MACHINE LEARNING AND DEEP LEARNING APPROACHES ON STUDENTS' SOCIAL MEDIA ADDICTION

INTRODUCTION

Many people are concerned that the role of social media in students' lives may lead to addiction and issues with their school work, mental health, and resting enough. Since more people use social media today, it is necessary to know and foresee social media addiction to create practical interventions. It meets this requirement by making use of ML and DL techniques to go through a dataset of 705 student records covering demographic information, how they use the system, and the results. Its value comes from providing helpful insights for both teachers and officials to prevent various difficulties and secure better outcomes for students.

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

Research Questions

1. Which model offers the highest accuracy and reliability in predicting social media addiction?
2. How do hybrid models (e.g., XGBoost+CNN) compare to standalone ML and DL approaches?
3. What is the impact of hyperparameter tuning on model performance? By addressing these questions, the study seeks to contribute to the growing field of educational data analytics, offering a foundation for future research and practical applications.

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.

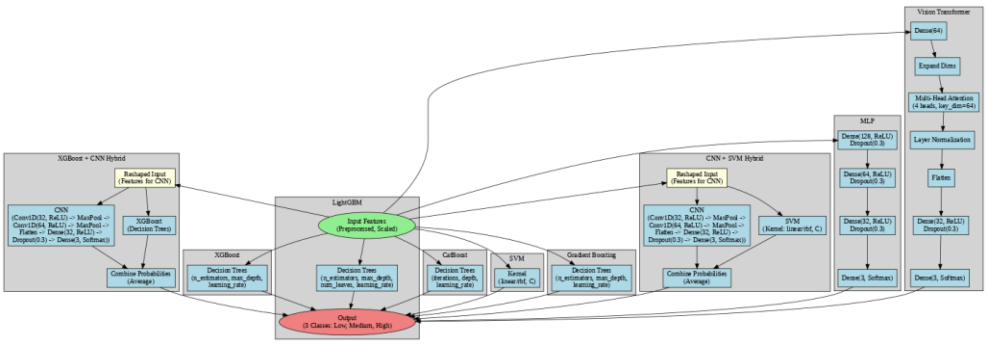

Figure 1: Model Architecture

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 ML and DL models

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

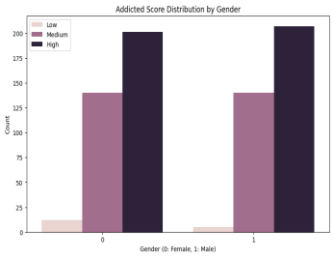

Figure 2: Addicted score by geneder

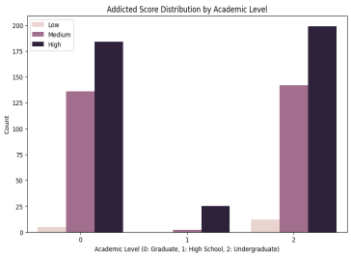

Figure 3: Addicted score by academic level

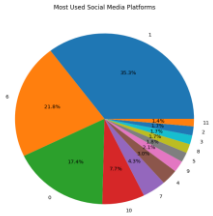

Figure 4: most used social media platform

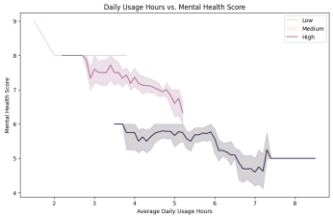

Figure 5: Sleeping hours with mental score

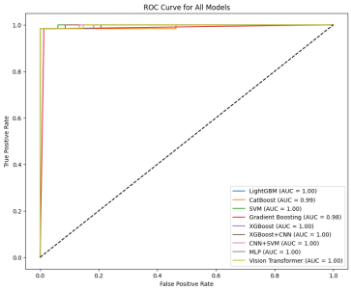

Figure 6: roc curve in ML and DL models

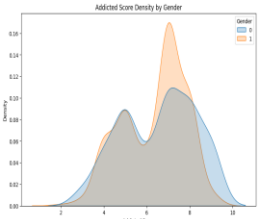

Figure 7: addicted score by gender

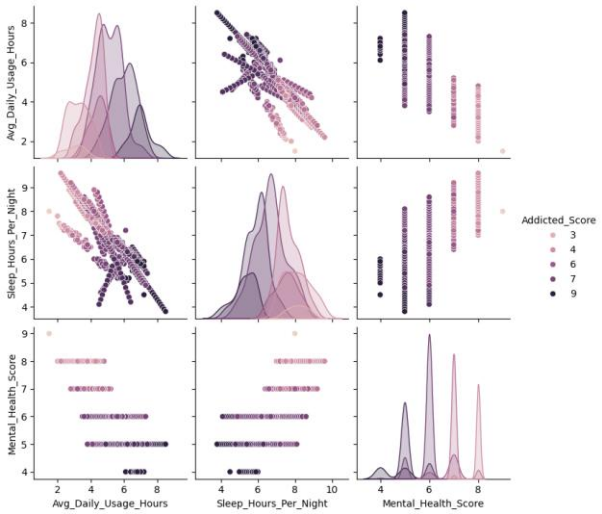

Figure 8: pair plot

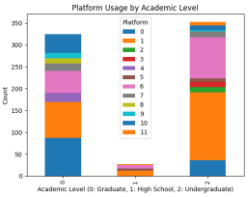

Figure 9: platform used in accadamic level

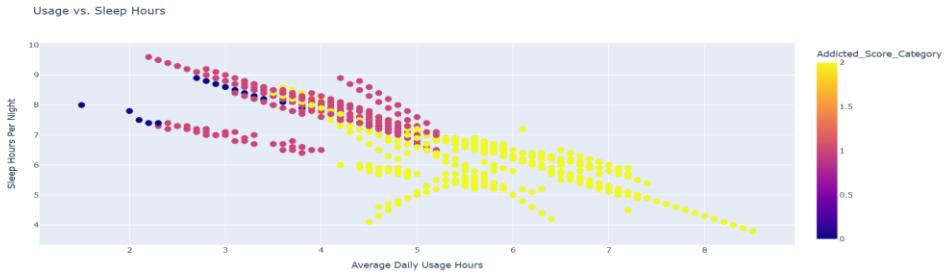

Figure 10: Sleeping hours vs usage of social media

CONCLUSIONS/ RECOMMENDATIONS

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

REFERENCES

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

Mim, M. N., et al. (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.

Savci, M., et al. (2022). "Prediction of problematic social media use (PSU) using machine learning approaches." Current Psychology **41**(5): 2755-2764.

Gómez-Galán, J., et al. (2020). Social Networks Consumption and Addiction in College Students During the COVID-19 Pandemic: Educational Approach to Responsible Use. Sustainability. 12 (8).

Alfaya, M. A., et al. (2023). Prevalence and determinants of social media addiction among medical students in a selected university in Saudi Arabia: a cross-sectional study. Healthcare, MDPI.

ACKNOWLEDGMENTS

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