A 3-Tier machine Learning Framework for Early Detection of Learning Difficulties in Basic setting

Samuel Odoom
Mampong Technical College of Education
Eric Opoku Osei
Kwame Nkrumah University of Science and Technology
Enock Quansah Effah
songreen18@gmail.com
Wiawso College of Education
Victoria Boafo
Mampong Technical College of Education
Seyram Dusu
Mampong Technical College of Education
Anthony Kweku Appiah
Mampong Technical College of Education

Research Article

Keywords: Learning Difficulties, Machine Learning in Education, Early Detection, Educational Data Mining, Ensemble Learning, Learning Analytics.

Posted Date: November 24th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7876870/v1

Additional Declarations: No competing interests reported.

Abstract

Early identification of learning difficulties (LDs) in elementary education is crucial for timely intervention, however, conventional approaches relying on subjective teacher assessments often result in delayed or inequitable support, particularly affecting underserved populations. This study proposes and evaluates a scalable machine learning (ML) based framework to support the early identification of learning difficulties (LDs) among upper primary school learners in Ghana's basic education system. Data were collected from 1,124 learners across public schools over two academic years (2021–2023), encompassing psychometric, behavioural, and demographic variables. Using validated instruments such as the Learning Difficulties Checklist (LDC), Conners' Teacher Rating Scale (CTRS), and WRAT subtests, learners were categorized into low, moderate, and high LD risk groups based on the composite score thresholds established through statistical analysis and expert input. A hybrid stacking ensemble model, which combines Random Forest, XGBoost, Support Vector Machine, and Logistic Regression outperformed individual models, achieving 95% accuracy, a macro – average F1 score of 93% and an ROC-AUC of 99.5%. Key predictors included working memory, rapid naming, math attitude, ADHD symptoms, and vocabulary. Construct reliability and validity were confirmed using Cronbach's alpha and ANOVA, with stratified 10 – fold cross validation ensuring model robustness. The proposed framework is designed to guide teachers' decision making and facilitate the early diagnosis of learning difficulties, ensuring inclusive education. Ethical considerations (data anonymization and ethical consent were sought and approved from designated institutions and participants respectively) were considered throughout the study to support responsible research conduct and future application. The findings underscore the potential of ML-driven tools to enhance special education by enabling proactive data-informed instructional planning in under -resourced contexts.

1. Introduction

The integration of machine learning (ML) into education is reshaping how institutions identify and support students with diverse learning needs. Learning Disabilities (LDs) are neurodevelopmental disorders that significantly impair the acquisition and application of academic skills such as reading, writing, or mathematics, despite adequate intelligence and educational opportunities. Common LDs include dyslexia (inability to read), dyscalculia (inability to calculate), dysgraphia (inability to write), attention-deficit/hyperactivity disorder (ADHD), and executive function deficits. These conditions, which often persist into adulthood, can negatively affect self-esteem, school engagement, and long-term socio-economic outcomes if not addressed early [1]; [2] & [3]). In many contexts, delayed identification intensifies learning gaps, leading to academic exclusion and increased dropout rates [4]. Early detection is critical because timely interventions, such as targeted instruction, assistive technologies, and socio-emotional support have been shown to mitigate the long-term impact of LDs ([5]; [6]). However, traditional screening in basic education often relies on informal teacher observations, which can be subjective and inconsistent, especially in large classes or resource-limited settings [7]. The result is frequent underdiagnosis, misclassification, and delayed support.

Recent research demonstrates that ML models can detect subtle cognitive-behavioural patterns linked to LDs with high predictive accuracy, enabling earlier and more objective identification. For example, ML approaches have been applied to predict ADHD traits [8], dyslexia [9], and general learning performance risks ([10]). Systematic reviews highlight a growing interest in intelligent systems for personalized learning support [11]; [12]. Yet, these tools are predominantly developed in technologically advanced or clinical environments, making them less feasible for low-resource schools where screening tools must be low-cost, explainable, and integrated into everyday teaching workflows [13]. A notable gap in prior work lies in the interpretability and operational integration of ML-based LD detection systems. Black-box models, even when accurate, often fail to earn the trust of non-technical educators, limiting their adoption in practice [14]. Scholars have therefore called for explainable AI (XAI) approaches that not only predict risk but also reveal contributing factors, allowing teachers to make informed and ethical decisions [15]. Furthermore, there is limited research on scalable, policy-aligned, and context-sensitive deployment strategies for ML-based LD screening in under-resourced environments.

This study addresses these challenges by proposing a machine learning-based decision-support framework tailored to the early identification of LD risk in basic schools. The framework draws on validated psychometric, behavioural, and demographic indicators and employs a stacking ensemble approach to classify learners into three risk tiers: Low, Moderate, and High. This three-tier structure is grounded in educational intervention models such as Response to Intervention (RTI), which emphasize differentiated levels of support based on assessed need [16]; [17]. By combining predictive accuracy with interpretability, the system delivers actionable insights that teachers can use for timely, targeted, and sustainable interventions, particularly in resource-constrained settings where efficient allocation of limited support services is essential.

The study was guided by the following objectives:

1. To develop a machine learning model for classifying learners into Learning Difficulty risk levels.
2. To evaluate the proposed model against related models for accuracy and interpretability.
3. To share insights that facilitate teacher decision-making and timely learner support.

To address the study objectives, the following research questions were formulated:

1. How can machine learning be used to classify learners into different levels of Learning Difficulty risk?
2. How does the proposed model compare with related machine learning models in terms of accuracy and interpretability?
3. How can the model’s insights support teachers in making timely and informed decisions for learner support?

2. Review of Related Literature

The intersection of machine learning (ML), educational diagnostics, and inclusive pedagogy has become a prominent focus in recent educational technology research, particularly about early screening for learning difficulties (LDs). Systematic reviews and bibliometric analyses have contextualized this growing interest. For example, [18] provided a comprehensive bibliometric analysis of AI applications in education, emphasizing the expanding role of ML tools in supporting early diagnostics and adaptive learning. Likewise, [19] highlight how gamified and personalized learning environments powered by AI enhance learner engagement and inclusivity, principles central to the effectiveness of ML-based screening systems for diverse student populations.

From a methodological standpoint, recent empirical works offer practical insights into how technology-enhanced strategies impact learner outcomes. [20] conducted a meta-analysis of flipped classroom models in mathematics, demonstrating that structured, tech-driven interventions can significantly enhance student performance, offering a precedent for AI-supported instruction. [21] employed process mining to analyse student behaviour in online learning environments, highlighting advanced analytic techniques relevant to predictive LD classification. Complementing this, [22] reviewed immersive and interactive technologies in digital learning, revealing their potential to improve inclusivity and accessibility. Additionally, [15] examined AI integration in educational settings, stressing the importance of institutional readiness, stakeholder collaboration, and effective evaluation metrics. These considerations are equally critical for deploying ML systems in basic education contexts.

Adoption of intelligent systems in schools also hinges on teacher readiness and institutional alignment. In line with this, [14] found that leadership support plays a mediating role in facilitating technology integration, especially when coupled with robust professional development for educators. In parallel, [23] proposed a task-technology-environment fit model that emphasizes how the congruence between instructional goals, digital tools, and learning contexts significantly affects student outcomes; offering a guiding framework for aligning machine learning (ML) tools with classroom needs. Further reinforcing this, [22] developed and validated a comprehensive instrument to assess digital competence among primary and secondary school teachers, underscoring the need to balance pedagogical fluency with technical skill. Similarly, [24] examined how behavioural intention and perceived ease of use shape digital adoption in educational contexts, highlighting the importance of intuitive, user-centred interfaces to support real-world uptake.

Recent innovations demonstrate a convergence of artificial intelligence, behavioural analytics, and inclusive instructional design. For instance, [10] leveraged ML algorithms to infer student attentiveness through multimodal behavioural and emotional cues offering methods analogous to those used in early screening for learning difficulties. Meanwhile, [25] explored the effectiveness of cybersecurity education platforms embedded with motivational design features, revealing the impact of psychological engagement on learner performance, principles highly relevant for enhancing the usability of ML-based diagnostic tools aimed at supporting diverse learners.

Collectively, the literature underscores a rapidly evolving, interdisciplinary field committed to developing explainable, equitable, and effective educational technologies. Yet, a notable research gap persists: few empirical studies offer school-embedded, explainable machine learning (ML) frameworks that translate learning difficulty (LD) risk predictions into pedagogically actionable insights, particularly within low-resource educational contexts [4]; [15]; [14]. While many existing ML models demonstrate technical potency, they often fall short in terms of interpretability, scalability, or integration with routine classroom practices, thus limiting their adoption in real-world educational settings. This study directly addresses these limitations by introducing a scalable, explainable ML-based decision-support model tailored for early LD screening in under-resourced basic education systems. Anchored in principles of inclusive pedagogy, interpretability, and user-centred design, the proposed framework seeks to empower teachers with timely, accessible, and actionable data to guide early interventions and individualized learner support.

3. Methodology

3.1 Research Design and Framework

This study employed multidimensional theoretical framework that synthesizes Cognitive Load Theory (CLT), Response to Intervention (RTI), and Universal Design for Learning (UDL) as the design of a machine learning (ML)-based tool for early detection of learning difficulties (LDs). The CLT provides foundational knowledge into the constraints of working memory and the negative impact of cognitive overload, particularly in learners exhibiting deficits in processing speed, emotional regulation, or executive function [26]; [27]). These cognitive constructs are operationalized using validated psychometric instruments, including the Wide Range Achievement Test-Fifth Edition (WRAT-5) for academic skills, the Behaviour Rating Inventory of Executive Function-Second Edition (BRIEF-2) for executive function, and a condensed version of the Difficulties in Emotion Regulation Scale (DERS) to assess emotional regulation. The ML model adopts RTI's tiered intervention structure, stratifying students into low, moderate, and high LD risk levels to support targeted, data-informed interventions [28]. UDL principles further enhance the framework by ensuring that the tool accommodates diverse learner profiles, including cognitive, behavioural, and demographic dimensions, thereby promoting equitable access, contextual relevance, and inclusive educational practices [29]; [30]

The operational architecture of the model was premised on four major domains: cognitive-academic performance, language and phonological processing, behavioural-emotional regulation, and demographic context. Academic performance was assessed through WRAT-5 indicators such as arithmetic, spelling, and reading, paired with cognitive variables like working memory and processing speed to capture learning capacity holistically. Language-based predictors, which include phonological awareness, vocabulary, rapid naming, and reading comprehension, align with current research on literacy-related LDs [2]. Emotional and behavioural regulation were assessed through components derived from BRIEF-2 and DERS scales, with multicollinearity checks confirming independence ($VIF < 3$). Additionally, sex, age, school level, and language spoken at home were integrated to capture relevant contextual moderators [31]; [32]. While Academic Achievement and Language Development variables showed low internal consistency (α < 0), they were retained due to their theoretical significance and demonstrated statistical relevance through significant ANOVA and multinomial logistic regression outcomes. This comprehensive structure ensures theoretically grounded, statistically validated, and practically deployable within educational settings.

3.2 Sampling and Data Structuring for Machine Learning

The stratified sampling technique was adopted for this study. This was guided by the Cochran formula for sample size determination, to ensure adequate representation across educational levels (P4 to JHS3), sex, and geographic zones - urban and rural [33]. This approach allowed the model development process to account for subgroup variability that may influence the risk of learning difficulties (LDs). Stratification was particularly useful for capturing underrepresented groups and controlling for contextual differences that could affect both assessment and intervention outcomes.

The Cochran formula provided a statistical estimate of the minimum required sample size based on the target population, desired confidence level, and acceptable margin of error:

$$\mathrm{n}_0 = \frac{\mathrm{Z}^2 \, . \, p \, . \, (1-p)}{e^2} \, . \, . \, . \, . \, . \, . \, . \, . \, . \, eqn \;\; (1)$$

Where:

Z = 1.96 (95% confidence level)

P = 0.5 (maximum variability)

E = 0.05 (margin of error)

$$\mathrm{n}_0 = \frac{(1.96)^2 \, . \, 0.5 \, . \, (0.5)}{(0.05)^2} = 384.16 \approx 385$$

3.3 Instrumentation and Assessment Constructs

A standardized questionnaires was administered to design a robust ML LDs prediction model. This was adapted from evidence-based psychometric instruments (as detailed in Section 3.2.1). These instruments were designed to capture a multidimensional profile of each learner, including cognitive-academic performance, language and phonological processing, behavioural-emotional regulation, and relevant demographic indicators. Data collection occurred across two full academic years **(2022–2024)** under the jurisdiction of the Ghana Education Service (GES). The instruments were administered to learners in both public and private basic schools located across multiple districts within the Asante Mampong municipal area in the Ashanti region of Ghana. This extended and regionally distributed data collection ensured a comprehensive dataset representative of varied school contexts and learner backgrounds.

3.3.1 Item Test Components

The study employed a comprehensive, multi-dimensional questionnaire to assess domains linked to learning difficulty (LD) risk, including cognitive-academic, behavioural, emotional, and demographic factors, summarized in appendix A. Early classroom indicators were gathered using an observational checklist inspired by local educational research on LD traits in Ghanaian learners (e.g., [34]). Behavioural screening for attention-related symptoms (e.g., inattention, hyperactivity) relied on the Disruptive Behaviour Disorders Rating Scale **(DBDRS)**, a validated and culturally adapted teacher-report tool used in Ghanaian school-based ADHD studies [35]. Also, Academic achievement was measured using WRAT-5 subtests, focusing on reading, spelling, and arithmetic; while cognitive constructs such as working memory and processing speed were adapted from simplified, classroom-friendly versions of WISC-V tasks [36]. Literacy-related predictors, which include phonological awareness and rapid naming, were operationalized based on evidence from early reading research [37]. Vocabulary and reading comprehension used locally adapted formats of the PPVT and GORT to suit Ghanaian instructional contexts. Numeracy was assessed with timed fluency tasks analogous to WIAT-III measures, ensuring alignment with recognized psychometric standards.

Emotional regulation and executive functioning, on the other hand, were assessed via adapted items derived from the BRIEF-2 and the Difficulties in Emotion Regulation Scale (DERS), both of which have strong construct validity in educational settings. Each instrument segment employed a 5-point Likert scale, following standard scoring protocols. A pilot involving 30 learners from diverse schools was conducted to ensure clarity, relevance, and preliminary reliability before full deployment.

3.4 Dataset Construction

The final dataset comprised 2,115 valid learner responses collected over two Ghana Education Service (GES) academic years (2022–2024). The LD risk category distribution comprised 1,024 Low risk learners (48.4%), 691 Moderate risk learners (32.7%), and 400 High risk learners (18.9%). Demographically, the sample included 1,052 males (49.8%) and 1,063 females (50.2%), with age ranging from 9 to 16 years (Mean = 12.4, SD = 1.7). The geographic split was urban: 1,112 (52.6%) and rural: 1,003 (47.4%). Class-level representation ranged from Primary 4 (n = 312) to Junior High School 3 (n = 295). Additional fields included languages spoken at home, medical history, and repetition history. Each record included psychometric ratings across cognitive, behavioural, emotional, and academic domains, alongside the LD risk label (Low, Moderate, High). The dataset contained 36 features representing a comprehensive profile of each learner. Stratified sampling during data collection ensured balanced subgroup representation. The Synthetic Minority Over-Sampling Technique (SMOTE) was later applied during preprocessing to correct any residual imbalances and enhance model generalizability.

3.5 Data Collection Procedure

Data collection was conducted in close collaboration with district education offices and school heads to ensure access, coordination, and contextual relevance. Trained assessors administered the questionnaires to learners during two cycles per academic year. This was done in Term 1 and Term 3 to capture academic and behavioural trends across different periods. Researchers obtained observational data from teachers through a structured checklist, enriching the dataset with professional insights into classroom behaviours. The entire data collection process strictly adhered to ethical guidelines for research involving human subjects. Informed consent was obtained from parents or legal guardians of all participating learners through Parent-Teacher Association (PTA) meetings, and verbal assent was secured from learners aged 12 years and above before participation. The study also received ethical clearance from the International Relations, Research and Professional Development (IRRPD) Unit of the Mampong College of Education, ensuring its compliance with national and institutional standards for educational research.

3.6 Data Preprocessing and Model Optimization

3.6.1 Preprocessing Techniques

The researchers employed standard preprocessing techniques to prepare the dataset for modelling. Missing values were addressed using listwise deletion after identifying incomplete records with the isna() method. Categorical variables (e.g., learner's sex, language spoken at home, class level) were encoded using one-hot encoding to prevent ordinal misinterpretation. Numerical features were standardized using StandardScaler() to normalize them to zero mean and unit variance; important for models such as SVM and Logistic Regression [38]. Class imbalance in the LD risk categories was addressed using SMOTE exclusively on the training set after the train-test split. This prevented data leakage and ensured fair evaluation of the model [39]

Hyperparameter tuning was employed using GridSearchCV within a 5-fold cross-validation loop on the training set. This approach systematically explored a predefined hyperparameter grid for each algorithm (e.g., number of trees and maximum depth for Random Forest; C, kernel, and gamma for SVM; learning rate and max depth for XGBoost) and selected the configuration that maximized mean cross-validation F1-score. GridSearchCV ensured reproducibility and optimal parameter selection while mitigating overfitting risk in imbalanced datasets. Collectively, these preprocessing steps yielded a clean, balanced, and machine learning-ready dataset in alignment with contemporary best practices.

3.6.2 Feature Engineering and Selection

The researchers retained all input features during the initial stage of the modelling process to support model transparency and interpretability. This approach aligns with the principles of Explainable Artificial Intelligence (XAI), which emphasize preserving domain-relevant variables in early iterations, particularly in educational and psychological research, where feature elimination may obscure meaningful learner characteristics [40]; [41]. Constructs such as working memory, phonological awareness, or emotional regulation hold significant theoretical and practical relevance. Therefore, excluding them prematurely based on statistical thresholds could compromise the depth of insight. After model training, Random Forest feature importance analysis was used to assess the predictive contribution of each feature to LD risk classification. Random Forest was chosen for its capacity to estimate feature importance via the mean decrease in impurity, providing a robust and interpretable ranking of variables [42]. This approach not only improved the transparency of the machine learning model but also enabled the extraction of actionable insights for educators and intervention specialists focused on targeted support.

3.7 Data Splitting and Modelling

Stratified 80/20 train-test split was implemented, ensuring the generalizability of the model's performance. This preserves the original class distribution across both subsets - a key approach to minimize evaluation bias in imbalanced educational datasets [43]. The 80% training set was used to develop and tune multiple classifiers, including Random Forest, Support Vector Machine (SVM), Logistic Regression, and XGBoost. The remaining 20% served exclusively for final performance validation. The study employed 5-fold cross-validation on the training set for hyper-parameter tuning. This method allows model selection while reducing overfitting risk and ensuring reproducibility, following best practices in educational predictive modelling [44]. Evaluation metrics included accuracy, F1-score, precision, recall, and ROC-AUC, providing a comprehensive basis for comparing models. Although 5-fold cross-validation was embedded in the training phase for hyper-parameter optimization, the single hold-out test set was retained to mimic real-world deployment scenarios, ensuring the reported metrics reflect out-of-sample performance and avoiding the optimistic bias that can occur when results are drawn solely from cross-validation.

Furthermore, class imbalance was addressed by applying SMOTE only to the training folds within the cross-validation pipeline, thereby preventing data leakage and preserving evaluation integrity. SMOTE's capability to synthetically generate minority class samples, without repeating existing ones, enhanced sensitivity and reliability in identifying underrepresented risk groups (Applied Sciences Educational Data Study, 2024). Such extensive preprocessing and validation practices reflect contemporary standards in educational machine learning, ensuring that model performance translates effectively into real-world use cases.

3.8 Model Development and Evaluation

The entire modelling was executed in two (2) main experiments to develop the proposed model framework. In the initial experiment, three standalone machine learning classifiers were implemented: A Random Forest Classifier configured with class weighting and 100 estimators; a Support Vector Machine (SVM) with balanced class weights and probability outputs enabled; and a Logistic Regression model using L2 regularization and class weight balancing. In Experiment 2, a hybrid stacking ensemble classifier was developed by combining the three standalone models as base learners with a regularized logistic regression meta-learner (C = 0.1, penalty='l2'). To mitigate bias and reduce overfitting, stratified 5-fold cross-validation was applied to generate meta-features within the stacking process. Model performance was assessed using a set of comprehensive evaluation metrics:

Accuracy

The proportion of total correct predictions, useful for general performance assessment across all classes. Mathematically expressed as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \dots \dots \dots \dots \quad eqn \ (2)$$

Where:

TP: True Positives

TN: True Negatives

FP: False Positives

FN: False Negatives

Precision

The ability of the model to correctly identify true positives among all predicted positives-critical for minimizing false positives. Mathematically, precision is represented as

$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$ .... Eq. (3)

Recall (Sensitivity)

Measures the proportion of actual positives that were correctly identified, highlighting the model's effectiveness in detecting at-risk learners. Mathematically expressed as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \dots \dots \dots \dots \dots \ \text{eqn} \ (4)$$

F1-Score

The harmonic mean of precision and recall, balancing both false positives and false negatives. Also, expressed as

$$\mathbf{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \dots \dots \dots \ eqn \ (5)$$

$$\text{True Positive Rate (TPR)} = \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN}$$

McFadden’s Pseudo R²

A goodness-of-fit measure used in classification models, particularly logistic regression, to assess the explanatory power of the model. Higher values (typically above 0.2) suggest a well-fitting model (McFadden, 1974). It can also be represented mathematically as

$$\mathrm{RMcFadden}^2 = 1 - \frac{\ln L_{model}}{\ln L_{null}} \ldots \ldots \ldots \; eqn \; (7)$$

Where:

$$\mathrm{lnL_{model}} : \ \mathrm{Log - likelihood\ of\ the\ fitted\ model}$$

$$\mathrm{lnL_{null}} : \ \mathrm{Log - likelihood\ of\ the\ model\ with\ only\ an\ intercept}$$

Note

Values closer to 1 indicate better fit. Values above 0.2 are considered good for logistic regression models ([45]*).*

The researchers also employed Visualization tools such as confusion matrices and ROC curves to provide deeper insight into class-level performance, enabling educators and system designers to understand which learner groups were most accurately identified and which required further tuning. In addition, researchers applied several techniques to detect and control overfitting, including the use of cross-validation during model training, stratified train-test splitting, monitoring the performance discrepancy between training and test phases, and applying regularization and class balancing in both base and ensemble models. This methodological approach ensured that the models were both robust and generalizable.

3.9 Educational Technology Integration

To enhance classroom applicability, model outputs were visualized through an interactive teacher-facing dashboard prototype, highlighting predicted LD risk levels, feature importance (e.g., low working memory or math fluency), and suggested action plans. This framework aligns with the technological dimension of inclusive education, ensuring that non-technical educators can interpret and act on data-driven insights (Ali et al., 2024; [18]. The dashboard design followed UDL principles and was tested for interpretability with a pilot group of basic schoolteachers.

3.10 Reproducibility and Transparency

Ensuring reproducibility and transparency was a core principle throughout the modelling and analysis phases of this study. The entire machine learning pipeline was implemented using Python version 3.11.5, with a suite of specialized libraries to support data processing, modelling, and visualization. **Scikit-learn** (version 1.3.0) was employed for model development, preprocessing, and evaluation, while **imbalanced-learn** (version 0.11.0) facilitated the implementation of the Synthetic Minority Over-Sampling Technique

(SMOTE) for class balancing. **XGBoost** (version 1.7.6) was used to implement the gradient boosting model. For data manipulation and preparation, **NumPy** (version 1.25.2) and **Pandas** (version 2.1.0) provided efficient array operations and structured data handling. Visualization and exploratory data analysis were conducted using **Matplotlib** (version 3.8.0) and **Seaborn** (version 0.13.0), enabling the creation of both descriptive and analytical plots to support the interpretation of results.

In addition, to promote full transparency and allow for independent verification of results, researchers ensured all essential artifacts of the modelling process were systematically logged and versioned. Also, the Jupyter Notebook was annotated with markdown cells that explain each modelling decision, justify methodological choices, and interpret key outputs. This not only aids replication but also enhances interpretability for other researchers, data scientists, and education stakeholders.

By adhering to the best practices in open and reproducible data science [46]; [47], this study contributes to the growing body of reproducible ML research in education and lays a transparent foundation for future benchmarking, extension, and contextual adaptation in diverse educational settings. The entire modelling process for the model's development has been depicted in Fig. 1. Files and source codes for the modelling of the LD detection ML ensemble stacking model can be obtained at https://github.com/samuelodoom-coder/LD-Detection-ML-Stacking/tree/main for the reproducibility section of the model.

3.11 Limitations of Methodology and Mitigation Approaches

The researchers acknowledged that the study faced several methodological limitations that could potentially affect the generalizability and interpretability of findings. While the dataset was robust in terms of psychometric and behavioural diversity, it was collected from a specific geographic and educational context-basic schools in a selected Ghanaian municipality. This localized scope may limit cross-context generalizability, particularly in more diverse or international educational settings. To mitigate this, the study prioritized transparent variable definitions, open data formatting (CSV), and a modular model design, enabling future researchers to adapt and retrain the framework with localized datasets.

In addition, certain psychometric constructs (e.g., WRAT and Language Development subscales) exhibited low internal consistency (Cronbach’s alpha < 0), which could introduce noise measurement. These indicators were retained due to their theoretical relevance; however, the study addressed this limitation by triangulating results through statistical significance testing (ANOVA, Chi-Square, and Multinomial Logistic Regression) and feature importance analyses, ensuring that conclusions were not dependent on unreliable constructs alone. Future refinements will involve revalidating or localizing these instruments for improved measurement reliability.

While machine learning models such as Random Forest and Stacking Ensemble achieved high predictive performance, they inherently present challenges in interpretability for non-technical stakeholders like teachers. To reduce the “black box” barrier, the study incorporated decision tree visualizations and feature importance plots to communicate classification logic clearly. This approach aligns with explainable AI (XAI) practices recommended in educational research [48], enhancing practical usability and stakeholder trust.

A further limitation was the absence of external validation, as the models were evaluated solely using an 80 – 20 train–test split within the same dataset. This raises questions about generalizability to different districts, regions, or countries. While reproducibility was ensured through comprehensive logging of preprocessing, training, and evaluation scripts in Jupyter Notebooks, future work will prioritize external validation and cross-context benchmarking.

Finally, ethical and privacy considerations were paramount due to the sensitive nature of learner data. These were addressed through IRPDU (Mampong Technical College of Education) approval, informed consent, strict anonymization, and encryption protocols, in accordance with ethical standards in educational data mining [49]. Hence, the researchers affirmed that these limitations were actively mitigated to strengthen methodological rigor and practical relevance. Future research will address potential biases in sample representativeness, expand validation to culturally and linguistically diverse contexts, and explore longitudinal prediction capabilities to better inform early intervention strategies.

3.12 Triangulation to Validate the Model

A triangulation strategy was employed, integrating **methodological, data**, and **theoretical** triangulation approaches. This approach strengthens the internal consistency of findings and supports the interpretability and practical value of the model in educational contexts. **Methodological triangulation** was achieved by combining machine learning classification with conventional statistical techniques. Specifically, the predictions generated by the stacking ensemble model were examined alongside the results of **multinomial logistic regression, Chi-square tests**, and **ANOVA**. These traditional inferential methods confirmed the predictive significance of key variables-such as working memory, math attitude, and ADHD symptoms, which were also ranked high in the ML model’s feature importance analysis. The consistency between statistical outcomes and model-derived insights reinforces the construct validity of the input features and the stability of the model’s logic.

In a similar vein, theoretical triangulation was embedded through the integration of multiple pedagogical and psychological frameworks: Cognitive Load Theory supported the cognitive basis for feature selection (e.g., working memory and executive function), while **Universal Design for Learning (UDL)** and **Response to Intervention (RTI)** informed the model’s classification architecture and decision thresholds. These theoretical lenses ensured that the model was not merely data-driven, but pedagogically grounded, thereby increasing its relevance for practitioners and policymakers alike. These triangulation strategies enhanced the **validity, reliability, and applicability** of the ML model and ensured that the findings are not artifacts of a single method or data type. This integrated approach also contributes to emerging best practices in **educational data science**, where combining AI-driven insights with human-centred theories is increasingly essential for responsible innovation [50].

3.13 Integration into Educational Technology Systems

The conceptual framework for the proposed model's integration (Fig. 2) illustrates how the proposed ML-based LD screening model can be integrated into a school's educational technology (EdTech) ecosystem. At the foundation of this framework is the *Data Collection Layer*, where psychometric, behavioural, and demographic data are gathered from teachers, caregivers, and learners via mobile forms, school portals, or paper-based tools digitized through teacher dashboards. These inputs are processed in the *Data Management Layer*, which includes anonymization, formatting into CSV structures, and storage within the school's learning information systems (LIS).

The *Model Processing Engine* receives this data and applies the trained stacking ensemble algorithm to classify learners into LD risk levels (Low, Moderate, High). This processing occurs either locally via school servers or remotely through cloud-based platforms, depending on infrastructure availability. The results are passed to the *Teacher Dashboard Interface*, a user-friendly visualization tool that translates predictions into actionable classroom recommendations. Teachers receive learner profiles with risk scores, suggested interventions, and historical tracking.

In parallel, the *Administrative Panel* provides school leaders and policymakers with aggregated reports and alerts, enabling strategic allocation of resources, referrals to specialists, and tracking of intervention effectiveness. The final layer, *Support Services Integration*, ensures that high-risk learners can be referred to school psychologists, health officers, or special educators through an automated flagging system.

This model underscores a closed-loop, data-driven support structure, emphasizing real-time feedback, teacher empowerment, and system-wide accountability. By embedding the model into existing EdTech workflows, such as student information systems (SIS), learning management systems (LMS), or RTI platforms, it promotes scalability and sustainable adoption, particularly in resource-constrained school environments [15].

4. Results and Discussion

This section presents the findings from the ML-based model developed for classifying learners into different risk levels of learning disabilities (LDs). The section is structured in line with the study's objectives.

4.1 Demographics Exploratory Analysis

The demographic analysis of learners across different levels of Learning Difficulty (LD) risk; low, moderate, and high, revealed distinct patterns associated with age, class level, and sex. These findings are essential for informing data-driven and equitable intervention strategies within educational settings. The age distribution (Fig. 3) shows that learners classified as low risk were slightly older on average (median age ≈ 13 years) than those in the moderate and high-risk groups (median age ≈ 12 years). This pattern suggests that increased age may confer cognitive or developmental maturity that enhances academic functioning. Research has shown that older learners tend to exhibit greater executive functioning and metacognitive skills, which can buffer against learning challenges [51]). Also, the class-level distribution (Fig. 4) further supports this trend. Learners in P6 and JHS3 were more frequently found in the low-risk category, while higher proportions of moderate and high-risk learners were observed in P4 and JHS1. These stages may represent key educational transitions where learners are particularly vulnerable due to changes in curriculum complexity, instructional practices, or assessment expectations. Prior studies have identified transition points, such as the shift from lower to upper primary or primary to junior high, as periods where students often struggle if adequate academic scaffolding is not in place.

Researchers further performed analysis on participants' sex distribution (Fig. 5), which revealed a nearly equal distribution of male and female learners across LD risk categories, with a slight predominance of females in the low-risk group. This finding diverges from previous literature that often reports higher LD prevalence among males, which is typically attributed to greater externalizing behaviours and referral biases [31]. The relative balance observed here may reflect improved screening tools and increased awareness among educators, resulting in more equitable identification across genders. Recent studies suggest that when behavioural and academic indicators are equitably measured, girls are nearly as likely as boys to experience learning difficulties, though they may present differently ([52]; [53]). Researchers then concluded that these demographic insights suggest that age, class level, and sex play significant roles in shaping LD risk. Integrating these variables into early warning systems and intervention planning can enhance precision in identifying at-risk learners. Moreover, tailoring support based on developmental stage and gender-sensitive approaches can improve educational outcomes and ensure that no subgroup is overlooked in LD diagnostics.

4.2 Psychometric Validity and Predictive Insights of Feature Variables

The evaluation of the academic skills (WRAT) and Language Development tests showed mixed results on their suitability for complex analysis ([54]). Their initial measure of suitability (Kaiser-Meyer-Olkin (KMO): 0.497 and 0.46) was slightly below the common standard of 0.50. However, another important statistical test (Bartlett’s Test) was highly significant for both ($p < 0.001$). This means that while the initial measure was low, the strong internal relationships among the test items and their established importance in learning difficulty (LD) research justified their inclusion.

The process of selecting the most important factors revealed that Working Memory, Rapid Naming, Math Fluency, and Vocabulary were consistently among the top predictors. This aligns with recent research on the critical role these skills play in identifying LD risk earlier [25]. Working Memory was especially powerful in distinguishing between risk groups, confirmed by a strong statistical result ($F = 86.91$, $p < 0.001$). This supports modern theories that see working memory deficits as a core part of LD.

Additionally, a significant relationship was found between sex and LD risk classification ($\chi^2 = 19.27$, $p < 0.001$), with male students more frequently classified in the Moderate and High-Risk groups. This finding supports recent studies that have documented higher LD diagnosis rates in males, potentially due to differences in neurobiological and behavioural development ([55]; [56]. To test the combined predictive power of all these tests and demographic factors, a special statistical model was used. This model showed a very good fit ($p < 0.001$) and had a strong explanatory power score (Pseudo $R^2 = 0.5639$), which is considered excellent in educational psychology by Hosmer-Lemeshow cited in [57]. This reinforces the value of using both theory-based and data-driven factors in LD risk models.

Furthermore, when these key features were combined into an advanced "ensemble" model, it achieved robust performance with 96.22% accuracy and a near-perfect discriminative score (AUC = 0.9967). This highlights the practical benefit of combining traditional psychometrics with modern computing for creating scalable screening tools in schools [58].

4.3 Objective 1: Development and Optimization of ML Models for LD Risk Classification

The first goal was to build and optimize computer models to classify students as Low, Moderate, or High Risk for learning difficulties (LD). Four different models were trained using data from tests, cognitive assessments, and demographic information. The data was prepared carefully to handle categories, scale numbers, and balance the groups.

From Table 1 below, Random Forest was the top-performing standalone model, achieving 95.04% accuracy. It was excellent at distinguishing between all risk groups (ROC AUC = 99.19%). This aligns with research on its strength for finding complex patterns in this type of data [59]. The XGBoost model performed nearly identically, also with 95.04% accuracy and the highest discriminative score (AUC = 0.9932) of the single models, reflecting its ability to model complex feature interactions ([60]; [61].

In contrast, the Support Vector Machine (SVM) model was less accurate (90.54%) and was not as sensitive at correctly identifying students in the Moderate and High-Risk groups, consistent with its known limitations [62]. The Logistic Regression model, while simple and fast, struggled the most with only 78.49% accuracy, as it could not capture the complex relationships between the predictors. A summary of the standalone model's performance is shown in Table 1 below. These results led to the creation of a combined Stacking Ensemble model using the top 10 predictive features, including Working Memory, Rapid Naming, Emotional Regulation, and ADHD Score ([63]; [64].

Table 1
Performance of standalone models as baseline for proposed Stacking model

Model	Accuracy (%)	F1-Score (%) L M H	Recall (%) L M H	AUC ROC (%)
RF	95	100 93 87	100 97 81	99
XGBoost	95	100 93 88	100 94 86	99
SVM	91	100 87 76	100 88 74	97
LR	78	90 68 71	90 64 78	91
Note: Low (L), Moderate (M), and High (H)				

4.4 Objective 2: Comparison of Performance – Ensemble vs. Baseline Models

A comparison of the baseline model and the proposed stacking ensemble model was performed using accuracy, precision, recall, f-score, AUC ROC, macro average, and weighted average.

4.4.1 Comparative Evaluation of Baseline and Stacking Ensemble Models for LD Risk Prediction

Researchers compared their new combined model, which integrates three other techniques (Logistic Regression, SVM, and Random Forest), against standard models using multiple performance measures (Table 2; Appendix D). The new model performed on par with the top-performing Random Forest, achieving a high overall accuracy of 96.21%. Its ability to distinguish between groups was exceptional at 99.67%, marginally outperforming Random Forest (99.19%). The model was also reliably accurate across all risk categories: correctly identifying 97.4% of Low Risk, 94.8% of Moderate Risk, and 93.1% of High-Risk cases. While the new model did not significantly exceed Random Forest in raw accuracy, it offered better precision, especially in distinguishing between the Moderate and High-Risk groups. These findings are consistent with prior studies [65], which show that combined models leverage the strengths of individual techniques to create a more balanced and reliable system. Despite the ensemble's slightly superior performance, Random Forest remains a highly viable and practical option, particularly where resources are limited. Conversely, the stacking ensemble is better suited for critical decision-making scenarios, where even modest gains in predictive power can lead to significant outcomes.

Table 2
Comparison of models' performance: Baseline vs. Stacking

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC ROC (%)	Macro Average (%)	Weighted Average (%)
RF	95	94	93	93	99.2	94	95
Stacking	96	96	94	95	99.7	96	96

4.4.2 Evaluation of Models' Performance Using a Confusion Matrix

To assess the effectiveness of the different computer models at predicting a student's risk level for learning difficulties, a detailed analysis was performed. This is shown in Table 3 below using a confusion matrix. The confusion matrix shows exactly where models make correct and incorrect predictions, focused on three risk categories: Low, Moderate, and High. The goal was to see how well models like Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and the new combined "Stacking Ensemble" model could identify at-risk students for early help. Among all models, the Stacking Ensemble was the most robust. It perfectly identified all 190 moderate-risk students (100% accuracy for this group), which is vital in education where missing these students can delay support. It was also excellent (99% accurate) at finding high-risk students. For low-risk students, it was 84% accurate, meaning 16% were mistakenly flagged as higher risk. This strong overall performance is due to the ensemble model combining the predictions of multiple models (RF, SVM, and LR) to balance their strengths and weaknesses, a finding supported by prior research [66]; [67] The Random Forest model was also very strong, nearly matching the ensemble's performance. It identified 97% of moderate-risk students and was perfect for low-risk students. However, it missed slightly more high-risk students (19% were misclassified) than the ensemble. Despite this, its consistency and ease of use make it a highly practical and efficient choice for real-school settings, as found in [68]

In contrast, Logistic Regression had significant difficulties. It struggled most with the higher-risk categories, correctly identifying only 78% of high-risk and 64% of moderate-risk students. This is because it assumes the data is simple and linear, which is often not the case with complex educational data, a limitation noted in other studies ([8]; [19]. The SVM model also showed suboptimal performance, finding only 74% of high-risk students, indicating it had trouble distinguishing between the complex risk categories, though it was perfect for low-risk students. The Stacking Ensemble took the topmost model position, especially where it matters most, that is identifying students at moderate and high risk of LD. Random Forest is a very capable and practical alternative, while Logistic Regression and SVM have notable limitations that could impact their use in planning critical educational interventions. These results highlight the advantage of using combined ensemble models for sensitive educational tasks where accuracy is essential.

Table 3
Evaluation of Models' performance using the confusion matrix

Model	True positives	False positives
	L (%) M (%) H (%)	L (%) M (%) H (%)
Stacking	**84 100 99**	**17 0 1**
RF	**81 100 97**	**19 0 3**
SVM	74 100 94	26 0 6
XGBoost	100 94 86	0 6 14
LR	72 84 64	28 16 36
Note: Low (L), Moderate (M), and High (H)		

4.4.3 Inter-Model Performance Comparison with Related Works

To contextualize the performance of the proposed models, a comparative analysis was conducted between the classifiers developed in this study and those reported in recent peer-reviewed literature on learning difficulty (LD) risk prediction and educational diagnostics (see Table 4). Among the baseline models, the Random Forest (RF) classifier demonstrated strong standalone performance, achieving 95% accuracy, an F1-score of 93%, and a ROC-AUC of 99.2%. The proposed Stacking Ensemble, which combined RF, Support Vector Machine (SVM), and Logistic Regression (LR) as base learners, further enhanced generalization capabilities, reaching 96% accuracy, 95% F1-score, and the highest ROC-AUC of 99.7%. These results underscore the ensemble model's robustness in capturing complex, overlapping characteristics typical of LD risk profiles. This is consistent with prior findings on ensemble-based educational diagnostics [65].

Deep learning approaches continue to show promise but face interpretability and data sufficiency challenges. [9] applied a Convolutional Neural Network (CNN) to detect dyslexia using multimodal speech-text data, achieving 92% accuracy and a 91% F1-score. Similarly, [69] implemented a hybrid CNN-LSTM model on spatio-temporal reading pattern data, yielding 91% accuracy and 90% F1-score, demonstrating high sensitivity to sequential learning behaviours, albeit with considerable computational overhead. A recent variant, the Bi-LSTM with attention mechanism, employed by [70], achieved 90% accuracy and 94% ROC-AUC using reading comprehension logs, offering better interpretability and temporal representation. Traditional classifiers continue to be explored in low-resource settings. [67] reported a Naive Bayes model with 77% accuracy and 84% ROC-AUC for LD screening in Indian schools, highlighting its simplicity but also its limitations with overlapping features. Similarly, k-Nearest Neighbors (KNN) showed relatively poor performance, with [19] recording only 70% accuracy and 79% ROC-AUC for behavioural LD detection in Malaysian learners, reaffirming its struggles in high-dimensional, noisy data contexts.

Gradient boosting models such as XGBoost and LightGBM remain competitive. In the current study, XGBoost achieved 95% accuracy, a macro F1-score of 93%, and a ROC-AUC of 99.3%, validating its utility in structured educational datasets and reinforcing the results of [21], who applied XGBoost in early ADHD screening with 89% accuracy and 96% ROC-AUC. Likewise, LightGBM achieved 88% accuracy and 93% ROC-AUC in predicting academic success among Ghanaian SHS students [71], though it trades off some interpretability for computational efficiency. AutoML approaches are also emerging; [8] applied TPOT to LD risk detection in e-learning contexts, yielding 93% accuracy and 95.5% ROC-AUC with minimal manual tuning, suggesting future potential for automated model optimization in educational diagnostics.

To cap it all, the proposed stacking ensemble model outperformed most existing models across all critical metrics. The close performance of XGBoost and Random Forest further illustrates that carefully selected ensemble and boosting models can offer strong alternatives in real-world LD detection systems. These results support the strategic integration of ensemble learning within scalable diagnostic frameworks, especially where early identification and intervention for at-risk learners are crucial.

Table 4
Comparative Performance of Proposed LD Risk Models and Related Machine Learning Approaches from Recent Studies

Model	Study / Source	Context / Dataset	Accuracy (%)	F1-Score (%)	ROC-AUC (%)	Key Remarks
Random Forest	*Current Study*	LD Risk Classification (Psychometric Data)	**95**	**93**	**99.2**	Strong standalone performance; robust to nonlinearities
Stacking Ensemble	*Current Study*	Combines RF, SVM, and LR	**96**	**95**	**99.7**	Best overall performance; excellent generalization
CNN (Deep Learning)	[9]	Dyslexia detection using multimodal speech-text inputs	92	91	94	Effective on unstructured data; less interpretable
Naive Bayes	[67]	LD screening in Indian schools (Survey Data)	77	76	84	Simple baseline; struggles with feature correlation
XGBoost	[21]	ADHD detection from behavioral and attention data	89	88	96	Competitive with RF; sensitive to hyperparameters
Hybrid CNN-LSTM	[72]	Spatio-temporal reading data from eye-tracking	91	90	95	High sensitivity; needs large, labeled datasets
LightGBM	[71]	Academic success prediction in Ghanaian SHS learners	88	87	93	Efficient and scalable; limited transparency
KNN (k = 5)	[19]	Behavioral traits for LD in Malaysian students	70	69	79	Poor performance with sparse, high-dimensional features
Bi-LSTM with Attention	[70]	Learning difficulty classification using reading comprehension logs	90	89	94	Captures long-term dependencies; interpretable attention weights
AutoML (TPOT)	[8]	Educational data mining (LD risk in e-learning systems)	93	92	95.5	Automated model search; slightly lower than custom-tuned ensembles

4.5 Objective 3: Decision-Support Framework for Early LD Detection

4.5.1 Insights from Decision Tree Analysis

Decision trees, created by a Random Forest model is used to develop a visual guide to easy understand students LDs in this study. These trees act as a practical framework, giving educators clear, step-by-step insights to assess a student's risk level for learning difficulties. As shown in Fig. 6 and Fig. 7, these visual guides illustrate how different factors, like cognitive skills, behaviour, and personal background - combine to determine risk level. This offers teachers a transparent and user-friendly tool to aid their professional judgment.

A key finding was that Working Memory was the most important factor. Students with weaker working memory were more likely to be classified as moderate or high risk, confirming its essential role in learning. This supports previous research showing that working memory is critical for success in reading and math. [46]; [73]. In practical terms, using short working memory screening tools in schools could improve early detection and intervention. Another important factor was Math Attitude. Students with more negative feelings about math were often directed toward higher-risk categories. This aligns with studies showing that math anxiety can lower engagement and performance [74]. Including simple teacher surveys on student attitudes could serve as an early warning system, allowing for support before academic challenges become more serious.

The analysis also highlighted the importance of processing skills like Rapid Naming and Processing Speed, especially for identifying high-risk students. These skills are closely tied to reading fluency and are often affected in conditions like dyslexia [42]; [75]. Students who were slower in these tasks were consistently classified as higher risk, suggesting that quick fluency assessments could help teachers identify candidates for literacy support. Other important factors included Vocabulary and Reading Comprehension, which helped distinguish between moderate and high-risk learners. This reinforces the importance of language skills for academic success, especially for multilingual students or those with limited language exposure [2]. Home Language also played a role, indicating that linguistic background interacts with other learning skills to shape risk, a pattern noted in research on bilingual development [76].

Behavioural factors were also significant. Students with higher ADHD traits or difficulties with Emotional Regulation were more often classified as high risk. This is consistent with studies linking attention and self-control to academic performance ([77]; [78]. Finally, contextual details such as medical history, grade level, and home language background also contributed to risk classification. For example, younger students with reading challenges were more likely to be considered high risk, possibly due to developmental delays. Similarly, students with medical conditions were often at higher risk, underscoring the need to consider health and social background in supporting learners [79].

In summary, the decision trees provided a clear, scalable, and practical tool for early detection of learning difficulties. By turning complex data into intuitive visual logic, this approach helps teachers make informed decisions without needing technical expertise. During training workshops, educators used simplified versions of these trees to understand how risk levels are determined for individual students, a major step toward building inclusive, data-aware education systems that support early and equitable intervention.

4.5.2 Feature Prioritization for Teacher-Led LD Risk Identification

To support the development of a scalable, teacher-led decision-support framework for LD risk detection, a Random Forest model was used to identify the most influential predictors of learning difficulty classification. The resulting features of importance visualization (Fig. 8) highlighted Working Memory as the top predictor, contributing over 10% to model decisions. This finding aligns with prior research [46]; [73], which underscores working memory as central to learners' academic functioning, particularly in reading, math, and attention regulation. From a classroom application standpoint, this suggests that teachers could administer simple memory-based tasks, such as listening span or mental math activities, to screen for early learning challenges.

Closely following in importance were Rapid Naming, Math Attitude, Math Fluency, and Processing Speed. These predictors emphasize the critical role of processing efficiency and affective engagement in learning. For instance, rapid naming and processing speed are integral to phonological fluency and have been widely used in dyslexia screening tools [75]. Similarly, math-related constructs reflect both cognitive skill and learner motivation, resonating with recent findings that attitudes toward numeracy strongly influence math outcomes [74]; [80]. Other top-ranking variables included Vocabulary, Emotional Regulation, Phonological Awareness, and ADHD Score, reinforcing the multidimensional nature of LD risk. These features point to the importance of language development, executive functioning, and emotional-behavioral regulation, which can be monitored in school settings through teacher observation, language-rich instruction, and socio-emotional screeners.

Conversely, reading-related variables such as Reading Habits and Comprehension, while relevant, ranked moderately suggesting that their predictive power may be influenced by instructional quality and language exposure. Demographic factors like sex, class level, home language, and medical history appeared least important in the model, indicating that surface-level attributes are less predictive of LD risk than cognitive and behavioral factors. This supports needs-based intervention frameworks promoted by inclusive education policies [79]. Overall, this feature prioritization provides an empirical foundation for developing teacher-friendly screening tools centered on the most diagnostically relevant indicators-such as working memory, naming speed, math fluency, and emotional regulations enabling proactive and equitable intervention strategies [81]; [82]).

4.5.3 Practical and Theoretical Implications

The study's findings align with established theories that identify processing speed, working memory, and phonological/language skills as core predictors of academic risk ([5]; [2], reinforcing the validity of using these domain-relevant features for early identification. By pairing a high-performing ensemble model with transparent, educator-oriented explanation tools such as SHAP summaries and simplified decision-tree diagrams ([42]; [15], the proposed framework translates complex predictive outputs into actionable classroom information; an approach shown to increase interpretability and trust in model-driven decisions. This makes the system well-suited to low-resource contexts where specialist assessment capacity is limited, as it leverages routinely collected indicators to identify learners with learning difficulty for timely, targeted support while remaining lightweight enough for school-level deployment [4]. Furthermore, the study strikes a balance between technical and educational applicability by combining robust preprocessing and validation practices, including stratification, SMOTE applied within training folds, and cross-validated hyperparameter search [83], with interpretability-first outputs, ensuring that predictive accuracy is aligned with transparency and responsible use in classroom decision-making.

4.6 Psychometric Reliability and Statistical Significance of Predictors

To ensure the robustness of the learning difficulty (LD) risk classification model, comprehensive psychometric and statistical validation was performed on the input variables (see Table 4). Reliability analysis revealed poor internal consistency for the Academic Achievement domain (WRAT: $\alpha = -0.176$) and Language Development ($\alpha = -0.022$), suggesting weak inter-item correlations and limited construct coherence. Although these values fall below the recommended Cronbach's alpha threshold of 0.70, the corresponding chi-square statistics were statistically significant ($\chi^2 = 43.02$ and 45.25, respectively, $p < .001$), indicating sufficient item correlation for inclusion in multivariate analysis. These results suggest that while scale refinement is warranted, the constructs still contribute meaningfully to LD classification ([84]).

Further inferential tests confirmed the statistical relevance of key predictors. A one-way ANOVA revealed significant group differences in Working Memory ($F = 86.91$, $p < .001$), confirming its role in LD differentiation. Additionally, the chi-square test examining the association between sex and LD risk ($\chi^2 = 19.27$, $p < .001$) highlighted gender disparities, consistent with recent findings that boys are more likely to be diagnosed with LD due to behavioral or referral biases [85]; [86]. These results validate the inclusion of both cognitive and demographic variables in the proposed model.

Multinomial logistic regression results further supported the model's predictive strength, achieving a chi-square of 1251.07 ($p < .001$) and a pseudo R^2 of 0.5639. This indicates that over half of the variation in LD risk classification is explained by the included variables. Among the strongest negative predictors were Working Memory ($\beta = -1.78$), Executive Function ($\beta = -1.64$), and Processing Speed, which align with recent studies emphasizing cognitive deficits as key diagnostic indicators of LD [73]; [87]. Medical history emerged as a dominant positive predictor ($\beta > +5.24$), underscoring the value of integrating health-related variables in early screening. ADHD symptoms were also positively associated with moderate and high-risk groups, while academic features like math fluency and reading comprehension significantly influenced classification outcomes [88]; [89]. These findings support the multidimensional nature of LD risk and advocate for holistic screening tools that integrate cognitive, academic, behavioral, and contextual data.

4.7 Cross-Validation and Feature Contribution Analysis

To assess model stability and generalizability, a 5-fold cross-validation (see Appendix 6) was conducted across four classifiers. The Support Vector Machine (SVM) demonstrated superior performance with the highest mean macro-F1 score (0.9657, SD = 0.0086), followed closely by Random Forest (0.9570, SD = 0.0118) and XGBoost (0.9548, SD = 0.0117). In contrast, Logistic Regression yielded a substantially lower macro-F1 (0.7772, SD = 0.0206), suggesting its limited suitability for complex, multi-feature classification tasks in learning difficulty (LD) detection. These results affirm the effectiveness of kernel- and tree-based models in capturing non-linear and high-dimensional relationships among cognitive, behavioral, and demographic features [90].

Ablation analysis (Fig. 10) further clarified the relative importance of feature groups. Removing cognitive features consistently led to the largest decline in macro-F1 across all models, Random Forest (Δ = −0.0333), SVM (Δ = −0.0513), and XGBoost (Δ = −0.0227); underscoring their foundational role in LD classification. Language-related features also had a substantial impact, while emotional and demographic attributes contributed less. Notably, in XGBoost, removal of emotional (Δ = +0.0053) and demographic (Δ = +0.0088) features slightly improved performance, indicating potential noise or multicollinearity in those variables. These findings align with recent studies emphasizing the diagnostic primacy of cognitive metrics like working memory and processing speed in LD risk modeling [88]; [89]. They also support emerging best practices in educational data science that advocate for evidence-based feature pruning to enhance model interpretability and computational efficiency.

Figure 8: Feature importance of predictors

5. Conclusion and Implications

This study sought to develop a scalable, context-sensitive machine learning (ML) decision-support framework to aid in the early identification of learning difficulties (LDs) among basic school learners. Recognizing that many teachers in low-resource settings lack access to specialized screening tools or training, the proposed model was designed to be both practical and interpretable. Through the integration of psychometric, cognitive, emotional, and demographic indicators, the framework aimed to deliver timely insights to support proactive rather than reactive educational interventions. Among the models evaluated, the stacking ensemble demonstrated the strongest overall performance, achieving an accuracy of 96%, a macro F1-score of 95%, and an ROC-AUC of 99.7%. Additionally, cross-validation confirmed the stability of top-performing models such as XGBoost (mean F1 = 93.56%) and Random Forest (mean F1 = 93.32%), while feature ablation experiments revealed the critical role of cognitive and language-related variables in accurate classification. These findings highlight the value of combining domain-relevant features with robust ensemble techniques in building generalizable LD risk classifiers. Importantly, the interpretability of models like Random Forest, combined with validated constructs such as working memory, rapid naming, vocabulary, and ADHD traits, supports educational relevance and practical adoption.

The implications of this study are significant but should be interpreted with caution. While the model is not a substitute for formal diagnosis, it can serve as a useful early-warning system to guide teachers in identifying learners who may benefit from targeted support. Integration into school dashboards or learning management systems could enhance usability and facilitate timely interventions. Furthermore, the observed inconsistencies in certain psychometric scales, such as low Cronbach's alpha for WRAT and Language Development, underscore the need to adapt and validate these tools within the local context. As education systems across sub-Saharan Africa strive toward inclusive education goals, this study contributes a replicable and ethically grounded example of how ML can support these efforts. Future work should focus on field testing, tool refinement, and teacher training to ensure responsible implementation and sustainability.

6. Recommendation

Based on the study's findings, researchers recommend that the model's integration into educational practice should adopt a phased deployment strategy. This would begin with pilot implementation in a limited number of schools to assess operational feasibility, determine necessary contextual adaptations, and collect feedback from educators. Following the pilot, targeted teacher training should be provided through interactive workshops for teachers and school psychologists, focusing on practical interpretation of SHAP visualizations, decision paths, and feature importance rankings, enabling them to relate these insights to classroom observations [15]. The final phase should emphasize policy integration, involving collaboration with educational authorities to develop guidelines that ensure ethical use, protect learner data privacy, and establish clear referral pathways for flagged learners.

For future research, it is recommended that the model undergo external validation using datasets from varied academic years, geographic regions, and linguistic backgrounds to confirm robustness across diverse contexts [83]. Additionally, longitudinal studies should be conducted to track learners over multiple academic terms, assessing the model's ability to predict emerging risks at earlier stages and thereby enable proactive interventions. Finally, incorporating adaptive learning analytics and periodic model retraining will be essential to maintaining the model's accuracy, interpretability, and relevance as educational environments and learner needs evolve.

Declarations

Authors' Contribution:

Eric Opoku Osei (PhD.) supervised the research, provided conceptual guidance, and reviewed the manuscript.

Samuel Odoom and Enock Quansah Effah conducted the investigation, performed data analysis, and drafted the manuscript.

Seyram Dusu, Victoria Boafo, Anthony Kweku Appaih reviewed the manuscript for final submission.

All authors contributed to editing and approved the final version.

Acknowledgement:

Special appreciation goes to participants as well as management and staff of the Mampong Technical College of Education as well as all the Basic schools selected for the study for their support and cooperation.

Statements and Declarations

All the authors agree to be accountable for all aspects of the work, ensuring that questions related to the accuracy or integrity of any part of it are appropriately investigated and resolved.

Competing Interest

None of the authors have any competing interests and had no funding from any organizations, corporate body(s) or individual(s).

Data availability statement

Dataset and model’s code log for this study is available and can be accessed from google drive with this link.

Funding Declaration

No funding was received from any individual, group or institution for the study.

Clinical Trial Number

Not applicable

Consent to Participate Declaration

All human participants employed in this study freely gave their informed consent before participation. The researchers vividly and comprehensively explained the procedures and purpose of the study to all participants as well as respective stakeholders. In addition, participants were made aware of their exclusive right to withdraw from the study at any period without any penalty. Participation was absolutely voluntary. Participants were assured of confidentiality of their responses. All ethical guidelines for human subjects’ participation in a study were duly adhered to by this study.

Human Ethics Approval

The academics affairs and research units of Mampong Technical College of Education (ethics committee) and the authorities of the selected schools reviewed and approved the study. Thus, ethical clearance was sought and approved by ethics committee of the colleges of education. All procedures performed in the study in relation to human participation were duly executed in accordance with institutional and national research committee ethical standards.

References

1. Corteilla C, Horowitz SH. The State of learning disabilities: Facts, trends and emerging issues. Volume 3. New York: National center for learning disabilities; 2014. pp. 2–45.
2. Snowling MJ, Hulme C. Annual research review: Reading disorders revisited - The critical importance of oral language. J Child Psychol Psychiatry. 2022;63(4):451–72.
3. Hulme PE. Unwelcome exchange: International trade as a direct and indirect driver of biological invations worldwide, *One Earth*, vol. 4, no. 5, pp. 666 – 79, 2021.
4. UNESCO. Global Education Monitoring Report 2023: The State of Inclusion in Education. UNESCO Publishing; 2023.
5. Fletcher-Watson S, Adams J, Brook K, Charma T, Crane L, Cusack J, Leekam S, Milton D, Parr JR, Pellicano E. Making the future together: Shaping autism research through meaningful participation, *Autism*, vol. 23, no. 4, pp. 943–953, 2019.
6. Vaughn M, Jang BG, Sotirovska V, Cooper-Novack G. Student agency in literacy: A systematic review of the literature. Read Psychol. 2020;41(7):712–34.
7. Chugh R, Gupta N, Kumar S. Machine learning approaches in identifying learning disabilities among school children: A comparative study. Educ Inform Technol. 2021;26(6):7003–25.
8. Ahmed H, Mensah IA, Boateng R. Automated machine learning for identifying learning difficulties in e-learning systems. Comput Educ. 2023;194:10469.
9. Zhao L, Chen J, Zhang H. Dyslexia detection using speech and text-based CNN models. Comput Educ. 2022;3:100072.
10. Elbawab M, Henriques R. Machine Learning applied to student attentiveness detection: Using emotional and non-emotional measures. Educ Inform. 2023;28(12):15717–37.
11. Rojas-Sanchez MA, Palos-Sanchez P, Folgado-Fernandez JA. Systematic literature review and bibliometric analysis on virtual reality and education. Educ Inform Technol. 2023;28(1):155–92.
12. Oliveira GMMD, Brant LCC, Polanczyk CA, Malta D, Biolo A, Nascimento BR, Souza MDFMD, Lorenzo ARD, Fagundes Junior AADP, Schaan BD, Castilho FMD. Cardiovascular Statistics–Brazil 2021. Arquivos brasileiros de cardiologia. 2022;118:115–373.
13. Turan Z, Karabey SC. The use of immersive technologies in distance education: A systematic review. Educ Inform Technol. 2023;28(12):16041–64.
14. Landa RJ, Gross R, Stuart E, Bauman ML. Developmental trajectores of early vocabulary in autism. J Child Psychol Psychiatry, pp. 254–63, 2023.

15. Ifenthaler D, Yau JYK. Artificial intelligence in education: Challenges and opportunities and sustainable implementation. Br J Educ Technol. 2023;54(3):437–54.
16. Fuchs D, Fuchs LS, Crompton DL. The enduring role of RTI in identifying students with learning disabilities. Learn Disabil Res Pract. 2021;36(2):84–93.
17. Fletcher JM, Savage R, Vaughn S. A commentary on Bowers (2020) and the role of phonics instruction in reading. Educational Psychol. 2021;33(3):1249–74.
18. Zawacki-Richter O, Jung I, Xiao J. Von St. Vieth, A bibliometric review of research on artificial intelligence applications in education. Br J Edu Technol. 2023;54(2):225–48.
19. Yusof N, Wahab MHA, Azizan SN. Enhancing learner engagement through personalized gamification and AI-driven feedback in virtual learning environments. Comput Education: Artif Intell. 2024;5:100155.
20. Guler G. A bibliometric analysis on power analysis studies. J Meas Evaluation Educ Evaluation. 2023;14(3):235–48.
21. Li Q, Wang Y, Chen M. Early ADHD screening using XGBoost and behavioural data from educational settings. Comput Educ, p. 104646, 2023.
22. Tzafilkou K, Perifanou M, Economides AA, Nikolaidou M. A systematic review of immersive and interactive technologies in higher education. Educ Inform Technol. 2023;28:1289–312.
23. Liu L, Iketani S, Guo Y, Chan JFW, Wang M, Liu L, Luo Y, Chu H, Huang Y, Nair MS, Yu J. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. Nature. 2022;602(7898):676–81.
24. Muller W, Leyer M. Lean and employee-driven digital process innovation behaviour: the mediating role of how employees perceive the lean degree. Prod Plann Control pp. 1–16, 2025.
25. Chen H, Zhang Y, Xu Q. Machine learning based screening of reading disabilities in primary education. J Educational Comput Res. 2023;61(1):5–25.
26. Sweller J, Van Merrienboer JJG, Paas FGWC. Cognitive architecture and instructional design: Twenty years later. Educational Psychol Rev. 2019;31:261–92.
27. Paas F, Sweller J. An evolutionary upgrade of cognitive load theory: Using the human moter system and collaboration to support the learning of complex cognitive tasks. Educational Psychol Rev pp. 1245–68, 2021.
28. Fuchs J, Gartner H, Hellwagner T, Hummel M, Hutter C, Wanger S, E. Weber and, Zika G. IAB-Prognose 2021/2022, *Arbeitsmarkt auf Erholungskurs (No. 20/2021). IAB-Kurzbericht.*, 2021.
29. CAST. Universal Design for Learning Guidelines version 3.0., 2024. [Online]. Available: https://udlguidelines.cast.org
30. Rao K, Ok MW, Bryant BR. A review of research on Universal Design for learning models. Remedial Special Educ. 2021;2:91–106.
31. Willcutt EG, Petrill SA, Wu S, Boada R, DeFries JC, Olson RK. Comorbidity between reading disability and math disability: Concurrent psychopathology, functional impairment, and neuropsychological

functioning. J Learn Disabil. 2020;53(5):343–57.

32. Arnett AB, Pennington BF, Willcutt EG, DeFries JC, Olson RK. Sex differences in ADHD symptom severity. J Child Psychol Psychiatry. 2017;58(6):632–9.
33. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 6th ed. SAGE; 2018.
34. Alhassan ARK, Abosi OC. Teacher effectiveness in adapting instruction to the needs of pupils with learning difficulties in regular primary schools in Ghana. Sage Open. 2014;4(1):2158244013518929.
35. Agyei AA, Owusu-Dabo E, Addo B, Quansah E. Screaning for ADHD among primary school children in Ghana using the Disruptive Behavior Disorders Rating Scale (DBDRS). J Atten Disord. 2017;24(12):1684–94.
36. Wechsler D. Wechsler Intelligence Scale for Children - Fifth Edition (WISC-V), Pearson Clinical Assessment, 2020.
37. Torgesen JK, Wagner RK. Assessing phonological awareness and its role in reading development. In: Parris SR, Headley K, editors. Reading assessment and instruction for all learners. 3rd ed. Parris: Guilford Press; 2021. pp. 43–62.
38. Geron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media Inc; 2022.
39. Abolfathi B, Aguado D, Aguilar G, Prieto C, Almeida A, Ananna T, Anders F, Anderson S, Andrews B, Anguiano B, Aragón-Salamanca A. The fourteenth data release of the Sloan digital sky survey: First spectroscopic data from the extended baryon oscillation spectroscopic survey and from the second phase of the Apache Point Observatory Galactic Evolution Experiment. Astrophys J Supplement Ser. 2018;235(2):42.
40. Adadi A, Berrada M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Acess, pp. 52138–60, 2018.
41. Arrieta AB, Diaz-Rodriguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Herrera F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. Inform Fusion. 2020;58:82–115.
42. Lundberg SM, Erion GG, Chen H, DeGrave A, Prutkin JM, Nair B, Lee SI. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2:56–7.
43. Keller JD, Brian MN. Fundamentals of machine learning for predictive analytics. London: MIR Press,: and A. D'Arcy; 2015.
44. Sanusi HT. Network Intrusion Detection Using Deep Reinforcement Learning, 2023.
45. McFadden D. The measurement of urban travel demand. J Public Econ. 1974;3(4):303–28.
46. Peng P, Wang C, Namkung J. A meta-analysis of working memory deficits in children with learning difficulties: Is there a difference between verbal and numerical domains? J Learn Disabil. 2018;51(2):155–68.

47. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. PLoS Comput Biol. 2013;9(10):e1003285.

48. Holstein K, Wortman Vaughan J, Daume H III, Dudik M, Wallach H. Improving fairness in machine learning systems: What do industry practitioners need? in *Proceedings of the* 2019 *CHI conference on human factors in computing systems*, 2019.

49. Pardo A, Siemens G. Ethical and privacy principles for learning analytics. Br J Edu Technol. 2014;45(3):438–50.

50. Holmes W, Tuomi I. State of the art and practice in AI in education. Eur J Educ. 2022;57(4):542–70.

51. McCormick C, Barry DN, Jafarian A, Barnes GR, Maguire EA. vmPFC drives hippocampal processing during autobiographical memory recall regardless of remoteness. Cereb Cortex. 2020;30(11):5972–87.

52. Sideridis GD, Tsaousis I, Al-Harbi K. Identifying student subgroups as a function of school level attributes: A multilevel latent class analysis. Front Psychol. 2021;12:624221.

53. Sobotka SA, Lynch E, Quinn MT, Awadalla SS, Agrawal RK, Peek ME. Unmet respite needs of children with medical technology dependence. Clin Pediatr. 2019;58:11–2.

54. Kaiser MO. Kaiser-Meyer-Olkin measure for identity correlation matrix. J Roy Stat Soc. 1974;52(1):296–8.

55. de Miranda Junior UJP, Novaes MRCG, Santos HBA, Tatsch JFS, Ferreira RS, W. P. D. F. M., Vieira, Lapao LV. Perspectives of organizational identity in a health higher education institution: a mixed-method analysis, *BMC medical education*, vol. 21, no. 1, p. 51, 2021.

56. Alvarez R, Kallio E. Neurodevelopment risk factors and gender disparities in learning disability diagnosis. Comtemporary Educational Psychol. 2022;69:102053.

57. Li Y, Jian K, Sun Z, Chen Y. Construction and Validation of a Risk Prediction Model for Postoperative Urinary Tract Infection in Intracranial Hemorrhage Patients. Front Med. 2023;12:1552191.

58. Zhang Y, Xu B, Huang Z. The effectiveness of ensemble models in predicting learning performance in MOOCs. Comput Educ, p. 104487, 2022.

59. Azarnik A, Khosravi A. Optimized Intrusion Detection Using Harris Hawks Feature Selection and Deep Learning: A High-Performance Approach, in 2025 *11th International Conference on Web Research (ICWR).*, IEEE, 2025.

60. Hakkal S, Lahcen AA. XGBoost to enhance learner performance prediction. Computers Education: Artif Intell. 2024;7:100254.

61. Chen Y, Zhai T. Machine learning screening for autism using health claims data. BMJ Health Care Inf. 2022;29:e100544.

62. Yeh IC, Hwang CL. Evaluating support vector machine performance in multi-class educational classification. Appl Soft Comput J. 2023;123:109049.

63. Sezgin F, Kilinc AC, Ozdemir S, Ozdemir N, Erdogan O. Exploring the moderation role of teacher cultural value profiles on the association between transformational leadership and teacher job

satisfaction: Evidence from Turkish education context. Educational Manage Adm Leadersh, p. 17411432241237374, 2024.

64. Xia Y, Chen L, Wang J. Ensemble learning for early academic risk detection in elementary education. Comput Educ. 2025;190:104789.
65. Zhang Y, Kim D, Lee K. Explainable AI for diagnosing learning diffícuties: A machine learning approach. J Educational Comput Res. 2022;60(8):2031–50.
66. Zhao Y, Yang Y, Li H. Developmental trends and school-level interventions for learning difficulties: Evidence from a national sample. Educational Stud. 2022;48(4):521–40.
67. Singh V, Kapoor R, Raj S. Machine learning-based identification of learning disabilities in school-aged children: A survey-based approach. Educ Inform Technol. 2021;26:5435–52.
68. Abdallah MS, Aboona BE, Adam J, Adamczyk L, Adkins JK, Agakishiev G. Measurements of H Λ 3 and H Λ 4 lifetimes and yields in Au + Au collisions in the high baryon density region. Phys Rev Lett. 2022;128(20):202301.
69. Kim J, Choe J, Lee K. Sequential field development plan through robust optimization coupling with CNN and LSTM-based proxy models. J Petrol Sci Eng. 2022;209:109887.
70. Rodriguez MA, Torres D, Luna E. Bi-LSTM with attention for classifying learning difficulties from comprehension logs. Br J Edu Technol, pp. 365–81, 2023.
71. Agyeman MO, Oppong C, Amoako I. Predicting the academic success of Ghanaian high school students using LightGBM. Educ Inform Technol. 2022;27(4):4823–42.
72. Kim H, Lee J, Park S. Detecting learning difficulties using a hybrid CNN-LSTM from eye-tracking data. Computers Hum Behav. 2021;121:106788.
73. Alloway TP, Carpenter RK. The relationship among children's learning disabilities. working memory, and problem behaviours in a classroom setting: Three case studies. Educational Developmenta Pychologiest. 2020;37(1):4–10.
74. Carey E, Hill F, Devince A, Szucs D. Understanding mathematics anxiety: Investigation of student profiles in primary and secondary schools in the UK. Front Psychol. 2022;13:829032.
75. Norton EW, Wolf M. Rapid automatized naming (RAN) and reading fluency: Implications for dyslexia screening. Ann N Y Acad Sci. 2020;1482(1):59–73.
76. Hammer CS, Jia G, Uchikoshi Y. Language status and vocabulary development for dual language learners. Dev Psychol pp. 1002–16, 2020.
77. Bussing R, Mason DM, Bell LS, Porter E, Garvan CW, King S. Impact of ADHD symptoms on academic performance in children and adolescents: A longitudinal study. J Atten Disord. 2021;25(10):1430–40.
78. McClelland MM, Ponitz CC, Messersmith EE, Tominey SL. Self-regulation: Integration of cognition and emotion in classroom success. Early Child Res Q, pp. 1–20, 2018.
79. OECD. Equity in Education: Breaking Down Barriers to Social Mobility. OECD Publishing; 2021.

80. Dowker A, Sarkar A, Looi CY. Mathematics anxiety: What have we learned in 60 years? Front Psychol. 2019;10:341.
81. Zelazo PD, Blair CB, Willoughby MT. Executive function development and differences in educational trajectories. Ann Rev Psychol. 2021;72:687–718.
82. Grigorenko EL, Crompton DL, Fuchs LS, Wagner RK, Willcutt EG, M FJ. Understanding, educating, and supporting children with specific learning disabilities: 50 years of science and practice. Am Psychol. 2020;75(1):37.
83. Fernandez A, Salvador G, Herrera F, Chawla NVC. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J Artif Intell Res. 2018;61:863–905.
84. Field SN, Miles RM, ER WD. Linking heart health and mental wellbeing: Centering Indigenous perspectives from across Canada. J Clin Med. 2022;11(21):6485.
85. Catts HW, Bridges MS. Early identification of reading and learning disabilities: Advances in screening approaches. Reading Res Q. 2023;58(1):10–25.
86. Boets B, De Smedt B. Sex differences in learning disorders: Evidence from cognitive profiles and educational outcomes. Dev Sci. 2020;23(6):12974.
87. Peterson RL, Pennington BF, Willcutt EG. Neurocognitive mechanisms of learning disabilities: Update and review. J Child Psychol Psychiatry. 2021;62(5):614–31.
88. McGill RJ, Styck KM, Harrison A. Predicting learning disability status using cognitive and academic indicators: A machine learning approach. School Psychol Rev. 2023;52(2):208–25.
89. Pimperton H, Nation K, Hulme C. A transdiagnostic approach to early reading and learning difficulties. Developmental Sci p e13412, 2024.
90. Koutras L, Ntoliou I, Doulgeri Z. Towards Passivity Based Nonprehensile Bimanual Manipulation of Large Objects, in 2023 *IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids).*, IEEE, 2023.
91. Bjureber J, Ljotsson B, Tull MT, Hedman E, Sahlin H, Lundh LG, Gratz KL. Development and validation of a brief version of the Difficuties in Emotion Regulation Scale: The DERS-16. J Psychopathol Behav Assess. 2016;38(2):284–96.
92. Chen Y, Chen Q, Kong L, Liu G. Early detection of autism spectrum disorder in young children with machine learning using medical claims data. BMJ Health Care Inf, p. e100544, 2022.
93. Diamond A. Executive functions. Ann Rev Psychol, pp. 135–68, 2020.
94. Gupta R, Sharma A, Singh R. The enduring role of RTI in identifying students with learning disabilities. Learn Disabil Res Pract. 2021;18(1):32.
95. Kuhn M, Johnson K. Applied Predictive Modelling. Springer, 2013.
96. Landerl K, Wimmer H, Frith U. Phonological processing and reading fluency in dyslexia. Cognit Neuropsychol. 2019;36(5):353–72.

Figures

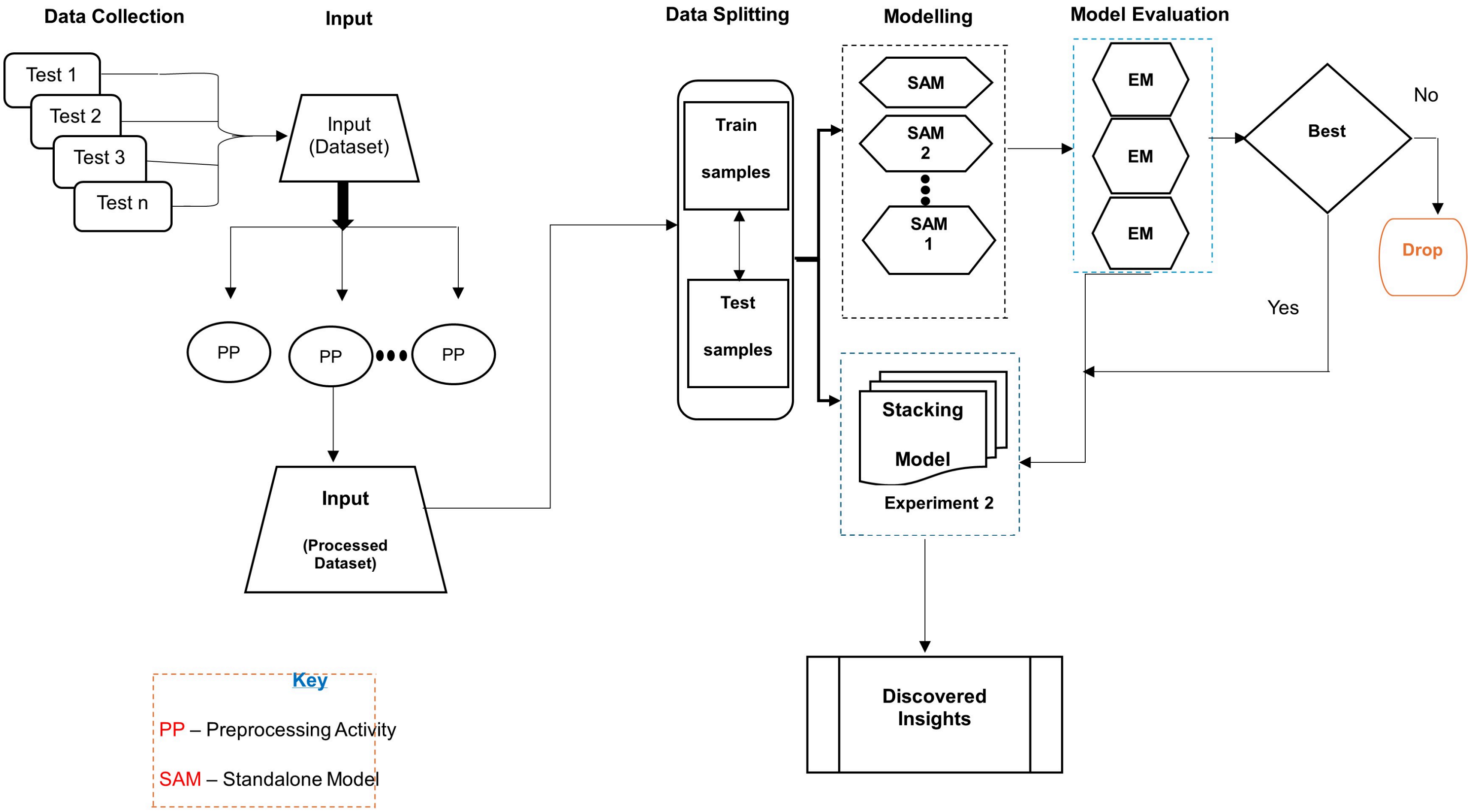

Fig 1: Framework of proposed model

Figure 1

See image above for figure legend.

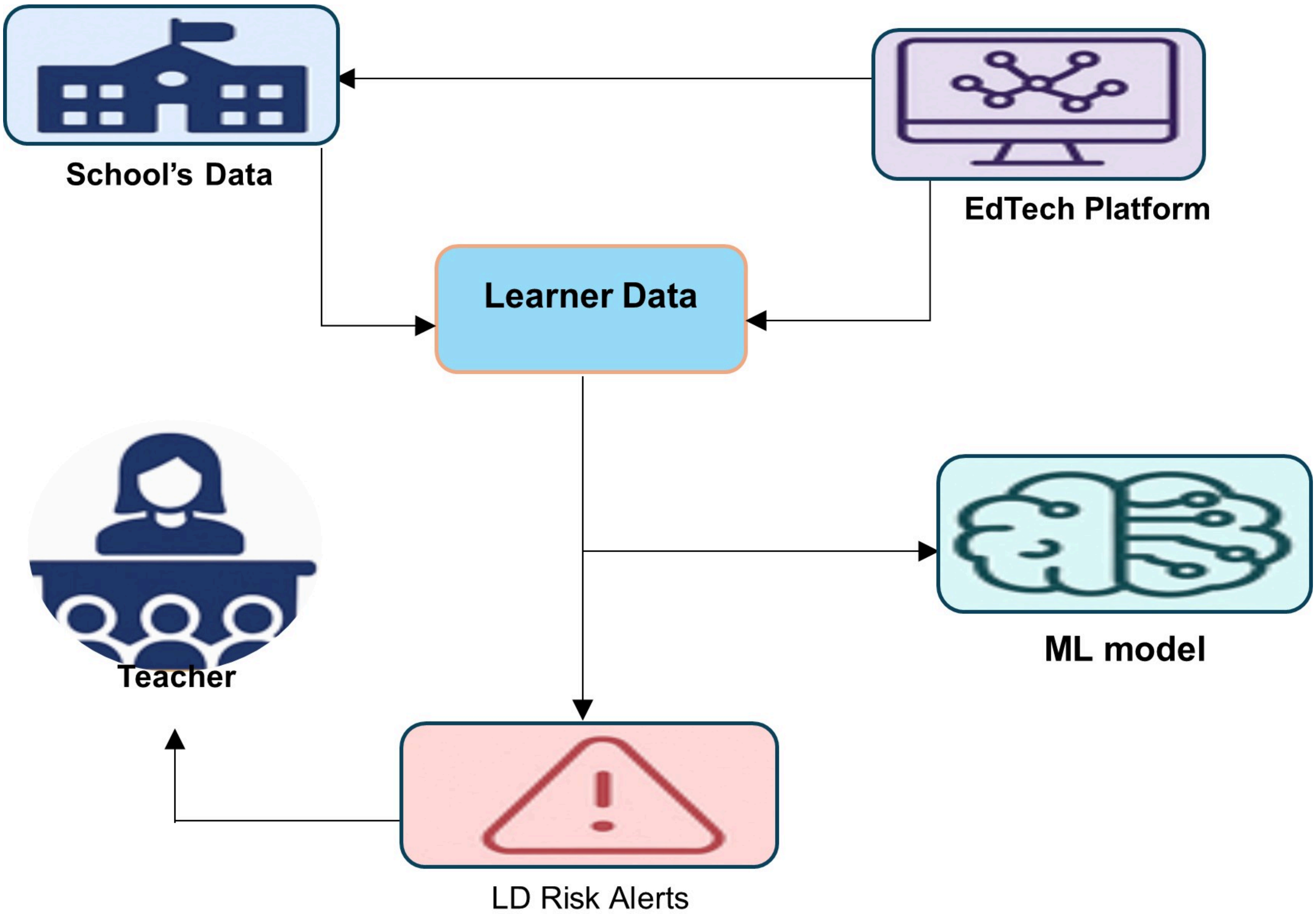

Fig 2: Proposed Model's integration framework

Figure 2

See image above for figure legend.

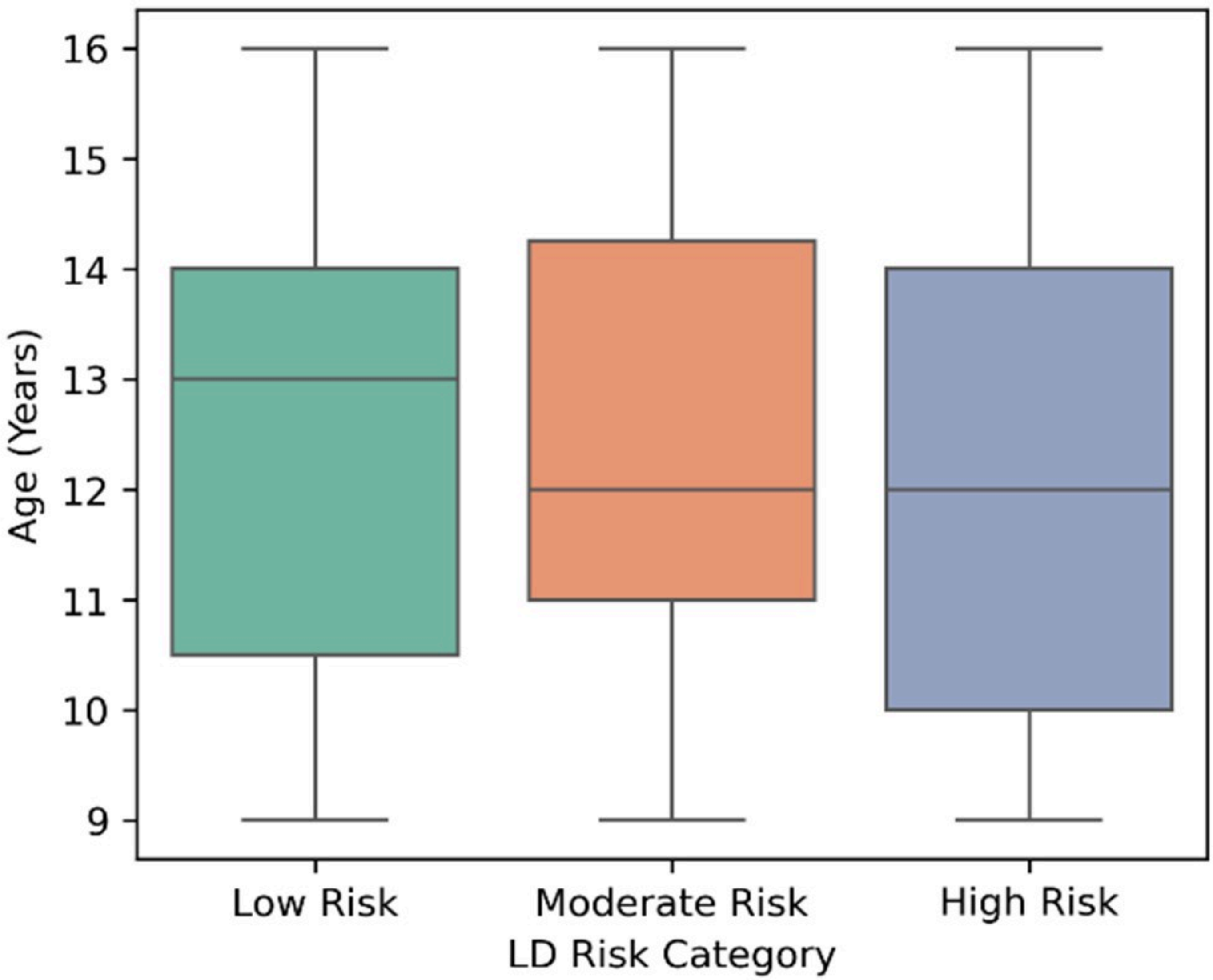

Fig 3: Age distribution by LD Risk

Figure 3

See image above for figure legend.

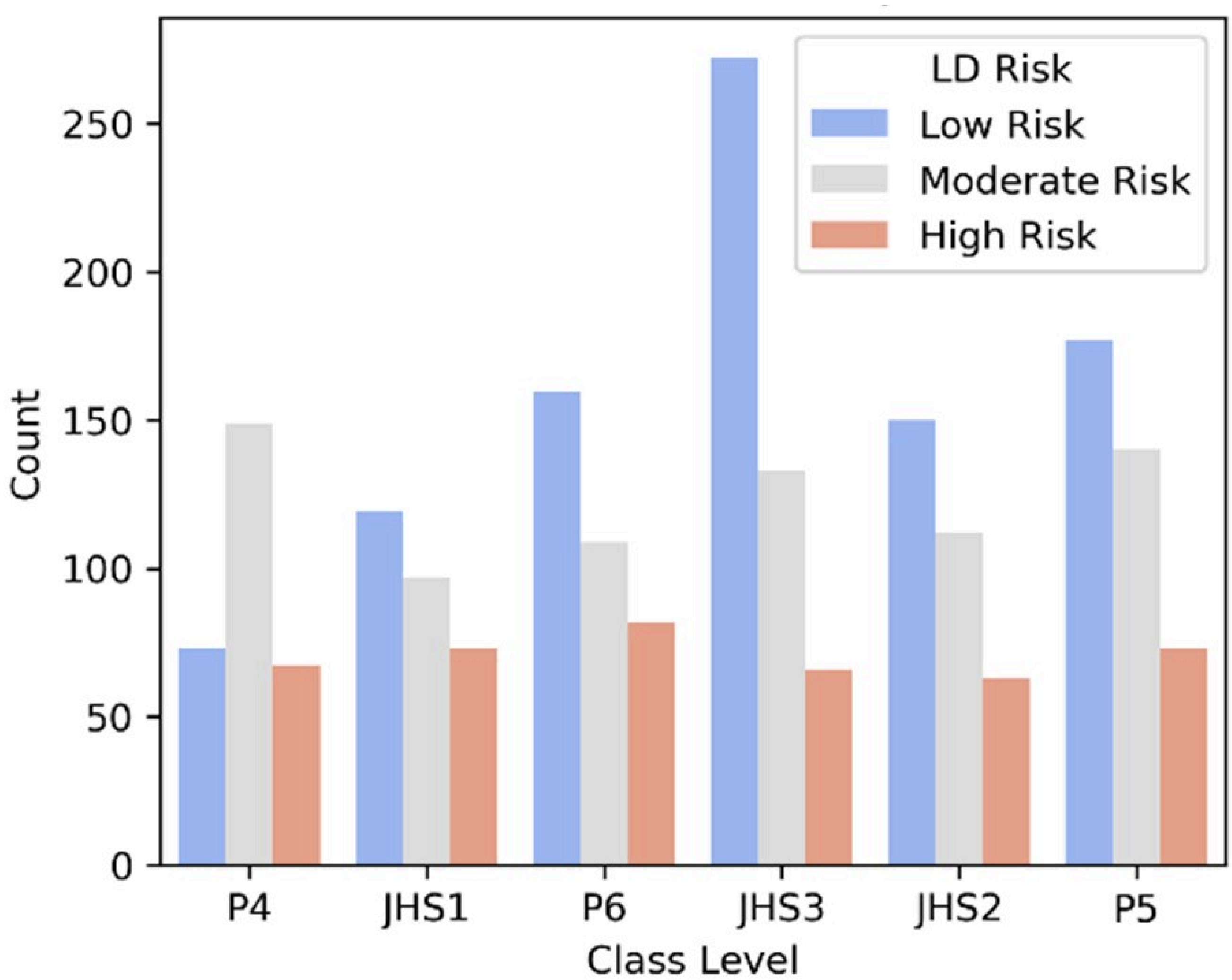

Fig 4: Class Level distribution by LD Risk

Figure 4

See image above for figure legend.

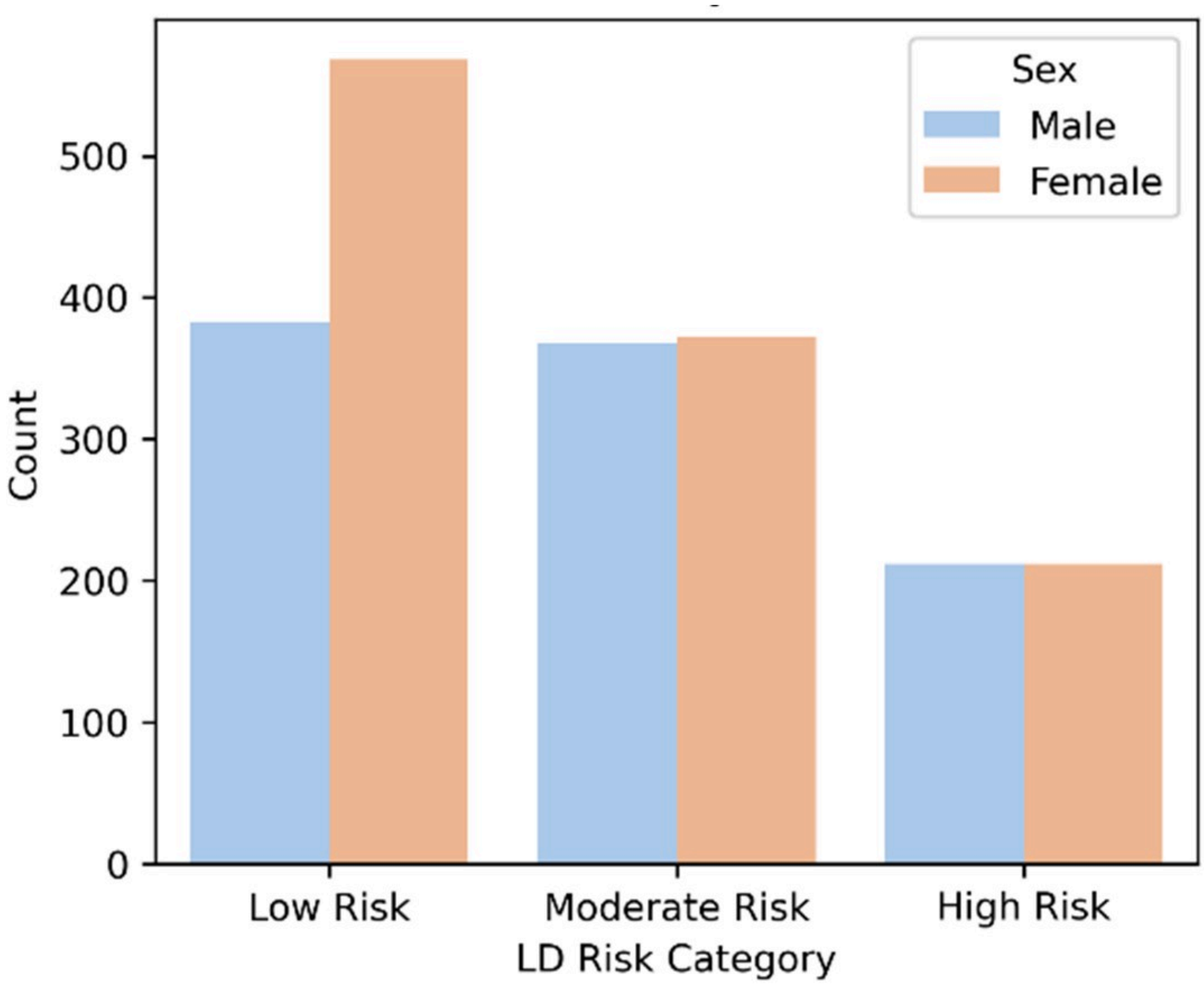

Fig 5: Sex distribution of participants by LD Risk

Figure 5

See image above for figure legend.

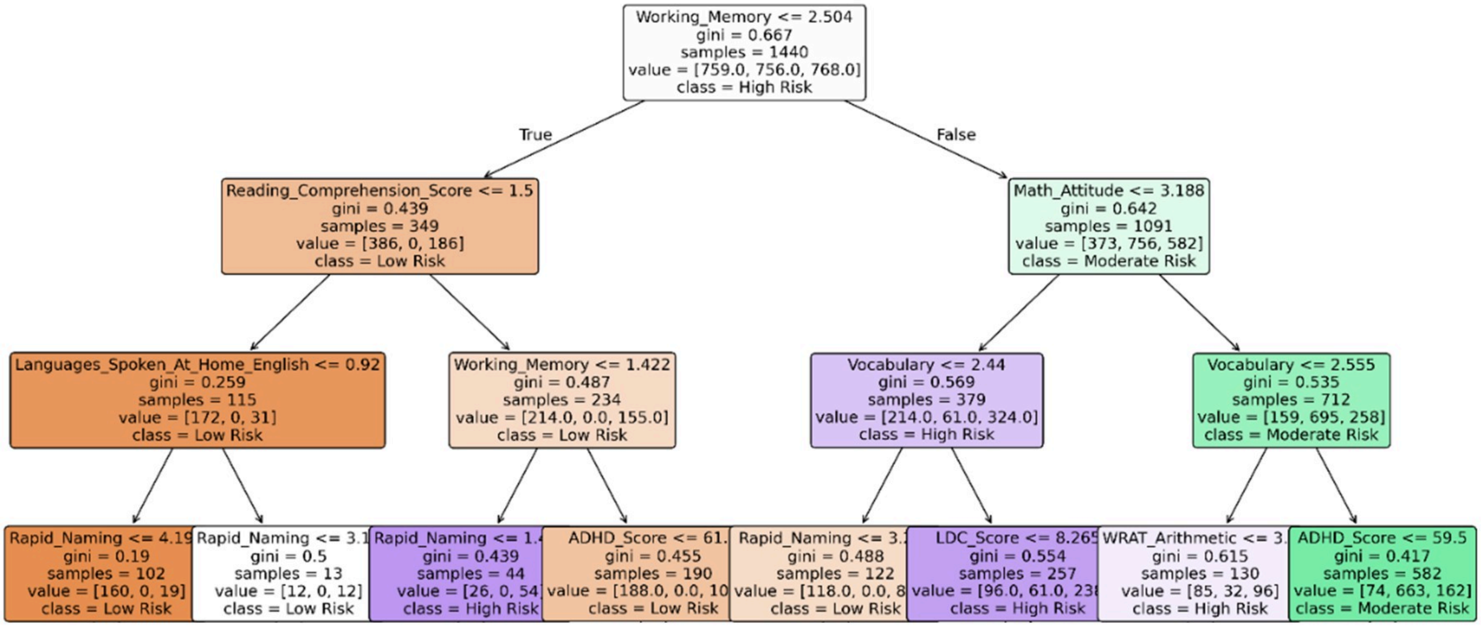

Fig 6: Decision tree showing relationship of input features for prediction

Figure 6

See image above for figure legend.

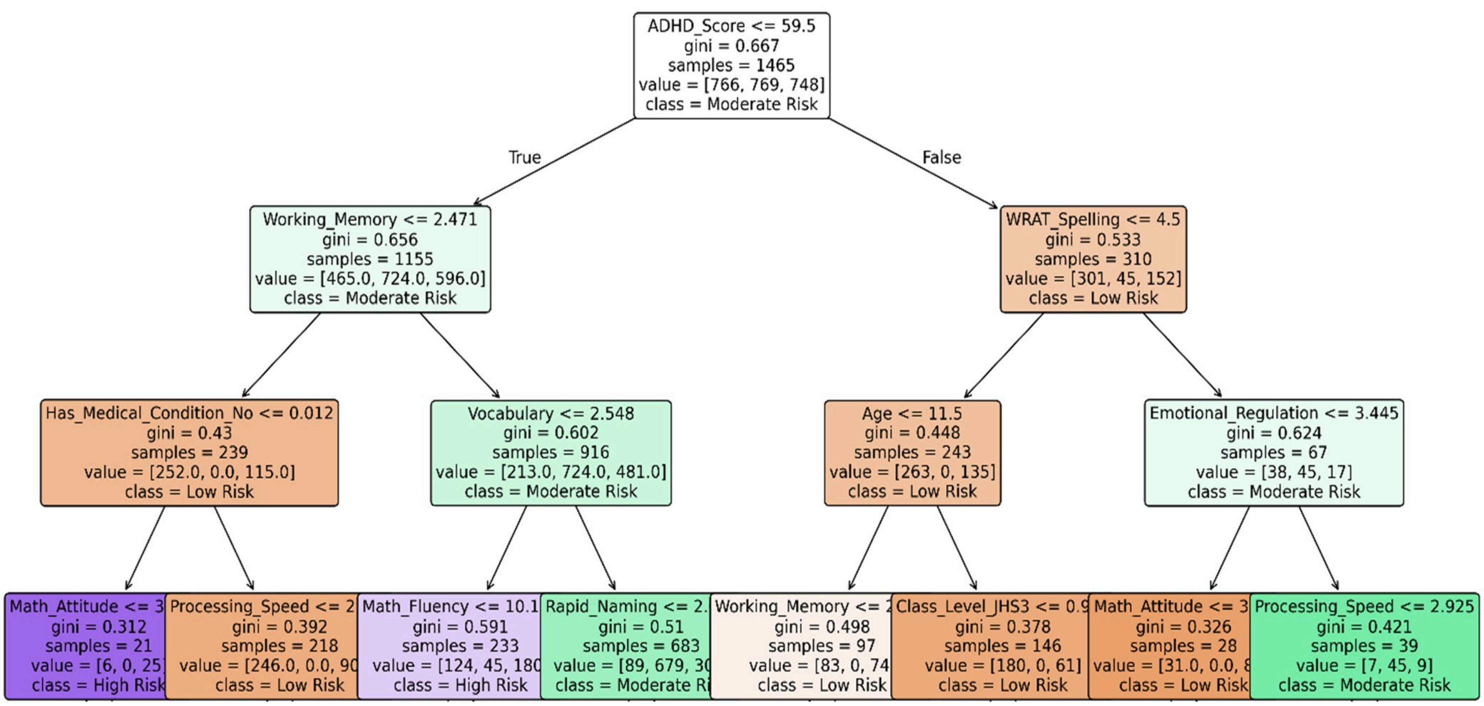

Fig 7: Relationship of input features for prediction from Decision tree

Figure 7

See image above for figure legend.

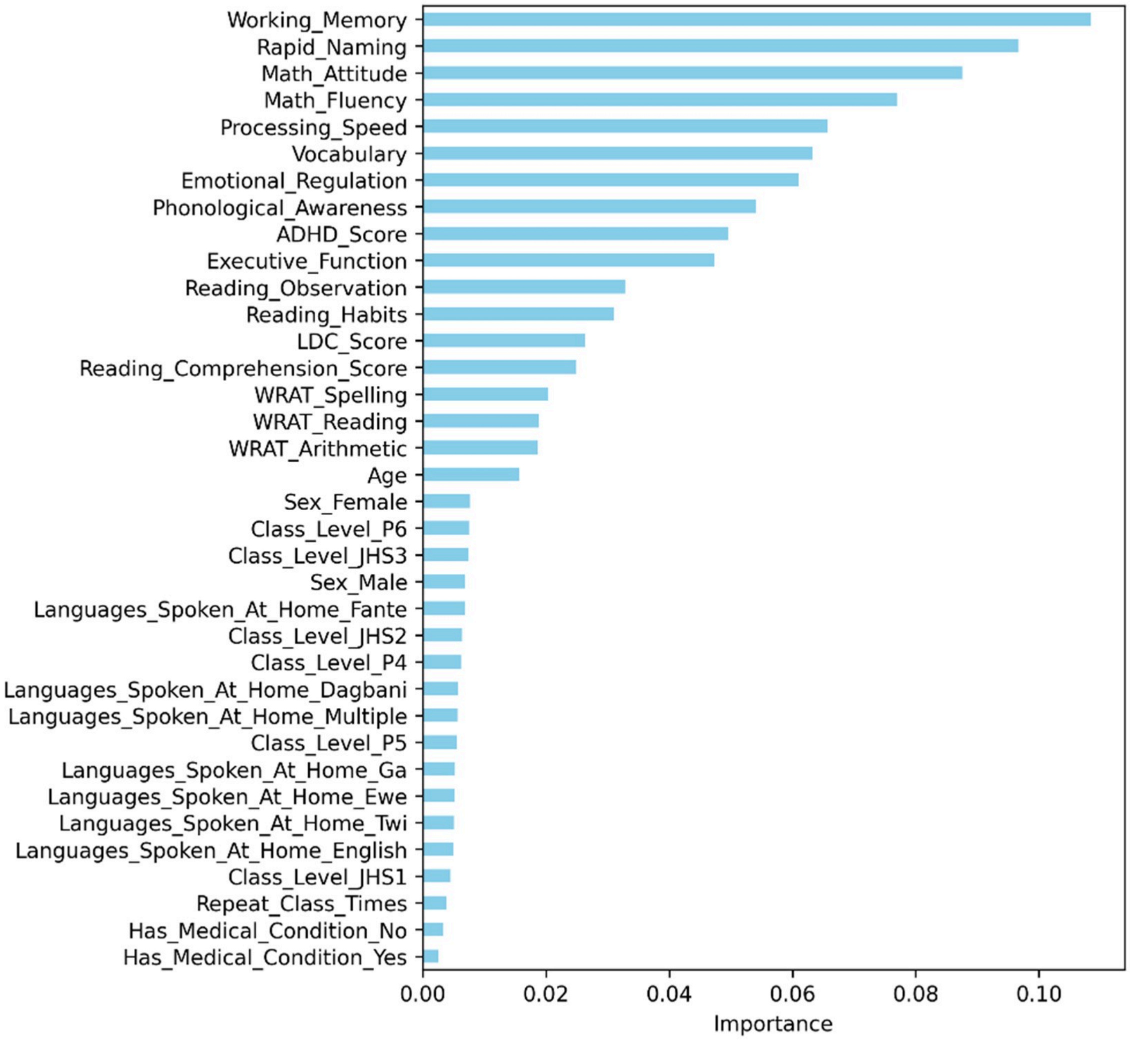

Fig 8: Feature importance of predictors

Figure 8

See image above for figure legend.

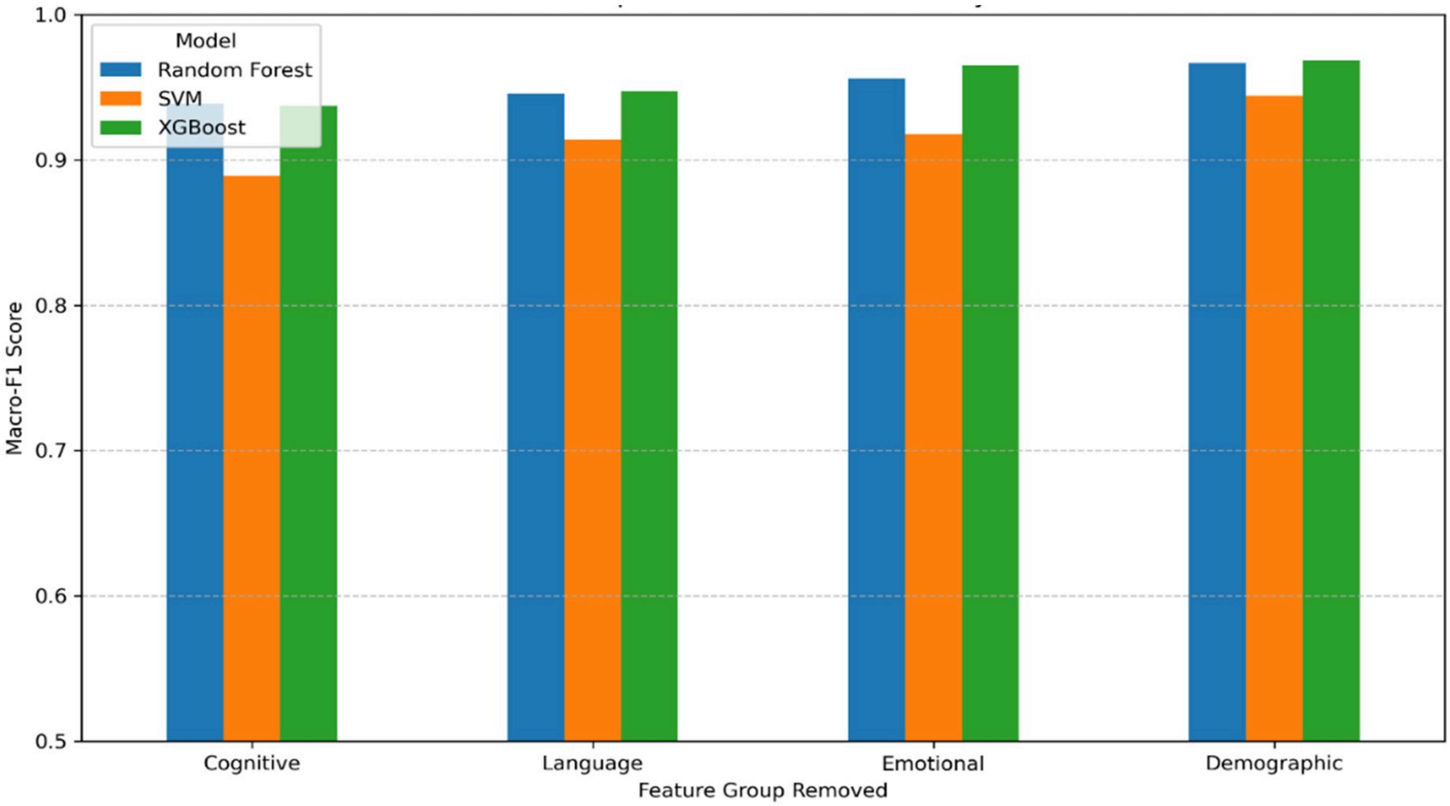

Fig. 9: Features Ablation Macro-F1 score by model

Figure 9

See image above for figure legend.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendices.docx