The Screen-Time Trap: An Explainable Machine Learning Framework for Pediatric Screen-Time Overuse Prediction

Kinga Tshering*[1], Mushfika Hossain Piya[2], Ashraf Sharif[3], Marzan Islam[4]

*[1,2,3,4]Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh

E-mail: kinga.tshering@northsouth.edu, mushfika.hossain@northsouth.edu, ashraf.sharif@northsouth.edu, marzan.islam@northsouth.edu

Received: 24 May 2026; Revised: 06 July 2026; Accepted: 08 August 2026; Published: 13 August 2026

Abstract

Excessive screen time in children and adolescents is associated with adverse outcomes such as poor sleep, eye strain, anxiety, and obesity risk. Existing studies have often focused on smartphone addiction or descriptive association studies rather than explainable predictive modeling of screen-time overuse in pediatric populations. In this study, we develop and evaluate an explainable machine learning framework to predict whether children exceed the recommended daily screen-time limit using a real-world dataset of 9,668 Indian children aged 8–18 years after preprocessing (4,928 male and 4,740 female). Screen-time overuse was observed in 86.7% of male respondents and 84.8% of female respondents. Our pipeline includes preprocessing, duplicate removal, encoding, imputation, feature scaling, SMOTE-based class balancing on the training set, comparative evaluation of multiple classifiers, hyperparameter optimization, ensemble learning, and SHAP-based explanation. Among the evaluated models, XGBoost with hyperparameter tuning achieved the best performance, with 85% accuracy, 94% precision, 89% recall, 91% F1-score, and 91% ROC-AUC. Explainability analysis showed that poor sleep and the educational-to-recreational screen-time ratio were the strongest predictors, while eye strain and anxiety also emerged as important factors. These findings show that explainable machine learning can support early identification of children at risk of screen-time overuse and inform targeted interventions.

Keywords: *SMOTE, Hyperparameter Optimization, Ensemble Learning, Explainable AI, Stacking Classifier, Hard and Soft Voting.*

1. INTRODUCTION

In the modern digital world, children's exposure to screen time has increased over the past decades, raising concern about potential health impacts. According to the Australia's physical activity and sedentary behaviour guidelines children and young people between the ages of 5 and 17 should limit recreational screen time to not more than 2 hours per day (Australian Government, Department of Health and Aged Care, 2024). However, many children exceed the recommended screen time, leading to various issues such as poor sleep, eye strain, anxiety, and an increased risk of obesity. Following the COVID-19 pandemic, the increase in digital screen time did not decrease, leading to increased anxiety, difficulty sleeping, addiction, and various sociobehavioral difficulties among children (Mesce et al., 2022). This period of time largely contributed to the development of a habit of increased exposure to screens to cope with boredom and social isolation among children and adolescents.

Therefore, this paper explores the application of machine learning techniques to address this issue of excessive screen time exposure by predicting whether the children exceeded the screen time limit based on various demographic, behavioral, and health-related factors. The research is based on a real-world dataset (Panday, 2025) containing records from 9,668 children in India, aged 8–18, which includes both the urban and rural demographics. The goal is to predict screen time overuse based on factors like device type, educational to recreational screen time ratio, and socio-demographic variables.

With the raw dataset, we have preprocessed it before feeding it to machine learning algorithms. To handle imbalanced class distribution in the target, We have used SMOTE (Synthetic Minority Oversampling Technique) (Chawla et al., 2002). We trained our model with suitable classification machine learning algorithms and used hyperparameter tuning and ensemble learning techniques like stacking and voting (Hard & Soft) to have a com-

parative analysis of the results. The comparative study among the different classification algorithms and with and without hyperparameter optimizations and ensemble learning, is analyzed in Section 4 for better representation and understanding of the efficiency and accuracy of the model trained with different algorithms. The proposed solution of this paper has been described in Section 3, including the detailed analysis of the dataset preprocessing and attribute selection. The key contributions of this paper are:

- We develop a machine learning framework to predict pediatric screen-time overuse using a real-world dataset of Indian children and adolescents.
- We compare multiple machine learning models, tuned models, and ensemble methods to identify the most effective approach for predicting screen-time overuse.
- We use explainable AI (SHAP) to interpret the model's predictions and identify the key factors associated with excessive screen time, such as poor sleep and screen-use patterns.

2. RELATED WORK

Several studies have examined the adverse effects of excessive screen-time exposure on children's physical, psychological, and developmental health. In parallel, machine learning techniques have been increasingly applied in healthcare and psychology to predict behavioral and health-related outcomes, including technology-use patterns and screen-related behaviors. Research has shown that exposure to screen devices among infants and toddlers is associated with expressive speech delays, with each additional 30 minutes of handheld screen exposure being reported to increase the likelihood of expressive speech delay by 49% (Martin et al., 2017). Other reported effects of excessive screen exposure include obesity due to reduced physical activity, sleep disturbance associated with blue-light exposure, behavioral problems, reduced academic performance, and desensitization to violent entertainment content (Twenge and Campbell, 2018; Reichel, 2019).

Twenge and Campbell (2018), in a large-scale population-based study, reported significant associations between excessive screen use and poor psychological health among children and adolescents. These associations included lower curiosity, reduced self-control, and higher prevalence of anxiety and depression. Such findings highlight the importance of examining screen-time exposure not only as a behavioral issue but also as a potential indicator of broader psychological and developmental concerns.

The development of computing power and the production of large amounts of data have made it possible to implement machine learning-based prediction, classification, and analysis of screen-related behaviors. Healthcare and psychology have increasingly adopted machine learning models to understand how technology use is associated with health risks and behavioral outcomes.

Abu-Taieh et al. (2022) analyzed the psychological predictors of smartphone use and social isolation among Jordanian young people based on the Technology Acceptance Model. Their findings showed that perceived usefulness, trust, and social influence were among the most important predictors, while Random Forest performed better than the other evaluated models.

Lin et al. (2022) applied regression-based machine learning models to explore the relationship between smartphone addiction and physical activity among Chinese students in Korea and demonstrated modest predictive power. Similarly, Lee and Kim (2021) investigated problematic smartphone use using a prospective Korean dataset and reported model accuracies of 82.59% for Random Forest, 80.77% for XGBoost, and 74.56% for Decision Tree.

In predicting internet addiction among Chinese adolescents, Liu et al. (2025) utilized Extra Random Forest with SHAP explainability and achieved an accuracy of 0.798, an AUC of 0.854, and an F1-score of 0.659. Their study identified peer influence, academic stress, and self-control as the most noteworthy predictors.

Contributing to this research area, Amora et al. (2025) focused on the effects of educational and recreational screen time on the academic and behavioral performance of children aged 5–15. They created composite measures such as the Academic Performance Score (APS), Behavioral Score (BS), and overall impact using pediatric screen-time guidelines. Using a Kaggle dataset of 66 records with categories such as age, gender, and day type, they tested Decision Tree, Random Forest, Logistic Regression, SVM, and XGBoost with a five-fold cross-validation strategy. The highest accuracy of 97.1% was obtained with Random Forest, followed by Logistic Regression, SVM, and XGBoost with

accuracies between 95–96%, while Decision Tree achieved 94.2%. Recreational screen time was found to be the strongest predictor.

Although these studies provide valuable insights into screen-time effects, smartphone addiction, and internet addiction, many of them focus on general technology-use behavior rather than pediatric screen-time overuse based on recommended daily limits. Moreover, limited work combines comparative machine learning evaluation with explainable AI to identify the factors associated with excessive screen-time behavior. Therefore, this study addresses this gap by developing an explainable machine learning framework for predicting pediatric screen-time overuse. The comparative study of the notable works and their limitations is shown in Table 1.

Table 1: *Comparative summary of related studies and research gaps.*

Study	Research Focus	Dataset / Context	Method and Main Finding	Limitation / Gap
Twenge and Campbell (2018)	Screen time and psychological well-being	Population-based US data	Statistical analysis linked screen exposure with psychological indicators.	No predictive ML model.
Abu-Taieh et al. (2022)	Smartphone addiction, social networking, anxiety, and depression	Survey data from 511 parents in Jordan	Applied ML and behavioral modeling; Random Forest performed well.	Not focused on pediatric screen-time overuse.
Lin et al. (2022)	Smartphone addiction and physical activity	Chinese students in Korea	Used regression-based models with moderate predictive performance.	Focused on addiction severity.
Lee and Kim (2021)	Problematic smartphone use prediction	Large-scale Korean dataset	Compared DT, RF, and XGBoost; RF achieved about 82.6% accuracy.	Limited explainability.
Liu et al. (2025)	Explainable internet addiction prediction	Chinese longitudinal dataset	Used Extra Random Forest with SHAP; achieved 0.798 accuracy and 0.854 AUC.	Focused on internet addiction.
Amora et al. (2025)	Screen time, academic performance, and behavior	Kaggle dataset with 66 records including age, gender, day type, and screen-time variables	Compared DT, RF, LR, SVM, and XGBoost; RF achieved about 97.1% accuracy.	Small dataset.
Our Study	**Explainable pediatric screen-time overuse prediction**	**Indian child screen-time dataset with demographic, device-use, behavioral, and health-impact features**	**Evaluates classifiers, HPO, ensembles, and SHAP; XGBoost achieved the best overall performance.**	**Limited by geographic and demographic scope.**

3. METHODOLOGY

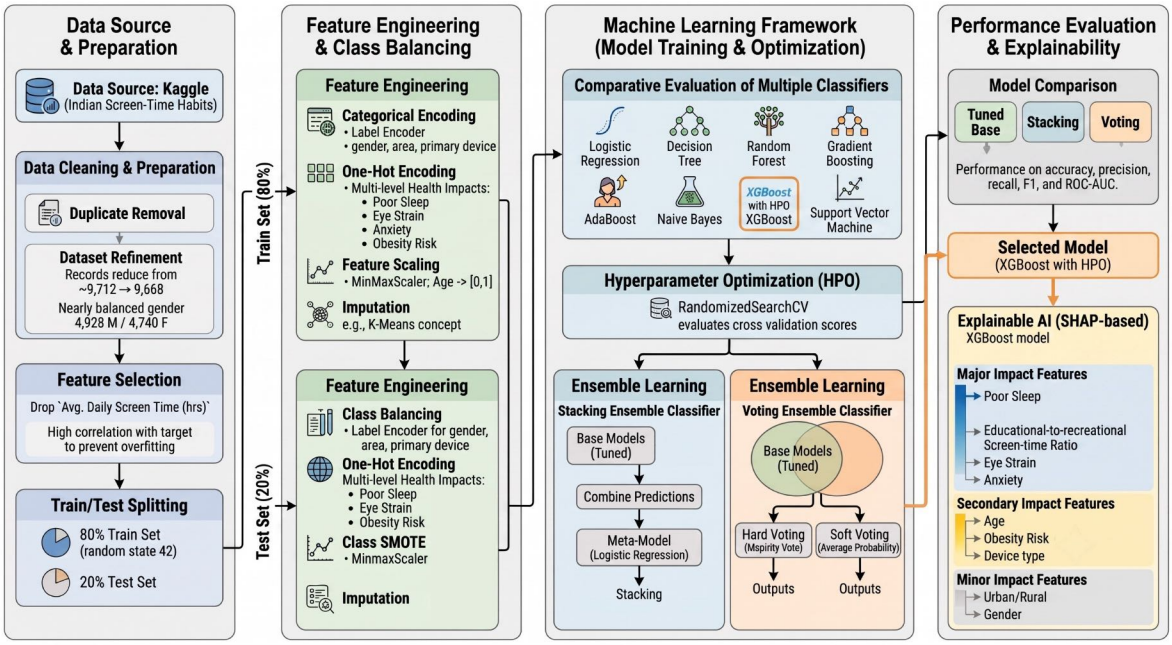

Fig. 1: *Block diagram of the entire work.*

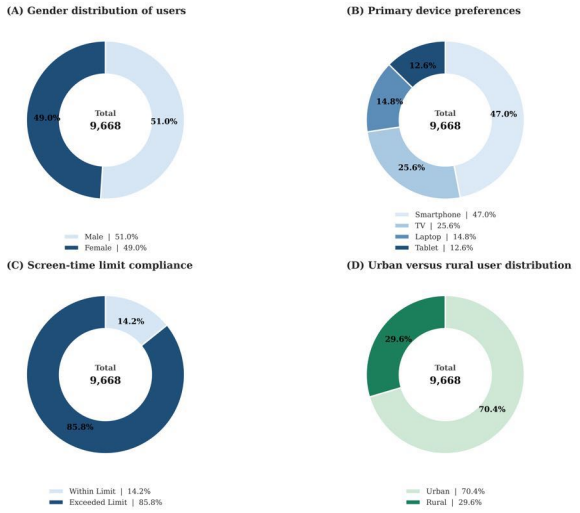

Fig. 2: *Digital Device Usage and Demographics Analysis*

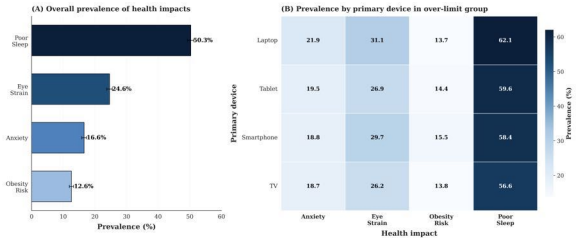

Fig. 3: *Distribution of health issues (anxiety, eye strain, obesity risk, and poor sleep), and device-usage patterns associated with these health issues.*

Figure 1 represents the workflow of the entire work. The main aim of this work is to create a model that will predict whether children have exceeded recommended screen-time limits. The methodology employed in this study involves several key steps: dataset description, exploratory data analysis, dataset preparation, feature engineering, data preprocessing, model selection, model evaluation, and explainability analysis.

3.1 Dataset Description and Exploratory Data Analysis

The data for this task is collected from Kaggle, originally gathered by Indian research and health studies to explore screen time habits in children. It comprises of 9,668 children aged 8-18 years. There is a total of 7 features in the dataset, excluding the target feature (Panday, 2025). As illustrated in Figure 2, we observe that the study cohort is nearly balanced by gender, is predominantly urban, and relies primarily on smartphones, while a remarkably large proportion of children exceed

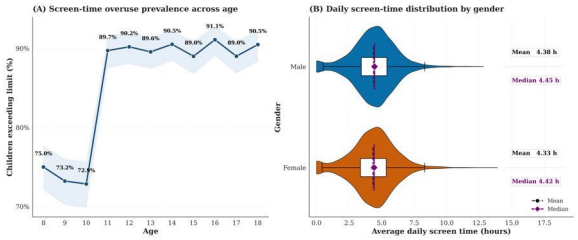

Fig. 4: *Distribution of age by exceeded recommended limit and screen-time distribution by gender*

the recommended screen-time limit, underscoring the extent of screen overuse in this population. As shown in Figure 3, we find that poor sleep is the most prevalent health impact overall and remains consistently dominant across all primary device categories, suggesting that sleep-related difficulties are the most prominent adverse outcome associated with excessive screen exposure in this dataset. Finally, Figure 4 demonstrates that the prevalence of screen-time overuse increases sharply after age 10 and remains persistently high across the later age groups, whereas male and female participants exhibit highly similar daily screen-time distributions, indicating that age appears to be a more important differentiating factor than gender in this cohort

3.2 Attribute Selection

We set an attribute "Exceeded Recommended Limit" as the target for the classification task. It is a Boolean indicator (True/False), Where True represents children who exceeded the recommended daily screen time limit, and False represents children who did not exceed the daily screen time limit. This is a variable that a machine learning models are trying to predict. We used a Pearson correlation heat map to get a clear understanding how strongly each feature correlates with others, especially before training the machine learning model. The Pearson correlation heat map has been shown in Appendix Figure 11.

Overall, We found out that one particular feature, Avg. Daily Screen Time (hrs) was highly correlated with the target feature. Therefore, we had to drop it so as to reduce model from overfitting.

3.3 Data Splitting

In our work, we have split the dataset into 80% trainset and a 20% test-set with a random state of 42. Random state ensures that the same randomization

Table 2: *Sample records from the screen-time dataset.*

Age	Gender	Daily Screen Time (hrs)	Primary Device	Exceeded Limit	Edu./Rec. Ratio	Health Impacts	Area
14	Male	3.99	Smartphone	True	0.42	Poor Sleep, Eye Strain	Urban
11	Female	4.61	Laptop	True	0.30	Poor Sleep	Urban
18	Female	3.73	TV	True	0.32	Poor Sleep	Urban
15	Female	1.21	Laptop	False	0.39	None	Urban
12	Female	5.89	Smartphone	True	0.49	Poor Sleep, Anxiety	Urban

Note: Edu./Rec. Ratio denotes the educational-to-recreational screen-time ratio.

is used each time you run the code, resulting in the same splits of the data.

Fig. 5: *Class distribution before SMOTE*

Since there is a class imbalance in the data, we used SMOTE (Synthetic Minority Over-sampling Technique) (Chawla et al., 2002) and applied it to the training set, thereby balancing the dataset and giving the model a better chance to learn from both classes. SMOTE is designed to increase the representation of the minority class in an imbalanced dataset.

Fig. 6: *Class distribution after SMOTE*

3.4 Feature Engineering

Several categorical values, including gender, urban or rural, and primary device, were converted into numerical values using a label encoder. At the same time, features with multi-level data for each instance, like health impacts, were handled with one-hot encoding, which creates a separate binary column for each possible value in a categorical variable. Each column represents the presence (1) or absence (0) of a particular level. Moreover, the Age feature was normalized using *MinMaxScaler*, which scales data to a range [0,1]. This ensures that all the features are within the same scale, allowing the model to train effectively, especially models like SVM and logistic regression.

The *MinMaxScaler* transformation is given by:

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

- X represents the original value of a data point (before normalization).
- X_{min} is the minimum value in the dataset/feature.
- X_{max} is the maximum value in the dataset/feature.

3.5 Model Selection

Supervised classification algorithms were applied on the train set of the Indian kids' screen time dataset to predict whether the child has exceeded the daily recommended screen time. In this work, we applied at least 8 classification algorithms to compare the predictive power of diverse models and picked the best one. The machine learning algorithms are Logistic regression, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, Naive Bayes, XGBoost, and Support Vector Machine (SVM).

Hyperparameter tuning is the process of selecting the optimal values for the machine learning model's hyperparameters to improve the performance and generalization of a machine learning model. For each model used in our experiment, we used *RandomizedSearchCV* to obtain the best-performing models. As the name suggests, *RandomizedSearchCV* picks a random combination of hyperparameters from the given ranges instead of checking every single combination like *GridSearchCV*. The *RandomizedSearchCV* technique was used to obtain the best hyperparameters for the models. The best hyperparameters for the different models are displayed in Table 3. There was a slight

improvement in the performance of the models after hyperparameter optimization compared to the default hyperparameters.

Table 3: *Parameter space and best parameters of the evaluated models.*

Model	Parameter Space	Best Parameters
Logistic Regression	C: [0.1, 1, 10, 100] penalty: [l2] solver: [liblinear, saga]	solver: liblinear penalty: l2 C: 0.1
Decision Tree	max_depth: [5, 10, 20, None] min_samples_split: [2, 5, 10] min_samples_leaf: [1, 2, 4] criterion: [gini, entropy]	max_depth: 20 min_samples_split: 2 min_samples_leaf: 1 criterion: gini
Random Forest	n_estimators: [50, 100, 150, 200] max_depth: [5, 10, 20, None] min_samples_split: [2, 5] min_samples_leaf: [1, 2, 4]	n_estimators: 100 max_depth: 20 min_samples_split: 5 min_samples_leaf: 1
Gradient Boosting	n_estimators: [50, 100, 200] learning_rate: [0.01, 0.05, 0.1, 0.2] max_depth: [3, 5, 10]	n_estimators: 200 max_depth: 10 learning_rate: 0.2
AdaBoost	n_estimators: [50, 100, 150] learning_rate: [0.1, 0.5, 1.0]	n_estimators: 50 learning_rate: 0.1
Naive Bayes	var_smoothing: [1e-9, 1e-8, 1e-7]	var_smoothing: 1e-9
SVM	C: [0.1, 1, 10] kernel: [linear, rbf] gamma: [scale, auto]	kernel: linear gamma: scale C: 0.1
XGBoost	n_estimators: [50, 100, 200] learning_rate: [0.01, 0.1, 0.2] max_depth: [3, 5, 10] subsample: [0.6, 0.8, 1.0] colsample_bytree: [0.6, 0.8, 1.0]	n_estimators: 100 max_depth: 10 learning_rate: 0.2 subsample: 0.8 colsample_bytree: 1.0

3.6 Ensemble Machine Learning Model Construction

Recent studies have shown that tree-based models and ensemble methods, such as XGBoost algorithms can perform better on tabular data (Shwartz-Ziv and Armon, 2022), with the problem of having clear decision boundaries and heterogeneous feature spaces. Driven by the recent findings and emphasis on ensemble models, in this research, 3 ensemble models were constructed to predict whether the children have exceeded a daily screen time limit.

3.6.1 Stacking Ensemble Classifier

The Stacking Classification Model (SCM) is an ensemble learning method that improves classification performance by integrating multiple base learners. The ensemble approach combines several models from the previous subsection to improve the overall performance and reliability of predictions. It involves the following steps:

Base Models: Training multiple models (level-0 models) on the same dataset.

Meta-Model: Training a new model (level-1 or meta-model) to combine the predictions of the base models. Using the predictions of the base models as input features for the meta-model.

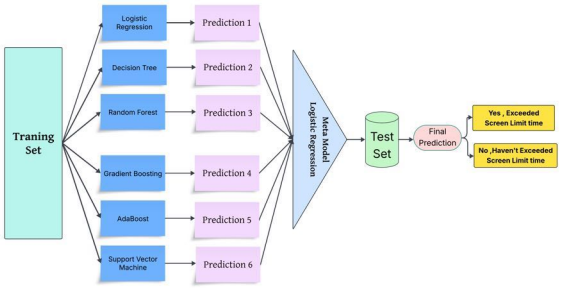

Fig. 7: *Stacking Ensemble Classifier Workflow*

3.6.2 Voting Ensemble Classifier

A voting classifier is a machine learning model that gains experience by training on a collection of several models and forecasts an output (class) based on the class with the highest likelihood of becoming the output. To forecast the output class based on the largest majority of votes, it averages the results of each classifier provided to the voting classifier.

Hard Voting: The final prediction is made based on the majority class. Each classifier "votes" for a class, and the class with the majority of votes becomes the final prediction.

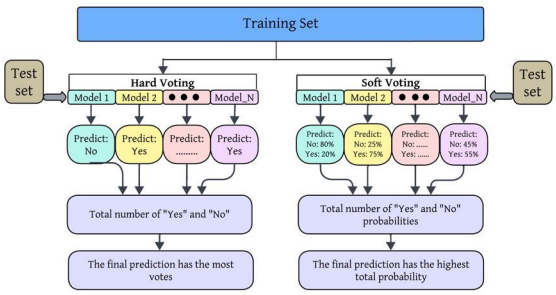

Fig. 8: *Voting Ensemble Classifier Workflow*

Soft Voting: The final prediction is made based on the average of the predicted probabilities. The class with the highest average probability across all classifiers is chosen as the final prediction.

3.7 Performance Metrics

The performance of the machine learning models was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. These metrics were calculated from the confusion matrix, which consists of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

Accuracy measures the proportion of correctly classified samples among all samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision measures the proportion of correctly predicted positive cases among all predicted positive cases:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall, also known as the True Positive Rate (TPR), measures the proportion of actual positive cases that were correctly identified:

$$\text{Recall} = \text{TPR} = \frac{TP}{TP + FN}$$

The F1-score is the harmonic mean of precision and recall:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The False Positive Rate (FPR) measures the proportion of negative cases incorrectly classified as positive:

$$\text{FPR} = \frac{FP}{FP + TN}$$

ROC-AUC summarizes the model's ability to distinguish between the positive and negative classes across different decision thresholds.

4. RESULTS

4.1 Performance on machine learning Algorithms

In this study, several machine learning models were evaluated for predicting whether a child exceeded the recommended screen-time limit. The models were trained using the preprocessed and SMOTE-balanced training data, and their performance was evaluated on the test set using accuracy, precision, recall, F1-score, and ROC-AUC.

Table 4 presents the comparative performance of the evaluated machine learning models without hyperparameter tuning. Decision Tree and XGBoost achieved the highest accuracy of 0.84, while Decision Tree obtained the highest recall of 0.89. Logistic Regression, AdaBoost, Naive Bayes, and SVM achieved the highest precision of 1.00, although their recall values were lower at 0.79. In terms of ROC-AUC, Logistic Regression, Gradient Boosting, AdaBoost, and Naive Bayes achieved the highest value of 0.91. Overall, XGBoost, Decision Tree, and Random Forest showed competitive performance across accuracy, recall, and F1-score, while no single model dominated all evaluation metrics.

Table 4: *Machine learning model evaluation.*

Model	Acc.	Prec.	Rec.	F1	AUC
Logistic Regression	0.82	**1.00**	0.79	0.88	**0.91**
Decision Tree	**0.84**	0.92	**0.89**	**0.90**	0.74
Random Forest	0.84	0.92	0.88	**0.90**	0.90
Gradient Boosting	0.82	0.99	0.79	0.88	**0.91**
AdaBoost	0.82	**1.00**	0.79	0.88	**0.91**
Naive Bayes	0.82	**1.00**	0.79	0.88	**0.91**
SVM	0.82	**1.00**	0.79	0.88	0.90
XGBoost	**0.84**	0.94	0.87	**0.90**	0.90

Figure 9 provides a visual comparison of the evaluated models across the key performance metrics.

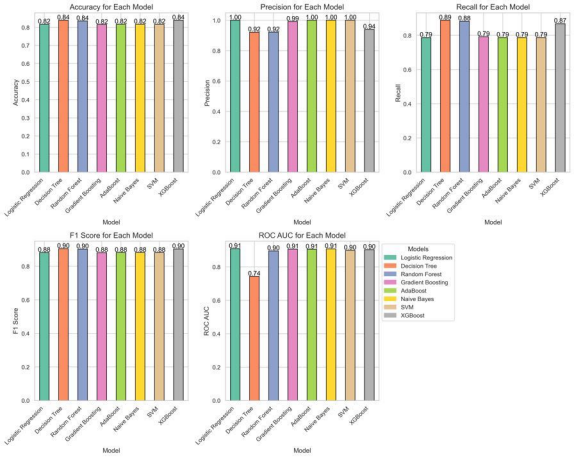

Fig. 9: *A comparison of machine learning models' performance on key metrics*

4.2 Performance on Machine Learning Algorithms with Hyperparameter Tuning

Table 5 presents the performance of the machine learning models after hyperparameter tuning. XGBoost and Gradient Boosting achieved the highest accuracy of 0.85 and the highest F1-score of 0.91. Gradient Boosting achieved the highest recall of 0.92, indicating stronger sensitivity in identifying children who exceeded the recommended screen-time limit. However, XGBoost achieved a higher precision of 0.94 and the highest ROC-AUC of 0.91, showing a strong balance between classification accuracy and discriminative ability. Based on its overall performance across accuracy, precision, recall, F1-score, and ROC-AUC, XGBoost

with hyperparameter tuning was selected as the best-performing model.

Table 5: *Model performance after hyperparameter tuning.*

Model	Acc.	Prec.	Rec.	F1	AUC
Logistic Regression	0.82	**1.00**	0.79	0.88	**0.91**
Decision Tree	0.83	0.96	0.84	0.89	0.85
Random Forest	0.83	0.94	0.86	0.90	0.90
Gradient Boosting	**0.85**	0.91	**0.92**	**0.91**	0.90
AdaBoost	0.82	**1.00**	0.79	0.88	**0.91**
Naive Bayes	0.82	**1.00**	0.79	0.88	**0.91**
SVM	0.82	**1.00**	0.79	0.88	**0.91**
XGBoost	**0.85**	0.94	0.89	**0.91**	**0.91**

4.3 Performance on Ensemble Learning

In this study, we compared the performance of three ensemble learning models: the Soft Voting Classifier, the Hard Voting Classifier, and the Stacking Classifier. Evaluation was based on several key metrics like Accuracy, Precision, Recall, F1 Score, and ROC AUC. The results, which can be seen in Table 6, show that all three models performed effectively in predicting whether a child exceeded the recommended screen-time limit.

Several machine learning algorithms like Logistic Regression, Decision Tree, Random Forest, XGBoost, AdaBoost, Gradient Boosting, and SVC were used as base models to create a stronger, more predictive model. In a Stacking ensemble Classifier, several machine learning algorithms were used as base models, and Logistic Regression is used as a meta model to combine their predictions. The Stacking Classifier achieved the highest recall of 86%, indicating stronger sensitivity in identifying children who exceeded the recommended screen-time limit. However, its ROC AUC score was slightly lower at 77% compared to the Voting Classifiers

Soft Voting achieved an accuracy of 83%, with an impressive precision score of 98%. However, its recall was a bit lower at 81%, meaning it missed some cases of children exceeding the screen time limit.

The Hard Voting Classifier had even higher precision at 99%, but its recall was slightly lower (79%). This suggests it was more cautious in its predictions but still reliably detected most of the cases.

Overall, among the ensemble classifiers, the Stacking Classifier achieved the highest accuracy and recall, while the Hard Voting Classifier achieved the highest precision and ROC-AUC. However, the ensemble models did not outperform the best tuned individual models, particularly XGBoost and Gradient Boosting. This suggests that the additional ensemble complexity did not provide a clear performance advantage for this dataset.

Table 6: *Ensemble learning model performance evaluation.*

Model	Acc.	Prec.	Rec.	F1	AUC
Soft Voting	0.83	0.98	0.81	0.89	0.85
Hard Voting	0.82	**0.99**	0.79	0.88	**0.88**
Stacking	**0.84**	0.94	**0.86**	**0.90**	0.77

4.4 Explainable AI

Although decisions made by machine learning models may be accurate but they often come with a question of how they made that prediction. This study uses explainable AI to give a better explanation for the decisions made by machine learning models.

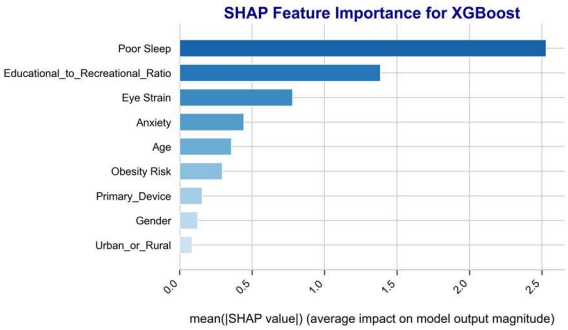

Fig. 10: *SHAP feature importance for XGBoost*

In Figure 10, SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017) is used to understand the factors affecting the XGBoost model's predictions of whether the child has exceeded the recommended screen-time limit. Poor sleep emerges as the biggest influencing feature, indicating that inadequate sleep has an important role in determining screen time behavior; likely, children with poor sleep may be more prone to excessive screen use. The educational to recreational ratio holds a significant weight in influencing the model's predictions, suggesting children who spend more time in recreational activities compared to educational activities tend to exceed screen time recommendations, possibly due to poorly structured routines. Eye strain and anxiety follow as important factors, indicating that children experiencing eye strain and anxiety might engage

in prolonged screen time as a coping mechanism. On the other hand, features like age, obesity risk, and device type have a moderate influence, indicating that while they may impact screen time decisions, they are secondary to lifestyle factors. Finally, urban or rural and gender have the least influence on screen time predictions, with their smaller impact suggesting the demographic and environment factors might not be crucial in predicting screen time compared to other health impacts. This analysis helps us understand the model's reliance on health, behavioral, and lifestyle traits rather than merely demographic ones, which could be essential for designing interventions aimed at managing children's screen time.

5. DISCUSSION

In an age where digital devices have become an integral part of children's lives, understanding the factors that contribute to excessive screen time is important for safeguarding their health and well-being. This research explores the predictive power of machine learning techniques to determine whether children exceed the recommended daily screen time limits, using a diverse dataset of 9668 children from urban and rural India. The findings of this study provide both a powerful insight into the factors influencing screen time behavior and a hopeful pathway for leveraging machine learning to mitigate its detrimental effects.

The study underscores the impressive capabilities of machine learning in predicting excessive screen time. Among the various algorithms explored, XGBoost with hyperparameter optimization stands out from all the models trained and tested, achieving an accuracy of 85%, F1-score of 91%, Recall of 89%, Precision of 94%, and ROC-AUC score of 91%. These models excelled not only in precision but also in recall, highlighting their robustness in accurately identifying children who are at risk of surpassing the screen time threshold. Such results are indicative of the adaptability and precision of tree-based algorithms, especially in complex, high-dimensional datasets such as this one.

The stronger performance of XGBoost and Gradient Boosting may be explained by the suitability of tree-based boosting methods for structured tabular data. Unlike linear models, boosting based models can capture nonlinear relationships and feature interactions among demographic, behavioral, health-related, and device use variables. This is important in the present dataset because screen-time overuse is unlikely to depend on a single factor alone; instead, it may reflect interactions among sleep quality, recreational screen use, device type, age, and health-related symptoms. The relatively weaker improvement from voting and stacking ensembles may be due to the limited diversity among base learners, as several models produced similar decision patterns after preprocessing and SMOTE-based class balancing.

What truly distinguishes this study, however, is the ability to interpret which factors play the most pivotal role in influencing screen time behaviors. The feature importance analysis revealed that the most influential predictor of excessive screen time is poor sleep. This suggests that children with sleep disturbances are more likely to engage in extended screen use, potentially as a coping mechanism for stress or anxiety. This aligns with the growing body of evidence linking screen time with sleep disruptions and mental health challenges in children.

Further enriching our understanding, the educational-to-recreational screen time ratio was identified as a significant factor. Children who allocate more time to recreational screen use, as opposed to educational activities, are more prone to exceeding the recommended screen time limit. This finding underscores the importance of fostering a balanced screen time routine, where educational screen use is prioritized, and recreational activities are moderated to prevent negative health outcomes.

6. CONCLUSIONS

This study has investigated how supervised machine learning can predict children at risk of spending excessive time on screens beyond recommended limits. We tested different models like Logistic Regression, Random Forest, Gradient Boosting, and XGBoost, as well as a combined ensemble model. By carefully doing the preprocessing of the data, which includes handling missing data, encoding categorical variables, scaling, and applying SMOTE to address the imbalance in class. Thus, the model achieved balanced and reliable results. The code for the experiments is available in https://anonymous.4open.science/r/Predictive-Modeling-of-Excessive-Pediatric-Screen-Time-6564

For our model, XGBoost with hyperparameter

optimization stands out and performed best among all models with an accuracy of 85%, F1-score of 91%, precision of 94%, and ROC-AUC score of 91%.

The explainability tools helped to show which factors matter most in predicting too much screen time. We found that poor sleep, high anxiety, and use of screens more for fun than for learning are important predictors based on SHAP-based explainability analysis. These results highlight a strong link between screen use and children's behavior and well-being, suggesting that this knowledge can be used to spot the risks early and support healthier habits.

7. FUTURE DIRECTIONS

Even though the model showed positive results, there are still many ways to improve it. The model would be more useful if the dataset had more diverse demographics like countries, cultures, age and economic backgrounds. Accuracy might improve further by using deep learning and other advanced AI techniques. Collecting data over time could also help to understand the cause-and-effect links between screen time use and their health or well-being. Moreover, if more experiments were carried out on the ensemble learning techniques with different groups of ML algorithms, there is high possibility to achieve a better result, as it goes same for hyperparameter optimizations. Lastly, the model could make a real impact if it is applied in practical tools like mobile apps or in schools. This way, parents or teachers, or even children can understand the impact in their life by using excessive screen time and how it is affecting their daily life and health.

8. REFERENCES

Abu-Taieh, E. M., AlHadid, I., Masa'deh, R., Alkhawaldeh, R. S., Khwaldeh, S., and Alrowwad, A. (2022). Factors affecting the use of social networks and its effect on anxiety and depression among parents and their children: Predictors using ml, sem and extended tam. *International Journal of Environmental Research and Public Health*, 19(21):13764. https://doi.org/10.3390/ijerph192113764.

Amora, E. N. O., Agoylo, J. C., Tagud, J. A., Olaybar, J. A., Sy, K. N. L., and Acaso, R. S. (2025). Analyzing the impact of screen time on child and adolescent behavior and academic performance using machine learning models. In *2025 3rd International Conference on Inventive Computing and Informatics (ICICI)*, pages 1097–1104. https://doi.org/10.1109/ICICI65870.2025.11069749.

Australian Government, Department of Health and Aged Care (2024). Physical activity and exercise guidelines for all australians. Children and young people (5 to 17 years): No more than 2 hours per day of sedentary screen time. Accessed: 22 May 2026. https://www.health.gov.au/topics/physical-activity-and-exercise/physical-activity-and-exercise-guidelines-for-all-australians.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357. https://doi.org/10.1613/jair.953.

Lee, J. and Kim, W. (2021). Prediction of problematic smartphone use: A machine learning approach. *International Journal of Environmental Research and Public Health*, 18(12):6458. https://doi.org/10.3390/ijerph18126458.

Lin, B., Teo, E. W., and Yan, T. (2022). The impact of smartphone addiction on chinese university students' physical activity: Exploring the role of motivation and self-efficacy. *Psychology Research and Behavior Management*, 15:2273–2290. https://doi.org/10.2147/PRBM.S375395.

Liu, J., Chen, L., Chen, Y., Luo, J., Yu, K., Fan, L., Yong, C., He, H., Liao, S., and Jiang, L. (2025). Explainable machine learning prediction of internet addiction among chinese primary and middle school children and adolescents: A longitudinal study based on positive youth development data (2019–2022). *Frontiers in Public Health*, 13:1590689. https://doi.org/10.3389/fpubh.2025.1590689.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

Martin, S. S., Black, L., and Alessio, L. (2017). Handheld screen time linked with speech delays in young children. In *Pediatric Academic Societies Meeting*. Conference news release/abstract. https://publications.aap.org/aapnews/news/10384/Handheld-screen-time-linked-with-speech-delays-in.

Mesce, M., Ragona, A., Cimino, S., and Cerniglia, L. (2022). The impact of media on children during the covid-19 pandemic: A narrative review. *Heliyon*, 8(12):e12489. https://doi.org/10.1016/j.heliyon.2022.e12489.

Panday, A. (2025). Indian kids screentime 2025. Kaggle dataset. Accessed: 22 May 2026. https://www.kaggle.com/datasets/ankushpanday2/indian-kids-screentime-2025.

Reichel, C. (2019). The health effects of screen time on children: A research roundup. *Journalist's Resource*. https://journalistsresource.org/education/screen-time-children-health-research/.

Shwartz-Ziv, R. and Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81:84–90. https://doi.org/10.1016/j.inffus.2021.11.011.

Twenge, J. M. and Campbell, W. K. (2018). Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study. *Preventive Medicine Reports*, 12:271–283. https://doi.org/10.1016/j.pmedr.2018.10.003.

A. APPENDIX

The Pearson correlation heatmap of the dataset features is shown here.

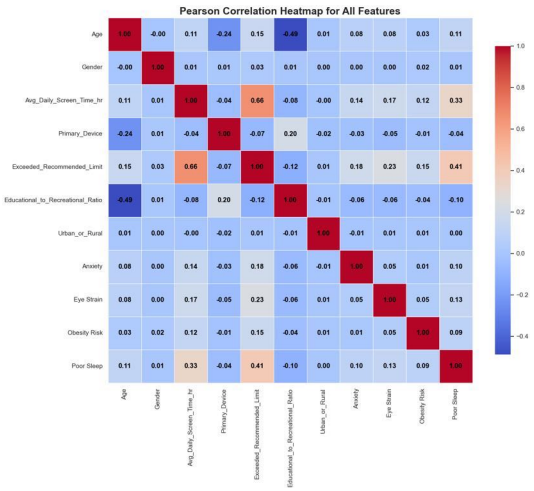

Fig. 11: *Pearson Correlation Heat Map*