Predicting Problematic Internet Use in Children Using Feature-Rich Structured Data with Ensemble Machine Learning and Bayesian Optimisation

Niteesh K R
Department of Computer Applications
PES University
100 ft. Ring Road, Bangalore, India
niteeshkr@pes.edu

Pooja T S
Department of Computer Applications
PES University
100 ft. Ring Road, Bangalore, India
poojats@pes.edu

***Abstract*—Problematic Internet Use (PIU) among children is a growing concern with cognitive, behavioural, and psychological implications. This work focuses on developing a robust machine learning framework to automatically predict the severity of PIU using a rich dataset comprising behavioural assessments, demographic data, and neurophysiological features over time series. The target variable, an ordinal score termed the Severity Index of Internet-use (SII), is predicted using multiple classical machine learning models, including XGBoost, LightGBM, CatBoost, and ExtraTreesClassifier.**

Bayesian Optimisation is applied to fine-tune hyperparameters for all base models and incorporate threshold optimisation via Quadratic Weighted Kappa (QWK), a metric well-suited for ordinal classification. An ensemble stacking model using Logistic Regression as a meta-learner is also explored. The best-performing model, XGBoost, achieves a QWK score of 0.72, indicating substantial agreement with expert-assigned labels. All models are evaluated using stratified k-fold cross-validation and compared across Accuracy, F1-score, and QWK.

This study demonstrates that interpretable, classical models can effectively capture ordinal patterns in health-related behavioural data, and the pipeline lays the groundwork for early identification of digital behaviour disorders in youth.

***Index Terms*—Machine Learning, Bayesian Optimisation, SII, PIU, Ensemble Learning**

I. Introduction

The pervasive integration of the internet into daily life has brought about significant shifts in how individuals, particularly adolescents and children, engage with information, entertainment, and social interaction. While this digital accessibility offers numerous advantages, concerns have emerged surrounding Problematic Internet Use (PIU)—a behavioural condition marked by excessive or uncontrolled online activity that negatively affects psychological, academic, and social functioning [1]. Numerous studies have linked PIU with symptoms such as anxiety, depression, attention-deficit behaviours, and social withdrawal, emphasising the need for early detection and intervention strategies [2]

Traditional methods for assessing PIU rely primarily on clinical interviews and psychometric evaluations [3], which are often labour-intensive, subjective, and not scalable for population-level screening. Machine learning approaches present a promising alternative, offering the capability to develop automated, data-driven systems that can detect behavioural patterns indicative of PIU from structured inputs such as demographic, physiological, and self-reported questionnaire data.

The Child Mind Institute's Problematic Internet Use dataset, made available publicly, provides a unique opportunity to investigate this problem using real-world data. The dataset includes various modalities such as baseline questionnaire responses, physiological time-series summaries, and demographic features, with the target variable labelled as the Severity Index of Internet-use (SII) —an ordinal label ranging from 0 (no concern) to 3 (high concern).

This work is situated within the context of building an effective and interpretable machine learning pipeline to predict the SII label from the available structured features. The core objective of the study is to determine whether ensemble-based ML models can reliably model PIU severity levels and generalise effectively across diverse validation folds. The study focuses specifically on XGBoost, LightGBM, CatBoost, and ExtraTreesClassifier, which are known for their ability to handle structured tabular data and non-linear relationships [4]. Furthermore, Bayesian Optimisation is employed to fine-tune hyperparameters, and a threshold optimisation technique is incorporated to align predicted scores with the ordinal nature of the SII target.

The significance of this study lies in its focus on early detection and risk profiling for PIU using scalable, algorithmic models. By benchmarking multiple classifiers and evaluating them through Quadratic Weighted Kappa (QWK), macro-averaged F1 score, and accuracy, the work aims to provide actionable insight into the relative strengths of different models in capturing the nuances of ordinal behavioural data.

While the machine learning techniques applied in this study are grounded in established methodologies, the novelty lies in their application to a complex, real-world behavioral health dataset. Unlike previous studies that rely heavily on psychometric models or single modality data, this work evaluates multiple ensemble-based models on multi-modal structured data comprising demographic, psychological, and physiological features. The inclusion of threshold tuning to optimize ordinal classification further aligns the model outputs with clinical interpretation needs. The practical utility of the pipeline is reinforced by its interpretability, making it potentially suitable for scalable deployment in early screening and risk stratification of problematic internet use in clinical settings.

The central research question driving this investigation is whether ensemble-based machine learning models accurately predict the severity level of problematic internet use, based on structured behavioural and physiological data, while addressing data imbalance and preserving the ordinal structure of the output.

II. Related Work

Recent literature has explored the psychological and behavioral underpinnings of Problematic Internet Use (PIU), as well as its neurocognitive and psychiatric associations.

Marino et al. [5] examined the influence of attachment styles on behavioral addictions among late adolescents. Their findings highlighted that anxious attachment correlates with problematic social network use, while avoidant attachment is linked to binge-watching behavior. The study emphasizes the importance of relational dynamics in predicting digital behavioral addictions.

Fineberg et al. [6] synthesized cross-sectional and longitudinal evidence on neurocognitive deficits associated with PIU. Inhibitory control and decision-making impairments were found to be prevalent, particularly in individuals with comorbid psychiatric disorders such as ADHD and OCD. The paper recommends longitudinal research to better establish causality and progression.

During the COVID-19 pandemic, Montag and Elhai [7] highlighted a global surge in PIU, particularly across behaviors like gaming, gambling, and pornography use. Their review underscores the mental health implications of pandemic-induced digital reliance and calls for increased awareness and preventive frameworks.

Busch and Frison [8] focused on social media usage patterns using smartphone tracking data. By clustering users based on their engagement with platforms like Facebook, Instagram, and WhatsApp, the study identified distinct risk profiles for PIU. The work is grounded in theories like the I-PACE model and Compensatory Internet Use Theory.

Orben et al. [9] explored PIU in children and adolescents, utilizing clinical assessments from the Healthy Brain Network. Their results show that PIU is significantly associated with psychiatric disorders, including depression and ADHD, as well as with sleep disturbances and daily impairment, reinforcing the need for early intervention.

Additional research has examined specific predictors, manifestations, and consequences of Problematic Internet Use (PIU) and related digital behavior disorders.

A recent systematic review by Antoniadou et al. [10] analyzed longitudinal studies investigating risk factors for problematic social media use in youth. It concluded that low self-esteem, high neuroticism, and poor peer relationships were consistent predictors, emphasizing the interaction between individual traits and environmental stressors.

Vigna-Taglianti et al. [11] conducted a clinical study on adolescent psychiatric patients and reported gender differences in PIU patterns. While male adolescents showed higher rates of online gaming-related compulsions, females were more affected by social media addiction, revealing the necessity for tailored interventions.

A predictive modeling study by Shaikh et al. [12] used machine learning to forecast academic performance from internet usage metrics. Their findings revealed that usage timing, not just volume, was a significant indicator, especially night-time browsing patterns.

Serrano-Puche et al. [13] provided a comprehensive review of problematic smartphone use and its overlap with specific PIU domains. They highlighted the lack of consensus on assessment tools but noted consistent associations between excessive smartphone use and poor academic, emotional, and physical health outcomes.

Yalcin et al. [14] explored the effect of excessive technology use on adolescent social media disorder. Their regression models showed that low self-control and lack of family belonging were significant predictors of addiction severity.

Kircaburun et al. [15] conducted a mixed-methods study analyzing motives behind problematic social media use. Quantitative results indicated that attention dysregulation and escapism were dominant risk factors, while qualitative responses suggested social validation as a recurring theme.

Lastly, Bezerra et al. [16] applied machine learning to explore digital inequality during remote learning adoption. Their study revealed that gender and regional disparity significantly influenced access and engagement, reinforcing the need for equitable digital infrastructure when deploying internet-based educational and behavioral assessments.

Recent studies have begun applying machine learning techniques to predict behavioral health outcomes using structured data in adolescent populations. For instance, Liu et al. [17] leveraged explainable models to predict Internet addiction in Chinese youth, identifying key temperament and usage variables. Gan et al. [18] utilized ensemble and regression models to forecast adolescent Internet addictive behaviors in a large sample from Chongqing. Similarly, ML models have successfully predicted loneliness in school-aged children, with Internet Addiction emerging as a significant feature. These studies underscore the relevance and novelty of the current work in modeling problematic Internet use through interpretable machine learning frameworks.

These studies demonstrate that PIU is a multifactorial condition influenced by psychological, cognitive, behavioral, and contextual factors. The present work builds upon these foundations by applying data-driven modeling approaches to predict PIU severity using clinical, behavioral, and demographic features.

Previous studies on Problematic Internet Use (PIU) show that it is influenced by many factors, including psychological traits like anxiety and impulsivity, as well as social aspects such as peer pressure and family dynamics. Research has also linked PIU with mental health conditions like ADHD, depression, and compulsive behaviors, highlighting the need for early detection. While some studies have used machine learning or statistical models to predict PIU, most rely heavily on self-reported surveys and small, localized datasets. These limitations affect how well the results can be applied to broader populations. In addition, many existing models treat PIU as a simple yes/no problem, without considering the different severity levels. Few works compare multiple models or apply advanced methods like Bayesian optimisation or threshold tuning. This work aims to bridge these gaps by using a publicly available, structured dataset and testing several interpretable machine learning models through rigorous cross-validation.

III. Detailed Methodology

A. Dataset Description

The dataset used [19] in this study was obtained from the Child Mind Institute's Problematic Internet Use Challenge hosted on the Kaggle platform. It contains structured data collected from children and adolescents to evaluate the severity of their internet use behaviour. The dataset is composed of tabular features derived from multiple sources, including demographic information, psychological questionnaire scores, and summarised time-series physiological measurements.

The primary objective is to predict a target variable referred to as the SII, which is an ordinal variable taking values from 0 to 3, where higher values correspond to more problematic usage patterns. As a classification task, this introduces not only a multi-class problem but also one with ordinal structure, which makes standard classification metrics insufficient on their own.

The dataset includes:

- Demographic Features: Age, gender, and related metadata.
- Psychological Assessment Scores: Derived from standardized clinical questionnaires measuring traits such as attention, mood, and behaviour.
- Aggregated Time-Series Features: Physiological signal summaries (e.g., heart rate variability), provided as processed numerical values across sessions.

During preprocessing, features with excessive missing values were dropped, while others were imputed using median imputation to retain robustness. The target variable sii was rounded and cast to an integer to ensure its proper interpretation as an ordinal class. All input features were kept in a numerical format, and categorical variables were encoded appropriately for tree-based models.

No raw time-series data was used directly; instead, only the pre-aggregated representations were incorporated, making the input format suitable for classical machine learning algorithms. The final dataset was split into a training and test set as per the competition's structure, with all modeling, validation, and optimisation conducted exclusively on the training portion.

This dataset provides a realistic and challenging benchmark for machine learning applications in digital behaviour analysis, combining multiple modalities with an ordinal classification objective.

B. Data Preprocessing

To ensure the reliability and robustness of the predictive models, several data preprocessing steps were applied to the raw training data. These steps were critical for handling missing values, preparing features for model compatibility, and aligning the input format with the requirements of machine learning algorithms.

1) Handling Missing Values: The raw dataset contained several features with missing or null values. Initially, a feature-wise inspection was performed to quantify missingness. Features with an excessively high percentage of missing entries were discarded, as their inclusion could introduce bias or instability in training. For the remaining features with moderate or low missingness, the median imputation strategy was employed. Median imputation was chosen over mean imputation due to its robustness against outliers and skewed distributions, which are common in clinical and behavioural data.

Imputation was conducted using SimpleImputer from the scikit-learn library. During cross-validation, imputation was applied independently within each fold to prevent information leakage from the validation set into the training data. The same fold-wise imputer was also applied to the test data within each fold to ensure consistency.

2) Target Variable Preparation: The target variable, SII, is an ordinal integer ranging from 0 (no concern) to 3 (high concern). The original values were cast to integers using round() and converted via astype(int) to ensure that all downstream models interpreted the target correctly as a classification label rather than a continuous value. No binning or additional transformation of the target variable was performed.

3) Feature Encoding and Format Standardization: The dataset primarily contains numerical features, and categorical encoding was minimal. Where necessary (e.g., for gender or binary flags), categorical variables were encoded numerically. Since the models selected (XGBoost, LightGBM, CatBoost, and ExtraTrees) can handle unscaled numerical input, no feature scaling or normalization was applied globally. However, in cases where models like Logistic Regression or deep learning were considered, StandardScaler was used selectively to ensure numerical stability.

4) Time-Series Feature Aggregation: Although the original dataset included physiological time-series data, this work made use of the pre-aggregated statistical summaries provided in the structured tabular format. These features include derived metrics such as heart rate variability measures, session-wise means, standard deviations, and other aggregate statistics. This choice ensured that all modeling could be conducted using traditional machine learning models without the need for deep time-series architectures.

5) Train-Test Separation and Fold Splitting: Following the structure of the Kaggle competition, only the provided training set was used for model development. The training data was internally split into folds using Stratified K-Fold Cross-Validation to preserve the class distribution of the ordinal target across all folds. A 5-fold and 10-fold stratified split were both used in different experiments to assess the model's generalisability.

All preprocessing operations, including imputation and target formatting, were repeated independently within each fold to ensure fair evaluation and avoid leakage.

C. Model Development

To address the ordinal multi-class classification task of predicting the SII, this work explores several ensemble-based machine learning models. These models were selected for their strong performance on structured tabular data, their robustness to outliers, and their capacity to capture non-linear relationships. All models were trained using stratified k-fold cross-validation, and their hyperparameters were optimised using Bayesian optimisation to maximise the Quadratic Weighted Kappa (QWK) metric [20]. Additionally, a threshold tuning mechanism was employed to better align model outputs with the ordinal nature of the target variable.

1) XGBoost: Extreme Gradient Boosting (XGBoost) is a widely-used, scalable tree boosting algorithm known for its speed and performance on structured data. It builds additive decision trees in a forward stage-wise manner, using second-order gradients to optimise an objective function [4].

In this work, XGBRegressor was used to model the SII target. While the task is classification-based, XGBoost's regressor variant was selected to produce continuous outputs, which could then be fine-tuned via threshold optimisation. The hyperparameters such as Learning Rate, Maximum Tree Depth, and the number of estimators were subject to Bayesian optimisation

2) LightGBM: Light Gradient Boosting Machine (LightGBM) is a gradient boosting framework that uses a histogram-based approach to build decision trees. It is optimised for memory efficiency and large-scale data [21].

LGBMRegressor was used with the same rationale as XGBoost—to output continuous values that can later be mapped to ordinal classes via threshold rounding.

3) CatBoost: CatBoost is a gradient boosting algorithm developed by Yandex, designed to handle categorical variables natively and prevent overfitting through Ordered Boosting [22].

In this study, CatBoostRegressor was used with GPU acceleration. Although the dataset had minimal categorical features, CatBoost was included for comparative evaluation.

4) ExtraTrees: Extremely randomised Trees (ExtraTrees) is an ensemble learning method that builds multiple decorrelated decision trees using random subsets of the data and features, and averages their predictions. Unlike gradient boosting models, ExtraTrees does not use boosting but relies on bagging and randomisation to reduce variance. [23]

In this work, ExtraTreesClassifier was used as a direct classifier (not regressor), and therefore returned integer class labels directly. While it did not require threshold tuning, it was still included in the ensemble and evaluated under the same validation strategy

5) Ensemble Model: An ensemble approach was also tested, where the predictions from XGBoost, LightGBM, and CatBoost were combined and passed to a meta-model (Logistic Regression) [24]. Each base model was trained independently using its optimised parameters, and their outputs were weighted and used as inputs to the logistic regression classifier [25].

Although the ensemble did not outperform XGBoost on its own, it provided balanced results and demonstrated the potential of model stacking, especially when class boundaries are ambiguous.

A tabulation of the Hyperparameters used is given in Table I.

Figure 1 illustrates the high-level flow of the machine learning pipeline used in this work, including preprocessing, model training, Bayesian optimisation, and threshold tuning stages. Each step is designed to accommodate the structured nature and ordinal target of the PIU dataset.

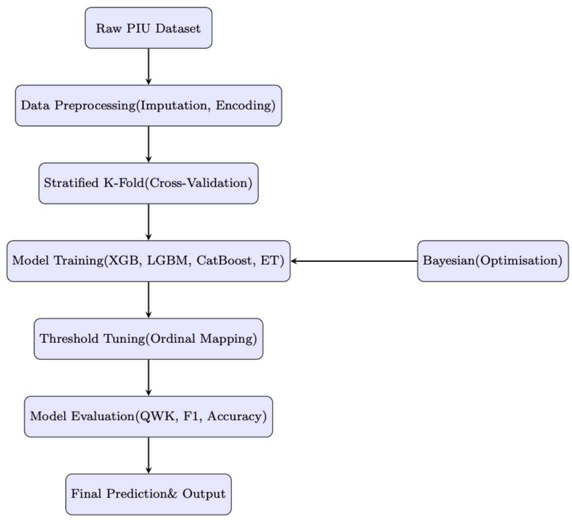

Fig. 1. Overview of the machine learning pipeline for ordinal PIU prediction.

D. Optimisation Techniques Used

1) Bayesian Optimisation: To fine-tune hyperparameters efficiently, Bayesian optimisation was applied using the bayes_opt library. Unlike grid search or random search, Bayesian optimisation models the objective function and intelligently chooses the next set of parameters to evaluate based on the probability of improvement [26].

Each model had a tailored parameter search space, and the objective function maximised during optimisation was the mean validation QWK over cross-validation folds. Integer hyperparameters (like max_depth or num_leaves) were type-cast to integers during each trial.

This approach significantly reduced training time and improved model generalisation.

2) Threshold Optimisation: Because many models (particularly the regressors) returned continuous predictions, a threshold rounding strategy was implemented to map real-valued outputs into the discrete ordinal classes (0, 1, 2, 3). The rounding thresholds were optimised using scipy.optimise.minimize, targeting maximum QWK [27].

This technique was especially beneficial for aligning predicted values with the ordinal nature of the SII target, and led to noticeable improvements in validation QWK scores across models.

TABLE I
HYPERPARAMETER CONFIGURATION FOR DIFFERENT MODELS

Model	Hyperparameter	Value / Range
XGBoost	learning_rate	0.01 – 0.1
	max_depth	3 – 15
	n_estimators	100 – 300
	subsample	0.5 – 1.0
	colsample_bytree	0.5 – 1.0
	reg_alpha	0 – 10
	reg_lambda	0 – 10
LightGBM	learning_rate	0.01 – 0.1
	max_depth	6 – 15
	num_leaves	50 – 500
	min_data_in_leaf	5 – 50
	feature_fraction	0.5 – 1.0
	bagging_fraction	0.5 – 1.0
	bagging_freq	1 – 10
	lambda_l1, lambda_l2	0 – 10
CatBoost	learning_rate	0.01 – 0.1
	depth	4 – 10
	iterations	100 – 500
	l2_leaf_reg	1 – 10
ExtraTrees	n_estimators	150 (fixed)
	max_depth	None (full growth)
	min_samples_split	2 (default)
	bootstrap	False

Experiments were conducted using Python 3.10, scikit-learn 1.3.2, XGBoost 1.7+, and LightGBM 3.3.5 on a machine equipped with an NVIDIA GPU (Tesla T4). Bayesian optimisation was implemented using the bayes_opt library.

IV. RESULTS AND ANALYSIS

A. Evaluation Metrics

The model performances were evaluated using three metrics: **Accuracy**, **Macro F1-Score**, and **Quadratic Weighted Kappa (QWK)**. Accuracy measures the overall correctness of predictions, while Macro F1-Score provides a balanced measure of precision and recall across all classes, treating each class equally. The QWK score, particularly relevant for ordinal classification tasks, measures agreement between predictions and ground truth with consideration for the severity of misclassification.

B. Cross-Validation Strategy

All models were evaluated using **Stratified K-Fold Cross-Validation** with $k = 5$ to ensure that each fold maintained the same label distribution as the full dataset. This strategy was chosen to mitigate class imbalance issues and to ensure robustness and generalisability of the models across different splits.

C. Model-wise Performance Comparison

The table below summarises the fold-wise validation performance of the models, with the average (mean) scores across all folds presented for each evaluation metric.

TABLE II
PERFORMANCE METRICS FOR DIFFERENT MODELS (5-FOLD CV)

Model	Mean Accuracy	Mean F1-Score	Mean QWK
XGBoost	**0.8111**	**0.4266**	**0.7218**
LightGBM	0.6866	0.3925	0.5510
CatBoost	0.6424	0.3421	0.4631
ExtraTrees	0.6617	0.4122	0.3980
Ensemble	0.6641	0.3518	0.4991

D. Discussion of Findings

The results in Table II indicate that **XGBoost** consistently achieved the best performance across all metrics, particularly excelling in QWK with a score of **0.7218**. This demonstrates its robustness and effectiveness in handling the ordinal nature of the target variable.

LightGBM and **CatBoost** also performed reasonably well, but lagged slightly in capturing inter-class relationships. **ExtraTreesClassifier**, a non-boosting model, performed competitively with a respectable QWK of **0.3980**, showcasing the effectiveness of randomised decision trees even without boosting.

The **Ensemble model**, which combined LightGBM, XGBoost, and CatBoost outputs with a logistic regression meta-learner, offered moderate performance gains but did not surpass XGBoost. This indicates that stacking needs more refined blending strategies or diverse model behaviours to yield significant improvements.

The incorporation of **Bayesian optimisation** played a significant role in fine-tuning hyperparameters, yielding marked improvements in each model's performance. Threshold tuning further helped in mapping continuous outputs to ordinal labels, optimising QWK scores.

Beyond performance metrics, this work emphasizes interpretability and clinical applicability. By utilizing classical ensemble models such as XGBoost and ExtraTrees, which support feature importance analysis, the modeling framework allows practitioners to identify the most influential behavioral and physiological predictors of PIU severity. The use of stratified cross-validation and ordinal-aware evaluation ensures that the results generalize to real-world settings, where early and explainable screening tools are critical for timely intervention.

V. Conclusion

This work investigated the application of classical machine learning models to the task of predicting Problematic Internet Use (PIU) severity, based on structured data provided by the Child Mind Institute. The target variable, SII, posed a unique challenge due to its ordinal multi-class nature and imbalanced class distribution.

A comprehensive modeling pipeline was developed, incorporating data imputation, stratified cross-validation, and evaluation using metrics suited for ordinal classification—namely Accuracy, Macro F1-Score, and Quadratic Weighted Kappa (QWK). The study systematically evaluated a suite of ensemble-based models, including XGBoost, LightGBM, CatBoost, and ExtraTreesClassifier. In addition, an ensemble model using logistic regression as a meta-learner was explored. Bayesian optimisation was utilized for hyperparameter tuning, and threshold optimisation techniques were applied to align model outputs with the ordinal target structure.

Among all the models tested, XGBoost achieved the highest predictive performance, attaining a QWK score of 0.7218 and a mean accuracy of 81.11%. LightGBM and CatBoost demonstrated moderate effectiveness, while the ExtraTreesClassifier showed competitive performance in terms of interpretability and computational efficiency. Although the ensemble model offered a balanced performance across metrics, it did not surpass XGBoost in predictive strength.

Overall, the results demonstrate that classical machine learning models, when appropriately optimised, can effectively predict behavioural risk profiles such as PIU. This approach provides a scalable and interpretable alternative to deep learning models, especially in scenarios where structured tabular data is available. The methodology and findings presented in this study can serve as a baseline for future work in automated behavioural health screening and related clinical prediction tasks.

References

[1] S. E. Caplan, "Problematic internet use and psychosocial well-being: Development of a theory-based cognitive–behavioral measurement instrument," *Computers in Human Behavior*, vol. 18, no. 5, pp. 553–575, 2002.

[2] V. Carli, T. Durkee, D. Wasserman *et al.*, "The association between pathological internet use and comorbid psychopathology: a systematic review," *Psychopathology*, vol. 46, no. 1, pp. 1–13, 2013.

[3] K. S. Young, "Internet addiction: The emergence of a new clinical disorder," *CyberPsychology & Behavior*, vol. 1, no. 3, pp. 237–244, 1998.

[4] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[5] C. Marino, G. Gini, A. Vieno, and M. M. Spada, "New addictions in late adolescence and emerging adulthood: How attachment style may predict problematic use of social networks and binge-watching," *Journal of Behavioral Addictions*, vol. 11, no. 1, pp. 23–34, 2022.

[6] N. A. Fineberg, Z. Demetrovics, and M. N. Potenza, "Problematic usage of the internet and cognition: A review of neurocognitive deficits in individuals with pui," *Current Addiction Reports*, vol. 9, no. 3, pp. 163–176, 2022.

[7] C. Montag and J. D. Elhai, "Problematic use of the internet during the covid-19 pandemic: Good practices and mental health recommendations," *Current Opinion in Psychology*, vol. 45, p. 101302, 2022.

[8] P. A. Busch and E. Frison, "The interplay between social media use and problematic internet usage: Four behavioral patterns," *Computers in Human Behavior*, vol. 123, p. 106878, 2021.

[9] A. Orben, A. K. Przybylski, and S.-J. Blakemore, "Problematic internet use in children and adolescents: Associations with psychiatric disorders and impairment," *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 59, no. 9, pp. 1036–1046, 2020.

[10] N. Antoniadou, C. Papatheodorou *et al.*, "Risk factors for problematic social media use in youth: A systematic review of longitudinal studies," *Current Psychology*, 2025.

[11] F. Vigna-Taglianti *et al.*, "Problematic internet use among adolescent male and female psychiatric patients: A comparative study," *Child Psychiatry & Human Development*, 2022.

[12] A. Shaikh, N. Khan *et al.*, "Predicting the impact of internet usage on students' academic performance using machine learning," *Social Network Analysis and Mining*, 2024.

[13] J. Serrano-Puche *et al.*, "Problematic smartphone use and specific problematic internet uses in university students: A review," *Education and Information Technologies*, 2022.

[14] A. Yalcin *et al.*, "Excessive technology use and social media disorder among adolescents: The role of self-control and social belonging," *Child and Adolescent Social Work Journal*, 2025.

[15] K. Kircaburun *et al.*, "A mixed-methods study of problematic social media use, attention dysregulation, and social media use motives," *Current Psychology*, 2022.

[16] C. Bezerra *et al.*, "Access to online learning: A machine learning analysis from a social justice perspective," *Education and Information Technologies*, 2022.

[17] J. Liu *et al.*, "Explainable machine learning prediction of internet addiction among adolescents," *Frontiers in Public Health*, 2025.

[18] Y. Gan *et al.*, "Application of machine learning in predicting adolescent internet behavioral addiction," *Frontiers in Psychiatry*, 2025.

[19] Child Mind Institute, "Problematic internet use prediction - child mind institute," https://www.kaggle.com/competitions/child-mind-institute-problematic-internet-use, 2023, accessed 2024-05-01.

[20] J. Cohen, "Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit," *Psychological Bulletin*, vol. 70, no. 4, pp. 213–220, 1968.

[21] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, 2017, pp. 3146–3154.

[22] A. V. Dorogush, V. Ershov, and A. Gulin, "Catboost: gradient boosting with categorical features support," *arXiv preprint arXiv:1810.11363*, 2018.

[23] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, no. 1, pp. 3–42, 2006.

[24] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013.

[25] Z.-H. Zhou, L. Rokach, and O. Maimon, Eds., *Ensemble Machine Learning: Methods and Applications*. Springer, 2012. [Online]. Available: https://link.springer.com/book/10.1007/978-1-4419-9326-7

[26] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems*, vol. 25. Curran Associates, Inc., 2012, pp. 2951–2959.

[27] A. A. Al-Shabi, M. A. Al-Razgan, and M. A. Al-Razgan, "Threshold optimization and random undersampling for imbalanced data classification," *Journal of Big Data*, vol. 10, no. 1, pp. 1–15, 2023.