the survey design, there were 4,690,876 students enrolled in both public and private universities [24]. The survey was structured using a combination of convenience sampling and simple random sampling because it was not feasible to perform probability sampling on this large population due to logistical and resource limitations. Following the combined sampling structure, universities were chosen based on their reachability, and classrooms from various departments were chosen at random within each university. In those classrooms, all present students were verbally invited to participate in the survey anonymously and voluntarily. Table I summarizes the feature set from the LML dataset, grouped by domain, data type, and corresponding value ranges. The dataset includes features from biological, psychological, and behavioral domains, encompassing both categorical and discrete data types. The corresponding descriptive statistics of the sample data are illustrated in Fig. 2. The panel in Fig. 2(a) shows the means and standard deviations (SD) for the discrete variables: Intelligence Score, Age, and Sleep Duration. There is a wide range of intelligence scores in the sample, as observed by the mean and SD values (97.43 ± 14.35), which

Algorithm 1 ML-based human intelligence prediction using SHAP-based RFE

```
Load and preprocess data by removing missing, null, and inconsistent values and by label encoding
Initialize base regressors: Decision Tree, Random Forest, Extra Trees, Gradient Boosting, XGBoost, LightGBM, CatBoost
Split dataset into predictors (X) and target variable (y)
  Assign predictors to X and target intelligence score to y
  Split into training and test sets: X_train, X_test, y_train, y_test
for each regressor in base regressors do
    Train the regressor on X_train, y_train
    Predict on X_test
    Compute performance metrics
end for
Select the best-performing model (BestBaseModel) based on metrics
Create ensemble models by combining BestBaseModel with one and two other base regressors
for each ensemble model do
    Train on X_train, y_train
    Predict on X_test
    Compute performance metrics
    if ensemble model outperforms BestBaseModel then
        add it to the optimization list
    end if
end for
Select BestModel from BestBaseModel and models in the optimization list
Initialize SHAP explainer using BestModel and X_train
Compute SHAP values on X_train
Rank features from least to most important based on mean absolute SHAP value
for each model in the optimization list do
    Call RunSHAP_RFE(model, SHAP ranking, X_train, X_test, y_train, y_test) where:
      RunSHAP_RFE = SHAP-based Recursive Feature Elimination procedure:
          Start with all features in SHAP importance order
          Iteratively remove the least important feature(s) based on SHAP ranking
          Retrain the model on the reduced X_train
          Predict on reduced X_test
          Compute performance metrics and record performance for each feature subset size
          Keep track of the subset giving maximum accuracy and store it as the optimized model configuration
end for
Store optimized model, best feature subset, and full performance log
Plot Accuracy vs. number of SHAP-ranked features (from RFE) for each optimized model
Plot SHAP summary plot
return trained models, performance metrics, and SHAP-based explanations
```

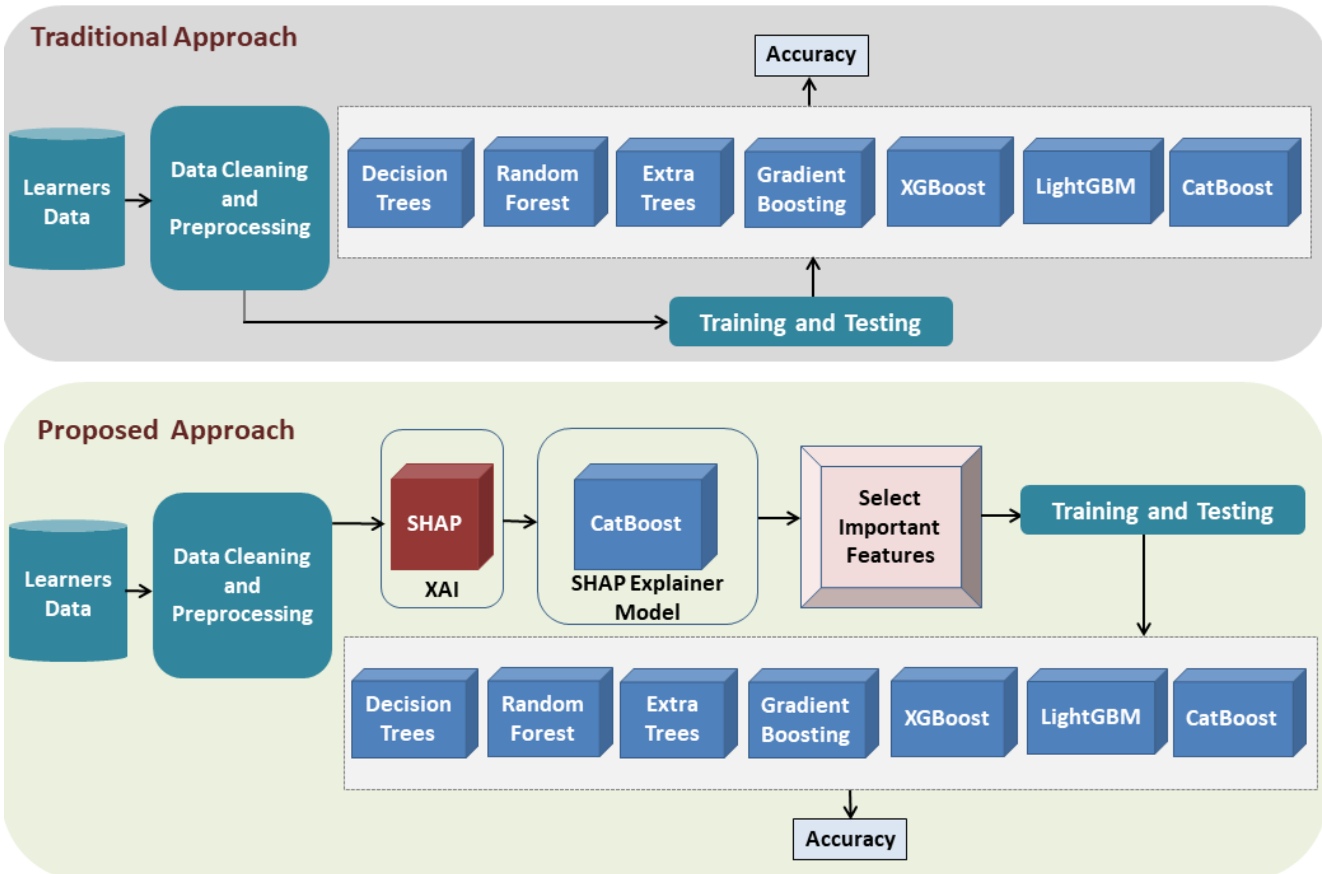


Fig. 1. Workflow of the proposed methodology.

ranges from 1 to 140. Participants' ages range from 18 to 27 years old, with an average age of 21.72 ± 1.76 years. The average Sleep Duration, which ranges from 5 to 11 hours, is 6.29 ± 1.14 hours. Fig. 2(b) panel depicts the distributions of categorical variables, Gender, Illusion of Competence, Chronotype, and Imposter Phenomenon, in terms of frequency and percentage of the total sample ($N = 1021$). The Gender distribution shows a higher proportion of male participants (638, 62.49%) than females (383, 37.51%). Although this appears skewed, it matches the existing gender disparity in higher education enrollment in Bangladesh [24].

The majority of students experienced moderate levels of Illusion of Competence (465 participants, 45.54%), followed by severe levels (315 participants, 30.85%) and mild levels (241 participants, 23.60%). The total proportion of moderate and severe levels exceeds 76%, implying that a large number of learners misestimate their competency. The Chronotype distribution is relatively balanced, with 313 participants (30.66%) identifying as Morningness types, 369 (36.14%) as Intermediates, and 339 (33.20%) as Eveningness types. This spread captures a well-represented diversity in biological rhythms, or sleep-wake preferences, among the university students. The majority of students reported that they experienced the Imposter Phenomenon on a moderate to frequent basis, with 324 participants (31.73%) and 557 participants (54.55%), respectively. This total percentage of 86.28% highlights how common impostor syndrome is among students, highlighting the significant psychological burden often associated with academic performance and self-perception.

B. ML Models

The Decision Tree regressor predicts continuous values by recursively partitioning the dataset based on feature conditions to reduce variance at each node, with final predictions based on the leaf nodes' mean values. It effectively models nonlinear relationships and is often employed in regression tasks.

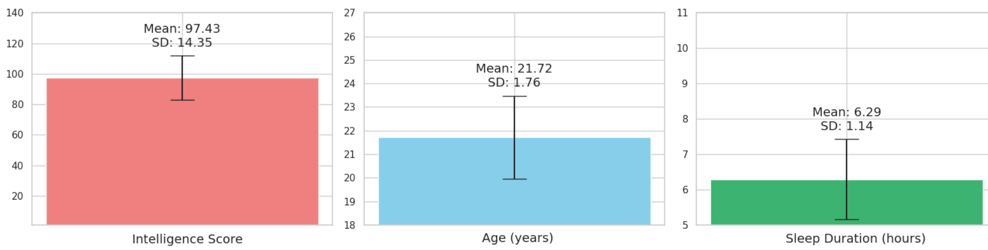
Random Forest regression improves on this by constructing an ensemble of Decision Trees, each trained on a random subset of data and features, and then averaging their outputs to reduce overfitting, boost stability, and improve robustness compared to a single tree.

The Extra Trees regressor is similar to Random Forest, but it adds more randomization by randomly generating split thresholds instead of optimizing them, which decreases variance and speeds up training while maintaining competitive accuracy.

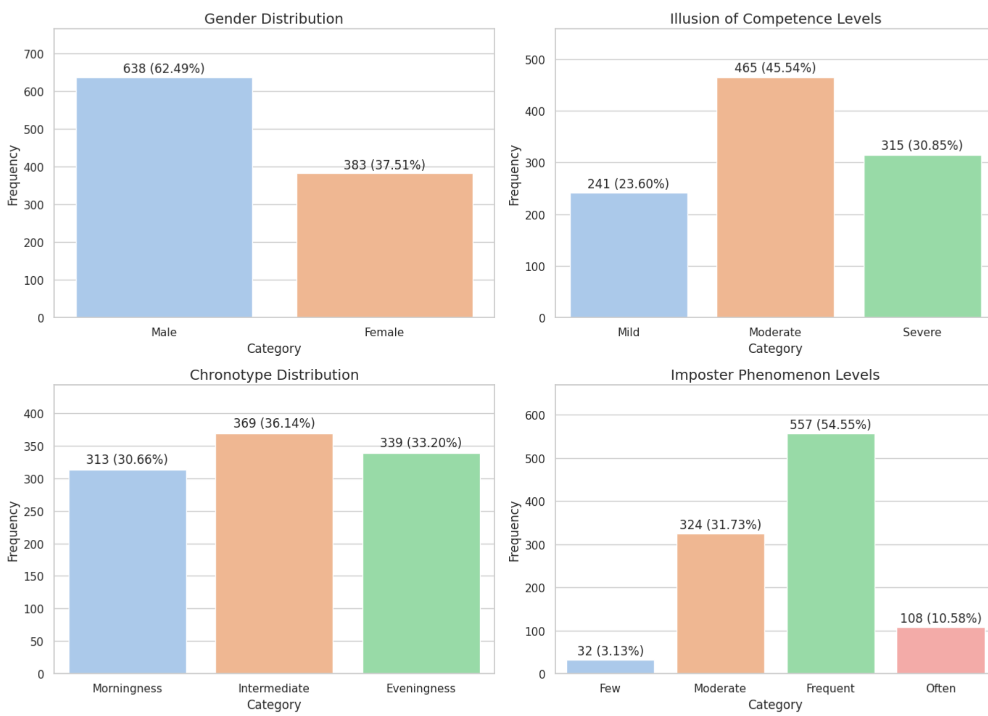
Gradient Boosting is a sequential ensemble method for creating an additive model by training every new tree to rectify the residual errors of the combined previous learners and minimizing a loss function using gradient descent.

TABLE I. Feature Set Description

Feature Group	Data Type	Feature Name	Value Range
Biological	Discrete	Age	18 – 27 years
Biological	Categorical	Gender	Male, Female
Psychological	Categorical	Illusion of Competence	Mild, Moderate, Severe
Behavioral	Discrete	Sleep Duration	5 – 11 hours
Biological	Categorical	Chronotype	Morningness, Intermediate, Eveningness
Psychological	Categorical	Imposter Phenomenon	Few, Moderate, Frequent, Often
Psychological	Discrete	Overall Intelligence Score	0 – 140



(a)



(b)

Fig. 2. Descriptive statistics of the seven learner parameters: (a) Mean and SD of continuous variables; (b) Frequency and percentage of categorical variables.

XGBoost is an effective Gradient Boosting solution that provides regularization, parallel processing, and out-of-core computation. It improves performance by using second-order gradients, efficient tree pruning, and built-in regularization to reduce overfitting. It also effectively handles missing data and produces cutting-edge results on large-scale, multidimensional datasets, making it widely utilized in prediction tasks.

LightGBM is another Gradient Boosting model that uses a histogram-based method and a leaf-wise tree development technique to improve training speed while retaining high precision. It also works well on large data sets and handles categorical features natively.

CatBoost uses ordered boosting to avoid prediction shift and efficiently handles categorical features without requiring extensive preprocessing. It is ideally suited for complicated regression problems involving heterogeneous data types due to its fast training, good generalization, and accuracy with minimal parameter adjustment.

C. Model Evaluation Metrics

To evaluate the prediction performance of the models, the following standard metrics are employed: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Accuracy [47], [48], [49].

The MAE quantifies the average magnitude of the errors in predictions relative to the actual outcomes. The formula for MAE is as follows.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

where, y_i denotes the actual value, $\hat{y}_i$ the predicted value, n the number of observations, and i indexes each observation from 1 to n.

The MSE calculates the average discrepancy between the actual values and the predictions. The process involves squaring the differences and then calculating the mean. The formula for MSE is as follows.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

RMSE is the square root of MSE, which represents the residuals' SD. It gives an interpretable measure in the same unit as the target variable. The representation is as follows.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

Accuracy is the percentage of variance explained by the model. Variance measures how far the actual values deviate from their mean. The more the model's predictions lower this spread (or error), the greater the explained variance, or the model's accuracy. A higher accuracy indicates predictions are very close to the actual values. This metric is calculated using the following equation.

$$\text{Accuracy} = \left(1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}\right) \times 100 \tag{4}$$

where, $\bar{y}_i$ denotes the mean of actual values.

IV. Results and Discussion

A. Model Performance

The performance of the regressors in the traditional approach using the complete set of features is presented in Table II. CatBoost emerged as the most effective model, attaining the highest accuracy (98.32%) and the lowest error rates across all evaluation metrics (MAE: 1.38, RMSE: 1.88), indicating its strong generalization capability and suitability. Gradient Boosting and LightGBM achieved accuracy levels of 94.96% and 94.38%, respectively. In contrast, XGBoost performed poorly in comparison to other boosting models. Extra Trees and Random Forest achieved moderate results indicating their limited predictive performance. The Decision Tree model achieved the lowest accuracy and the highest error rates, underscoring its inadequacy.

TABLE II. Model Performance with the Traditional Approaches

Model	MAE	MSE	RMSE	Accuracy (%)
CatBoost	1.38	3.55	1.88	98.32
Gradient Boosting	2.53	10.63	3.26	94.96
LightGBM	2.68	11.87	3.45	94.38
XGBoost	3.83	23.60	4.86	88.82
Extra Trees	4.39	30.44	5.52	85.58
Random Forest	4.54	32.03	5.66	84.83
Decision Tree	7.18	85.43	9.24	59.52

CatBoost, as an individual model, consistently outperformed all ensemble combinations as well, that include CatBoost combined with Decision Tree, Random Forest, Extra Trees, Gradient Boosting, XGBoost, and LightGBM (ensemble-2), and CatBoost combined with any two of the individual models (ensemble-2). Despite ensemble methods generally being expected to enhance performance by combining model strengths, in this case they introduce more weaknesses by combining model weaknesses [46]. This highlights that ensembling is not always effective, contrary to the findings reported in [45].

The accuracy curves of the regressors, presented in Fig. 3, initially stay flat or slightly rise for most of the models, indicating that the early-eliminated lower-ranked features have little contribution to prediction performance. CatBoost consistently outperforms all other models across nearly the entire range of feature numbers. The CatBoost curve peaks at 98.53% accuracy, which is obtained with 35 features. Its accuracy increases from 98.32% as the number of features is progressively decreased from 54 to 36. But after 35, the accuracy starts to drop, implying that more feature removal eliminates significant predictors. Additionally, it implies that features outside of the top 35 are redundant and have minimal contribution to the prediction.

Similar trends are observed for Gradient Boosting, LightGBM, XGBoost, Extra Trees, and Random Forest. In contrast, the Decision Tree model consistently exhibits the lowest overall accuracy. Its curve stays comparatively flat with a few oscillations.

For better interpretability, the SHAP summary plot is presented in Fig. 4. The magnitude and direction of the top features' contributions to the model's predictions are clearly shown in this plot. The feature label 'MIQ#' corresponds to specific MI-related question numbers as sequenced in the dataset. Each point on the plot represents a single observation. The horizontal position represents the SHAP value (the feature's contribution to the model output), and the colors represent the actual values of the feature. Red color means high values, and blue color means low values. The continuous gradient from blue to red graphically displays the range of feature values, making it easier to determine whether higher or lower feature values contribute positively or negatively to the predictions. For instance, for the most important feature, MIQ#32, higher feature values (red points) have positive contributions, as indicated by points located on the right side

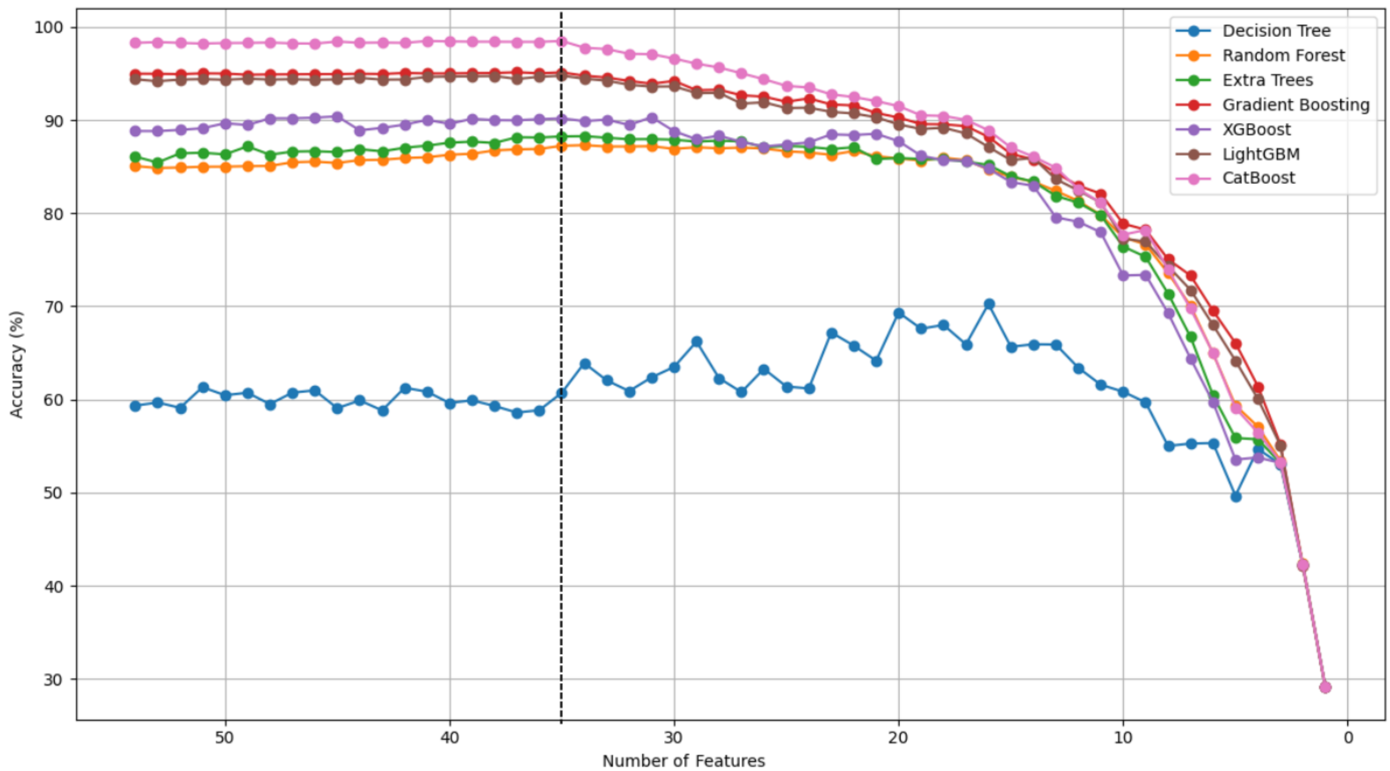


Fig. 3. Accuracy curves for all regression models.

of the axis, whereas lower values (blue points) correspond to negative contributions on the left side. Similarly, MIQ#5 and MIQ#9 display comparable trends, where higher feature values increase the predicted outcome, while lower values lessen it. MIQ#24 shows a wider distribution of SHAP values across mid- to high-range feature values (represented by purple to red points), suggesting that its influence on the prediction may vary in interaction with other variables. The SHAP values of features that are listed lower in the plot have a smaller range and variation, indicating a relatively modest, yet statistically significant, effect on the models' predictions.

Table III shows the performance metrics of the regressors following feature selection with SHAP-based RFE. When compared to the baseline results in Table II, which utilized all 54 features, Table III demonstrates consistent performance gains across all models after feature selection. CatBoost remained the top performer, increasing accuracy from 98.32% to 98.53%. Gradient Boosting's accuracy increased from 94.96% to 95.11%. The accuracy of LightGBM also increased from 94.38% to 94.76%. XGBoost, showed a greater improvement. Its accuracy increased from 88.82% to 90.19. Extra Trees and Random Forest, achieved improved accuracies from 85.58% to 88.14% and 84.83% to 87.30%, respectively. Relatively, the highest improvement is observed in the Decision Tree model. Its accuracy rose sharply from 59.52% to 60.82. Although it remains the least accurate among all models.

In summary, feature selection using SHAP-based RFE have enhanced all regressors' predictive precision (improvement in accuracy and reduction in error metrics). Notably, external learner factors showed no measurable influence on the prediction of innate intelligence scores within the current ML framework. These findings reinforce previous research, which emphasizes the use of SHAP values and RFE for improving model accuracy, transparency, and interpretability in multi-domain prediction tasks [45], [46], [49], [50].

TABLE III. Model Performance on Selected Features

Model	MAE	MSE	RMSE	Accuracy (%)
CatBoost	1.25	3.09	1.76	98.53
Gradient Boosting	2.46	10.32	3.21	95.11
LightGBM	2.55	11.06	3.33	94.76
XGBoost	3.63	20.70	4.55	90.19
Extra Trees	3.92	25.03	5.00	88.14
Random Forest	4.13	26.80	5.18	87.30
Decision Tree	7.34	82.68	9.09	60.82

B. Future Work and Applicability Scope

The findings of this study offer various potential avenues for advancement in educational design, learner assistance, and institutional policy and decision making. As presented in Fig. 5, based on the proposed intelligence prediction model, the following future research prospects and application areas are suggested for further exploration and practical implementation.

1) Data-driven admissions or placement decisions: Innate intelligence prediction could be used to complement standard admission criteria, allocate scholarships, or place students in advanced learning programs. By including a measure of learning capacity in addition to prior performance or exam results, this method would allow for more equitable selection.

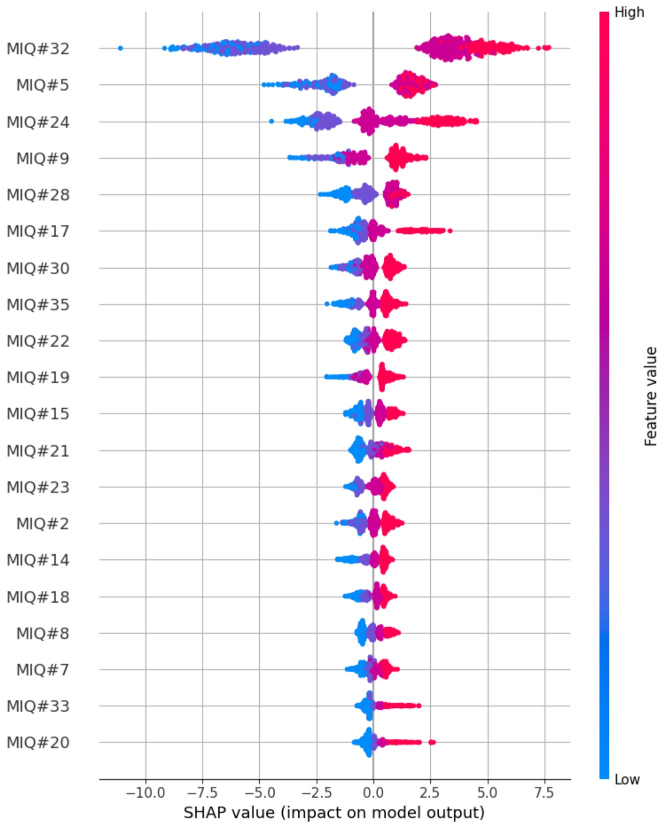


Fig. 4. Visualization of the impact of the top 20 features on the CatBoost model using SHAP.

2) *Adaptive institutional practice and policy:* The predicted intelligence score can serve as a good indicator of a learner's overall ability to receive, process, and apply newly acquired knowledge. Institutions could use it to forecast general academic performance or learning outcomes and group learners to allocate mentoring or resources accordingly.

3) *Personalized teaching strategy design:* Intelligence scores can be used as a basis for personalizing teaching strategies. In order to achieve equitable results, learners with higher intelligence scores can be given accelerated modules or enrichment activities; those with medium intelligence scores can benefit from regular instruction with occasional support; and those with lower intelligence scores can be assisted with structured and practice-based learning strategies.

4) *Strategic curriculum planning:* At the course or program level, average intelligence scores across groups of students could be used to inform syllabus pace and complexity design. For instance, to better match students' innate abilities, educational institutions may design differentiated tracks (basic, intermediate, and advanced).

5) *Intelligent tutoring systems' personalization:* Learners' intelligence scores could be used by intelligent tutoring system platforms to adapt the way content is delivered. For example, the complexity of challenges or tasks, variation in examples, and the use of simplified instructions according to learners' ability ranges can be incorporated. This would improve learning efficiency and synchronize involvement in educational settings.

6) *Learning analytics dashboards for monitoring and planning:* Innate intelligence can be integrated with learning analytics systems to deliver actionable and anticipatory information. Based on students' intelligence scores, these dashboards may provide focused intervention plans, specialized support services, or alternate educational techniques.

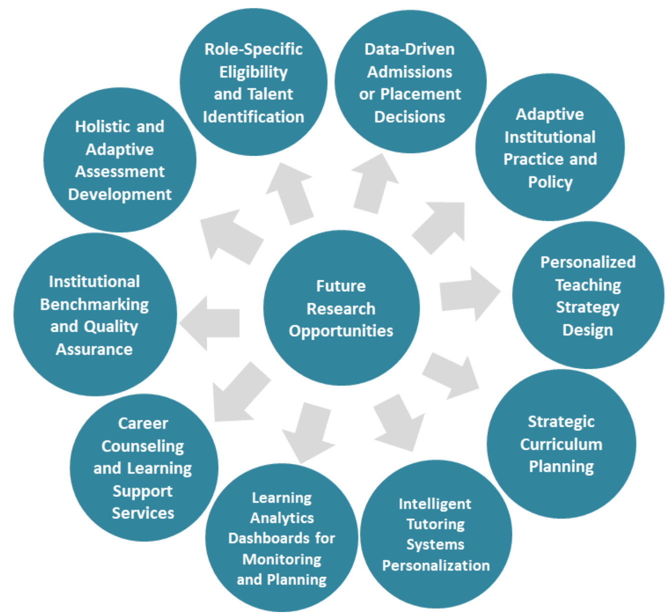


Fig. 5. Potential application areas of ML-based human intelligence modeling.

7) *Career counseling and learning support services:* Counselors can use intelligence scores to help students choose occupation possibilities that are compatible with their cognitive strengths. Students with high intelligence scores, for example, might do well in multidisciplinary studies, professions demanding quick decisions, or research-oriented positions requiring critical thinking.

8) *Institutional benchmarking and quality assurance:* Institutions could analyze the distribution of intelligence scores among the students to determine whether instruction standards, resource allocation, or students' learning outcomes are in line with learner potential. This can assist with accreditation, policy adjustments, and educational reform.

9) *Holistic and adaptive assessment development:* Knowing a student's intrinsic intelligence would allow teachers to apply assessments other than written tests. For instance, integrative projects could be used to assess students with high innate intelligence, while assessments for other students might use more scaffolding or different formats, like hands-on activities. While this study predicts intelligence as a composite indication, subsequent future research will concentrate on predicting intelligence type-wise scores independently in order to better aid specialized applications.

10) *Role-specific eligibility and talent identification:* Overall intelligence scores, like CGPA in competitive fields, could be used as benchmarks for roles requiring cognitive rigor. For example, programs may set minimum intelligence score requirements in place of or beside the CGPA. Likewise, just

as subject-wise grades determine eligibility for certain careers, intelligence-type specific scores could inform role suitability. For instance, military or sports training institutes might require high bodily-kinesthetic and logical-mathematical intelligences. These intelligence scores could be used as minimum marks or GPA in specific subjects (e.g., needing a minimum GPA in Math and Physics) for eligibility.

Along with these application areas and prospects, this research has some limitations that provide more scope for future work. The LML dataset utilized in this study only includes data from students in the age group of 18 to 27. Therefore, the results may not be directly applicable to pedagogical (child-centered) applications and are most relevant to andragogy and heutagogy. Future research could include data from a wider range of age groups, cultural backgrounds, geographic regions, and educational systems to broaden the generalizability of the findings. Furthermore, while this study used seven well-known ML models with SHAP and RFE, these are not the only potential models and methods. Subsequent investigations may examine more advanced models and approaches like deep neural networks, graph-based models, or hybrid frameworks to explore more about the results and possibly enhance interpretability. A direct comparison with similar approaches within a shared evaluation framework was not possible to conduct due to the absence of comparable methods, but can be performed once such methods become available.

V. Conclusion

In conclusion, this study demonstrated the effectiveness of ML models, enhanced by SHAP-based feature analysis, in accurately predicting innate human intelligence. The findings provide interpretable and actionable insights, paving the way for the development of personalized educational strategies, advancing educational AI practices, and enabling policy-relevant interventions aligned with the fourth SDG: Quality Education. In particular, they highlight that intelligence can be modeled holistically and reliably without relying on conventional subject-specific assessments or IQ tests. Real-world educational outcomes can be enhanced by complementing existing educational practices with the proposed method for predicting learners' intelligences. This study has implications for educators, institutions, policymakers, and future AI-driven educational systems, and promotes the use of AI-powered, data-driven, learner-centric techniques to address learner diversity and support educational decision making.

Funding

This research was funded by Multimedia University, Cyberjaya, Selangor, Malaysia (Grant Number: Post-Doc(MMUI/240029)).

Institutional Review Board Statement

The study is conducted in accordance with relevant guidelines and national regulations. The Biosafety, Biosecurity, and Ethical Clearance Committee at Jahangirnagar University in Bangladesh provided ethical approval for the creation and analysis of the LML dataset involving human data (Reference No.: BBEC, JU/M 2022/01 (18); Date of approval: 04 July 2022). Since the dataset contains non-medical data, the Declaration of Helsinki does not apply when utilizing it for non-medical purposes.

Informed Consent Statement

All participants were well informed of their anonymity, the goal of the study, and how their responses would be utilized. Survey responses were collected using a Google form. Consent to participate in the survey and consent to publish anonymized data were obtained using the same Google form. After reading the information consent form, participants had to agree to participate before they could continue on to the survey. Before submitting the completed questionnaire, participants were requested for their further approval to have their anonymized data published collectively for research purposes via an online platform. Only those who selected "Agree" for publishing consent were able to submit survey responses; no information was collected for those who did not give both consents.

Data Availability Statement

The data that support the findings of this study are openly available in the Zenodo repository at zenodo.org/records/8382187.

References

[1] J. S. Chall, *The academic achievement challenge: what really works in the classroom?*, Guilford Press, New York London, 2000.

[2] Y. Jiang and J.-E. Peng, "Exploring the relationships between learners' engagement, autonomy, and academic performance in an English language MOOC," *Computer Assisted Language Learning*, pp. 1–26, 2023, doi:10.1080/09588221.2022.2164777.

[3] G. Kuh, "What Matters to Student Success," The University of British Columbia, 2009, doi:10.14288/1.0228432.

[4] L. M. Nunn, *Defining Student Success: The Role of School and Culture*, Rutgers University Press, 2019, doi:10.36019/9780813563633.

[5] Z. Swiecki *et al.*, "Assessment in the age of artificial intelligence," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100075, 2022, doi:10.1016/j.caeai.2022.100075.

[6] K. M. Pitychoutis and A. Al Rawahi, "Smart Teaching: The Synergy of Multiple Intelligences and Artificial Intelligence in English as a Foreign Language Instruction," *Forum Linguist. Stud.*, vol. 6, no. 6, pp. 249–260, Dec 2024, doi:10.30564/fls.v6i6.7297.

[7] D. Zalesne and D. Nadvorney, "Why Don't They Get It?: Academic Intelligence and the Under-Prepared Student as 'Other'," 2000.

[8] R. Al-Ali, K. Alhumaid, M. Khalifa, S. A. Salloum, R. Shishakly, and M. A. Almaiah, "Analyzing Socio-Academic Factors and Predictive Modeling of Student Performance Using Machine Learning Techniques," *Emerg Sci J*, vol. 8, no. 4, pp. 1304–1319, Aug 2024, doi:10.28991/esj-2024-08-04-05.

[9] M. N. Yakubu and A. M. Abubakar, "Applying machine learning approach to predict students' performance in higher educational institutions," *K*, vol. 51, no. 2, pp. 916–934, Feb 2022, doi:10.1108/k-12-2020-0865.

[10] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Educ Inf Technol*, vol. 28, no. 1, pp. 905–971, Jan 2023, doi:10.1007/s10639-022-11152-y.

[11] M. Bilal, M. Omar, W. Anwar, R. H. Bokhari, and G. S. Choi, "The role of demographic and academic features in a student performance prediction," *Sci Rep*, vol. 12, no. 1, Jul 2022, doi:10.1038/s41598-022-15880-6.

[12] S. Mohd Zaki, S. Razali, M. A. R. Awang Kader, M. Z. Laton, M. Ishak, and N. Mohd Burhan, "Predicting students' performance at higher education institutions using a machine learning approach," *K*, Aug 2024, doi:10.1108/k-12-2023-2742.

[13] T. Wilkinson, M. Boohan, and M. Stevenson, "Does learning style influence academic performance in different forms of assessment?," *Journal of Anatomy*, vol. 224, no. 3, pp. 304–308, Mar 2014, doi:10.1111/joa.12126.

[14] H. Gardner (ed.), *Frames of mind: the theory of multiple intelligences*, Basic Books, New York, NY, 2011.

[15] H. Gardner, *Intelligence reframed: multiple intelligences for the 21st century*, Basic Books, New York, 1999.

[16] K.-C. Cheung, "Reforming Teaching and Learning Using Theory of Multiple Intelligences: The Macao Experiences," in *Reforming Learning*, C. Ng and P. D. Renshaw, Eds., vol. 5, Dordrecht: Springer Netherlands, pp. 159–179, 2009, doi:10.1007/978-1-4020-3024-6_8.

[17] R. J. Sternberg and E. L. Grigorenko, *Teaching for successful intelligence: to increase student learning and achievement*, 2nd ed., Corwin Press, Thousand Oaks, CA, 2007.

[18] C. A. Tomlinson, *Assessment and student success in a differentiated classroom*, ASCD, Alexandria, Va, 2013.

[19] M. Weatherton and E. E. Schussler, "Success for All? A Call to Re-examine How Student Success Is Defined in Higher Education," *LSE*, vol. 20, no. 1, p. es3, Mar 2021, doi:10.1187/cbe.20-09-0223.

[20] J. W. Santrock, *Educational psychology*, 5th ed., McGraw-Hill Education, New York, NY, 2011.

[21] B. Berrios Aguayo, C. Arazola Ruano, and A. Pantoja Vallejo, "Multiple intelligences: Educational and cognitive development with a guiding focus," *SAJE*, vol. 41, no. 2, pp. 1–10, May 2021, doi:10.15700/saje.v41n2a1828.

[22] Baanurekha K Ravihchandran and Hasnah Binti Toran, "MULTIPLE INTELLIGENCE THEORY BY HOWARD GARDNER AND TEACHING STRATEGIES FOR ADHD," *SE*, vol. 3, no. 1, p. e0044, Mar 2025, doi:10.59055/se.v3i1.44.

[23] S. Corraya, S. A. Mamun, and M. S. Kaiser, "LML-Learning Meta-Learning Dataset_update version," Zenodo, Jul 2023, doi:10.5281/ZENODO.8382187.

[24] S. Corraya, S. A. Mamun, and M. S. Kaiser, "Learning Meta-Learning (LML) dataset: Survey data of meta-learning parameters," *Data in Brief*, vol. 51, p. 109777, Dec 2023, doi:10.1016/j.dib.2023.109777.

[25] A. Kumar, T. DiJohnson, R. Edwards, and L. Walker, "The application of adaptive minimum match k-nearest neighbors to identify at-risk students in health professions education," 2021, doi:10.48550/ARXIV.2108.07709.

[26] S. Salehi *et al.*, "Context Matters: Social Psychological Factors That Underlie Academic Performance across Seven Institutions," *LSE*, vol. 20, no. 4, Dec 2021, doi:10.1187/cbe.21-01-0012.

[27] Q. Jiang, H. Horta, and M. Yuen, "High- and low-achieving international medical students' perceptions of the factors influencing their academic performance at Chinese universities," *Medical Education Online*, vol. 29, no. 1, Dec 2024, doi:10.1080/10872981.2023.2300194.

[28] V. Bhardwaj, S. Zhang, Y. Q. Tan, and V. Pandey, "Redefining learning: student-centered strategies for academic and personal growth," *Front. Educ.*, vol. 10, Feb 2025, doi:10.3389/feduc.2025.1518602.

[29] "The Learner-Centered Model: From the Vision to the Future," in *Interdisciplinary Applications of the Person-Centered Approach*, New York, NY: Springer New York, pp. 83–113, 2013, doi:10.1007/978-1-4614-7144-8_9.

[30] D. Yeh *et al.*, "Cognitive impairment in the schizophrenia spectrum: exploring the relationships of the g-factor with sociodemography, psychopathology, neurodevelopment, and genetics," *Psychol. Med.*, vol. 54, no. 15, pp. 4243–4254, Nov 2024, doi:10.1017/s0033291724002538.

[31] L. Anunciação *et al.*, "Psychometric properties of a brief non-verbal test of g factor intelligence," *J. Psychol. Perspect.*, vol. 6, no. 1, pp. 47–62, May 2024, doi:10.47679/jopp.617532024.

[32] L. V. Shavinina, "Beyond IQ: a new perspective on the psychological assessment of intellectual abilities," *New Ideas in Psychology*, vol. 19, no. 1, pp. 27–47, Apr 2001, doi:10.1016/s0732-118x(00)00015-5.

[33] "Logical fallacies used to dismiss the evidence on intelligence testing," in *Correcting fallacies about educational and psychological testing*, Washington: American Psychological Association, pp. 11–65, doi:10.1037/11861-001.

[34] K. E. Stanovich (ed.), *What intelligence tests miss: the psychology of rational thought*, Yale University Press, New Haven, 2009.

[35] "IQ, EQ, and Multiple Intelligences: A Brief Review of the Discussion," in *The Oxford Handbook of Individual Differences in Organizational Contexts*, 1st ed., Oxford University Press, pp. 53–63, 2024, doi:10.1093/oxfordhb/9780192897114.013.23.

[36] "A Brief History of IQ Testing: Fixed vs. Malleable Intelligence," in *Intelligence in Context*, Cham: Springer International Publishing, pp. 59–92, 2022, doi:10.1007/978-3-030-92798-1_4.

[37] L. S. Almeida, M. D. Prieto, A. I. Ferreira, M. R. Bermejo, M. Ferrando, and C. Ferrándiz, "Intelligence assessment: Gardner multiple intelligence theory as an alternative," *Learning and Individual Differences*, vol. 20, no. 3, pp. 225–230, Jun 2010, doi:10.1016/j.lindif.2009.12.010.

[38] N. Goyibova, N. Muslimov, G. Sabirova, N. Kadirova, and B. Samatova, "Differentiation approach in education: Tailoring instruction for diverse learner needs," *MethodsX*, vol. 14, p. 103163, Jun 2025, doi:10.1016/j.mex.2025.103163.

[39] A. Alam, "Improving Learning Outcomes through Predictive Analytics: Enhancing Teaching and Learning with Educational Data Mining," in *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, pp. 249–257, May 2023, IEEE, doi:10.1109/iciccs56967.2023.10142392.

[40] G. J. Maulany, P. I. Santosa, and I. Hidayah, "Multiple Intelligence Learning Style Detection in E-learning using Support Vector Machine (SVM) Algorithm," in *2024 7th International Conference on Informatics and Computational Sciences (ICICoS)*, Semarang, Indonesia, pp. 250–255, Jul 2024, IEEE, doi:10.1109/icicos62600.2024.10636915.

[41] 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," *Decision Analytics Journal*, vol. 7, p. 100204, Jun 2023, doi:10.1016/j.dajour.2023.100204.

[42] Y. Ji, H. Shang, J. Yi, W. Zang, and W. Cao, "Machine learning-based models to predict type 2 diabetes combined with coronary heart disease and feature analysis-based on interpretable SHAP," *Acta Diabetol*, Apr 2025, doi:10.1007/s00592-025-02496-1.

[43] K. Ileri, "Comparative analysis of CatBoost, LightGBM, XGBoost, RF, and DT methods optimised with PSO to estimate the number of k-barriers for intrusion detection in wireless sensor networks," *Int. J. Mach. Learn. & Cyber.*, May 2025, doi:10.1007/s13042-025-02654-5.

[44] M. Begum, M. H. Shuvo, and J. Uddin, "MLRec: A Machine Learning-Based Recommendation System for High School Students Context of Bangladesh," *Information*, vol. 16, no. 4, p. 280, Mar 2025, doi:10.3390/info16040280.

[45] M. F. Mahmoud, M. Arabi, and S. Pallickara, "Harnessing ensemble Machine learning models for improved salinity prediction in large river basin scales," *Journal of Hydrology*, vol. 652, p. 132691, May 2025, doi:10.1016/j.jhydrol.2025.132691.

[46] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, 0th ed., Chapman and Hall/CRC, 2012, doi:10.1201/b12207.

[47] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning: with applications in R*, 2nd ed., Springer texts in statistics, Springer, New York, 2021.

[48] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*, vol. 7, p. e623, Jul 2021, doi:10.7717/peerj-cs.623.

[49] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," 2012, doi:10.48550/ARXIV.1201.0490.

[50] A. M. Salih *et al.*, "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," *Advanced Intelligent Systems*, vol. 7, no. 1, p. 2400304, Jan 2025, doi:10.1002/aisy.202400304.