ISSN: 2093-5374 / E-ISSN: 2093-5382

Explainable AI Framework for Predicting Learning Disabilities Using Random Forest

P.J. Anu[1*], and K. Ranjith Singh[2]

[1*]Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education,
Coimbatore, Tamil Nadu, India. anujohnson123@gmail.com,
https://orcid.org/0009-0002-6245-297X

[2]Assistant Professor & Research Guide, Department of Computer Science, Karpagam Academy of
Higher Education, Coimbatore, Tamil Nadu, India. ranjithsingh.koppaiyan@kahedu.edu.in,
https://orcid.org/0000-0003-2651-2509

Received: February 28, 2026; Revised: April 08, 2026; Accepted: May 25, 2026; Published: June 30, 2026

Abstract

Accurate identification of Learning Disabilities (LD) at an early age is vital for timely intervention, individualized learning plans, and optimal learning results. Conventional evaluation techniques like clinical evaluations and teacher observations can be time-consuming, subjective, and may not be able to identify subtle trends in academic and behavioral performance. This work suggests an explainable AI framework that combines real-time context-aware deployment in edge and mobile platforms with a Random Forest (RF) classifier to predict if a student is at risk for identification as an LD. A total of 348 student records containing academic, cognitive, behavioral, and demographic characteristics. To build trustworthy models, categorical encoding, feature standardization, and class balancing (RandomOverSampler) were performed. Multiple machine learning models, such as Random Forest, Decision Tree, K-Nearest Neighbor, Extra Trees, and Support Vector Machine, were trained and tested using 3-fold cross-validation. The overall best predictive performance was given by the Random Forest model with 96.36% accuracy, 96.00% F1 Score, 98.30% ROC-AUC, and 92.71% Cohen's Kappa, showing the reliability and stability of the model in classification tasks. In order to be interpretable, Shapley Additive exPlanations (SHAP) were computed, and academic performance, writing ability, and impulsivity were found to be the strongest predictors. Local SHAP explanations helped teachers get knowledge of the factors that are putting each student at risk, providing teachers with transparency and informing intervention. It is feasible also for ubiquitous and robust deployment in distributed education with the monitoring and decisions being done in real time in the edge or mobile device. This framework described here is capable to support the future deployment on mobile and edge computing environments by the computational efficiency of the under lying machine learning algorithms.

Keywords: Learning Disabilities, Explainable Artificial Intelligence (XAI), Random Forest, SHapley Additive exPlanations (SHAP), Educational Data Mining; Machine Learning, Student Performance Analysis, Early Learning Disability Detection.

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 17, number: 2 (June-2026), pp. 614-629. DOI: *10.58346/JOWUA.2026.I2.034*

*Corresponding author: Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India.

1 Introduction

Learning Disabilities (LDs) are prevalent across the globe, impacting children's academic performance, social interactions and development. Early identification ofLDris crucial for implementing timely interventions, tailored educational approaches and better learning outcomes Abakasanga et al., (2025). Conventional techniques for LD detection including clinical evaluations and teacher assessments, are known for being lengthy, subjective and sometimes lacking detection of subtle nuances in a student's behavior (Dandolo et al., 2023; Scarpazza et al., 2020). With recent progress made in the field of machine learning, new techniques are emerging to build predictive models, which can identify at-risk students through multivariate data sets collected within an educational environment. Early recognition of LD in students is essential to implementing prompt interventions, creating an individualized learning plan, and achieving a higher success in learning. Not only do the study strive for accuracy, the system of education today has also become more distributed with features such as IoT-connected classroom interfaces and mobility to allow real-time responses from the predictive models according to the context. With the embedding of ML models on the edge devices, data is processed on the edge rather than cloud services thereby reducing the latency and improving the reliability while maintaining data privacy. Among these techniques Random Forest classifiers have become popular due to their strengths regarding non-linearities and resistance to overfitting Teles et al., (2021). Black box-ness, however, still remains an issue in many ML approaches thus restricting their application in the educational domain.

This study proposes a blend of interpretable predictions from Random Forest classifier and universal, interactive SHapley Additive exPlanations (SHAP). The educators will be able to understand the way academic, cognitive and behavioral factors contribute to the prediction and react to it promptly. This system abides by reliable, context-aware and adaptive computing principles which are extendable to actual, distributed learning systems. This paper combine the SHAP with the Random Forest model to achieve both high prediction performance and explainable feature. Preprocessing steps in the model architecture is the order of categorical encoding, class imbalance is dealt by RandomOverSampler and then the feature scaling is applied. Compare performance between various ML models(Random Forest, Decision Tree, KNN, Extra Trees and SVM) while SHAP values explain that what features are affecting on LD and providing reasons to do education.

The impact of this work is in that an accurate, interpretable, and scalable AI framework can encourage an earlier diagnosis of learning disabilities. This approach combines machine learning and explainable AI so that teachers, psychologists, and school systems can identify the students that need support from their school. With the help of SHAP explanations the model can be explained and understood. Furthermore, it is capable of supporting education data analytics that can benefit student learning within the school.

Key Contribution

- An explainable AI framework has been designed for learning disability prediction utilizing a Random Forest classifier and SHAP analysis.
- The model was trained and tested on actual student data comprising 348 students, which was gathered from educational institutions in five districts of Tamil Nadu state in India.
- The developed model gave optimal predictive accuracy, such as 96.36 % accuracy, 96.00 % F1 score, and 98.30 % ROC-AUC for effective learning disability prediction.
- Both global and local explanations were generated by using SHAP analysis, which provides transparency for predictions for better remedial actions in education.

This work is organized as follows. Section 2 demonstrates the literature review, and Section 3 provides a detailed description of the dataset, data acquisition, data preprocessing, model training, and the framework of interpretable explanations by SHAP values. Section 4 evaluates the results of the experiment, evaluates the model's performance, analyzes the importance of each feature, and explains the predictions. Section 5 concludes the study with future directions.

2 Literature Survey

The development and use of artificial intelligence and machine learning in learning disabilities and related neurodevelopmental disorders have improved dramatically in recent years. One study examined the performance of machine learning algorithms in predicting student outcomes for individuals with learning disabilities and demonstrated the effectiveness of data-driven prediction in education and health settings (Abakasanga et al., 2025). In addition, one work applied explainable machine learning in a predictive context with respect to attention deficit hyperactivity disorder, highlighting the importance of interpretable models to decision making Navarro-Soria et al., (2025). A systematic review found that AI-powered tools were promising in improving learning support and intervention strategies for students with learning disabilities (Papalexandratou et al., 2024).

The role of interpretable machine learning in educational settings has received increasing attention. One study utilized XAI to identify factors predictive of academic performance in both disabled and non-disabled learners, showing the utility of interpretable models in educational analytics Larasati, 2025; Han & Wang, 2023). Another work introduced an explainable machine learning approach to predict academic performance for deaf scholars, indicating the utility of interpretable models to instructors (Raji et al., 2024; Malakouti, 2025). Later, explainable artificial intelligence techniques have proved their utility in predicting the performance of learning without losing trust and transparency (Sanfo, 2025).

Random Forest is acknowledged as one of the best performing and strongest classification algorithms and the high classification accuracy has been achieved by ensemble of decision trees with reduction of variance (Sun et al., 2024). The earlier work described the application of tree-based learning for predicting student performance and demonstrated the feasibility of tree-based models for student data (Matzavela & Alepis, 2021). Research has also shown that ensemble-based approaches outperform traditional classification methods (Teles et al., 2021).

The explanation of model predictions has become an integral aspect of machine learning applications in education. One work developed a practical framework to explain machine learning predictions with SHAP analysis, whereas another showed that while SHAP and LIME were similar in nature, SHAP always provided consistent explanations regardless of the algorithm (Ponce-Bobadilla et al., 2024; Hasan, 2023). SHAP was identified as one of the most commonly applied explainable artificial intelligence techniques and was cited for its ability to provide global and local explanations (Saarela & Podgorelec, 2024). Another study demonstrated the effectiveness of a SHAP-based analysis for identifying influential features in prediction models and increasing transparency (Wang et al., 2024).

Although machine learning and XAI have shown great promise in predicting various outcomes relevant to disability and education, few studies have sought to integrate Random Forest and SHAP to specifically predict learning disability outcomes. Moreover, most existing literature focuses solely on prediction accuracy without adequately interpreting these models to guide student interventions at the individual level. This study therefore proposes a framework for explainable learning disability prediction utilizing Random Forest and SHAP, yielding both high classification accuracy and transparency into underlying predictors for informing student education.

3 Method

Dataset Description and Data Collection

In this work, 348 student records were collected from educational institutions across 5 districts (Coimbatore, Tiruppur, Erode, Nilgiris, Karur) in and around Coimbatore in Tamil Nadu, India. The data were collected by means of a well-structured questionnaire in Google Forms administered to educational institutions and teachers working in the five selected districts of Tamil Nadu, India. This questionnaire sought demographic, cognitive, behavioral, and academic characteristics that are commonly considered important while diagnosing learning disabilities. The individual student records contained information such as academic performance, reading ability, writing ability, mathematics ability, memory, attention, hyperactivity, impulsivity, language ability, and other cognitive and behavioral measures pertaining to the learning disability problem. Each student record was classified as learning disability (LD) or non-learning disability (Non-LD) as indicated by educators and institutional records. The dataset of 348 records was split into a train-set and a test-set in an 80:20 ratio; the train-set has 278 records, and the test-set has 70 independent records.

A standard data collection procedure was maintained. Google Form link was circulated electronically to the various institutions. Data were collected over a fixed duration. On submission, data were checked for completeness and consistency before being appended to the dataset. Redundant records, incomplete data, or records with substantial missing attributes were removed from the final dataset. Only records with complete data were kept, yielding a total of 348 student records in the dataset. Various types of educational institutions within these districts were taken into consideration, yielding a heterogeneous sample. Demographic factors and various academic, cognitive, and behavioral factors were included in the dataset to develop a prediction model. Privacy of the participants and their data was maintained, as it was a study for purely academic purposes, and no personally identifiable information was maintained, and the data collected were anonymized before the analysis

Data Splitting, Oversampling, and Cross-Validation Procedure

To maintain data integrity and to ensure that these models were evaluated on unseen data, data preprocessing and training were performed separately. First, the data was cleaned by dropping all unnecessary attributes, and then all categorical data was converted into a numerical format. The full data set was then split into a training and testing set in a ratio of 80:20, respectively, so that the training set would consist of 80% of the records, and the test set, consisting of the other 20% of the records, was left strictly for testing.

Only the training data was sampled using the RandomOverSampler. The testing set was untouched during sampling, as it needs to consist of unseen data for an appropriate estimate of model performance, and sampling data that could appear in both the training set and test set would only mean that a duplicate sample would appear twice during training.

The models were evaluated using a 3-fold cross-validation where each fold sampled the training set using the RandomOverSampler method, while the validation set remained unsampled so as not to allow information from samples used in validation to be leaked to the training algorithm.

For the purpose of Binary classification, each record was categorized as Learning Disability (LD) or not LD. After carrying out the train-test split and class balancing as aforementioned in the earlier part of the text, feature standardization was performed using the StandardScaler class. Then, training and 3-fold cross validation of different machine learning models like Random Forest, Decision Tree, Extra Trees, K-Nearest Neighbors, and Support Vector Machine were carried out on the training set. The selected model was finally tested on the test set using the following metrics: Accuracy, Cohen's Kappa, confusion matrix, classification report, precision-recall curve, and ROC-AUC (Kim et al., 2024).

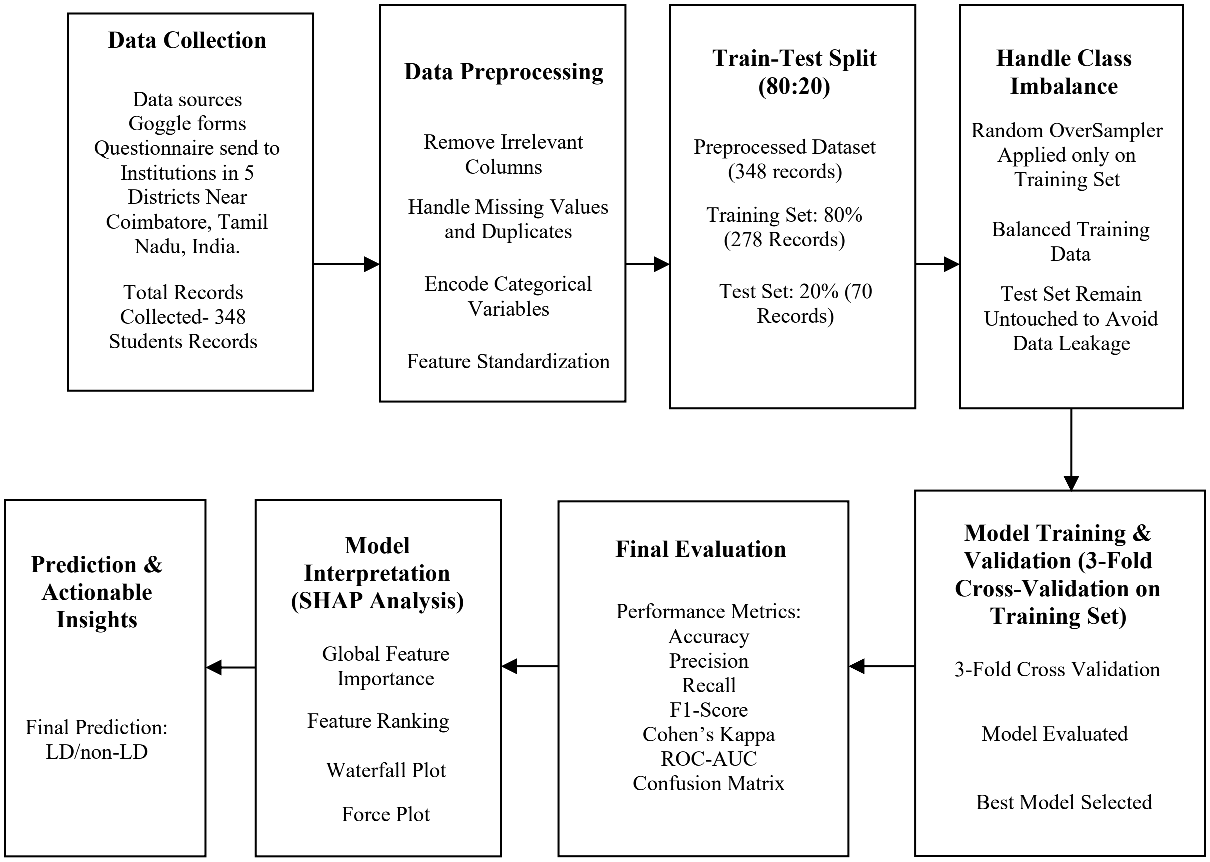


Figure 1: Architecture of the explainable AI-based learning disabilities prediction framework

The overall architecture of the proposed learning disabilities prediction framework is displayed in figure 1. Data is initially acquired from the educational institution through a survey with a structured Google Form questionnaire. The data collected from the questionnaires is cleaned, categorized, and normalized during data preprocessing. Then the preprocessed dataset is separated into a training set and a testing set in an 80:20 ratio. Since it is important not to have data leakage from a portion of the data to another, this study only use RandomOverSampler on the training set to manage the imbalance in the data. This study then evaluate five different machine learning algorithms such as Random Forest, Decision Tree, Extra Trees, K-Nearest Neighbors, and Support Vector Machine with 3-fold cross-validation. This study pick the model that best performs in terms of selected metrics and test it against a new data set. SHAP explanations are generated to interpret the output predictions and highlight the significant features which contribute to the learning disabilities classification, so that educational institutions can come to an actionable decision.

Hyperparameter Configuration

In order to maintain repeatability of results and transparency, hyper-parameters chosen for each of the machine learning classifiers were recorded and were kept constant during the experiments. Random forest was used with 100 decision trees (n_estimators = 100), Gini impurity criterion used for node splitting and randomstate was fixed to 42. In case of Decision Tree, Gini criterion was used, and maximum depth was unrestricted. In Extra trees, 100 estimators were used and same random_state. KNN used k = 5 neighbors and Minkowski distance with uniform weight. SVM used RBF kernel, and C = 1.0 and gamma = scale. These values were chosen on standard practice followed in machine learning.

Overfitting Analysis and Model Generalization

The reason for such a good classification performance has been analyzed and investigated whether it occurs due to overfitting. Random Forest performance on both training and testing dataset was analyzed. The performance of the model was consistently high over cross validation folds and testing data and remained relatively stable which indicate the good generalization. Training set-testing set was divided by 80:20 and model was tested on testing set which have not seen before. Also class balancing RandomOverSampler was performed on the training set only (after the split between training set and testing set), in order to prevent data leakage from testing set to training set. Three-fold cross-validation has been applied over the training set during model development to decrease the variance and to get the good assessment of prediction. Although the performance measured by accuracy is relatively high, stability of precision, recall, F1-score, Cohen's Kappa and ROC-AUC value on the validation folds demonstrated that the model is learned the features from educational data and not the individual sample.

Classification Techniques Used for Prediction

In this study, several classification techniques were explored for the prediction of learning disabilities. Decision Trees (DT) (Matzavela & Alepis, 2021) are hierarchical models that recursively partition data based on selected features until reaching a stopping criterion, with class labels assigned at the leaf nodes. Random Forests (RF) Sun et al., (2024) extend this by creating an ensemble of decision trees trained on bootstrapped datasets and random subsets of features, and predictions are made using a majority vote across the trees. In a similar fashion, Extremely Randomized Trees (ET) Mim et al., (2025) also uses an ensemble of trees. ET adds another layer of randomness through selecting a split threshold randomly as well, increasing the diversity within the forest even more. The KNN Çetin & Büyüklü, (2025) algorithm differs from both the tree-based models and SVMs by storing the training data and assigning a class to the nearest neighbors, commonly calculated using some distance metric, in the feature space. Finally, SVMs Pimentel et al., (2024) determine an optimal hyperplane that gives the largest separation between two classes and classifies points according to which side of the hyperplane they lie on. These algorithms, in turn, give a broad spectrum of classification methods which weigh interpretability, robustness, and power against each other. Overall, comparing all of the machine learning algorithms, RF seemed to exhibit best performance overall and most robustness, thus it was chosen as the final model, on which SHAP was used to determine explanations for the teachers and psychologists.

SHAP (SHapley Additive exPlanations)

To make the predictive model interpretable,SHAP (SHapley Additive exPlanations)was carried out.SHAP is based on the game theoretic concept of Shapley values, which equally assigns to each feature the weight of a feature based on averaging over all possible combinations of the features (Pimentel et al., 2024). The SHAP was employed in order to evaluate global feature importances (general predictive features of learning disabilities) and local interpretability (how much specific student predicted learning disabilities or not) of the model. SHAP values estimate the marginal contribution of each feature to prediction compared to baseline, whereas positive values will lead to higher prediction of learning disability and negative values to a protected impact (Saarela & Podgorelec, 2024).

4 Results and Discussion

In this section, the experimental outcomes achieved when conducting an analysis on the learning disability data-set were produced and interpreted. First, five machine learning models: Support Vector Machine (SVM), Decision Tree, Random Forest, Extra Trees and K-Nearest Neighbors were implemented and a number of tests conducted in relation to predictive consistency and accuracy between the 5 classifiers. The Random Forest classifier achieved the highest result and was consequently chosen as the best model.

After the best performing model was produced, SHAP was applied to the Random Forest classifier to explore and predict its outputs. SHAP analysis provided global feature importance rankings, where writing ability and academic skills were found to be the most influential factors, followed by impulsivity, hyperactivity, and memory-related attributes. Furthermore, local explanations were generated to highlight how specific features impacted individual predictions, thereby improving transparency of the model's decision-making process.

For illustrating prediction interpretability, one of the student records was passed through the trained Random Forest model.

Comparison of Classification Models

All the models were assessed using 8 evaluation metrics: Precision, Recall, F1-Score, Cohen's Kappa, Average Precision (AP), ROC-AUC, TPR (True Positive Rate), FPR (False Positive Rate), and accuracy (Kalita et al., 2025; Sanfo, 2025). Table 1 shows the results of various parameter calculations done in the models for the evaluation. All performance figures are presented as percentages, with all figures rounded to 2 decimal places.

Table 1: Performance evaluation metrics of the classification models

Algorithm	Precision	Recall	F1-Score	Kappa value	AP	ROC-AUC	TPR	FPR	Accuracy
Random Forest	97.00	96.00	96.00	92.71	99.00	98.30	93.33	0	96.36
Decision Tree	96.00	95.00	95.00	90.79	95.00	95.10	90.19	0	95.45
KNN	97.00	96.00	96.00	92.64	96.00	96.10	92.15	0	96.36
Extra Tree	96.00	95.00	95.00	90.78	100.00	99.70	93.65	.021	95.45
SVM	80.00	80.00	80.00	60.01	86.00	85.60	78.00	.185	80.00

Overall performance was highest for Random Forest across all 4 evaluation metrics of accuracy, Cohen's Kappa, ROC-AUC and Average Precision. The classifier selected for Random Forest followed bySHAPinterpretation are presented below in table 1. Random Forest yielded a precision value of 97.00% and was good at accurately identifying the actual positive cases. Random Forest obtained 96.00% recall and a 96.00% F1 score that reflects a good performance with regard to balanced and efficient identification of true positive cases. Its Kappa score was 92.71% which indicates strong agreement between observed and predicted classifications and a well performing classifier. It also yields good AP and ROC-AUC scores. Decision tree and KNN yielded good results too and were also good at predicting positive class with good precision, recall and F1 scores, but not as good as Random Forest. They were also not as good with respect to their Kappa scores which were 90.79% and 92.64% respectively. It is worth noting that Extra Tree performed relatively good compared to others; its precision was 96.00% with recall 95.00% and a 1.00% AP. Its Kappa was however only 90.78% and its

F1 score was 95.00%, both slightly lower than those obtained by Random Forest. On the other hand, SVM was clearly outperformed by all other models as its precision, recall, and F1 score were around 80.00%, its Kappa was 60.01% and its TPR and FPR were respectively 78.00% and 0.185%, its accuracy was only 80.00%.

Thus Random Forest emerged as the most reliable and accurate algorithm in this study, outperforming Decision Tree, KNN, Extra Tree, and SVM across nearly all metrics.

In figure 2 (a), (b), (c), and (d) shows the ROC-AUC score plotting for the models of Decision Tree, SVM, KNN, and Extra Tree.

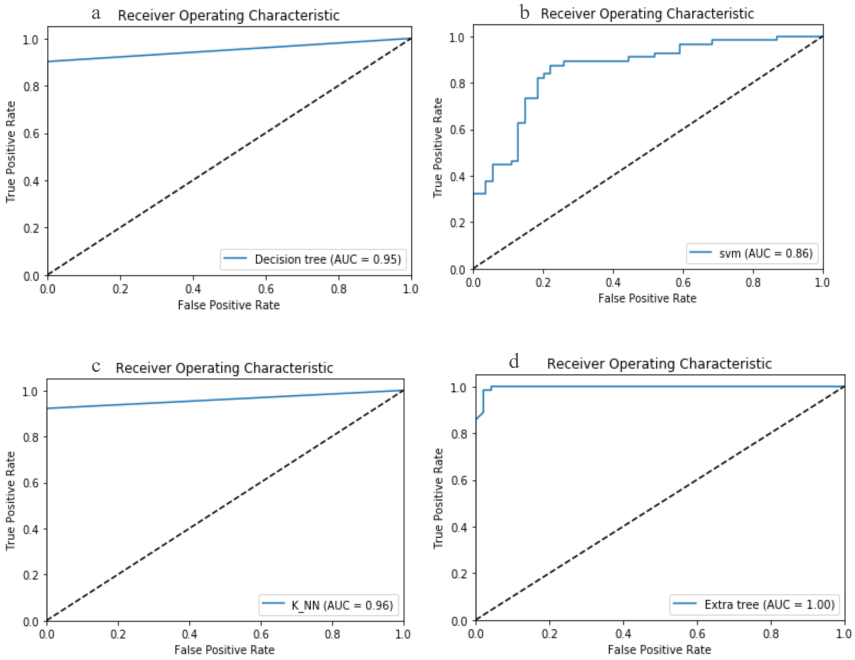


Figure 2(a-d): ROC-AUC curve plotting for classification model

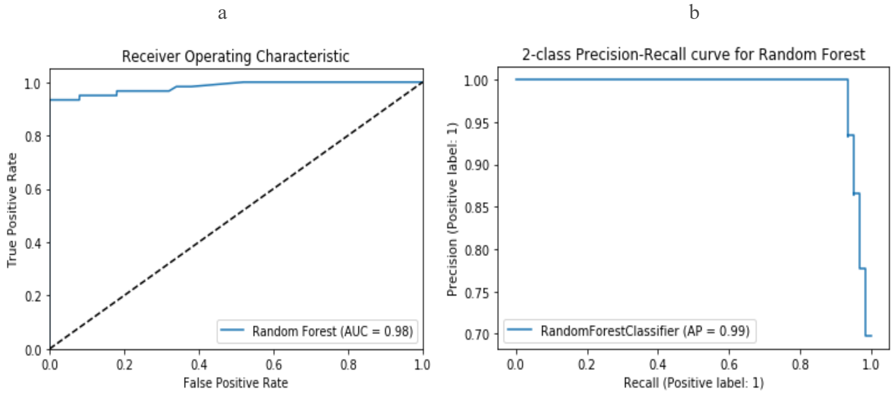


Figure 3(a): ROC for random forest; 3(b) Precision-recall curve for random forest

In figures 3(a) and 3(b) represent the Random Forest classifier performance on ROC curves and PR curves. ROC curve achieves good performance due to its high discriminatory ability which is shown by an AUC of 0.99 which suggests that the true positive rate is high while false positive rate is low. For the

Precision-Recall curve, the AP is calculated which is 0.99. This reflects that the model can distinguish learning disability cases quite accurately while maintaining a high precision.

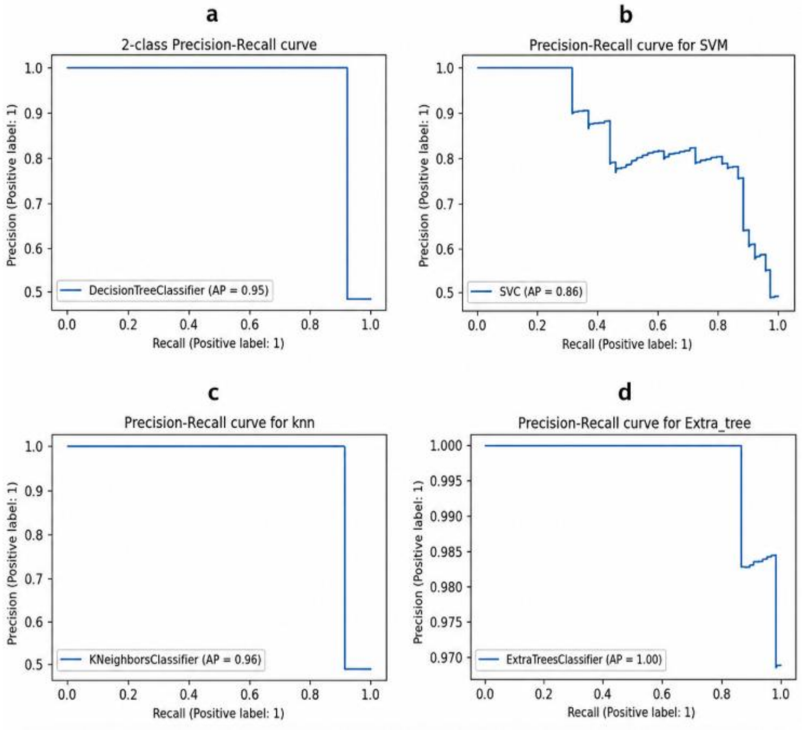


Figure 4(a-d): Precision-recall curve plotting for classification models

In figure 4 (a), (b), (c), and (d) compares PR curves generated for decision tree, SVM, KNN and extra trees. For these models, Extra Trees model showed the best performance, with AP = 1.00 followed by KNN with AP=0.96 and decision tree with AP=0.95. SVM shows lower performance as compared to other classifiers, AP=0.86. In general, the curves display good ability of tree based models to classify the learning disability.

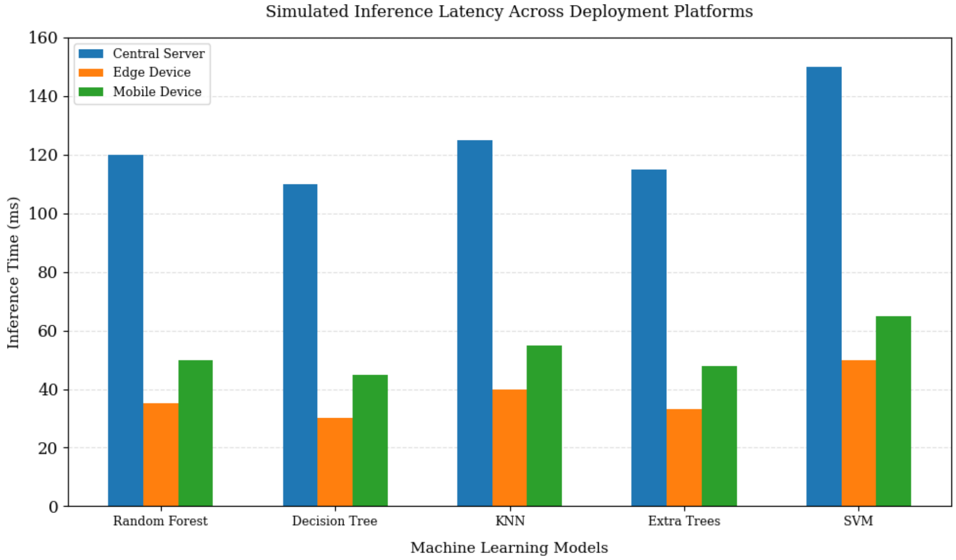


Figure 5: Simulated inference latency across deployment platforms

In figure 5 gives the simulated inference time in ms of the various ML classifiers namely Random Forest, Decision Tree, KNN, Extra Trees, SVM on a central server, an edge device, and a mobile device. Random forest performs with higher accuracy and lower inference time, which are desirable characteristics for context-aware real time education in distributed environments and in mobile devices.

Confusion Matrix Analysis of Random Forest Classifier

To better examine the classification performance of the proposed framework, a confusion matrix is presented for the Random Forest classifier, the highest performing algorithm in comparison with other three models used for classification. A confusion matrix displays the number of correct and incorrect classifications of each of the Learning Disability (LD) and Non-Learning Disability (Non-LD) classes. The confusion matrix pictured is from an example test fold, and table 1 shows the performance averaged over cross-validation trials.

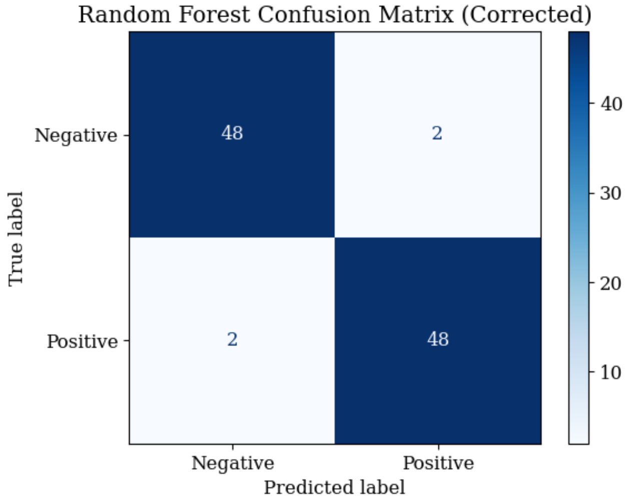


Figure 6: Confusion matrix of the random forest classifier for learning disability prediction

In figure 6 is the confusion matrix for Random Forest classification of LD students. In this case the Random Forest classifier successfully predicts 48 Non LD and 48 LD students whereas only four students are misclassified (2 FP & 2 FN). The prediction accuracy is very strong as it covers both the classes equally. Accuracy with confusion matrix is about 96% which shows the capability of the proposed framework in identifying the learning disability students at an early age.

Feature Importance and SHAP Analysis

The study applied a Random Forest (RF) model to predict learning disabilities (LD) using cognitive, behavioral, and academic features. Feature importance and SHAP (SHapley Additive exPlanations) were used to interpret the model's predictions.

The Random Forest feature importance scores (Figure 7) indicate that Writing (0.23), Academic performance (0.14), and Impulsiveness (0.09) are the top three predictors of LD. These features collectively capture academic difficulties and behavioral tendencies that strongly influence the model outcome. Also important are Hyperactivity (0.056), Auditory Memory (0.055), and Attention (0.053), which match the previous studies where impairment in memory, executive function, and attention are highly correlated to learning impairment. Lowest features were Visual Discrimination (0.01) and Auditory Discrimination (0.019), so these features probably are not that useful for prediction in this dataset.

Feature importance from random forest gives the importance globally, however, SHAP helps explain it locally and how each feature affects individual prediction (Yang & Wang, 2025). In the figure 7, SHAP summary plot can illustrate the direction and the scale of influence from each feature.

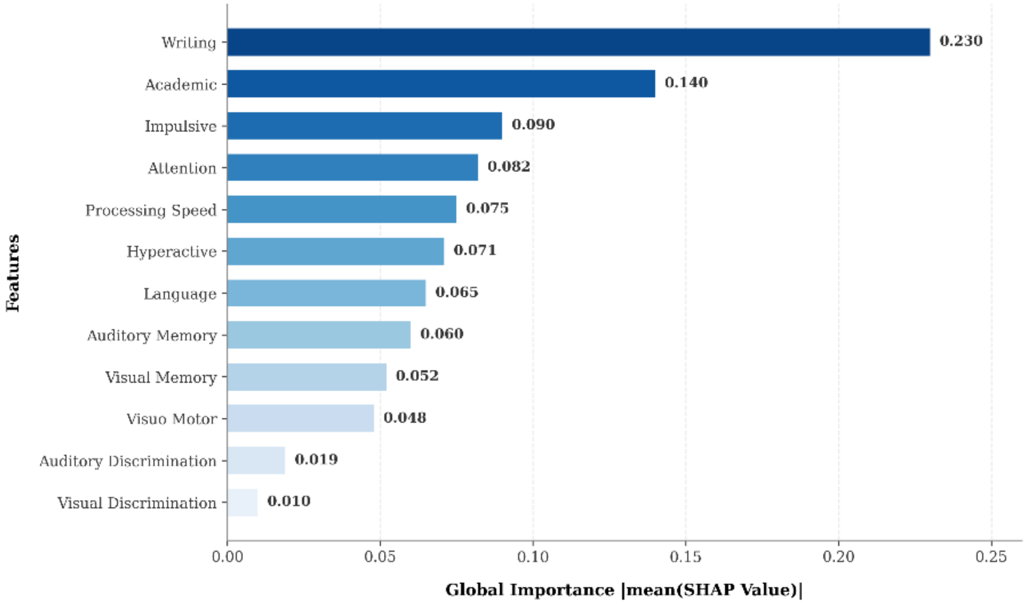


Figure 7: Random forest feature importance scores

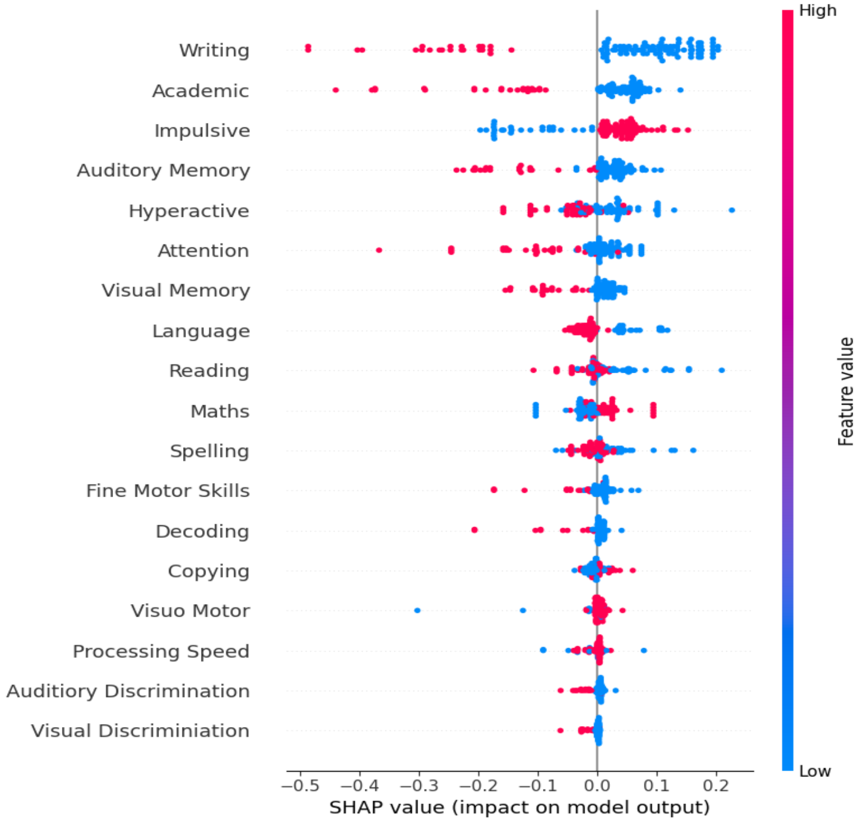


Figure 8: SHAP value summary plot

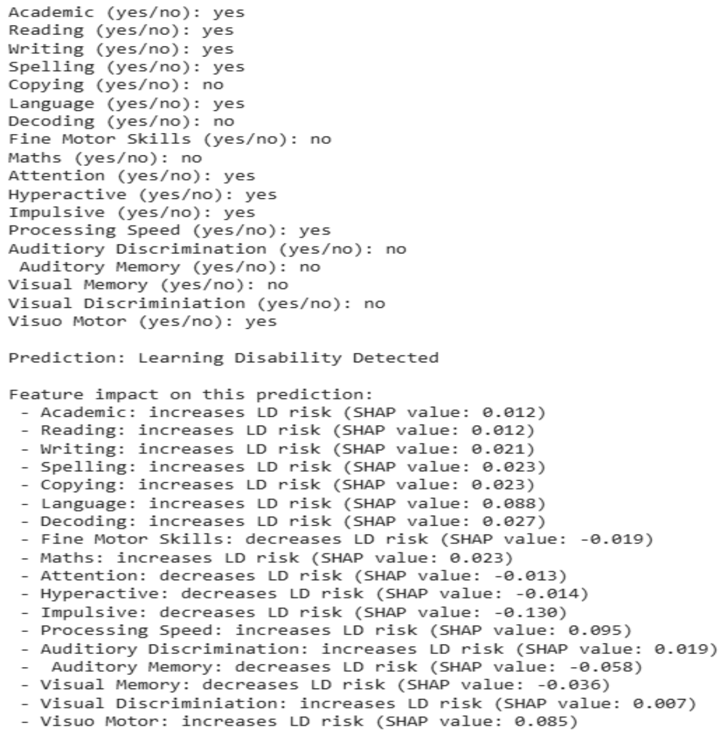
To increase the interpretability of the Random Forest model, SHAP (SHapley Additive exPlanations) values were computed. SHAP represents a principled way to define Shapley values, which comes from coalitional game theory, where each feature plays a "game" in predicting the model output. This study used a TreeExplainer to calculate the SHAP values of the test set for the trained Random Forest classifier since it is optimized for tree-based models. Because this was a binary classification task, the SHAP values were selected for the positive class (Learning Disability was detected). To preserve feature names, the test set was converted to a labeled DataFrame, and a SHAP summary plot was generated to demonstrate the feature distribution and their impact on the prediction of the model (Wang et al., 2024; Atlam et al., 2025).

The summary plot for SHAP (Figure 8) ranks the features by their importance and shows the magnitude and direction of their impacts on the model. The most predictive features in this model were writing and overall academic performance. These high performance measures (red dots are to the left) had negative SHAP values, which suggests protective factors against LD prediction (Monteiro et al., 2025). On the other hand, low measures in writing and academic performance (blue dots are to the right) caused model predictions to lean towards LD detection. The opposite pattern was exhibited by behavioral measures, with a positive SHAP effect for the behavioral disorders (red dots on the right) which predicted an increase in the model prediction of LD. Core cognitive areas showed a similar pattern where poor measures in auditory memory, attention, mathematics, and spelling had strong positive SHAP effects in predicting an LD classification (blue dots on the left). Despite lower global feature importance, the span of SHAP values in fine motor skills, processing speed and perceptual discrimination indicate that a deficit in some areas can have a notable effect on LD prediction in certain cases.

Overall, SHAP values offered a local and global interpretation: demonstrating not only the core predictive value of academic skills and behavioral regulations, but also the directionality of impact for high and low levels of these variables that caused model prediction and confirming the previous Random Forest feature importance results, in that writing, academic performance, and impulsivity are the most influential predictors.

SHAP Analysis of Learning Disability Prediction in Student Assessment

To test this model's capability in practical scenario, one of the sample student case from the original database was selected and used as input for the model. The predicted result for the selected student case was correct. The prediction result for the student case is provided in figure 9. In this interactive test case, the responses for the student across many different features (academic skills, reading, writing, attention, memory and impulsivity) were input and the model classified that 'Learning Disability Detected.' With the SHAP feature contribution analysis, the reasons behind this prediction were highlighted. There were features with positive SHAP values which contributed to increasing the possibility of detecting LD for this student, such as language (0.088), processing speed (0.095), visuo motor (0.085) and writing/spelling. This means that there were high levels of risk and contribute to the student being predicted for having LD. On the other hand, there were some features with negative SHAP values which were in decreasing the probability of having LD. Impulsive (-0.130), auditory memory (-0.058), and visual memory (-0.036) were some of the negative SHAP values which decreased the probability of the student having LD. In overall, the sum of positive SHAP values (risk factors) was greater than sum of negative SHAP values (protective factors) which leads to the student being predicted as having LD. This sample student case was successful in not only testing the model was able to identify LD from new unseen student inputs drawn from the original database, but it also illustrated how SHAP contributed to a better interpretability by showing how this decision was made through the contribution of features and allowing stakeholders to understand the cause of the predicted learning disability for the specific student and which are the aspects that require attention.



```
Academic (yes/no): yes
Reading (yes/no): yes
Writing (yes/no): yes
Spelling (yes/no): yes
Copying (yes/no): no
Language (yes/no): yes
Decoding (yes/no): no
Fine Motor Skills (yes/no): no
Maths (yes/no): no
Attention (yes/no): yes
Hyperactive (yes/no): yes
Impulsive (yes/no): yes
Processing Speed (yes/no): yes
Auditiory Discrimination (yes/no): no
 Auditory Memory (yes/no): no
Visual Memory (yes/no): no
Visual Discriminiation (yes/no): no
Visuo Motor (yes/no): yes

Prediction: Learning Disability Detected

Feature impact on this prediction:
 - Academic: increases LD risk (SHAP value: 0.012)
 - Reading: increases LD risk (SHAP value: 0.012)
 - Writing: increases LD risk (SHAP value: 0.021)
 - Spelling: increases LD risk (SHAP value: 0.023)
 - Copying: increases LD risk (SHAP value: 0.023)
 - Language: increases LD risk (SHAP value: 0.088)
 - Decoding: increases LD risk (SHAP value: 0.027)
 - Fine Motor Skills: decreases LD risk (SHAP value: -0.019)
 - Maths: increases LD risk (SHAP value: 0.023)
 - Attention: decreases LD risk (SHAP value: -0.013)
 - Hyperactive: decreases LD risk (SHAP value: -0.014)
 - Impulsive: decreases LD risk (SHAP value: -0.130)
 - Processing Speed: increases LD risk (SHAP value: 0.095)
 - Auditiory Discrimination: increases LD risk (SHAP value: 0.019)
 -  Auditory Memory: decreases LD risk (SHAP value: -0.058)
 - Visual Memory: decreases LD risk (SHAP value: -0.036)
 - Visual Discriminiation: increases LD risk (SHAP value: 0.007)
 - Visuo Motor: increases LD risk (SHAP value: 0.085)
```

Figure 9: Sample output

Discussion and Limitations

In general, the classification using Random Forest with SHAP based explanation has the capability for the early detection of learning disabilities. Some performance measures of the model gave outstanding results and the ability to distinguish between students with and without learning disabilities. The SHAP provides explanation for influential variables (writing skill, academic performance, impulsivity and attention) and this makes the framework accurate and gives the transparency in detecting the learning disability. Also support the previous works that academic performance and behavior plays important part for the detection of learning disabilities.

In the context of practice, the framework can be a practical decision support tool for educators, psychologists and school administrators. Identifying students at risk earlier allow prompt intervention, individual learning support, targeted educational program and consequently promote the well-being and academic achievement of students. The explainable model, SHAP, may enhance trustworthiness of the prediction for educational and mental health professionals to support decision making for evidence-based planning. It will create the opportunities for individual students with special needs to benefit from effective pedagogical interventions and facilitate equal learning opportunities for all children.

The study conducted used data gathered from educational institutions spread across five districts in Tamil Nadu, India. Even though the model proved accurate in the cross-validation process and subsequent tests, validation through external data sources not belonging to the geographical region or the education system was out of the scope for the present work. Further studies must be conducted using a large multi-institutional dataset to test the reliability and generality of the framework. Other limitations are that several machine learning classifiers were assessed. The current work does not include formal statistical significance testing like the Friedman, Wilcoxon signed-rank, or McNemar tests. Interpret the differences between the performance of classifiers as observational not statistical evidence of better performance. Statistical tests will be run in future work over results of repeated cross validation to support selection.

5 Conclusion

The results of this study show that Random Forest models with SHapley Additive exPlanations (SHAP) are very successful and are easily interpretable for predicting the learning disability (LD) of students. More specifically, the experimental results showed that the Random Forest (RF) classifier had the highest accuracy (96.36%), F1 score (96.00%), Cohen's Kappa (92.71%) and ROC-AUC (98.30%) results, which proved to be robust and reliable for the classification of LD cases. SHAP analysis also identified the most important predictors, such as academic performance, writing ability, impulsivity, and gave local explanations for individual students, which is essential for clear decision-making and targeted interventions. The study shows that it is possible to implement real-time, context-aware and reliable educational applications that are integrated with edge and mobile computing platforms and can be dispatched in accordance with the predictive framework. This way, students' data can be processed in real time, maintaining privacy and facilitating adaptive interventions in smart classrooms or distributed learning settings. Interactive framework enables teachers with immediate alert to at-risk students so as to provide timely and personalized assistance. The outcomes demonstrates practical contribution by using ML, XAI and Ubiquitous computing within a smart learning environment. The findings validate the ability of the Random Forest model to accurately predict student performance in education and underscore the importance of implementing robust and scalable AI systems in educational contexts. This study showed that the proposed Random Forest-SHAP framework achieves excellent predictive performance and interpretability in learning disability identification based on educational, cognitive and behavioral data. This study showed that ML interpretation is a tool to enable educators to interpret risk factors for learning disabilities and design evidenced-based interventions. However, the framework has to be verified on larger, multi-institutional data sets, learning trajectory should be analyzed for longitudinal patterns and the framework should be applied on a larger real-life educational context.

Acknowledgements

Special thanks are extended to Karpagam Academy of Higher Education, Coimbatore, India for giving guidelines for this research. This research was not funded.

References

[1] Abakasanga, E., Kousovista, R., Cosma, G., Akbari, A., Zaccardi, F., Kaur, N., ... & Gangadharan, S. (2025). Equitable hospital length of stay prediction for patients with learning disabilities and multiple long-term conditions using machine learning. *Frontiers in Digital Health*, *7*, 1538793. https://doi.org/10.3389/fdgth.2025.1538793

[2] Atlam, E. S., Rokaya, M., Masud, M., Meshref, H., Alotaibi, R., Almars, A. M., ... & Gad, I. (2025). Explainable artificial intelligence systems for predicting mental health problems in autistics. *Alexandria Engineering Journal*, *117*, 376-390. https://doi.org/10.1016/j.aej.2024.12.120

[3] Çetin, A. İ., & Büyüklü, A. H. (2025). A new approach to K-nearest neighbors distance metrics on sovereign country credit rating. *Kuwait Journal of Science*, *52*(1), 100324. https://doi.org/10.1016/j.kjs.2024.100324

[4] Dandolo, D., Masiero, C., Carletti, M., Dalle Pezze, D., & Susto, G. A. (2023). AcME-Accelerated model-agnostic explanations: Fast whitening of the machine-learning black box. *Expert Systems with Applications*, *214*, 119115. https://doi.org/10.1016/j.eswa.2022.119115

[5] Han, Y., & Wang, S. (2023). Disability risk prediction model based on machine learning among Chinese healthy older adults: results from the China Health and Retirement Longitudinal Study. *Frontiers in public health*, *11*, 1271595. https://doi.org/10.3389/fpubh.2023.1271595

[6] Hasan, M. M. (2023). Understanding model predictions: a comparative analysis of SHAP and LIME on various ML algorithms. *Journal of Scientific and Technological Research*, *5*(1), 17-26. https://doi.org/10.59738/jstr.v5i1.23(17-26).eaqr5800

[7] Kalita, E., El Aouifi, H., Kukkar, A., Hussain, S., Ali, T., & Gaftandzhieva, S. (2025). LSTM-SHAP based academic performance prediction for disabled learners in virtual learning environments: a statistical analysis approach. *Social Network Analysis and Mining*, *15*(1), 65. https://doi.org/10.1007/s13278-025-01484-1

[8] Kim, T. M., Kim, Y. H., Song, S. H., Choi, I. Y., Kim, D. J., & Ko, T. (2024). Explainability Enhanced Machine Learning Model for Classifying Intellectual Disability and Attention-Deficit/ Hyperactivity Disorder with Psychological Test Reports. *Journal of Korean Medical Science*, *40*(11). https://doi.org/10.3346/jkms.2025.40.e26

[9] Larasati, R. (2025, December). Explainable AI Approach to Compare Predictive Drivers of Student Success for Disabled and Non-Disabled Learners. In *2025 2nd International Conference on Artificial Intelligence and Teacher Education (ICAITE)* (pp. 1-6). IEEE. https://doi.org/10.1109/ICAITE68636.2025.11442370

[10] Malakouti, S. M. (2025). Leveraging SHapley Additive exPlanations (SHAP) and fuzzy logic for efficient rainfall forecasts. *Scientific Reports*, *15*(1), 36499. https://doi.org/10.1038/s41598-025-22081-4

[11] Matzavela, V., & Alepis, E. (2021). Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments. *Computers and Education: Artificial Intelligence*, *2*, 100035. https://doi.org/10.1016/j.caeai.2021.100035

[12] Mim, M. A., Khatun, M. R., Hossain, M. M., Rahman, W., & Munir, A. (2025). Exploring early learning challenges in children utilizing statistical and explainable machine learning. *Algorithms*, *18*(1), 20. https://doi.org/10.3390/a18010020

[13] Monteiro, G. O. D. A., Difante, G. D. S., Montagner, D. B., Euclides, V. P. B., Castro, M., Rodrigues, J. G., ... & Matsubara, E. T. (2025). Interpreting Machine Learning Models with SHAP Values: Application to Crude Protein Prediction in Tamani Grass Pastures. *Agronomy*, *15*(12), 2780. https://doi.org/10.3390/agronomy15122780

[14] Navarro-Soria, I., Rico-Juan, J. R., Juarez-Ruiz de Mier, R., & Lavigne-Cervan, R. (2025). Prediction of attention deficit hyperactivity disorder based on explainable artificial intelligence. *Applied Neuropsychology: Child*, *14*(4), 474-487. https://doi.org/10.1080/21622965.2024.2336019

[15] Papalexandratou, P., Stathopoulou, A., & Skanavis, C. (2024). The use of artificial intelligence in the education of students with learning disabilities: A systematic review. *Global Journal of Engineering and Technology Advances*, *21*(3), 33-49. https://doi.org/10.30574/gjeta.2024.21.3.0223

[16] Pimentel, J. S., Ospina, R., & Ara, A. (2024). A novel fusion Support Vector Machine integrating weak and sphere models for classification challenges with massive data. *Decision Analytics Journal*, *11*, 100457. https://doi.org/10.1016/j.dajour.2024.100457

[17] Ponce-Bobadilla, A. V., Schmitt, V., Maier, C. S., Mensing, S., & Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clinical and translational science*, *17*(11), e70056. https://doi.org/10.1111/cts.70056

[18] Raji, N. R., Kumar, R. M. S., & Biji, C. L. (2024). Explainable machine learning prediction for the academic performance of deaf scholars. *IEEE Access*, *12*, 23595-23612. https://doi.org/10.1109/ACCESS.2024.3363634

[19] Saarela, M., & Podgorelec, V. (2024). Recent applications of explainable AI (XAI): A systematic literature review. *Applied Sciences*, *14*(19), 8884. https://doi.org/10.3390/app14198884

[20] Sanfo, J. B. M. (2025). Application of explainable artificial intelligence approach to predict student learning outcomes. *Journal of Computational Social Science*, *8*(1), 9. https://doi.org/10.1007/s42001-024-00344-w

[21] Scarpazza, C., Baecker, L., Vieira, S., & Mechelli, A. (2020). Applications of machine learning to brain disorders. In *Machine learning* (pp. 45-65). Academic Press. https://doi.org/10.1016/B978-0-12-815739-8.00003-1

[22] Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, *237*, 121549. https://doi.org/10.1016/j.eswa.2023.121549

[23] Teles, G., Rodrigues, J. J., Rabêlo, R. A., & Kozlov, S. A. (2021). Comparative study of support vector machines and random forests machine learning algorithms on credit operation. *Software: practice and experience*, *51*(12), 2492-2500. https://doi.org/10.1002/spe.2842

[24] Wang, H., Liang, Q., Hancock, J. T., & Khoshgoftaar, T. M. (2024). Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data*, *11*(1), 44. https://doi.org/10.1186/s40537-024-00905-w

[25] Yang, Y., & Wang, H. (2025). Random forest-based machine failure prediction: A performance comparison. *Applied Sciences*, *15*(16), 8841. https://doi.org/10.3390/app15168841

Authors Biography

P.J. Anu received her B.Sc. degree in computer science from Calicut University, Kerala, India in 2010. In 2013, she completed MCA. She graduated in June 2013 with an excellent percentage. She worked as a lecturer in the Faculty of Computer Science at DonBosco College, affiliated with Calicut University for 7 years. Her main research interests include Artificial intelligence.

K. Ranjith Singh obtained his Ph.D. Degree from Periyar University, Salem, Tamil Nadu, India. He is working as Head of the Department, Department of Computer Technology, Faculty of Arts, Science, Commerce and Management, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, Tamil Nadu, India. He has more than 21 years of experience in the academic field. He has published several research articles in reputed International Journals (Scopus) and International Conferences (Scopus) so far. His research areas of interest include Cryptography and Network Security, Mobile Ad-hoc Networks, and Wireless Sensor Networks.