Machine Learning and Deep Learning Approaches on Students' Social Media Addiction

Abstract

Using social media too much by students has become a concern because it might hurt their progress in school, their minds, and general health. It examines how many machine learning and deep learning models can predict and classify a student's social media addiction using demographic details, social media actions, and the record's outcome measuring sleep and grades. The data came from many countries to different levels of education. This was processed further to handle gaps and change categorical data into numerical. The addicted_score was divided into three groups called low, medium, and high. The data was scaled for features and divided into 80% for training and 20% for testing to get it ready for model training. The analysis of used models, such as LightGBM, CatBoost, SVM, gradient boosting, XGBoost, XGBoost+CNN, CNN+SVM, MLP, and vision transformer. CatBoost, CNN_SVM, and MLP was reached highest accuracy (0.9929) and training time 133.81, 24.23 and 11.57 seconds. Consider the training accuracy values in MLP (0.9965) was highest rather than other DL models. Integrating various layers in these hybrid approaches helped them process the data more effectively and extract essential features from the tabular information. Hyperparameter optimization increased the achievements of the ML, DL and hybrid models. The results reveal that using an ensemble or a hybrid model works better than relying on DL models. The study points out how ML and DL may assist in determining which students are likely to struggle in their learning and how to help them. To improve the model and resolve some of its drawbacks, future studies may study very large datasets or look at real-time data.

Key words: Social media addiction, ML and DL, Student Performance, Academic Impact, Mental Health

Introduction

Many people worry that social media's role in students' lives could lead to addiction and problems with their schoolwork, mental health, and rest (Savci et al., 2022). As social media use increases, it becomes important to understand and predict social media addiction to develop effective solutions (Ehsan & Basit). This is achieved by using ML and DL techniques to analyze a dataset of 705 student records, which includes demographic information, their usage patterns, and outcomes. Its value lies in offering useful insights for teachers and officials to prevent difficulties and improve student success (Mim et al., 2024).

The purpose is to see how different ML and DL models (such as LightGBM, CatBoost, XGBoost, CNN, SVM, and Vision Transformer) handle classifying students based on their social media use into various addiction categories (low, medium, and high). The study aims to identify the approach that provides the best and most efficient results in predicting medicine side effects.

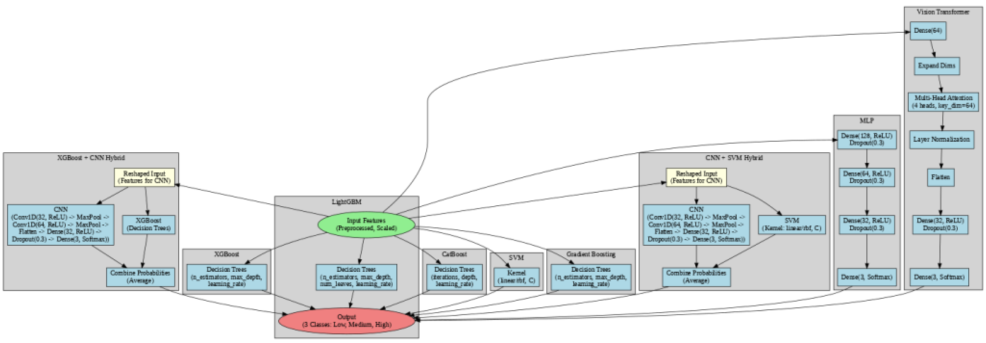
Literature Review

The study (Mim et al., 2024) investigates social media addiction in Bangladesh, linking excessive use of platforms like Messenger to mental health risks. Using machine learning, it achieves 82% accuracy with Random Forest, identifying higher addiction rates among those under 18. This study (Joseph & Maheswari, 2025) examines the relationship between social media addiction and academic dissatisfaction among 943 Bangladeshi students, utilizing machine learning algorithms (XGBoost, KNN, and Gradient Boosting) and LIME. XGBoost excelled with 95% recall, identifying academic stress and social comparisons as key addiction drivers, suggesting targeted interventions. The study (Parasar et al., 2025) uses the TMVM framework, integrating Theory of Mind AI and video modeling, to detect and reduce smartphone addiction in 786 Chennai students (aged 11–19) via facial emotion analysis. Using CNN, TLBO, and CSO algorithms, it achieved significant behavioral improvements ($p<0.000$) in social identity, self-awareness, and responsibility, reducing addiction and enhancing school/social engagement.

Methodology

This study utilized a dataset of 705 student records to predict social media addiction, containing variables like gender, academic level, country, and Addicted_Score. Preprocessing involved removing missing values, encoding categorical variables (e.g., gender, country) using LabelEncoder, discretizing Addicted_Score into low, medium, and high categories, and scaling features with StandardScaler after splitting the data into 80% training and 20% testing sets (random state 42). Training was conducted using hyperparameter tuning. ML models LightGBM, CatBoost, SVM, GradientBoosting, XGBoost, XGBoost+CNN, CNN+SVM, MLP, and Vision Transformer were developed, incorporating hybrid approaches with convolutional layers. Evaluation metrics included train/test accuracy, precision, recall, F1-score, specificity (via custom confusion matrix), sensitivity, and training time, ensuring a comprehensive assessment of model performance.

Figure 1: Model Architecture



Results and Discussion

The study evaluated the performance of eight machine learning (ML) and deep learning (DL) models LightGBM, CatBoost, SVM, GradientBoosting, XGBoost, XGBoost+CNN, CNN+SVM, MLP, and Vision Transformer on a dataset of 705 student records to predict social media addiction levels. LightGBM achieved the highest test accuracy of 0.9858 with a training time of 40.1489 seconds, demonstrating robust performance with precision, recall, and F1-score all at 0.9859, and specificity and sensitivity at 0.9859 and 0.9858, respectively. CatBoost followed closely with a test accuracy of 0.9929, a precision of 0.9930, and identical recall and F1-score, though its training time was significantly longer at 133.8106 seconds. Hybrid models like XGBoost+CNN (0.9858 accuracy, 32.5334 seconds) and CNN+SVM (0.9929 accuracy, 24.2304 seconds) showcased competitive results, leveraging convolutional layers to enhance feature extraction. The Vision Transformer, with a test accuracy of 0.9787 and the shortest training time of 10.5932 seconds, exhibited slightly lower specificity (0.9808), suggesting a trade-off between speed and precision.

Table 1: Results of social media addiction

Model	Train Accuracy	Test Accuracy	Precision	Recall	F1-Score	Time (s)	Specificity	Sensitivity
LightGBM	1.0000	0.9858	0.9859	0.9858	0.9851	40.1489	0.9859	0.9858
CatBoost	1.0000	0.9929	0.9930	0.9929	0.9929	133.8106	0.9910	0.9929
SVM	0.9965	0.9858	0.9858	0.9858	0.9858	13.0687	0.9859	0.9858
Gradient Boosting	1.0000	0.9858	0.9859	0.9858	0.9851	95.8585	0.9859	0.9858
XGBoost	1.0000	0.9858	0.9859	0.9858	0.9851	21.2935	0.9859	0.9858
XGBoost+CNN	0.9947	0.9858	0.9859	0.9858	0.9851	32.5334	0.9859	0.9858
CNN+SVM	0.9929	0.9929	0.9930	0.9929	0.9929	24.2304	0.9910	0.9929
MLP	0.9965	0.9929	0.9930	0.9929	0.9929	11.5783	0.9910	0.9929
Vision Transformer	0.9929	0.9787	0.9790	0.9787	0.9781	10.5932	0.9808	0.9787

All the images (Figure 2, Figure 3, Figure 4, Figure 5) show how an "Addicted Score" is distributed and correlated with various demographic and behavioral variables. In the first figure, the number of people with Low, Medium, and High addicted scores is identified based on gender, where Medium and High scores are higher in both males and females, with females showing slightly more results. The second image compares the scores according to education level (Graduate, High school, Undergraduate), showing that undergraduates and graduates have the highest values for Medium and High scores, with undergraduates leading in High scores. The third image presents a line graph illustrating the negative correlation between hours of daily usage and mental health, indicating that mental health declines when usage exceeds 3-4 hours. In the fourth set of density plots, we analyze how the average daily hours of use, sleep hours per night, and mental health scores relate to one another and overlay these with addicted scores (3-9), revealing differences in their correlations. Finally, the fifth density plot by gender shows that both females and males tend to have the highest score distribution around an addicted level of about 4-6, with males exhibiting a slightly broader range than females, providing an overall view of addiction patterns across these variables.

Figure 2: (a) Addicted score by gender (b) Addicted score distribution by academic staff (c) Daily usage hours vs mental score

(a) (b) (c)

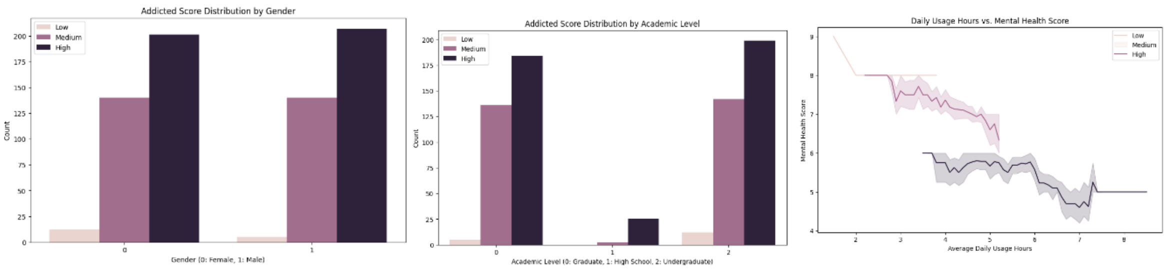


Figure 4: The graph of mental health

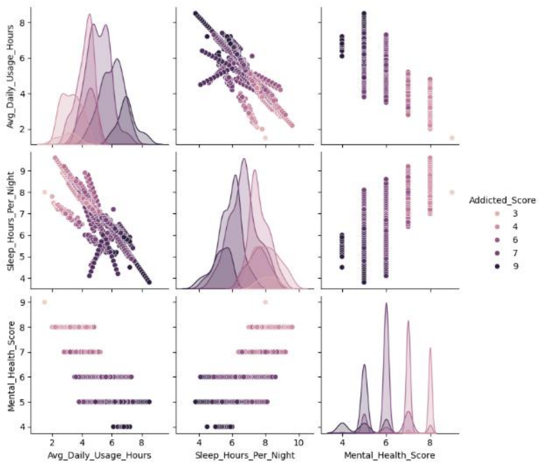
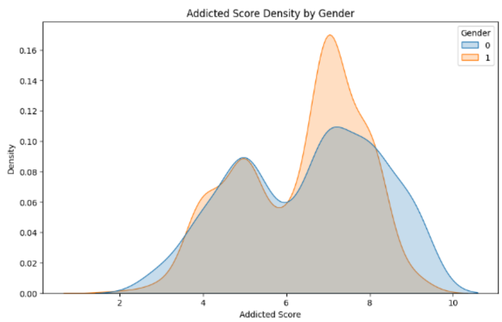


Figure 3: Addicted score density by gender



Implications/Conclusions

The analysis of social media addiction using various machine learning models reveals that CatBoost and CNN+SVM hybrids achieve the highest test accuracy (0.9929), with excellent precision, recall, and F1-scores. LightGBM, Gradient Boosting, XGBoost, and their CNN hybrid also perform well (0.9858 test accuracy), while SVM and MLP show slightly lower but still robust results. The Vision Transformer has a lower accuracy (0.9787), suggesting room for optimization. Training times vary, with Gradient Boosting and XGBoost being faster, while hybrids like XGBoost+CNN take longer. Overall, ensemble and hybrid models, particularly CatBoost and CNN+SVM, are the most effective for this classification task.

Based on the analysis of social media addiction using various machine learning models, it is recommended to prioritize CatBoost and CNN+SVM for deployment, given their impressive test accuracy of 0.9929 and strong performance across precision, recall, and F1-scores. The

Vision Transformer, with a lower accuracy of 0.9787, should undergo parameter optimization to enhance its effectiveness. For scenarios requiring faster training, Gradient Boosting or XGBoost are viable options due to their efficiency. Additionally, exploring further data preprocessing or feature engineering could boost the performance of SVM and MLP. Lastly, when utilizing hybrid models like XGBoost+CNN, it's advisable to monitor training times to strike a balance between accuracy and computational resource usage, ensuring optimal model selection

Acknowledgment

Appreciate our peers, mentors, and academic institution for their support and resources, which were crucial to this research effort.

References

Ehsan, T., & Basit, J. Machine Learning for Detecting Social Media Addiction Patterns: Analyzing User Behavior and Mental Health Data.

Joseph, C., & Maheswari, P. U. (2025). Facial emotion based smartphone addiction detection and prevention using deep learning and video based learning. *Scientific Reports*, *15*(1), 18025.

Mim, M. N., Firoz, M., Islam, M. M., Hasan, M., & Habib, M. T. (2024). A study on social media addiction analysis on the people of Bangladesh using machine learning algorithms. *Bulletin of Electrical Engineering and Informatics*, *13*(5), 3493-3502.

Parasar, A., Sujatha, N., Priya, S., Pramod, S., Kumar, A. S., & Palanimuthu, K. (2025). Machine Learning-Based Detection of Stress Levels in Students with Social Media Addiction. 2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS),

Savci, M., Tekin, A., & Elhai, J. D. (2022). Prediction of problematic social media use (PSU) using machine learning approaches. *Current Psychology*, *41*(5), 2755-2764.