Exploring Early Learning Challenges in Children Utilizing Statistical and Explainable Machine Learning

Article *in* Algorithms · January 2025

DOI: 10.3390/a18010020

CITATIONS
0

READS
246

5 authors, including:

Mithila Akter Mim
Bangladesh University
2 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Mst. Rokeya Khatun
Bangladesh University
14 PUBLICATIONS **14** CITATIONS
SEE PROFILE

Muhammad Minoar Hossain
Bangladesh University
22 PUBLICATIONS **232** CITATIONS
SEE PROFILE

Wahidur Rahman
Uttara University
53 PUBLICATIONS **576** CITATIONS
SEE PROFILE

algorithms

MDPI

Article

Exploring Early Learning Challenges in Children Utilizing Statistical and Explainable Machine Learning

Mithila Akter Mim [1], M. R. Khatun [1], Muhammad Minoar Hossain [1,2], Wahidur Rahman [2,3] and Arslan Munir [4,*]

[1] Department of Computer Science and Engineering, Bangladesh University, Dhaka 1000, Bangladesh; mithila.bu.75@gmail.com (M.A.M.); rokeya.khatun@bu.edu.bd (M.R.K.); minoar.hossain@bu.edu.bd (M.M.H.)
[2] Department of Computer Science and Engineering, Mawlana Bhashani Science and Technology University, Tangail 1902, Bangladesh; wahidtuhin0@gmail.com
[3] Department of Computer Science and Engineering, Uttara University, Dhaka 1230, Bangladesh
[4] Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA
* Correspondence: arslanm@fau.edu

Abstract: To mitigate future educational challenges, the early childhood period is critical for cognitive development, so understanding the factors influencing child learning abilities is essential. This study investigates the impact of parenting techniques, sociodemographic characteristics, and health conditions on the learning abilities of children under five years old. Our primary goal is to explore the key factors that influence children's learning abilities. For our study, we utilized the 2019 Multiple Indicator Cluster Surveys (MICS) dataset in Bangladesh. Using statistical analysis, we identified the key factors that affect children's learning capability. To ensure proper analysis, we used extensive data preprocessing, feature manipulation, and model evaluation. Furthermore, we explored robust machine learning (ML) models to analyze and predict the learning challenges faced by children. These include logistic regression (LRC), decision tree (DT), k-nearest neighbor (KNN), random forest (RF), gradient boosting (GB), extreme gradient boosting (XGB), and bagging classification models. Out of these, GB and XGB, with 10-fold cross-validation, achieved an impressive accuracy of 95%, F1-score of 95%, and receiver operating characteristic area under the curve (ROC AUC) of 95%. Additionally, to interpret the model outputs and explore influencing factors, we used explainable AI (XAI) techniques like SHAP and LIME. Both statistical analysis and XAI interpretation revealed key factors that influence children's learning difficulties. These include harsh disciplinary practices, low socioeconomic status, limited maternal education, and health-related issues. These findings offer valuable insights to guide policy measures to improve educational outcomes and promote holistic child development in Bangladesh and similar contexts.

Keywords: child development; Bangladesh multiple indicator cluster surveys (MICS); machine learning; explainable AI; SHAP; LIME

Academic Editors: Antonio Sarasa Cabezuelo and María Estefanía Avilés Mariño

Received: 24 November 2024
Revised: 24 December 2024
Accepted: 31 December 2024
Published: 4 January 2025

Citation: Mim, M.A.; Khatun, M.R.; Hossain, M.M.; Rahman, W.; Munir, A. Exploring Early Learning Challenges in Children Utilizing Statistical and Explainable Machine Learning. *Algorithms* **2025**, *18*, 20. https://doi.org/10.3390/a18010020

1. Introduction

Children's learning capacities and psychological development are significantly influenced by the parenting practices and sociodemographic features in their environment. The impact of good relationships, engagement, and participation between parents and children on academic achievement is positive [1]. This engagement involves things like

reading aloud to kids, guiding studies, and staying in touch with them, all of which are essential for creating a positive learning environment. The long-term effects of parenting practices such as friendly manners and flexibility significantly influence children's cognitive development and academic performance [2]. Furthermore, different racial and ethnic communities have distinct policies concerning the extent of parental engagement with children, which greatly impacts the growth and development of young children [3]. Sociodemographic factors, such as parental education level and socioeconomic status (SES), also significantly influence children's learning patterns. Moreover, the provision of mental and physical healthcare for children has a direct impact on their brain and cognitive development, highlighting the importance of home environments [4,5]. Additionally, households of higher SES generally enjoy superior access to educational resources and opportunities as well as good healthcare services, leading to enhanced learning outcomes for children [6]. Researchers are extensively exploring the influence of parenting styles, ensuring privilege, mental and physical care, and awareness of children's learning abilities.

Previous studies have demonstrated the application of a variety of methods in child development research. Researchers diagnose mental health problems and analyze parenting techniques to identify learning difficulties. Several research works have examined the application of machine learning (ML) algorithms to predict learning difficulties and variables influencing school-aged children's and college students' academic performance. To diagnose learning difficulties in kids, another study [7] used artificial neural network (ANN) and support vector machine (SVM) classifiers. The study emphasized the innovative application of ML to accurately identify learning disabilities. Researchers used supervised ML algorithms [8] to uncover the reasons behind college students' poor academic performance. That provides advice to help struggling students to do better. Additionally, researchers have used ML approaches to identify mental health issues in children and adolescents, including attention deficit hyperactivity disorder (ADHD) and depression. Multiple researchers [9,10] have employed ML models to forecast depression and ADHD in children, attaining superior classification accuracy and emphasizing the possibility of timely identification and remediation. Additionally, researchers [11] developed and evaluated ML-based prognosis tools to identify learning difficulties among students, aiming to improve early detection and intervention strategies. Furthermore, researchers used ML models to predict ADHD by analyzing prefrontal brain activity, underscoring the importance of neurophysiological data in diagnostic prediction [12]. Early detection is essential for prompt assistance and intervention in cases of autism spectrum disorder (ASD). ML techniques for early ASD detection have been the subject of several studies [13,14]. These studies used a variety of classifiers, including random forest (RF), logistic regression (LR), and support vector machine (SVM), to create precise prediction models based on physiological and behavioral data. The results imply that ML methods can improve the accuracy and effectiveness of ASD diagnosis.

Recent advancements in ML have significantly improved ASD detection by using SVM, and LR achieved 100% accuracy for children and 97.14% for adults, respectively, and ANN also showed promise, reaching 94.24% accuracy with optimized hyperparameters [15]. Additionally, a study analyzing toddlers using six classification models, including LR, reported a perfect accuracy score of 100% in ASD detection [16]. Beyond ASD detection, research has also examined the influence of parenting practices and other factors on children's development. Studies highlight that parenting style, gender, and socioeconomic background play a significant role in academic achievement, with authoritative parenting being particularly crucial for fostering successful educational outcomes [17]. Again, comparative studies have investigated the effectiveness of ML techniques in predicting children's educational activities compared to traditional methods [18,19]. These

analyses highlight the benefits of ML algorithms for identifying important factors and improving prediction accuracy.

The Bangladesh Bureau of Statistics and UNICEF conducted the Multiple Indicator Cluster Surveys (MICS) in 2019. This survey collected a comprehensive dataset that includes information on child welfare, household characteristics, and parenting practices in Bangladesh [20]. Employing this information provides an opportunity to explore the multifaceted relationships between the sociodemographic characteristics, parenting styles, and educational achievement of Bangladeshi children, which can then be applied to similar studies. In addition, to analyze and forecast a wide range of outcomes regarding child development and education, ML approaches have become an increasingly powerful tool [21,22]. Researchers have the potential to use ML algorithms on the MICS 2019 dataset to gain valuable insights into the factors affecting children's learning skill comparisons. Moreover, they can apply these insights to shape interventions and policies that aim to enhance opportunities for children in Bangladesh.

The literature review shows how ML can predict academic performances, learning difficulties, mental health illnesses, and parenting styles in child development studies. These studies highlight the complicated relationship between individual, familial, and environmental influences on children. Our research is motivated by the critical need to address and understand challenges in early childhood development and education using advanced computational techniques. Our literature review revealed a research gap, as no previous studies have concentrated on identifying and forecasting the influencing factors for learning difficulties among Bangladeshi children. While ML has been widely used in child development studies in other countries, such as Oman, Australia, and the U.S., there has been limited exploration of early childhood learning outcomes in the Bangladeshi context. Our study addresses this gap by applying ML techniques to the MICS 2019 dataset, which uniquely captures the influence of sociodemographic factors, parenting styles, and children's learning abilities in Bangladesh. This research highlights the importance of leveraging local data to uncover insights specific to Bangladesh's context and social landscape, providing the foundation for targeted interventions. In addition to applying ML techniques, our study integrates explainable AI (XAI) methods like SHAP and LIME, which offer a novel transparent approach to understanding early childhood learning difficulties in Bangladesh. By incorporating XAI methods, our approach not only improves the interpretability of ML models but also identifies underlying causes of learning difficulties. These findings offer valuable insights for policymaking to improve early childhood education. The highlights of our objective are as follows:

- Employ statistical analysis to investigate the key factors influencing early learning among children under the age of five;
- Apply different ML models to forecast learning difficulties in children based on the selected features;
- Utilize XAI methods like SHAP and LIME to explain how different factors influence children learning difficulties. These insights will help guide policies to improve education and overall development in early childhood.

The rest of this paper is organized as follows. Section 2 provides a comprehensive overview of the materials and methods used in this study. It includes details about the dataset, statistical analysis, and ML algorithms that we employed and the criteria used for evaluating performance. Section 3 presents the results and discussion. Section 4 presents a comparison with previous works, while Section 5 concludes this study.

2. Materials and Methods

In this research, we investigated parenting practices and sociodemographic factors that influence children's learning skills using the MICS 2019 dataset. At first, we

preprocessed the dataset to prepare it for statistical analysis and machine learning model fit. Afterward, we evaluated the results based on the performance parameters. Figure 1 shows the step-by-step process involved in analyzing the impact of various factors on children's learning abilities, starting from data collection and preprocessing, moving through specific analytical steps, and ending with an integrated analysis.

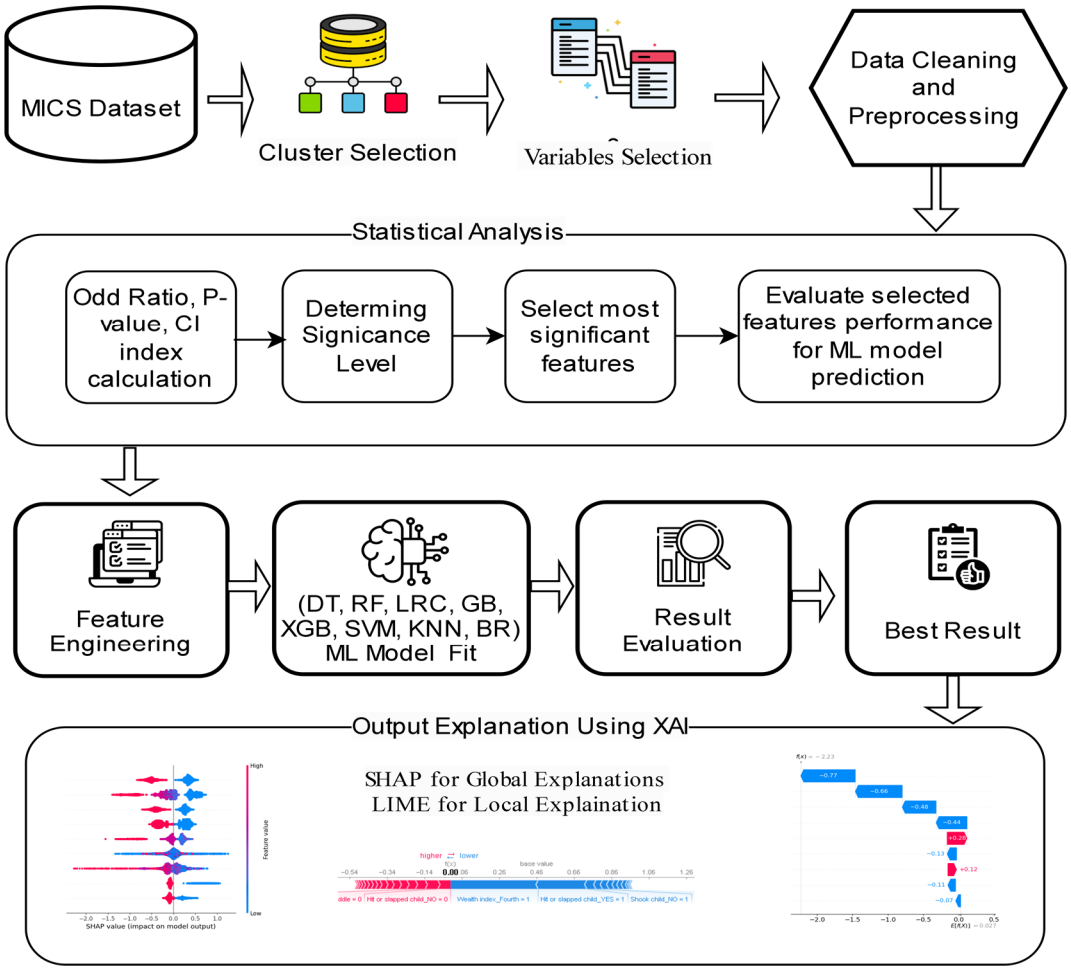


Figure 1. Workflow diagram for analyzing parenting practices and sociodemographic factors on children's learning abilities.

2.1. MICS Dataset

The Bangladesh Bureau of Statistics and the Ministry of Planning, with assistance from UNICEF, conducted the MICS 2019 in Bangladesh. The survey comprises clusters covering health, education, nutrition, water and sanitation, child health, and socioeconomic factors. The research population estimated several indicators on the status of children and women nationwide, in both urban and rural areas, in eight divisions and sixty-four districts. Initially, the total number of household samples was 64,000. The report BBS and UNICEF Bangladesh 2019 provides a comprehensive explanation of the survey methodology, which includes details on the calculation of the sample size. The dataset used in this research is publicly available from the MICS 2019, with permission from the UNICEF MICS team, and adheres to ethical standards, ensuring participant anonymity through de-identification and aggregation. To obtain access to the dataset initially, we had to obtain permission from the UNICEF MICS team. Thus, we sent them an email explaining our study plans, and in response, they gave us a link to access the dataset. Furthermore we addressed potential model biases by selecting features based on statistical relevance and

employing XAI techniques for transparency, fairness, and interpretability. This approach helps identify and mitigate biases, ensuring robustness in the findings.

2.2. Data Cleaning and Preprocessing

The MICS 2019 dataset encompasses various clusters with different domains. Among these, we chose the children's health (ch) dataset because it supported all the necessary domains for determining the learning skills of children under five years old. Initially, the dataset contained 24,686 samples with 346 features covering various sociodemographic, health, and behavioral aspects related to children and their families. However, not all features were directly relevant to understanding early childhood learning difficulties. Initially, data cleaning and preprocessing were performed to exclude irrelevant variables, such as identifiers and redundant features, as well as features with a high percentage of missing values. Then, we selected all possible features that linked to early childhood learning outcomes, such as parenting behaviors, maternal education, and child health metrics. We also focused on features that were most relevant to the target variable to ensure the inclusion. After careful consideration, we identified 19 features that demonstrated statistical significance, predictive power, and alignment with domain-specific knowledge. This process involved evaluating the importance of each feature, checking Pearson correlations, and of the most significant ones for model accuracy [23,24]. Subsequently, we removed samples with missing values in the target variable, resulting in a dataset of 14,058 samples. Further refinement involved dropping samples with "Don't know" or "No response" values, resulting in 14,044 samples. Furthermore, because addressing missing data points is critical for reliable model predictions, we employed different strategies to fill in missing values based on the feature data type [25]. For categorical features that represent non-numeric data, we filled missing values with the mode value. We imputed missing values for numerical features using the mean value. Afterward, duplicate entries were identified and removed from the dataset, resulting in a reduction to 13,821 unique samples. Next, we converted features with mixed or inconsistent data types to appropriate numerical formats (float, integer, or string) according to their intended meaning. Outliers, which can skew model results significantly, were identified and removed using the interquartile range (IQR) method [26]. Finally, we obtained a clean and well-organized dataset comprising 13,274 samples.

After repeated statistical analysis and data preprocessing, we chose the 10 most significant features out of 19. Table 1 describes the lists of selected variables or features. Out of 10 variables, 9 are independent variables, and 1 is a dependent variable. The independent variables include different factors related to the child, household, and parenting practices. The information collected includes numerical values like the child's weight (Child's_weight) and weight-for-age percentile (Weight_age_P). It also includes categorical variables that indicate disciplinary actions taken by parents (Took_away_privileges, Shook_child, Beat_hard, Called_dumb_lazy), the child's sex, wealth index quintile (Wealth_index), and the mother's education level (Mother's_education). The disciplinary action variables (Removed_privileges, Shook_child, Beat_hard, Called_dumb_lazy) represent different behaviors relevant to parenting practices. However, the dependent variable "Learning_skill" indicates whether or not the child has learning difficulties compared to other children of the same age. We combined these variables to create a comprehensive dataset to explore the child's learning difficulties based on the relationship between parenting practices, sociodemographic factors, and children's learning outcomes.

The MICS 2019 dataset is unbalanced because learning difficulties in children under five are relatively lower compared to those without such difficulties. This imbalance is a common characteristic of large-scale datasets dealing with minority outcomes, as such cases naturally occur less frequently. Factors like underreporting by parents, limited

awareness of early developmental challenges, and disparities in access to health and education services further contribute to this imbalance. The implications of this imbalance have been recognized, and appropriate measures were considered to address its effects on analysis and predictions.

Table 1. Definition of variables or features list.

Independent Variables				
MICS variables	**Variables**	**Description**	**Values**	**Data Type**
UCD2A	Removed_privileges	Took away privileges	YES, NO	Categorical
UCD2C	Shook_child	Shook child	YES, NO	Categorical
UCD2K	Beat_ hard	Beat child as hard as one could	YES, NO	Categorical
HL4	Sex	Sex	MALE, FEMALE	Categorical
windex5	Wealth_index	Wealth index quintile	Richest, Fourth, Second, Poorest, Middle	Categorical
welevel	Mother's_education	Mother's education	Higher secondary+, Primary, Secondary, Pre-primary or none	Categorical
UCD2H	Called_dumb_lazy	Called child dumb and lazy	YES, NO	Categorical
AN4	Child's_weight	Child's weight (kilograms)	1.1–35.8 kg	Numerical
WAP	Weight_age_P	Weight-for-age percentile	0.0–89.5	Numerical
Dependent Variables				
MICS variables	**Variable**	**Description**	**Values**	**Data Type**
UCF17	Learning_skill	Compared with children of the same age, child has difficulty learning things	NO DIFFICULTY, HAS DIFFICULTY	Categorical

Figure 2 presents six grouped bar charts comparing the count of children with and without learning difficulties across various categories. Blue bars represent children without learning difficulties, while red bars indicate children with learning difficulties. The charts show differences in experiences and demographics, highlighting factors and patterns that might contribute to learning difficulties.

The distribution of the variables Child's_weight and Weight_age_P was examined using density plots and boxplots to identify skewness, central tendencies, and outliers in Figure 3a,b. For Weight_age_P, the density plot revealed a highly right-skewed distribution, with values concentrated near zero and a long tail extending toward higher values. The boxplot confirmed this pattern, showing a narrow interquartile range (IQR) and numerous outliers above the upper whisker. To address this skewness and reduce the influence of outliers, a log transformation was applied, which stabilized variance and improved data symmetry. In contrast, the Child's_weight variable exhibited an approximately symmetric distribution, as seen in both the density plot and boxplot. The IQR captured most of the data, with the median centrally located. Although a few outliers were observed outside the whiskers, their impact on the overall distribution was minimal, and no transformation was required

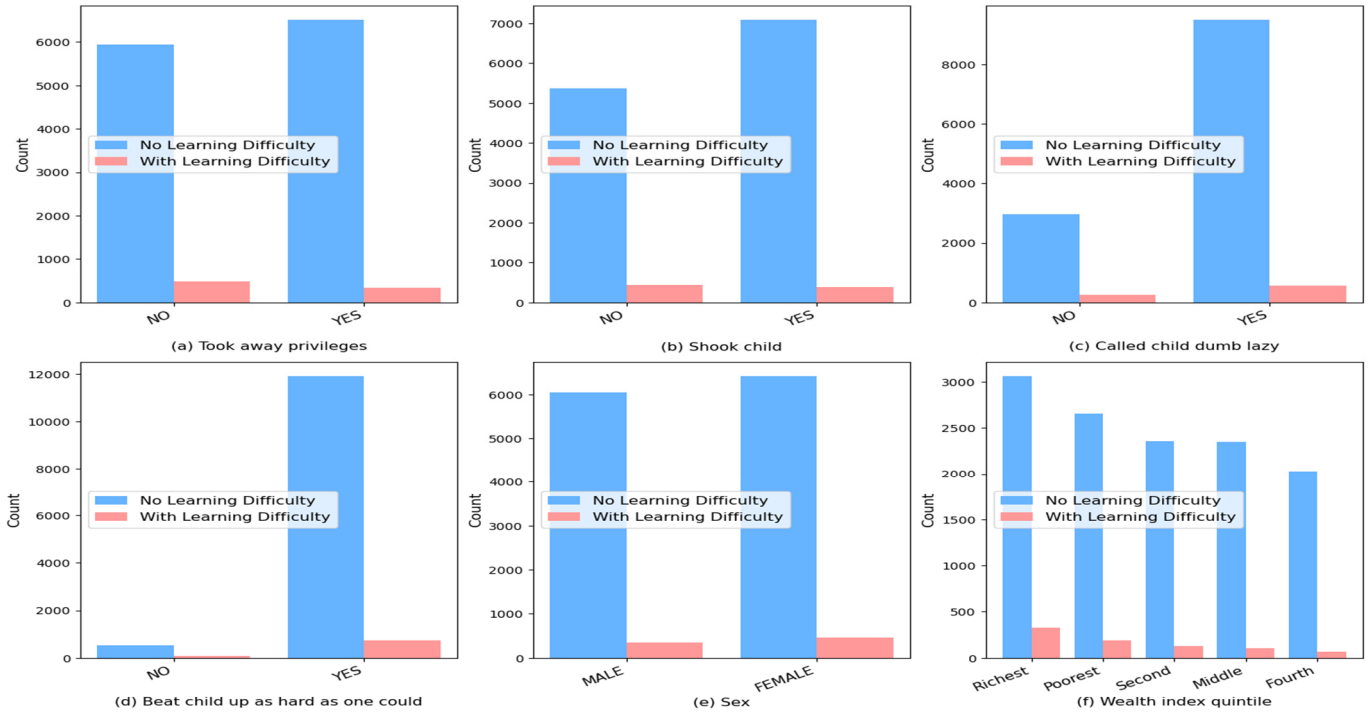
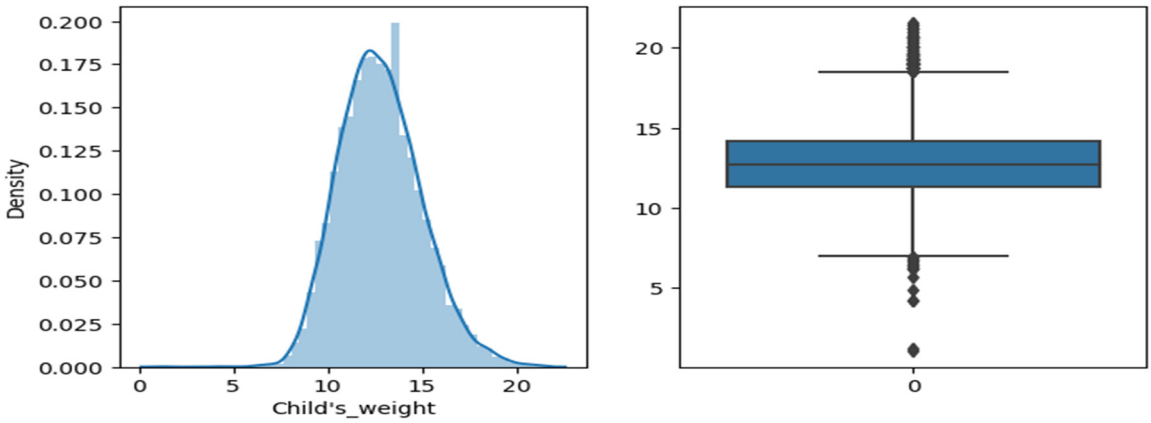
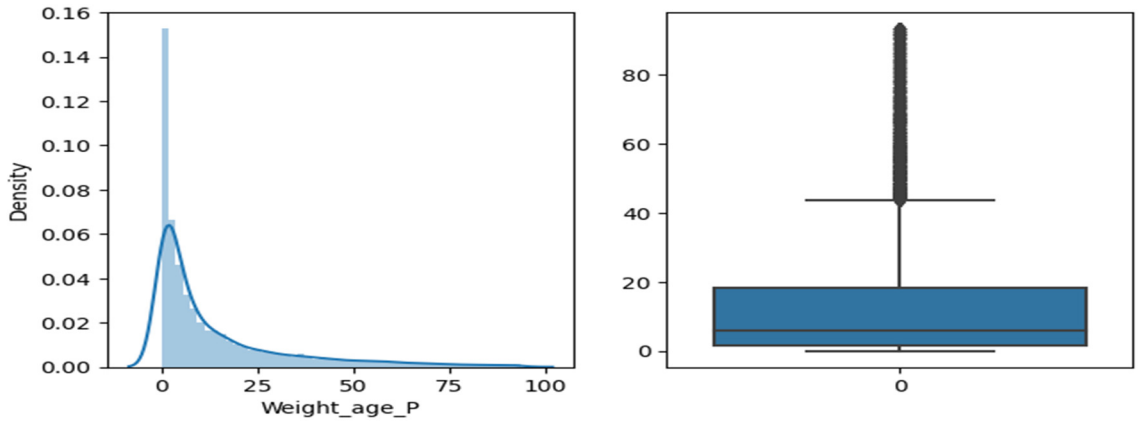


Figure 2. Analysis of categorical features.



(**a**)



(**b**)

Figure 3. Analysis of numerical features (**a**) Child's_weight and (**b**) Weight_age_P.

2.3. Feature Importance

The Figure 4, derived using the RF model, highlights the most influential factors for prediction of Learning_skill. "Weight_age_P" and "Child's_weight" are the most significant features, showing the highest impact. "Mother's_education" and "Wealth_index" also contribute notably but to a lesser extent. Other factors such as "Sex", "Shook_child", "Called_dumb_lazy", "Removed_privileges", and "Beat_hard" show relatively lower importance in influencing the model's output.

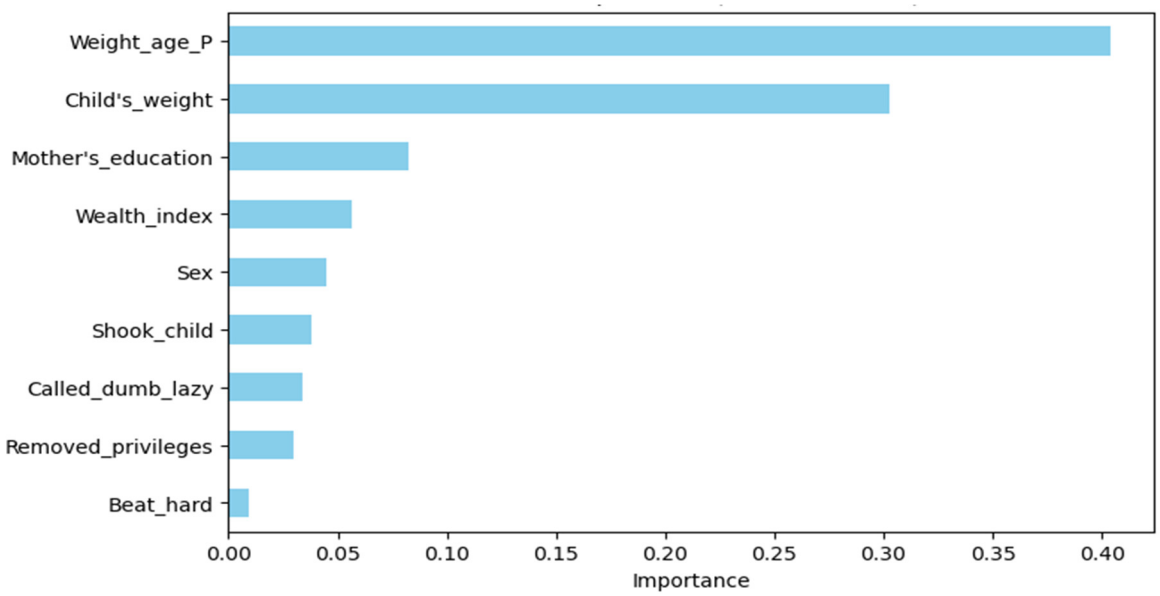


Figure 4. Feature importance for predicting learning skill.

2.4. Feature Engineering

Feature engineering is an important part of getting data ready for analysis and modeling. During this step, we used different methods to explore important patterns in data and make our predictions robust. First, we labeled or binned [27] the UCF17 feature, which encapsulates assessments regarding a child's learning difficulties. We carefully partitioned this categorical feature into two distinct classes: "NO DIFFICULTY" and "HAS DIFFICULTY", which provided a clear delineation for subsequent analysis. Furthermore, the categorical features were subjected to label encoding using a technique [28].

To prepare the data for fitting into an ML model, Table 2 displays the numerical representation of categorical variables. The table shows that we labeled features like taking away privileges, shaking or beating the child, and calling the child dumb or lazy with YES = 1 and NO = 2. In addition, we classified gender as follows: male is represented by the number 1, and female is represented by the number 2. The wealth index is coded on a scale of 1 to 5, with 1 representing the poorest and 5 representing the richest. The mother's education level is classified on a scale of 0 to 3, ranging from pre-primary or no education to higher secondary education. We also labeled the child's learning skills as NO DIFFICULTY = 0 and HAVE DIFFICULTY = 1.

Table 2. Numerical Labeling of Categorical Variables.

Variables	Assigned Numerical Values
Removed_privileges	YES = 1, NO = 2
Shook_child	YES = 1, NO = 2
Beat_ hard	YES = 1, NO = 2
Sex	MALE = 1, FEMALE = 2

Wealth_index	Poorest = 1, Second = 2, Middle =3, Fourth = 4, Richest = 5
Mother's_education	Pre-primary or none = 0, Primary = 1, Secondary = 2, Higher Secondary = 3
Called_dumb_lazy	YES = 1, NO = 2
Learning_skill	NO DIFFICULTY = 0, HAVE DIFFICULTY = 1

Furthermore, to handle the problem of class imbalance in our target variable, we used a mix of techniques. We balanced the representation of different classes in our dataset by adding more samples to the minority class using SMOTE (Synthetic Minority Oversampling Technique) and reducing the number of samples in the majority class with RandomUnderSampler [29,30]. This balancing made our analysis and modeling more reliable as well as helped to prevent any biases that might result from one class being more dominant than the others.

2.5. Statistical Measurements

In statistical analysis, understanding the relationship between variables depends upon several key metrics [31]. Together, the metrics odds ratio (OR), *p*-value, and 95% confidence interval (CI) provide a comprehensive framework for evaluating the strength, significance, and precision of the relationships observed in the data. The *p*-value is a statistical measure that indicates whether the relationship between an independent variable and whether the dependent variable is statistically significant. It helps to decide whether to reject the null hypothesis and thus plays a fundamental role in hypothesis testing. A low *p*-value (typically ≤ 0.05) indicates a significant association between the independent variable and the dependent variable. On the other hand, a high *p*-value (>0.05) indicates no significant association between the independent variable and the dependent variable. The OR is a crucial measure of association, indicating the odds of an event occurring in one group compared to another, whereas odds help to elucidate how likely an event is to happen compared to not happening. An OR greater than 1 indicates increased odds, an OR less than 1 indicates decreased odds, and an OR equal to 1 indicates no effect. The 95% CI further aids in understanding this precision. If the OR's 95% CI does not include 1, it suggests a significant effect of the independent variable on the dependent variable. If the 95% CI for an OR does not include 1, it suggests the result is statistically significant.

2.6. ML Models

Each ML model offers unique methodologies, algorithms, and mathematical formulations. We employed traditional machine learning models combined with ensemble techniques. A brief description of each model is presented below.

Decision trees (DT) are hierarchical structures that recursively split the data based on informative features at each node, aiming to maximize information gain or minimize impurity. At each node, the decision rule selects the attribute that best divides the data. The process continues until a stopping criterion is met, resulting in leaf nodes representing the final class predictions. The splitting criteria often include measures like Gini impurity or information gain [32].

Random forests (RF) are ensemble learning methods that aggregate predictions from multiple decision trees, each trained on random subsets of the data and features. By averaging or voting the predictions of individual trees, random forests improve generalization and reduce overfitting. Each decision tree in the forest operates independently, making predictions based on its subset of the data [33].

The logistic regression (LRC) model predicts the probability of the positive class using the logistic function. Logistic regression is a linear model used for binary classification,

estimating the probability that an instance belongs to a particular class. The logistic function, also known as the sigmoid function, is used to map the output to a probability value between 0 and 1. The model's coefficients are estimated using maximum likelihood estimation [34].

K-nearest neighbors (KNN) is a non-parametric classification algorithm that makes predictions based on the majority class of its *k* nearest neighbors in the feature space. KNN is simple yet effective, as it relies on local information to classify instances. The predicted class for a new instance is determined by a majority vote or weighted vote of its nearest neighbors [35].

Gradient boosting (GB) is an ensemble learning method that sequentially builds a strong model by combining the predictions of weak learners, typically decision trees. Each new tree corrects the errors of the previous ones, resulting in a highly accurate ensemble model. The prediction of the ensemble is a weighted sum of the predictions of the individual weak learners [36].

Extreme gradient boosting (XGBoost) is a powerful ensemble learning method that utilizes gradient boosting to sequentially train decision trees and minimize a differentiable loss function. By iteratively adding weak learners, XGBoost improves the model's performance while reducing bias and variance. The objective function in XGBoost comprises a combination of a differentiable loss function and a regularizer term, which is optimized during training [37].

The hyperparameter settings used to optimize the performance of the ML models and achieve better accuracy are detailed in the Table 3 below.

Table 3. The values for hyperparameters tuning.

Classifier	Parameters
DT	Max_depth = 6
RF	n_estimators = 42 random_state = 42
LRC	random_state = 42
KNN	n_neighbors = 5, metric = minkowski, p = 2
GB	n_estimators = 100, learning_rate = 1.0, max_depth = 5
XGB	booster = gbtree, max_depth = 6, gamma = 0, min_child_weight = 1, subsample = 1.0, colsample_bytree = 1.0

These hyperparameter configurations were fine-tuned to achieve optimal performance. For instance, the max_depth parameter in DT and XGB prevents overfitting by limiting the tree depth, while the n_estimators in RF, GB, and bagging enhances predictive stability. Additionally, using random_state ensures reproducibility, and parameters like learning_rate and min_child_weight help balance model complexity and learning efficiency.

2.7. Performance Metrics

A confusion matrix is a valuable tool for evaluating the performance of a classification model by comparing the model's predicted outcomes with the actual outcomes [38]. It provides a detailed summary of prediction results, which is especially useful in scenarios with imbalanced classes or when different classification errors have varying consequences. By analyzing the matrix, we can gain insights into the types and frequencies of

errors the model makes, aiding in understanding its overall performance. Figure 5 depicts the matrix, which includes key terms such as true positive (TP), true negative (TN), false negative (FN), and false positive (FP). Table 4 provides a summary of the performance measurement metrics used in this study. Accuracy represents the percentage of overall correct predictions, reflecting the model's ability to classify all cases correctly. Precision measures the rate of correct positive predictions among all predicted positive cases, indicating the model's reliability in identifying true positives. Recall highlights the percentage of correctly identified positive cases out of all actual positive cases, showcasing the model's sensitivity. Lastly, the F1-score combines precision and recall to provide a balanced measure of the classifier's precision and robustness, offering insight into the model's overall performance.

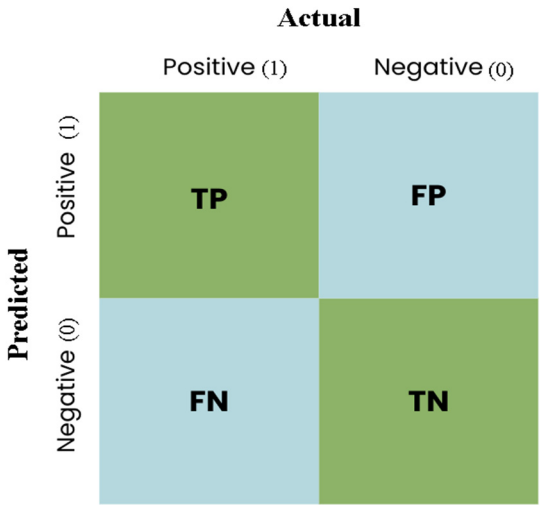


Figure 5. Format of confusion matrix.

Table 4. Description of performance measurement metrics:.

Metrics	Formula	Meaning
Accuracy	$\frac{TP+TN}{FP+TP+FN+TN} \times 100$	Rate of overall correct prediction
Precision	$\frac{TP}{TP+FP} \times 100$	Rate of correct positive predictions among all predicted positive cases
Recall	$\frac{TP}{TP+FN} \times 100$	Rate of correctly identified positive cases among all actual positive cases
F1-score	$2 \times \frac{Recall \times Precision}{Recall + Precision} \times 100$	Describe how precise and robust a classifier is

Receiver operating characteristic area under curve (ROC AUC): The ROC AUC is used to measure the area under the receiver operating characteristic (ROC) curve, which plots the true-positive rate (sensitivity) against the false-positive rate (1-specificity) for different threshold values [39]. It quantifies the model's ability to discriminate between

positive and negative classes across all possible thresholds. The ROC AUC score ranges from 0 to 1, where a score closer to 1 indicates better model performance.

K-fold cross-validation: We use the K-fold cross-validation method to assess the performance and validity of an ML model [40]. We divide the dataset into k equal parts or folds. The model is then trained and validated k times, using a different fold as the validation set and the remaining K−1 folds as the training set. This process generates K performance scores, which are averaged to produce a single performance estimate. K-fold cross-validation helps reduce overfitting and efficient use of data. This method provides a robust evaluation by leveraging various data subsets for training and validation.

2.8. Explainable AI (XAI)

XAI is a field focused on making the decision-making processes of AI systems transparent and understandable to humans [41]. As AI models become more complex, particularly with the rise of deep learning, their internal workings can often appear as "black boxes", making it difficult to understand how they produce results. XAI seeks to demystify these processes, providing clear explanations that build trust and allow users to challenge or correct the decisions made by AI systems.

2.8.1. Shapley Additive Explanations (SHAP)

SHAP is a methodology based on cooperative game theory that assigns a unique value to each feature, known as the SHAP value [42]. This value represents the contribution of that feature to the model's prediction for a specific instance. SHAP is versatile, offering both global explanations, which provide insights into the overall behavior of the model across the entire dataset, and local explanations that clarify how individual features influence specific predictions. SHAP values are consistent, ensuring that the sum of contributions from all features matches the difference between the model's output for a given sample and the average output. While SHAP offers a comprehensive understanding of feature importance and interactions, it can be computationally intensive, particularly for complex models and large datasets.

2.8.2. Local Interpretable Model–Agnostic Explanations (LIME)

LIME is a methodology that provides localized explanations by approximating the behavior of a complex model around a specific prediction [43]. It works by generating perturbed versions of the original data point with slight modifications to feature values and then training a simple, interpretable model (such as linear regression) on these modified data. This simple model serves as a local approximation of the complex model's behavior in close proximity to the prediction. Any ML model can apply LIME, and its computational efficiency makes it suitable for real-time analysis. However, LIME's explanations are valid only within the local region around the specific prediction and may not generalize well to the entire model, as it assumes local linearity, which might not always hold true.

3. Result and Discussion

3.1. Statistical Analysis

We performed a logistic regression analysis on the dataset using Python's statistics models [44] and the smf.logit function. The model predicts the likelihood of having learning difficulties based on several independent variables, including categorical and numerical ones. After fitting the model, the model summary provides us with coefficients, *p*-values, OR, and CI. The interpretation of results involves understanding the impact of each independent variable on the likelihood of having learning difficulties. Table 5

displays each variable's *p*-value, OR, and 95% CI. Below is a detailed analysis of all variables to understand their impact:

- Took away privileges: Children who had privileges taken away are 1.490 times more likely to have learning difficulties compared to those who did not. This association is statistically significant ($p < 0.001$), with a confidence interval of 1.284 to 1.728, indicating a strong effect;
- Shook child: The odds of having learning difficulties are 1.260 times higher for children who were shaken compared to those who were not. This finding is statistically significant ($p = 0.002$), with a confidence interval of 1.086 to 1.462;
- Beat child as hard as one could: Children who were severely physically punished (beaten as hard as one could) have 2.265 times higher odds of learning difficulties compared to those who were not. This effect is highly statistically significant ($p < 0.001$), with a confidence interval of 1.776 to 2.888;
- Sex: Male children are 1.215 times more likely to have learning difficulties compared to female children. This effect is statistically significant ($p = 0.009$), with a confidence interval of 1.049 to 1.406;
- Wealth index and mother's education: Children from the poorest households are 2.612 times more likely to have learning difficulties compared to the richest households. The odds decrease with increasing wealth, remaining significant in the second and middle quintiles but not in the fourth. Lower maternal education also increases the odds of learning difficulties, though not all levels are statistically significant;
- Called child dumb/lazy: The odds of learning difficulties for children who were called dumb or lazy are 1.077 times higher compared to those who were not, but this effect is not statistically significant ($p = 0.376$), with a confidence interval of 0.913 to 1.271;
- Weight-for-age percentile: The weight-for-age percentile does not show a statistically significant association with learning difficulties (OR = 1.000, $p = 0.963$), suggesting it has a negligible impact;
- Child's weight: Lower child weight is associated with slightly reduced odds of learning difficulties (OR = 0.954), and this effect is statistically significant ($p = 0.040$).

The analysis highlights several key factors associated with learning difficulties in children. Disciplinary practices, such as taking away privileges, shaking, and severe physical punishment, significantly increase the odds of learning difficulties. Socioeconomic status plays a crucial role, with children from the poorest households being at a much higher risk, and lower maternal education also contributing to this increased likelihood. Additionally, male children are more likely to experience learning difficulties compared to females.

Table 5. Association between the dependent variable and independent variables.

Factors	Samples with No Learning Difficulty	Samples with Having Learning Difficulty	OR	*p*-value	95% CI Interval	
					Low	High
Categorical						
Took away privileges						
NO (Reference)	5941	480	1			
YES	6516	337	1.490	0.000	1.284	1.728
Shook child						
NO (Reference)	5364	436	1			
YES	7093	381	1.260	0.002	1.086	1.462
Called child dumb lazy						
NO (Reference)	2955	244	1			
YES	9502	573	1.077	0.376	0.913	1.271
Beat child as hard as one could						

NO (Reference)	540	92	1			
YES	11,917	725	2.265	0.000	1.776	2.888
Sex						
FEMALE (Reference)	6418	465	1			
MALE	6039	352	1.215	0.009	1.049	1.406
Wealth index quintile						
Richest (Reference)	3064	331	1			
Poorest	2655	186	2.612	0.000	1.953	3.492
Second	2359	125	1.841	0.000	1.364	2.485
Middle	2351	108	1.401	0.033	1.027	1.911
Fourth	2028	67	1.275	0.128	0.933	1.743
Mother's education						
Higher secondary+ (Reference)	1493	138	1			
Pre-primary or none	3068	249	1.303	0.103	0.948	1.791
Primary	6130	358	1.216	0.188	0.909	1.627
Secondary	1766	72	1.110	0.449	0.848	1.452
Numerical	**Total samples**					
Weight-for-age percentile	13,274		1.000	0.963	0.994	1.006
Child's weight	13,274		0.954	0.040	0.912	0.998

The heatmap shown in Figure 6 reveals a moderately strong positive correlation (0.663) between "Child's_weight" and "Weight_age_P", indicating that as a child's weight increases, their weight-for-age percentile also tends to rise. In contrast, both "Child's_weight" and "Weight_age_P" have weak negative correlations with "Learning_skill" (−0.030 and −0.032, respectively), suggesting that these physical growth indicators have a minimal relationship with the child's learning skills compared to others.

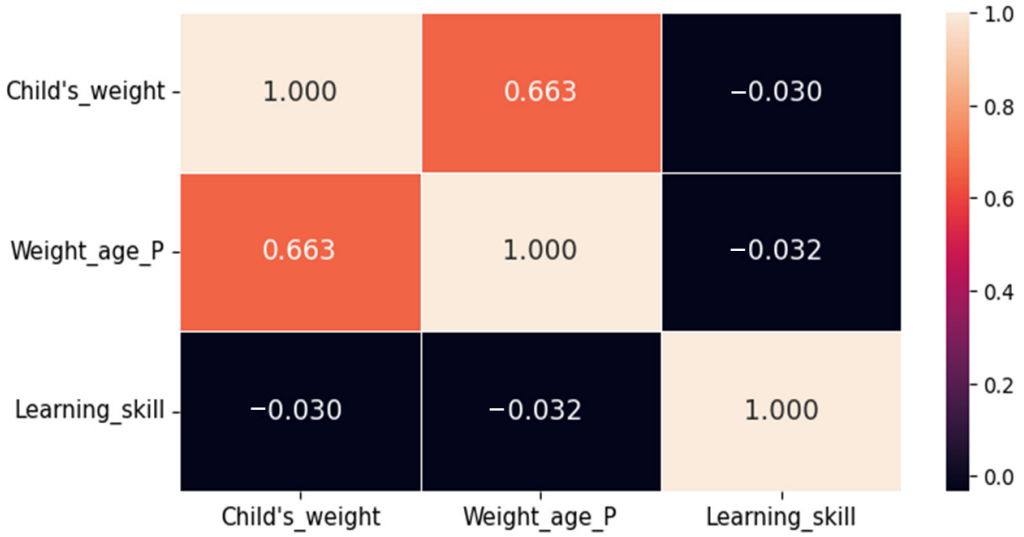


Figure 6. Correlation between child's weight, weight-for-age percentile, and learning skill compared to others.

3.2. ML Model Performance

These models were chosen to effectively capture the complex and multi-faceted nature of the dataset, which includes various socio-economic, health, and educational indicators relevant to learning difficulties. KNN was used for its simplicity and effectiveness in handling non-linear relationships within the dataset. LRC was chosen for its interpretability and ability to model linear relationships. DT and RF were selected for their

flexibility in managing hierarchical data, while GB and XGB were employed to improve predictive accuracy through their iterative error-correcting process and capacity to handle complex interactions in the data. By utilizing these models, the study aimed to validate its findings across different ML algorithms, ensuring the robustness of the results within the context of the MICS 2019 dataset.

Figure 7 presents a comprehensive comparison of several machine learning algorithms applied to the MICS 2019 dataset for predicting child learning abilities. We evaluated the models using various performance metrics and assessed them through 10-fold cross-validation. Furthermore, we analyzed the performance metrics of various classification models to assess their effectiveness in predicting child learning difficulties. The DT model demonstrated moderate performance, with an accuracy of 0.70, precision of 0.67, and recall of 0.80, resulting in an F1-score of 0.73. In contrast, the KNN model showed better results with an accuracy of 0.86, a high recall of 0.97, and an F1-score of 0.87. LRC performed similarly across most metrics, with values around 0.69. The RF model performed well, achieving an accuracy of 0.91, precision of 0.89, recall of 0.93, and an F1-score of 0.91. GB and XGB yielded the best results, achieving exceptional performance metrics such as accuracy of 0.95, precision of 0.96–0.97, recall of 0.93, and an F1-score of 0.95. These models also had high NPV and ROC AUC scores of 0.93 and 0.95, respectively. This analysis suggests that GB and XGB are the most effective models for this application, outperforming other classifiers in nearly all metrics. The results show how well different ML models predict learning difficulties in children. It compares their performance using various measures that assess how accurately the models identify children with and without difficulties and how balanced their predictions are. RF, XGB, KNN, and GB perform very well, with results close to perfect, meaning they make highly reliable predictions and are good at correctly identifying both children with and without learning difficulties. In contrast, models like DT and LRC perform slightly less effectively. Their predictions are still useful but less consistent and accurate compared to the top-performing models.

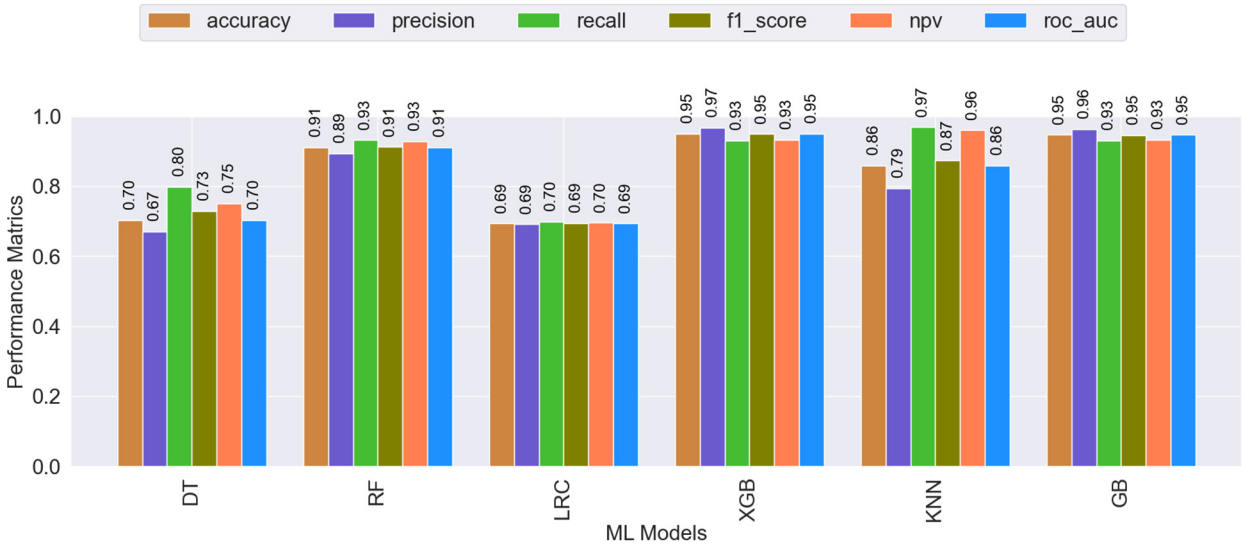


Figure 7. Performance metrics of various ML algorithms for predicting child learning difficulties.

The ML models in this study were trained on a workstation equipped with an Intel Core i3 CPU @ 2.3 GHz and 12 GB of RAM. Training times varied across models, reflecting their complexity and the chosen hyperparameters. The Table 6 shows that the DT model was the fastest, with a training time of 0.06 s and a testing time of 0.0007 s. Following DT, LRC had the best testing time at 0.0005 s, and XGB came next with a testing time of 0.0094 s. Other algorithms, while requiring more time, still provided valuable results with good

accuracy and performance. These training times demonstrate the computational efficiency of each model within the constraints of the available resources, allowing for effective model evaluation and tuning without significant delays.

Table 6. Time Requirements for Training ML Models.

Models	Train_Time (s)	Test_Time (s)
DT	0.063787	0.000712
KNN	2.099552	0.037495
LRC	0.243012	0.00052
RF	2.256923	0.261399
GB	5.796678	0.011094
XGB	1.841549	0.009374

Furthermore, the Table 7 presents the performance metrics of the XGB model, evaluated using 1- to 10-fold cross-validation for predicting child learning difficulties. Each fold displayed consistent results, with high accuracy precision, recall, F1-score, NPV, and ROC AUC score. Overall, the table highlights the robust performance of the XGB model across multiple evaluation metrics.

Table 7. K-fold cross-validation performance metrics for XGB model predicting child learning difficulties.

K_Fold	Accuracy	Precision	Recall	F1_Score	npv	roc_auc
1_Fold	0.95	0.97	0.92	0.94	0.94	0.95
2_Fold	0.94	0.95	0.92	0.94	0.94	0.94
3_Fold	0.94	0.96	0.92	0.94	0.94	0.94
4_Fold	0.94	0.96	0.92	0.94	0.94	0.94
5_Fold	0.95	0.97	0.93	0.95	0.94	0.95
6_Fold	0.95	0.97	0.93	0.95	0.94	0.95
7_Fold	0.95	0.97	0.93	0.95	0.94	0.95
8_Fold	0.95	0.97	0.93	0.95	0.94	0.95
9_Fold	0.95	0.97	0.93	0.95	0.94	0.95
10_Fold	0.95	0.97	0.93	0.95	0.94	0.95

The confusion matrices in Figure 8 generated by the six different ML models DT, RF, LRC, KNN, GB, and XGB provide key insights into their effectiveness in predicting early child learning difficulties, where the outcomes are classified as either "NO DIFFICULTY" (0) or "HAS DIFFICULTY" (1). The XGB model stands out as a very good performer, with the highest number of true positives (1168) and true negatives (1208), coupled with the lowest number of false positives (36) and false negatives (79), indicating its high accuracy in identifying both children who have and those who do not have learning difficulties. Similarly, the GB model demonstrates strong performance, with slightly fewer true positives (1170), more true negatives (1212), and very few false positives (32), making it another reliable option. The RF model also performs well, with a favorable balance of true positives (1154) and true negatives (1100), though it has a moderate number of false positives (144) and false negatives (93), indicating some room for improvement in reducing misclassifications. The KNN model, despite its strong identification of true positives (1199), struggles with a higher number of false positives (324), making it less reliable in correctly identifying children without learning difficulties. The LRC, however, has a significant number of false negatives (397), posing a risk of under-detection. The DT model, with a substantial number of false positives (500), indicates a tendency to over-predict

learning difficulties. Overall, XGB and GB emerge as the most effective models, offering the best balance between accurately predicting both classes and minimizing errors.

The ROC AUC curves in Figure 9 present a comparative analysis of six ML models applied to a dataset on child learning difficulties. The XGB and GB models, indicated by the orange and purple curves, respectively, both achieved the highest ROC AUC of 0.95, demonstrating their exceptional ability to differentiate between classes. The RF model, represented by the red curve, also performed well with an ROC AUC of 0.91, though it lagged slightly behind XGB and GB. In contrast, the DT, LRC, and KNN models, represented by blue, cyan, and green curves, respectively, each achieved an ROC AUC of 0.70, 0.69, and 0.86. This suggests that these models are less effective compared to the top-performing XGB, GB, and RF models. Overall, the analysis indicates that XGB and GB are the most effective models for predicting child learning difficulties, with RF following closely behind, while DT, LRC, and KNN exhibit lower classification performance.

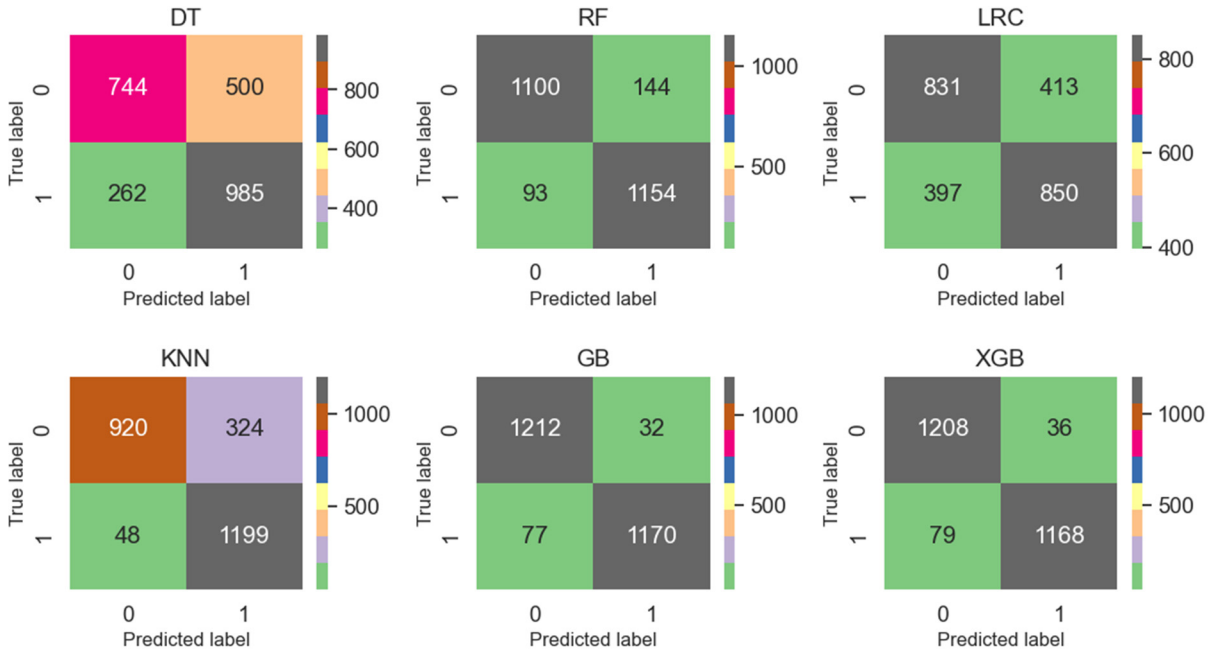


Figure 8. Confusion metrics for various machine learning algorithms for child learning difficulty prediction.

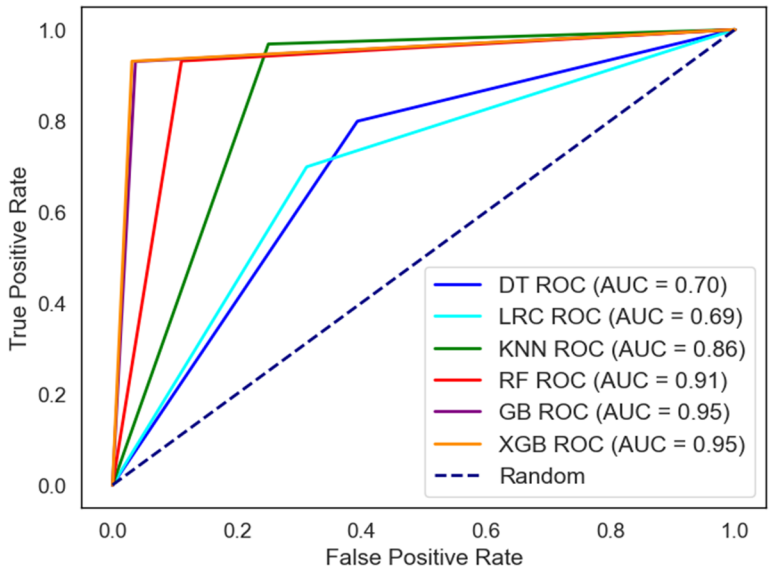


Figure 9. ROC AUC Curves for Various ML Algorithms for Child learning difficulties prediction.

3.3. SHAP Explanation

According to the ML model performance matrices, it is clear that the XGB and GB models perform best. Therefore, we utilized the XGB model in our XAI analysis to interpret the outputs based on the influence of the input features.

In Figure 10, the SHAP summary plot provides a detailed analysis of how various factors influence the probability of learning difficulties in children. Here, the negative SHAP value indicates the absence of learning difficulties, while the positive value indicates the presence of difficulties. The figure shows that removing privileges is a significant risk factor, increasing the probability of learning challenges. In contrast, a higher wealth index serves as a strong protective factor, reducing this risk by providing a more supportive environment for learning. Sex has a relatively smaller impact, suggesting that while gender differences exist, they are not as influential as socioeconomic factors. Negative behaviors, such as shaking the child, physical punishment (e.g., beating the child hard), and verbal abuse (e.g., calling the child dumb or lazy), are strongly associated with an increased likelihood of learning difficulties. These factors highlight the detrimental effects of harsh disciplinary actions on a child's cognitive development. On the other hand, higher levels of maternal education and better physical health indicators, such as a higher weight-for-age percentile and child's weight, act as protective factors, reducing the risk of learning difficulties. This emphasizes the importance of supportive environments in lowering the risk of learning difficulties and demonstrates how harmful negative behaviors and low socioeconomic status can be.

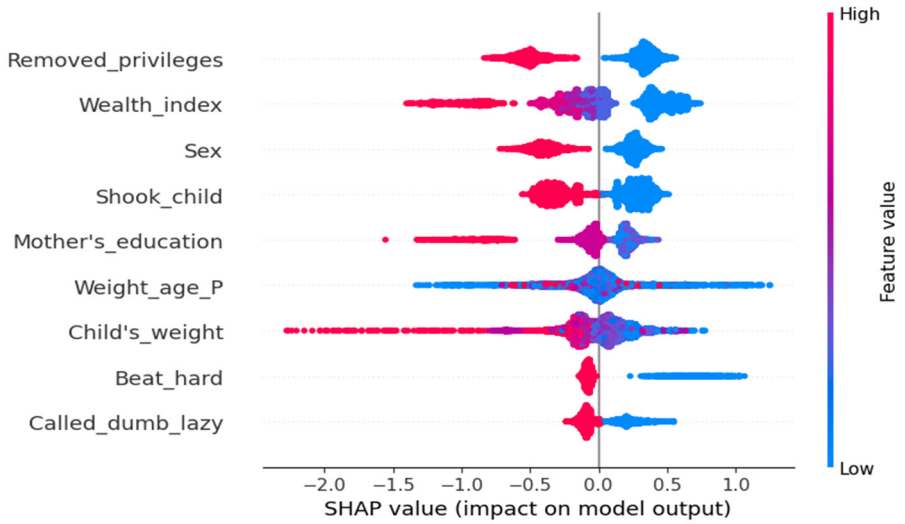


Figure 10. SHAP summary plot showing the influence of input features on children learning difficulties.

Furthermore, the SHAP waterfall plots reveal how various factors contribute to predicting the learning skill difficulties for two specific children. Both plots emphasize the strong influence of parenting practices, physical health, and socio-economic factors on the likelihood of learning difficulties in children.

In Figure 11, the model predicts a higher probability that the child has learning difficulties, as the output is f(x) = 1.229. The primary factors contributing to this prediction are the following:

- Beating hard (YES = 1) significantly increases the likelihood of learning difficulties (+0.51). Harsh physical punishment is strongly correlated with emotional and cognitive challenges in children (Beating or physically punishing a child is a crime

punishable by law in many countries. Such abusive behavior has been proven to severely hinder a child's emotional, cognitive, and educational development, including their learning capabilities);

- Removal of privileges (YES = 1) adds positively to the prediction (+0.39), as this negative disciplinary action can affect a child's mental well-being and learning ability;
- Shaking the child (NO = 2) contributes negatively (−0.34), implying that the absence of shaking reduces the likelihood of learning difficulties;
- Sex (MALE = 1) increases the likelihood of difficulties (+0.2), indicating that male children may face a higher probability of learning issues in this dataset;
- Mother's education (Primary = 1) has a moderate positive influence (+0.18), showing that lower maternal education correlates with a higher risk of learning difficulties;
- Weight–age percentile (0.8) and called dumb/lazy (YES = 1) have smaller positive contributions (+0.12 and +0.1, respectively), indicating that poor physical health and verbal abuse contribute to learning difficulties;
- Child's weight (11.9) and wealth index (Second = 2) have minimal effects, with wealth index showing a slight negative impact (−0.02), suggesting that lower socio-economic status marginally increases learning difficulties.

Again, in Figure 12, the model predicts a lower probability of the child having learning difficulties, as the output is f(x) = −2.23. The primary factors contributing to this prediction are the following:

- Removed privileges (NO = 2) significantly reduce the likelihood of learning difficulties (−0.77). This suggests that not using this negative disciplinary tactic supports a positive outcome;
- Child's weight (13.6) and weight–age percentile (29.9) both have substantial negative contributions (−0.66 and −0.48, respectively), implying that good physical health strongly correlates with better cognitive outcomes;
- Shaking the child (NO = 2) also contributes negatively (−0.44), reinforcing that avoiding physical mistreatment reduces learning challenges
- Sex (MALE = 1) has a smaller positive impact (+0.28), suggesting that being male slightly increases the likelihood of learning difficulties, though its effect is less significant in this prediction;
- Called dumb/lazy (NO = 2) contributes negatively (−0.13), indicating that not verbally labeling the child helps reduce the likelihood of learning difficulties;
- Mother's education (None = 0) and wealth index (Second = 2) provide smaller positive and negative influences (+0.12 and −0.11, respectively), showing a marginal effect of socio-economic and educational factors;
- Beating hard (NO = 2) has a slight negative contribution (−0.07), suggesting that avoiding physical punishment supports lower chances of learning difficulties.

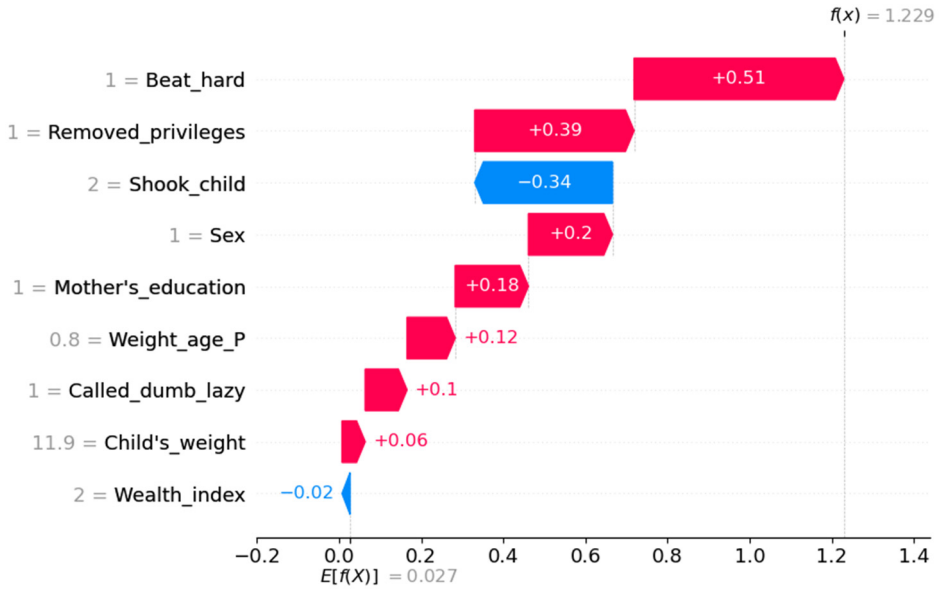


Figure 11. SHAP waterfall plot for predicting the presence of learning difficulties in a child.

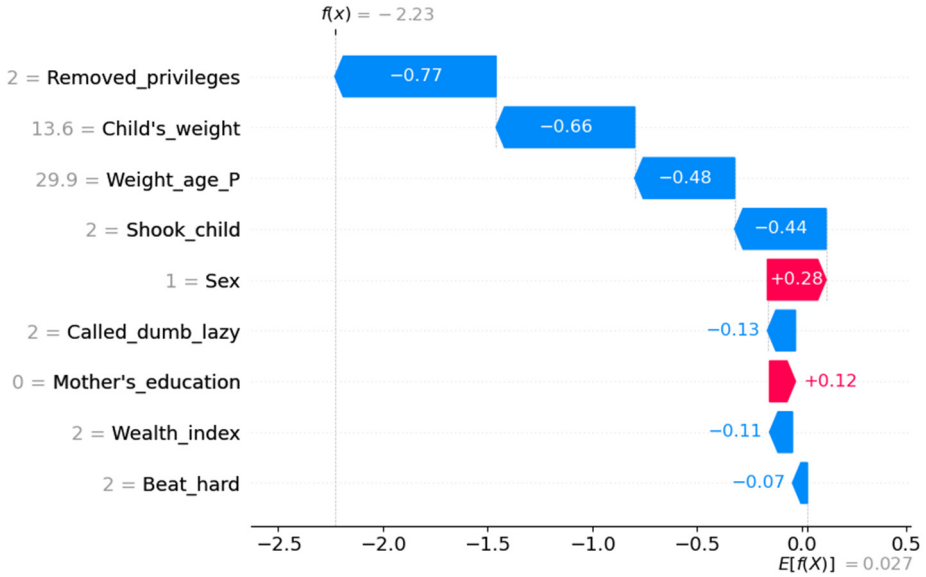


Figure 12. SHAP waterfall plot for predicting the absence of learning difficulties in a child.

3.4. LIME Explanation

The LIME plots offer a comprehensive explanation of the model's predictions on whether a child has learning difficulties based on various input features. These explanations, which were generated using the best-performing XGB model, emphasize the critical role of factors such as health, economic conditions, and parenting practices in predicting learning difficulties. They also identify areas that require improvement, especially in addressing factors that adversely affect a child's learning abilities.

In Figure 13, the model predicts a 0.93 probability that the child does not have learning difficulties. The primary factors contributing to this prediction are as follows:

- The factors of child's weight (14.30) and weight–age percentile (37.50) significantly influence the prediction that the child does not have learning difficulties. Higher values indicate good physical health, which is strongly associated with better cognitive development and learning capabilities;
- A higher wealth index (Richest = 5), which often provides access to better resources and educational opportunities, contributes to a lower likelihood of learning

difficulties. This positive socio-economic status helps to minimize the negative impact of removing privileges (YES = 1), which otherwise increases the risk of learning difficulties;

- A lower level of maternal education (Primary = 1) is often associated with a higher likelihood of experiencing learning difficulties, which has some impact on the prediction;
- Negative parenting behaviors, such as shaking the child (YES = 1), increase the likelihood of learning difficulties. But beating hard (NO = 2) and calling the child dumb/lazy (NO = 2) indicate that the absence of severe physical punishment and abuse decreases the likelihood of learning difficulties;
- The factor Sex (FEMALE = 2) contributes to the prediction that female children have a lower probability of experiencing learning difficulties.

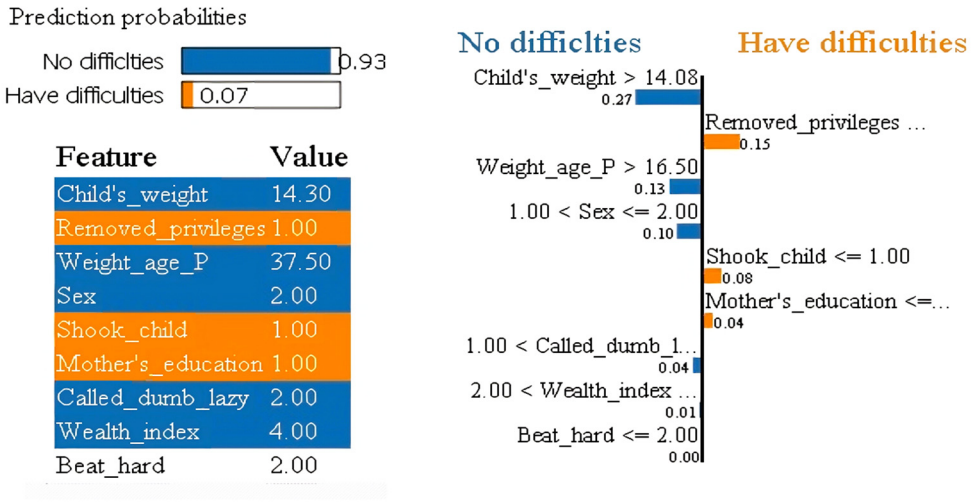


Figure 13. LIME plot explaining prediction of the absence of learning difficulties in a child.

Furthermore, in Figure 14, the model predicts with a high probability (1.00) that the child has learning difficulties. The key contributing factors to this prediction are as follows:

- The factors of the child's weight (11.28) and weight–age percentile (1.30) significantly contribute to the prediction of learning difficulties because the child's low weight and weight-for-age percentile suggest poor health, which is often associated with developmental challenges;
- The factors like removing privileges (YES = 1) along with wealth index (Poorest = 1) conditions are linked to a higher likelihood of learning difficulties;
- The mother's education (Primary = 1) factor slightly influences the prediction, as a lower level of maternal education is often associated with an increased risk of learning difficulties;
- The factor sex (FEMALE = 2) contributes to the "No difficulties" prediction in this case;
- Negative parenting behaviors: Shaking the child (NO = 2) and beating hard (NO = 2) slightly decreases the likelihood of learning difficulties, but calling the child dumb/lazy (YES = 1) contributes to the prediction of have difficulties, reinforcing the model's overall prediction.

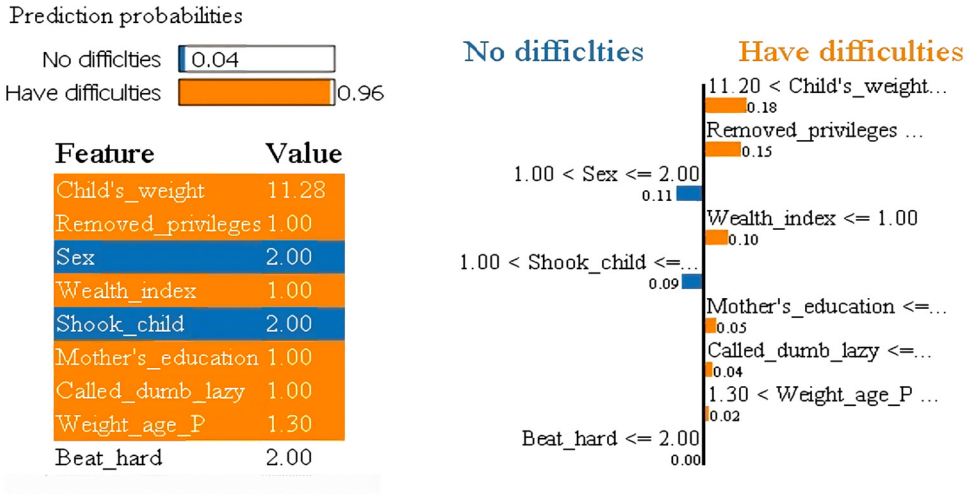


Figure 14. LIME plot explaining prediction of the presence of learning difficulties in a child.

4. Comparison with Other Works

This study compared its findings with research from Oman, Australia, and the U.S., focusing on early childhood learning outcomes. While previous studies used diverse datasets and methods, our approach uniquely applies the MICS 2019 dataset to the Bangladeshi context. Unlike prior research, which often lacked explainability, we integrated SHAP and LIME to enhance model interpretability. This comparison highlights our study's contribution to understanding learning difficulties through a region-specific lens, offering actionable insights for policymaking and education strategies.

Table 8 presents a detailed comparison of our work with other research studies, emphasizing methodologies, datasets, statistical applications of ML, and XAI techniques. The table highlights performance metrics like accuracy, precision, recall, and specificity to demonstrate the effectiveness of different approaches. Additionally, it outlines the integration of statistical ML and XAI tools, providing a comprehensive overview of how our work aligns with or advances existing research in the field.

One study explored factors impacting academic performance using statistical ML techniques, specifically the J48 algorithm, applied to a dataset of 6514 students over 11 years at a public university in Oman. The study achieved an accuracy of 82.4% but did not integrate any XAI methods or statistical explanations to interpret the results [8]. Another study applied ML to identify risk factors for ADHD in children using data from 45,779 participants in the 2018–2019 National Survey of Children's Health (NSCH). While logistic regression and eight ML classifiers were tested, the RF classifier stood out with an accuracy of 85.5% and an AUC of 0.94. However, the study focused only on statistical ML performance without employing XAI to explain the models [9].

Similarly, the study [10] predicted depression in children and adolescents using mental health and socio-environmental features from 6310 Australian participants (2013–2014). ML models like RF, XGB, DT, and Gaussian naive Bayes were used, with the RF model achieving 95% accuracy and 99% precision. Despite the robust performance, the study did not utilize XAI tools for model interpretability or statistical analysis of feature importance. Another research [11] evaluated ML techniques for predicting academic performance factors using data from 5084 students aged 10–17 in West Bank schools. Among the models tested (RF, ANN, SVM, DT, NB, and LR), LR achieved the best results with a sensitivity of 94.3%, specificity of 96.3%, and an AUC of 99.3%. However, no XAI techniques were used to explain the predictive factors.

In studies focused on autism spectrum disorder (ASD), research [15] analyzed data from 292 children and 704 adults using multiple ML algorithms. The ANN model achieved 94.24% accuracy and a high specificity of 95.05%, but no XAI methods were applied to interpret the findings statistically. Another study [16] focused on toddler ASD diagnosis, achieving a perfect performance of 100% accuracy, precision, recall, and F1-score using LR but without incorporating statistical analysis or XAI tools.

Table 8. Summary of research outcomes.

Ref.	Findings	Data Source	ML Application	Statistical Analysis	XAI Uses	ML Performance
[8]	Identified factors impacting academic performance using ML algorithms	Data from 6514 students over 11 years at a public university in Oman	J48 algorithm	×	×	J48 achieved 82.4% accuracy
[9]	Identified risk factors for ADHD in children using ML classification	2018–2019 NSCH data from 45,779 children aged 3–17 years	Logistic regression and eight ML classifiers	✓	×	RF classifier achieved 85.5% accuracy, sensitivity of 84.4%, specificity of 86.4%, and AUC of 0.94
[10]	Predicted depression in children and adolescents using ML models	Australian child and adolescent responses from 6310 participants (2013–2014)	RF, XGB, DT, and Gaussian naive Bayes algorithms	×	×	RF model achieved 95% accuracy and 99% precision
[11]	Evaluated ML techniques for predicting academic performance factors	Data from 5084 students aged 10–17 in the West Bank	RF, ANN, SVM, DT, NB, and LR models	×	×	LR model achieved sensitivity of 94.3%, specificity of 96.3%, and AUC of 99.3%
[15]	Improved early ASD diagnosis using machine learning	Obtained from UCI Repository; 292 children and 704 adults	RF, KNN,XGB, SVM, DT, NB, LR, and ANN	×	×	ANN achieved 94.24% accuracy, precision 89.80% recall 92%, F1-score, 91%, specificity 95.05%, and AUC of 93.84%

[16]	Enhanced toddler ASD diagnosis using machine learning.	From UCI ML Repository; 1054 toddlers (12–36 months) via ASDTests app	SVM, RF, LR, DT, KNN, and NB	×	×	LR achieved 100% accuracy, 100% precision, 100% recall, and 100% F1-score
This work	Identified key factors and predicted learning difficulties in Bangladeshi children under 5 years old using ML models	Refined dataset of 13,274 samples and 19 features from MICS 2019	DT, LR, KNN, RF, XGB, GB algorithms, and SHAP and LIME explanations	✓	✓	XGB classifier achieved 95% accuracy, 97% precision, 93% recall, 95% F1-score, 93% NPV, and 95% ROC

In contrast, our work integrated both statistical ML and XAI to identify key factors and predict learning difficulties in Bangladeshi children under five years of age. Using a refined dataset from MICS 2019, we employed ML algorithms like DT, LR, KNN, RF, XGB, and GB, alongside SHAP and LIME for interpretability. This combination allowed us to explain the significance of individual features statistically while enhancing the transparency of our models. The XGB classifier demonstrated superior performance, achieving 95% accuracy, 97% precision, 93% recall, 95% F1-score, 93% NPV, and an ROC of 95%. Our work not only matches or exceeds the performance metrics of prior studies but also provides a deeper understanding of the underlying factors through the integration of XAI, setting it apart from existing research.

5. Policy Implications

Figure 15 illustrates a structured framework that outlines a data-driven approach for predicting and addressing early childhood learning difficulties. It integrates family data collected by NGOs or health organizations, focusing on health, nutrition, family guidance, education, and socioeconomic status. Data were entered, processed, and analyzed through ML models to make predictions. XAI techniques were applied to identify key factors influencing learning difficulties, enhancing the interpretability of results. The identified learning gaps inform targeted policy implications and interventions, such as early childhood education, health services, parenting support, nutritional programs, and economic assistance. This framework provides actionable insights to address disparities and improve educational outcomes through informed policymaking.

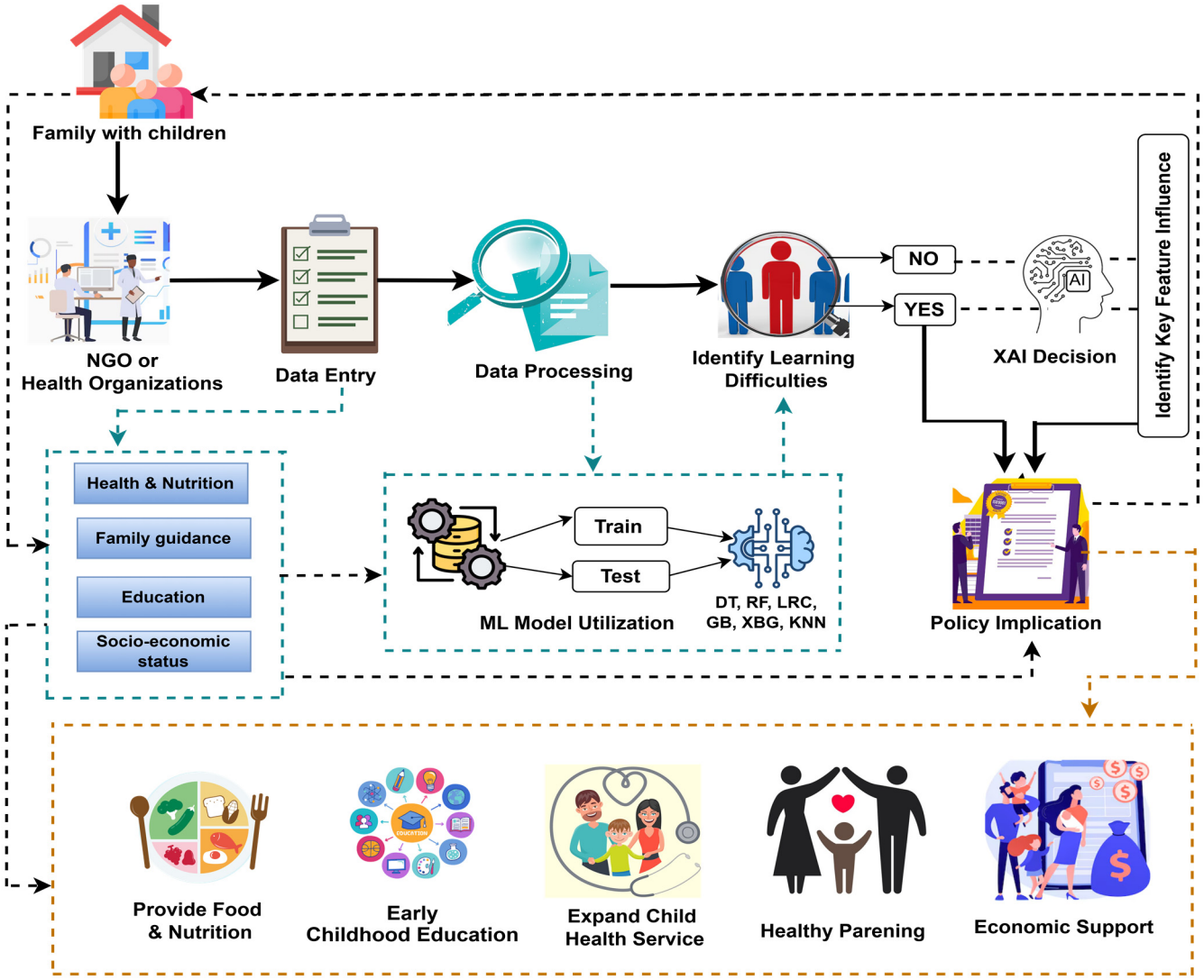


Figure 15. A data-driven framework for early childhood learning difficulties prediction and prevention policy implications.

The following steps will help in policy implications for addressing child learning disabilities:

- **Health and Nutrition Interventions:** To support children's health and development, it is essential to provide nutritional supplements or balanced meals to promote proper growth. Alongside this, organizing regular health check-ups and immunization drives helps prevent illnesses and ensures overall well-being. Additionally, educating parents on the importance of maintaining a healthy diet and hygiene creates a strong foundation for sustained child health;
- **Parenting Support:** Workshops should be conducted to raise awareness of early childhood care, provide counseling for behavioral issues, and encourage activities that strengthen parent–child bonds;
- **Education Enhancements:** Remedial or special education programs should be implemented to address the individual learning needs of children. At the same time, teachers should be trained in inclusive education practices and early detection of learning challenges, ensuring that resources are accessible for children with special needs;
- **Socio-economic Support:** Scholarships or financial aid should be provided to underprivileged families, and they should be linked to social welfare schemes to improve their living standards. This would be complemented by ensuring access to technology and learning aids in under-resourced communities, promoting equal opportunities for all children.

Generalizability and Limitations

This study, which uses the MICS 2019 dataset, provides useful insights into predicting early learning issues in children under the age of five years old. However, there are some limitations to consider. The use of self-reported data presents potential reporting biases. Guardians and parents might understate developmental issues due to cultural shame, a lack of understanding, or subjective impressions, hampering the dataset's credibility. Also, the dataset does not include all the important factors that can indicate learning difficulties, like family history, prenatal conditions, or socio-environmental factors. Furthermore, while the MICS dataset represents the Bangladeshi population, the socioeconomic and cultural diversity of other regions may restrict the findings' applicability to other populations. Despite these limitations, the proposed work demonstrates substantial potential for adaptation and application across other similar contexts (ASD, ADHD, etc.). By retraining and validating the models with region-specific data, the approach could be extended to address early learning challenges in diverse populations.

6. Conclusions

This study conducted both statistical and ML approaches to analyze the learning abilities of children under five years old, using the MICS 2019 dataset. The findings show that parental education, income, and parenting practices strongly affect children's learning outcomes. Additional factors contribute to learning difficulties, such as deprivation, intense physical punishment, and verbal reprimands. Socioeconomic factors exacerbate these issues, particularly urban living and poorer households. We then applied ML models to predict these difficulties and found that the GB and XGB classifiers performed best, achieving high accuracy, precision, recall, and F1-scores. Moreover, these findings underscore the importance of addressing socioeconomic disparities and improving parenting practices to reduce learning difficulties. Also, XAI techniques like SHAP and LIME provided more information about how different factors affect these predictions, which makes the results easier to understand and more open. The dataset may not have adequately captured some important variables, which is one of the study's limitations. Additionally, the findings are specific to Bangladesh and may not directly apply to other socio-economic and cultural contexts without further validation. Despite these limitations, the study emphasizes the importance of addressing socioeconomic disparities and promoting positive parenting practices to improve learning outcomes. The insights gained here can inform targeted policy interventions to reduce learning difficulties and foster comprehensive child development. In the future, researchers should focus on using longitudinal datasets to determine more factors to obtain a better picture of more complex factors, making sure that the results are valid to make them more useful.

Author Contributions: Conceptualization, M.A.M. and M.R.K.; methodology, M.M.H.; validation, M.A.M. and W.R.; formal analysis, M.R.K.; investigation, A.M.; resources, M.M.H.; writing—original draft preparation, M.A.M. and M.R.K.; writing—review and editing, A.M.; visualization, M.R.K.; supervision, M.M.H. and A.M.; funding acquisition, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The source of dataset is mentioned in Section 2.1.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. López-Torres, L.; Hidalgo-Montesinos, M.D. Parental involvement and children's academic achievement: A meta-analysis. *Front. Psychol.* **2020**, *11*, 206.
2. Turney, K.; Kao, G. Parental involvement and academic achievement: A meta-analysis. *Soc. Forces* **2023**, *84*, 841–876.
3. Cabrera, N.J.; Hofferth, S.L.; Chae, S. Patterns and predictors of father-infant engagement across race/ethnic groups. *Early Child. Res. Q.* **2022**, *43*, 44–58.
4. Han, W.J.; Ruhm, C.J. Maternal nonstandard work schedules and child cognitive outcomes. *Child Dev.* **2024**, *95*, 394–403.
5. Liu, D.; Diorio, J.; Tannenbaum, B.; Caldji, C.; Francis, D.; Freedman, A.; Meaney, M.J. Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. *Science* **2023**, *277*, 1659–1662.
6. Driessen, J. Socioeconomic status and parenting. In *The Encyclopedia of Child and Adolescent Development*; S. Hupp, S., Jewell, J., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2016; pp. 1–8.
7. Englund, M.M.; Luckner, A.E.; Whaley, G.J.L.; Egeland, B. Children's achievement in early elementary school: Longitudinal effects of parental involvement, expectations, and quality of assistance. *J. Educ. Psychol.* **2020**, *96*, 723–730.
8. Al-Alawi, L.; Al Shaqsi, J.; Tarhini, A.; Al-Busaidi, A.S. Using Machine Learning to Predict Factors Affecting Academic Performance: The Case of College Students on Academic Probation. *Educ. Inf. Technol.* **2023**, *28*, 12407–12432.
9. Maniruzzaman, M.; Shin, J.; Hasan, M.A.M. Predicting Children with ADHD Using Behavioral Activity: A Machine Learning Analysis. *Appl. Sci.* **2022**, *12*, 2737. https://doi.org/10.3390/app12052737.
10. Haque, U.M.; Kabir, E.; Khanam, R. Detection of child depression using machine learning methods. *PLoS ONE* **2021**, *16*, e0261131. https://doi.org/10.1371/journal.pone.0261131.
11. Qasrawi, R.; VicunaPolo, S.; Al-Halawa, D.A.; Hallaq, S.; Abdeen, Z. Predicting School Children Academic Performance Using Machine Learning Techniques. *Adv. Sci. Technol. Eng. Syst. J.* **2021**, *6*, 8–15. https://doi.org/10.25046/aj060502.
12. Yasumura, A.; Omori, M.; Fukuda, A.; Takahashi, J.; Yasumura, Y.; Nakagawa, E.; Koike, T.; Yamashita, Y.; Miyajima, T.; Koeda, T.; et al. Applied Machine Learning Method to Predict Children With ADHD Using Prefrontal Cortex Activity: A Multicenter Study in Japan. *J. Atten. Disord.* **2017**, *24*, 2012–2020. https://doi.org/10.1177/1087054717740632.
13. Akter, T.; Khan, M.I.; Ali, M.H.; Satu, M.S.; Uddin, M.J.; Moni, M.A. Improved Machine Learning based Classification Model for Early Autism Detection. In Proceedings of the 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, 5–7 January 2021. https://doi.org/10.1109/icrest51555.2021.9331013.
14. Vakadkar, K.; Purkayastha, D.; Krishnan, D. Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques. *SN Comput. Sci.* **2021**, *2*, 386. https://doi.org/10.1007/s42979-021-00776-5.
15. Rasul, R.A.; Saha, P.; Bala, D.; Karim, S.R.U.; Abdullah, I.; Saha, B. An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder. *Healthc. Anal.* **2024**, *5*, 100293.
16. Shambour, Q. Artificial Intelligence Techniques for Early Autism Detection in Toddlers: A Comparative Analysis. *J. Appl. Data Sci.* **2024**, *5*, 1754–1764.
17. Ali N, Ullah A, Khan AM, Khan Y, Ali S, Khan A; et al. Academic performance of children in relation to gender, parenting styles, and socioeconomic status: What attributes are important. *PLoS ONE* **2023**, *18*, e0286823. https://doi.org/10.1371/journal.pone.0286823.
18. Vasiou, A.; Kassis, W.; Krasanaki, A.; Aksoy, D.; Favre, C.A.; Tantaros, S. Exploring Parenting Styles Patterns and Children's Socio-Emotional Skills. *Children* **2023**, *10*, 1126. https://doi.org/10.3390/children10071126.
19. Armstrong, R.; Symons, M.; Scott, J.G.; Arnott, W.L.; Copland, D.A.; McMahon, K.L.; Whitehouse, A.J.O. Predicting Language Difficulties in Middle Childhood From Early Developmental Milestones: A Comparison of Traditional Regression and Machine Learning Techniques. *J. Speech Lang. Hear. Res.* **2018**, *61*, 1926–1944. https://doi.org/10.1044/2018_jslhr-l-17-0210.
20. Bangladesh Bureau of Statistics (BBS); UNICEF. *Bangladesh Multiple Indicator Cluster Survey 2019: Final Report*; BBS and UNICEF: Dhaka, Bangladesh, 2019.
21. Bhuvaneswari, K.; Kanimozhi, G.; Shanmuganeethi, V. Prediction of student learning difficulties and performance using regression in machine learning. *Distance Educ. E-Learn.* **2023**, *11*, 2455–2460.
22. Wu, T.K.; Huang, S.C.; Meng, Y.R. Evaluation of ANN and SVM classifiers as predictors to the diagnosis of students with learning disabilities. *Expert Syst. Appl.* **2008**, *34*, 1846–1856. https://doi.org/10.1016/j.eswa.2007.02.026.
23. Heinze, G.; Wallisch, C.; Dunkler, D. Variable selection-A review and recommendations for the practicing statistician. *Biom. J.* **2018**, *60*, 431–449.

24. Pryke, A.; Mostaghim, S.; Nazemi, A. Heatmap visualization of population based multi objective algorithms. In Proceedings of the Evolutionary Multi-Criterion Optimization: 4th International Conference, EMO 2007, Matsushima, Japan, 5–8 March 2007; Proceedings 4; Springer: Berlin/Heidelberg, Germany, 2007; pp. 361–375.
25. Little, R.J.A.; Rubin, D.B. *Statistical Analysis with Missing Data*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2019. https://doi.org/10.1002/9781119482260.
26. Johnson, M.; Davis, E. A Comparative Study of Outlier Detection Algorithms for Data Cleansing. *Data Sci. J.* **2021**, *20*, 1–15. https://doi.org/10.5334/dsj-2021-021.
27. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed.; Morgan Kaufmann: Burlington, MA, USA, 2016.
28. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: Berlin/Heidelberg, Germany, 2009.
29. Chawla, N.V.; Lazarevic, A.; Hall, L.O.; Bowyer, K.W.; (Eds.). SMOTEBoost: Improving prediction of the minority class in boosting. In *European Conference on Principles of Data Mining and Knowledge Discovery*; Springer: Berlin/Heidelberg, Germany, 2003.
30. Guo, X.; Yin, Y.; Dong, C.; Yang, G.; Zhou, G.; (Eds.). On the class imbalance problem. Natural Computation. In Proceedings of the 2008 ICNC'08 Fourth International Conference on IEEE, 18–20 October 2008, China.
31. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2013; Volume 112.
32. Zhang, L. Using Decision Tree Classification Algorithm to Predict Learner Outcomes. *J. Educ. Data Sci.* **2023**, *7*, 45–60.
33. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
34. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. Applied Logistic Regression. *John Wiley & Sons.:* 2013.
35. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27.
36. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **2001**, *29*, 1189–1232.
37. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
38. Townsend, J.T. Theoretical analysis of an alphabetic confusion matrix. *Percept. Psychophys.* **1971**, *9*, 40–50.
39. Park, S.H.; Goo, J.M.; Jo, C.H. Receiver operating characteristic (ROC) curve: Practical review for radiologists. *Korean J. Radiol.* **2004**, *5*, 11.
40. Aghbalou, A.; Sabourin, A.; Portier, F. On the bias of K-fold cross-validation with stable learners. Proceedings of The 26th International Conference on Artificial Intelligence and Statistics. *Proc. Mach. Learn. Res.* **2023**, *206*, 3775–3794.
41. Arrieta, A.B.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-López, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115.
42. Roshan, K.; Zafar, A. Utilizing XAI technique to improve autoencoder based model for computer network anomaly detection with shapley additive explanation (SHAP). *arXiv* **2021**, arXiv:2112.08442.
43. Gramegna, A.; Giudici, P. SHAP and LIME: An evaluation of discriminative power in credit risk. *Front. Artif. Intell.* **2021**, *4*, 752558.
44. MICS|unicef. Retrieved from Survey. 2019. Available online: https://mics.unicef.org/surveys?display=card&f[0]=region:2521&f[1]=datatype:0&f[2]=year:2019 (accessed on 22 November 2024).