Evaluating the Effectiveness of a Machine Learning–Based Approach for Detecting Learning Disabilities in Elementary Through Game Interactions

Arlah F. Licay
Southern Leyte State University – Tomas Oppus Campus, Philippines
Email: arlahfrenchlicay@gmail.com

Efren I. Balaba*
Southern Leyte State University – Tomas Oppus Campus, Philippines
Email: ebalaba@southernleytestateu.edu.ph
*Corresponding Author

Received: *01/12/2025* ***Revised:*** *19/12/2025* ***Accepted:*** *19/12/2025*

Abstract

This study explores the use of machine learning (ML) to assist in identifying potential learning disabilities in elementary students by analyzing their interactions with educational games. While traditional approaches such as teacher observations and standardized assessments are valuable, they may take time and occasionally overlook early conditions like dyslexia, ADHD, or dyscalculia. By examining gameplay data including response time, accuracy, and error patterns, the study applies supervised ML models like Decision Trees and Support Vector Machines to classify students' learning needs. Findings suggest that ML tools could offer meaningful support to educators in recognizing students who may benefit from early intervention.

Keywords: Machine Learning, Learning Disabilities, Educational Games, Early Detection, Student Behavior.

Introduction

Learning disabilities, such as dyslexia, ADHD, and dyscalculia, can significantly influence the learning and performance of children in elementary schools. These disabilities may lead to problems in reading, attention, and reasoning, among others, thus making it difficult for the children to achieve the desired level in their classes (American Psychiatric Association, 2013; Fletcher et al., 2018). Therefore, it is very important to diagnose the condition at the earliest because, in this case, the interventions will be more effective as the support is provided before the learning gaps become very wide. Unfortunately, the most common method of diagnosing learning disabilities is still very much dependent on teacher observation and the use of standardized tests which take long to yield results and may not be so accurate in detecting early or mild cases of learning difficulties (Kovas et al., 2007).

In the Philippines, the detection of learning disabilities still poses a significant problem. Several public schools are deprived of the necessary tools for formal screening, and the teachers' opinions often are the basis for the evaluation. The Department of Education (DepEd) reports that the screening of dyslexia, dyscalculia, and ADHD and other such disorders is still in its infancy, which causes delays in the giving of proper treatment. This absence of screening can result in the students concerned becoming less engaged in class and performing worse academically.

Machine learning (ML) has been proposed by recent studies as a way to mitigate this situation by helping in the systematic identification of learning difficulties based on the analysis of students' gameplay behavior. It has been reported that ML models are capable of pinpointing the presence of learning challenges in their very early stages through the application of pattern recognition of students' behavior and the monitoring of performance indicators (Khan et al., 2020; Hernández-Blanco & Herrera-Flores, 2019). The field of research on educational games also stresses the aspect of gaming being one of the revealing factors of students' thinking processes, strategizing, and even their way of responding to assignments (Chung, 2014; Geary, 2011). Even though the aforementioned circumstances are present and the knowledge is out there, only a few studies have applied and/or combined ML with game-based learning specifically to identify learning disabilities of very young learners (Luckin et al., 2016).

One limitation of previous research is that many studies focus on only a few types of learning disabilities or rely on general detection methods that may not be suitable for younger children. Additionally, real-time data from gameplay is rarely integrated with ML models to support early and accurate detection of learning challenges.

This study aims to address these gaps by using ML to analyze how elementary students interact with educational games. By examining features such as response time, accuracy, decision-making patterns, and error behaviors, this research explores how ML can help teachers and specialists detect learning disabilities earlier. This approach seeks to make early intervention more efficient, support inclusive education, and provide students with the assistance they need to succeed.

Objectives

To identify gameplay behavior patterns linked to learning disabilities.
To use existing ML tools on publicly available or simulated data.
To evaluate how well ML detects signs of learning difficulties.

Methodology

Variables and Sampling Design

This study uses an exploratory quantitative research design to investigate how machine learning (ML) can detect potential learning disabilities among elementary students based on their educational gameplay behavior. To avoid ethical concerns associated with data collection from young children, the study uses a publicly available secondary dataset the Museum Game Interaction Dataset (Ziya07, 2023) from Kaggle. This dataset contains recorded gameplay interactions of children completing tasks within an educational game environment.

The dataset includes behavioral variables that research has associated with underlying cognitive processes, such as accuracy, reaction time, mistake patterns, and decision delays (Munshi & Mohan, 2017). These indicators were selected for their relevance to learning behaviors and because they align well with ML techniques commonly applied in learning disability detection, including Decision Trees and Support Vector Machines (Khan et al., 2020).

Table 1: Table: Definitions of Learning Behavior and Performance Metrics

Feature	Description
Reaction Time	Average time taken by the student to respond to actions
Correct Responses	Frequency of accurate actions performed
Incorrect Responses	Number of errors committed
Time Spent	Total time the student spent completing the activity
Total Actions	Total number of gameplay interactions
Hint Usage	Number of times hints were used
Focus Duration	Length of time attention was sustained
Eye Tracking	Visual focus and movement patterns

Instrument Design

To ensure consistency and improve model performance, several preprocessing steps were applied:

- *Handling Missing Values*
  Missing or incomplete entries were removed or replaced using mean substitution.
- *Normalization*
  All numerical features were scaled using Min–Max normalization to prevent large-value features from dominating the ML model.
- *Class Combination*
  The dataset originally contained multiple student categories. Students labeled Struggling and Disengaged were merged into a single class to address imbalance and simplify classification.
- *Train–Test Split*

Data was divided into 80% training and 20% testing, ensuring the model could be evaluated on unseen data.

These preprocessing steps help standardize the dataset and improve the reliability of ML predictions (Cheng & Tsai, 2019).

Statistical Analysis

Two supervised ML models were selected based on their common use in educational and behavioral classification studies (Khan et al., 2020):

- Decision Tree Classifier
- Support Vector Machine (SVM)

The models were implemented using Python and the Scikit-learn library. Each model was trained using the standardized features from the dataset, with the goal of classifying students into either Struggling or Not Struggling categories. Model hyperparameters were tuned through iterative adjustments to optimize classification accuracy.

Model Evaluation

To assess the performance and reliability of the ML models, the following evaluation metrics were used:

Accuracy – percentage of correctly predicted labels
Precision – reliability of positive predictions
Recall – ability to identify struggling students
F1-Score – balance between precision and recall

Additionally, 5-fold cross-validation was applied to verify that each model performed consistently across different data subsets. Sensitivity checks were also performed by slightly modifying the dataset to ensure that model predictions remained stable, improving robustness and reducing the risk of overfitting (Baker & Inventado, 2014).

Research Results

This study explored how machine learning (ML) can help identify learning difficulties among elementary students by analyzing their gameplay behavior using the Museum Game Interaction Dataset (Ziya07, 2023). Students labeled as “Struggling” and “Disengaged” were combined into one class to simplify classification. Two supervised ML models Decision Tree and Support Vector Machine (SVM) were used to classify student performance based on features such as reaction time, total actions, hint usage, and error rates.

Table2: Performance comparison between Decision Tree and SVM models.

Decision Tree Results:				
	precision	recall	f1-score	support
0	0.27	0.33	0.30	18
1	0.93	0.91	0.92	182
accuracy			0.86	200
macro avg	0.60	0.62	0.61	200
weighted avg	0.87	0.86	0.87	200

SVM Results:				
	precision	recall	f1-score	support
0	0.23	0.89	0.37	18
1	0.98	0.71	0.82	182
accuracy			0.72	200
macro avg	0.61	0.80	0.60	200
weighted avg	0.92	0.72	0.78	200

The Decision Tree model achieved a higher overall accuracy (86%) and produced a stronger F1-score for Class 1 (Struggling/Disengaged). This indicates that the Decision Tree was better at detecting students who showed signs of difficulty. However, it performed poorly for Class 0 (Not Struggling), with lower precision and recall. This suggests the model was biased toward identifying struggling learners and less capable of recognizing students who were performing well. These results are consistent with findings from Khan et al. (2020), who noted that tree-based models often work well in identifying irregular learning patterns but may overfit to one class.

In contrast, the SVM model achieved a lower accuracy (72%) but performed better in identifying non-struggling students, with a recall of 0.89 for Class 0. This suggests that SVM generalized better for students without difficulties but struggled to correctly classify those who needed additional support. This aligns with Baker & Inventado (2014), who observed that linear-based models often generalize well for stable learner behaviors but may miss subtle signs of struggle.

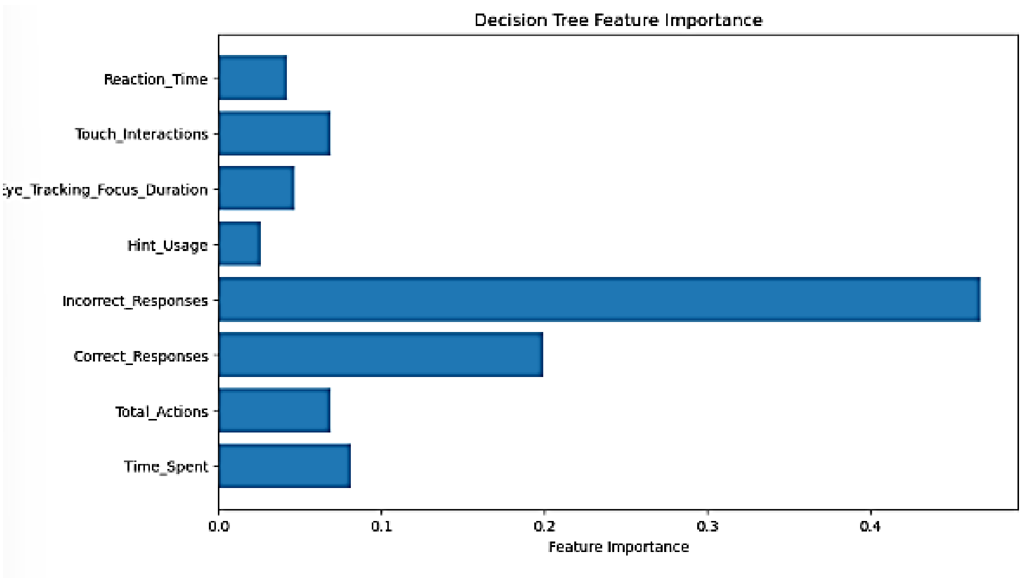

Figure 1. Feature importance from the Decision Tree model.

The feature importance results showed that Incorrect Responses was the most influential predictor of learning difficulty. Students who frequently responded incorrectly were more likely to be classified as struggling. This supports previous research by Chung (2014), who noted that error patterns often reflect underlying cognitive challenges. Correct Responses, Time Spent, and Total Actions also contributed significantly to classification, suggesting that both performance and behavioral engagement play important roles in identifying learning difficulties.

In contrast, features such as Hint Usage, Eye Tracking, Focus Duration, and Reaction Time showed lower importance. While these features may still be meaningful, their lower contribution suggests they require deeper contextual analysis or more refined measurement to be effective indicators of learning difficulty.

Model Trade-offs

- Decision Tree:
  - Strong at detecting struggling learners
  - Weaker with non-struggling learners
  - More interpretable but prone to class bias
- SVM:
  - Better at detecting stable or non-struggling patterns
  - Weaker at identifying struggling students
  - More balanced but less sensitive to irregular behaviors

Practical Implications

The results suggest that ML techniques can be integrated into educational games to help teachers identify students who may benefit from early intervention. A Decision Tree may be more suitable when the goal is to flag struggling students as early as possible, while an SVM may be better for monitoring general learner performance. These insights can support more informed decisions in the classroom, helping teachers provide targeted support and make learning more inclusive.

Discussion

The machine learning (ML) models used in this study have assisted to uncover the early signs of learning troubles of the elementary school students through their playing style. The Decision Tree and SVM models were the two choices that proved their worth differently but still had their own pros and cons. The Decision Tree model was able to pinpoint the students who were having difficulties, and this is a very helpful feature for early detection and intervention. But, on the other side, the model through its detection presented a bias of struggling students. It has been shown that the Decision Tree model's performance is indeed influence by the class imbalance that Baker and Inventado (2014) noted in their work that tree-based models can easily overfit to irregular behavioral patterns when class imbalance exists.

On the flip side, the overall accuracy of SVM model was lower but it was better at spotting non-difficult students. This means that SVM is good at generalizing for stable learner behaviors but might not be able to detect the very subtle patterns that are connected with learning problems. This is in agreement with Khan et al. (2020) who pointed out that linear-based models have better precision in separating consistent behavior but may falter in high variability data cases.

Furthermore, the feature importance analysis has deepened the results interpretation. Among all the factors, the Incorrect Responses stood out as the principal indicator of learning difficulty, thus confirming the idea that error patterns convey the already mentioned (cognitive) processes (Chung, 2014). Input from features like Correct Responses, Time Spent, and Total Actions was also considerable, thus the involvement of accuracy and even the engagement behavior are through the indicators of importance. But the features like Hint Usage, Eye Tracking, and Focus Duration showed lesser importance and thus the need

Conclusion and Suggestions

The research has shown that machine learning (ML) can be used to spot early signs of learning problems in elementary school kids based on their educational game play behavior. The study presented the two ML models Decision Tree and SVM- and compared their power in classifying the students as either struggling or not struggling by analyzing features such as the number of wrong answers given, response time, and total actions taken.

The Decision Tree model produced more correct predictions and proved to be more effective in spotting the struggling students, thus making it very appropriate for early detection and intervention. On the other hand, the SVM model was less precise in total accuracy but nonetheless revealed stronger results in recognizing the non-struggling students. These opposing findings point to the fact that ML-based techniques can help educate the teachers by providing them with data-driven perspectives on the learning behavior of the students.

Furthermore, the research demonstrates the potential of using game interactions as an early behavioral indication of possible cognitive difficulties. These difficulties might include disorders like dyslexia, ADHD, or dyscalculia. Moreover, the incorporation of ML tools in educational games would enable educators to receive immediate warnings and thus deal with the learning difficulties in time before they become more severe.

It is advisable that the future studies aimed at developing the models' generalization power should consider working with more diverse datasets that include live classroom data. Besides, the use of multimodal data like eye-tracking, keystroke patterns or adaptive gameplay feedback could contribute to the establishment of more accurate and personalized detection systems.

Limitations

This study is limited by the nature of the dataset used. The dataset contained a class imbalance, with more struggling students than non-struggling ones, which affected model fairness and contributed to the Decision Tree's bias toward Class 1 (Khan et al., 2020). Future studies should explore balanced datasets or apply techniques such as oversampling or class-weight adjustments.

Another limitation is that the dataset was secondary and did not include real-time classroom data. Important contextual factors such as students' emotional state, motivation, prior knowledge, or environmental distractions were not captured. These real-world factors may influence gameplay behavior and affect model predictions.

Additionally, the behavioral features available were limited to in-game actions, which may not fully represent the complexity of each student's learning process. Features such as eye tracking and focus duration also had low importance scores, possibly due to inconsistent data quality or lack of contextual interpretation.

Lastly, the ML models were tested on a single dataset, which limits generalizability. Different educational games or learning tasks may produce different patterns of behavior. Future studies should validate multiple ML models across diverse datasets to strengthen the reliability and practical utility of ML-based early detection systems

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). *American Psychiatric Publishing*. https://doi.org/10.1176/appi.books.9780890425596

Baker, R. S. J. d., & Inventado, P. S. (2014). Educational data mining and learning analytics.

In J. A. Larusson & B. White (Eds.), Learning analytics (pp. 61–75). *Springer*. https://doi.org/10.1007/978-1-4614-3305-7_5

Cheng, L., & Tsai, C.-C. (2019). The interaction of child characteristics and learning environment in the development of early literacy. *Journal of Educational Psychology*, 111(3), 481–492. https://doi.org/10.1037/edu0000305

Chung, G. K. W. K. (2014). Toward a framework for game-based assessment: Opportunities and challenges for future research. *ETS Research Report Series*, 2014(2), 1–18. https://doi.org/10.1002/ets2.12028

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2018). Learning disabilities: From identification to intervention (2nd ed.). *Guilford Press*.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, 47(6), 1539–1552. https://doi.org/10.1037/a0025510

Hernández-Blanco, A., & Herrera-Flores, B. (2019). A systematic review of deep learning approaches to educational data mining. *Complexity*, 2019, Article 492564. https://doi.org/10.1155/2019/492564

Khan, S., Nabi, M. K., Ahmad, M., & Almehmadi, M. (2020). Machine learning-based approaches for predicting learning disabilities: A survey. *IEEE Access*, 8, 137918–137938. https://doi.org/10.1109/ACCESS.2020.3011741

Kovas, Y., Haworth, C. M. A., Dale, P. S., & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*, 72(3), vii–144. https://doi.org/10.1111/j.1540-5834.2007.00453.x

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. *Pearson Education*. https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/Intelligence-Unleashed-Publication.pdf

Munshi, A., & Mohan, V. (2017). Data mining techniques for early diagnosis of learning disabilities. *Procedia Computer Science*, 115, 547–554. https://doi.org/10.1016/j.procs.2017.09.139

Ziya07. (2023). Museum game interaction dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/ziya07/museum-game-interaction-dataset