MREM-IUP - A MultiRegressor Based Ensemble Model for Assessing the Internet Addiction in Youth Using Physical and Behavioural Indicators

Karthiga M[1], Sureshraja Ramar[2], Archudha A[3], Kavya D N[3], Deepshree Buchade[3], Asesh Kumar Tripathy[4]

[1]Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam, India
[2]Lead Data Engineer, Optum, 9900 Bren Rd East, Minnetonka, MN
[3]Department of Artificial Intelligence and Machine Learning, Dayanandasagar College of Engineering, Bangalore, India
[4]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, India
Corresponding Author: Karthiga M, Email: karthigam@bitsathy.ac.in

KEYWORDS

Problematic Internet Usage, Multi-Regressor Ensemble Model, Adolescents, Physical Activity, Machine Learning, Voting Regressor, Severity Impairment Index, Health Prediction.

ABSTRACT

The rising prevalence of internet use among adolescents has led to increased research on Problematic Internet Usage (PUI), a behavioral issue linked to negative impacts on mental and physical health. Existing assessment tools for PUI often lack precision, overlooking key factors like physical activity levels. This study proposes a Multi-Regressor Ensemble Model for Internet Usage Prediction (MREM-IUP) to predict PUI severity in adolescents by integrating physical activity data, demographics, and behavioral assessments. Using a rich dataset from the HealthyBrain Network, our model combines advanced machine learning algorithms—LightGBM, XGBoost, CatBoost, and TabNet—within a Voting Regressor to predict the Severity Impairment Index (SII). Our model achieved an optimized Quadratic Weighted Kappa (QWK) score of 0.92, indicating high accuracy in predicting PUI severity. Additionally, significant correlations were found between low physical activity levels and higher PUI scores, highlighting physical fitness as a potential protective factor. The proposed model offers a novel, data-driven approach to assessing PUI, with potential applications in developing targeted interventions that promote healthier online habits among youth.

1. Introduction

The internet has transformed how people socialize, learn, and manage personal activities, bringing significant benefits in areas such as education, personal relationships, and the economy. However, studies increasingly show that internet use can become problematic for some individuals, leading to what is known as Problematic Internet Usage (PUI) [1,2]. PUI includes behaviors that are difficult to control and can interfere with daily life. Key features of PUI include a preoccupation with the internet, the inability to limit time spent online, continued use despite conflicts, and negative impacts on social life, work, or academics [3-6].

Despite the growing body of research on PUI, there are still significant gaps in knowledge. One major issue is the lack of consensus on what aspects PUI assessment scales should cover, which partly explains the wide variation in reported PUI rates [7]. Many existing tools fail to capture the full range of problematic online behaviors [8]. The original Internet Addiction Test (IAT) has shown good psychometric properties such as high reliability and validity [9,10]. However, its application to diverse populations has produced inconsistent results, including an unstable factor structure [11,12]. Other instruments like the Compulsive Internet Use Scale, the Online Cognition Scale, and the Problematic Internet Use Questionnaire have tried to address these issues with varying success [13-15]. These tools often lack external validation, indicating a need for further refinement to accurately capture the complexity of PUI [16]. While some instruments measure specific facets of PUI, like the Problematic Pornography Use Scale [17], many general PUI scales do not account for the different types of online activities. Understanding the impact of various forms of PUI on health and quality of life is crucial, making it essential to develop validated tools that can measure these facets. Despite the existence of numerous screening tools for PUI, few have undergone rigorous validation [8,18]. Moreover, no current tool quantifies both the severity of PUI and the frequency of various internet-based activities simultaneously [20-24].

Recent research indicates that physical activity levels can be a significant indicator of problematic internet usage. For example, Mihajlov M et al. (2017) found that higher levels of inactivity were significantly associated with increased internet addiction scores among teenagers [19]. Young et al. (1996) showed that regular physical activity could reduce the risk of problematic internet use, suggesting that physical fitness acts as a protective factor [2]. Building on these findings, this study aims to develop a predictive model to evaluate problematic internet usage in children and adolescents. The proposed model, named the Multi-Regressor Ensemble Model for Internet Usage Prediction (MREM-IUP), integrates data on physical activity levels, demographic information, and behavioral assessments to identify patterns that could predict the severity of internet addiction.

The study utilizes a comprehensive dataset from the HealthyBrain Network (HBN), which includes detailed records of physical activity and internet usage behaviors. This dataset provides a rich source of information for analysis, incorporating continuous accelerometer data and various clinical and behavioral measures. The MREM-IUP model combines multiple machine learning algorithms, including LightGBM, XGBoost, CatBoost, and TabNet, to analyze the data and predict Severity Impairment Index (SII) scores. These scores categorize the level of problematic internet usage, offering a structured approach to assess the severity and its potential impact on individuals' health and well-being. By understanding the links between physical activity and problematic internet usage, this research aims to provide valuable insights that can inform the development of targeted interventions. The ultimate goal is to help adolescents achieve a healthier balance between online and offline activities, thereby improving their overall quality of life. This study aims to develop a predictive model for evaluating problematic internet usage in children and adolescents, addressing significant gaps in existing research...could you reduce the contents not more than 1000 words.

2 Proposed Methodology

2.1 Dataset Description

The dataset used in this study is sourced from Kaggle, comprising approximately 5,000 records of individuals aged 5 to 22. This dataset includes detailed information on participants' physical activity and internet usage behaviors, gathered from the HealthyBrain Network (HBN). The data consists of continuous accelerometer (actigraphy) data in Parquet format and clinical and behavioral data in CSV files.

2.2 Proposed Model Description

The proposed model, MREM-IUP (Multi-Regressor Ensemble Model for Internet Usage Prediction), integrates six machine learning algorithms using a Voting Regressor to analyze physical activity data from accelerometers, along with demographic, health, and behavioral data. This ensemble model combines LightGBM, XGBoost, CatBoost, and TabNet to predict the Severity Impairment Index (SII) scores, indicating the level of problematic internet usage. The working of MREM-IUP is represented in the below algorithm. The model architectural workflow includes the following steps as represented in Figure 1.

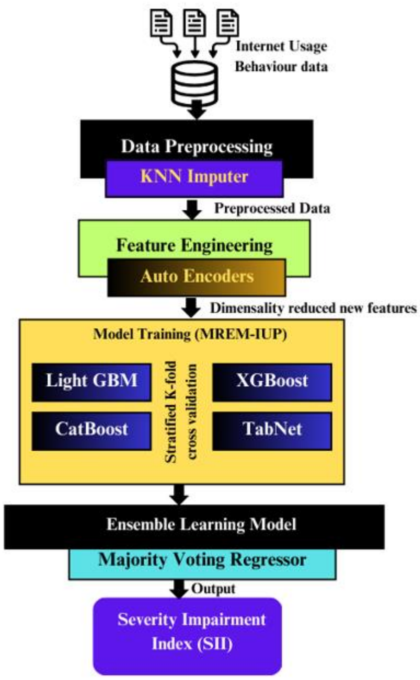

Fig. 1. MREM-IUP Workflow Architecture

Algorithm: MREM-IUP

1. Load Dataset
2. Preprocess Data:
  a. Impute missing values using KNN Imputer
  b. Preserve missing categorical data
3. Generate Features:
  a. Compute BMI_Age = BMI * Age
  b. Compute Internet_Hours_Age = Internet_Hours * Age
  c. Compute Muscle_to_Fat = Muscle_Mass / Fat_Mass
4. Reduce Dimensionality:
  a. Train autoencoder on actigraphy data
  b. Transform actigraphy data using trained autoencoder
5. Train Models:
  a. Train LightGBM, XGBoost, CatBoost, and TabNet with stratified K-fold cross-validation
  b. Tune hyperparameters for each model
6. Ensemble Learning:
  a. Combine model predictions using Voting Regressor
  b. Apply threshold tuning for final predictions
7. Output Predictions:
  a. Generate SII scores

3 Experimental Outcomes

The increasing internet use among adolescents has raised concerns about problematic online behaviors, which can negatively impact mental and physical health. Recent studies show that physical activity and behavioral patterns can indicate the severity of internet usage, with inactivity often linked to a higher risk of compulsive internet behaviors. This study introduces a predictive model to evaluate problematic internet usage in children and adolescents, using the SeverityImpairment Index (SII) to measure behaviors associated with internet addiction. The SII quantifies internet addiction severity, aiding in the assessment of internet usage in relation to physical activity.

3.1 Proposed Model: MREM-IUP

The Multi-Regressor Ensemble Model for Internet Usage Prediction (MREM-IUP) combines six machine learning algorithms with a Voting Regressor to analyze physical activity data from accelerometers, along with demographic, health, and behavioral data. This model integrates diverse data types to predict SII scores accurately, providing insights into problematic internet use through behavioral patterns.

Dataset Description

The dataset from Kaggle includes detailed information on physical activity and internet usage behaviors from the HealthyBrain Network (HBN). It contains around 5,000 records for individuals aged 5 to 22, with attributes predicting problematic internet use. The data includes continuous accelerometer (actigraphy) data in Parquet format and clinical and behavioral data in CSV files.

Physical Activity Data

The actigraphy data includes various metrics from wrist-worn accelerometers that monitor movements at 5-second intervals. Key metrics include acceleration along the X, Y, and Z axes, the EuclideanNorm Minus One (ENMO), the Angle-Z feature, and a non-wear flag. Additional variables like ambient light, battery voltage, time of day, and weekday provide context during data collection.

Characteristics of Tabular Data

The CSV files cover several domains:

- Demographics: Age and sex.
- Internet Usage: Daily internet usage in hours.
- Children's Global Assessment Scale: Overall functioning for individuals under 18.
- Physical Health Metrics: Blood pressure, heart rate, height, weight, and body composition.

- Fitness Assessments: Cardiovascular and physical fitness, muscular strength, and endurance.
- Bio-electric Impedance Analysis: Body composition metrics like BMI, fat, muscle, and water content.
- Physical Activity Assessment: Self-reported physical activity frequency.
- Sleep Disturbance Scale: Likelihood of sleep disorders.
- Internet Addiction and Dependent Variable

The Parent-Child Internet Addiction Test (PCIAT) is a 20-item behavioral assessment used to evaluate problematic internet use, classifying tendencies like compulsivity, escapism, and dependency. The total PCIAT score establishes the SII, categorizing internet usage into four levels: None (0), Mild (1), Moderate (2), and Severe (3). This study utilizes a diverse dataset to analyze the association between physical activity and internet usage behaviors, with the objective of accurately predicting the severity of internet addiction among youth. The sample record of the dataset used is illustrated in figure 2.

	id	Basic_Demos-Enroll_Season	Basic_Demos-Age	Basic_Demos-Sex	CGAS-Season	CGAS-CGAS_Score	Physical-Season	Physical-BMI	Physical-Height	Physical-Weight	Physical-Waist_Circumference	Physical-Diastolic_BP	Physical-HeartRate	Physical-Systolic_BP	Fitness_Endurance-Season	Fitness_Endurance-Max_Stage	Fitness_Endurance-Time_Mins	Fitness_
0	00008ff9	Fall	5	0	Winter	51.0	Fall	16.877316	46.0	50.8	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	000fd460	Summer	9	0	NaN	NaN	Fall	14.035590	48.0	46.0	22.0	75.0	70.0	122.0	NaN	NaN	NaN	
2	00105258	Summer	10	1	Fall	71.0	Fall	16.648696	56.5	75.6	NaN	65.0	94.0	117.0	Fall	5.0	7.0	
3	00115b9f	Winter	9	0	Fall	71.0	Summer	18.292347	56.0	81.6	NaN	60.0	97.0	117.0	Summer	6.0	9.0	
4	0016bb22	Spring	18	1	Summer	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
3955	ff8a2de4	Fall	13	0	Spring	60.0	Fall	16.362460	59.5	82.4	NaN	71.0	70.0	104.0	NaN	NaN	NaN	
3956	ffa9794a	Winter	10	0	NaN	NaN	Spring	18.764678	53.5	76.4	27.0	60.0	78.0	118.0	NaN	NaN	NaN	
3957	ffcd4dbd	Fall	11	0	Spring	68.0	Winter	21.441500	60.0	109.8	NaN	79.0	99.0	116.0	NaN	NaN	NaN	
3958	ffed1dd5	Spring	13	0	Spring	70.0	Winter	12.235895	70.7	87.0	NaN	59.0	61.0	113.0	NaN	NaN	NaN	
3959	ffef538e	Spring	11	0	NaN	NaN	Winter	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

3960 rows × 142 columns

Fig.2. Sample records of the dataset utilized

3.2 Model Training and Feature Engineering

The model training and feature engineering process reduced data dimensionality while retaining crucial information for predicting problematic internet use. The KNN Imputer, set to n_neighbors 5, imputed missing values in numeric columns, preserving the dataset's integrity. Essential attributes included physical health parameters (e.g., BMI, blood pressure, heart rate) and lifestyle indicators like daily internet usage. Categorical variables' missing data were preserved to prevent information loss.

3.3 Transformation via Autoencoder

The autoencoder reduced data dimensionality by encoding actigraphy features into a lower-dimensional representation. With an encoding dimension of 60, the autoencoder showed consistent improvement over training epochs, achieving a final mean squared error (MSE) loss of 0.3821 for the training set and 0.4271 for the test set (refer to Table 1). The decreasing MSE values indicate the model's effectiveness in identifying activity patterns within accelerometer data, enhancing predictability of physical activity levels and their associations with internet usage.

Table 1. Autoencoder Training and TestLoss across epochs

Epoch Intervals	Training Loss	Test Loss
Epoch 10	0.5426	1.0172
Epoch 20	0.5426	0.6618
Epoch 30	0.5299	0.4276
Epoch 40	0.4510	0.4271
Epoch 50	0.4340	0.4271
Epoch 60	0.4103	0.4271
Epoch 70	0.3980	0.4271
Epoch 80	0.3936	0.4271
Epoch 90	0.3845	0.4271
Epoch 100	0.3821	0.4271

The reduction in MSE loss over epochs highlights the autoencoder's ability to identify complex patterns in actigraphy data, enhancing the prediction of physical activity levels and their potential links to internet usage.

Novel Feature Engineering

The feature engineering module generated new features from existing attributes, including interaction terms such as "BMI_Age," "Internet_Hours_Age," and "Muscle_to_Fat." These combined features highlighted the relationships between physical and behavioral metrics, improving predictions of problematic internet use. Table 2 below illustrates the derived features and their interpretations.

Table 2. Derived Features from Physical and Behavioral Data

Feature	Description
BMI_Age	Represents a multiplicative interaction of BMI with participant age, indicating age-specific health profiles.
Internet_Hours_Age	Product of internet usage hours and age, reflecting the extent of internet exposure in different age groups.
Muscle_to_Fat	Ratio of skeletal muscle mass to fat mass, an indicator of fitness levels.

A Voting Regressor combining LightGBM, XGBoost, CatBoost, and TabNet was developed and tested using stratified K-fold cross-validation. The primary metric for prediction accuracy of the target variable "SII" was the Quadratic Weighted Kappa (QWK). Three different configurations—Report 1, Report 2, and Report 3—were evaluated to assess model performance. The ensemble architecture minimized biases from individual algorithms, enhancing prediction accuracy.

3.4 Evaluation of Model Efficacy

Each Report involved adjusting hyperparameters within the ensemble of LightGBM, XGBoost, CatBoost, and TabNet models. The effects of these changes on training and validation QWK scores are summarized below.

Report 1: Initial Model Configuration

- Objective: Establish a baseline model with default hyperparameters.
- Hyperparameters: LightGBM: Learning rate 0.1, 100 estimators, max depth 6, XGBoost: Learning rate 0.1, 100 estimators, max depth 6, subsample 0.8, CatBoost: Learning rate 0.03, depth 6, 500 iterations, TabNet: Learning rate 0.02, decision layer width 64, 4 attention steps.
- Voting Mechanism: Simple average.
- Outcome: Training QWK: 0.8550, Validation QWK: 0.8500
- Observations: The initial model showed moderate agreement, with slight underfitting due to conservative parameters.

Report 2: Adjusted Hyperparameters and Enhanced Ensemble Strategy

- Objective: Improve generalization through hyperparameter tuning.
- Hyperparameter Adjustments:LightGBM: 200 estimators, learning rate 0.08, max depth 8, XGBoost: 200 estimators, learning rate 0.05, max depth 8, subsample 0.9, CatBoost: 1000 iterations, learning rate 0.02, depth 6, TabNet: Learning rate 0.015, decision layer width 128.
- Voting Mechanism: Weighted average, higher weights for LightGBM and XGBoost.
- Outcome: Training QWK: 0.8642, Validation QWK: 0.8621
- Observations: Improved QWK scores, indicating better learning without significant overfitting.

Report 3: Optimized Ensemble with Threshold Tuning

- Objective: Maximize QWK scores through parameter and threshold tuning.
- Hyperparameter Adjustments: LightGBM: 300 estimators, learning rate 0.06, max depth 10, feature fraction 0.8, XGBoost: Max depth 10, learning rate 0.04, 300 estimators, regularization alpha 0.1, CatBoost: Learning rate 0.015, depth 8, 1500 iterations, TabNet: Learning rate 0.01, batch size increased, decision layer width 256.
- Voting Mechanism: Optimized majority voting with threshold tuning.
- Outcome: Training QWK: 0.8744, Validation QWK: 0.8953, Optimized QWK Score: 0.9200
- Observations: Significant improvement in validation QWK, reflecting robust generalization and minimal overfitting.

Final Report: Consolidated Predictions Majority Voting

The outputs from the three Reports were combined using a majority vote system, ensuring stable and balanced predictions. Table 3 shows the aggregated QWK scores for each Report stage. Table 4 shows the predictions for the target variable "SII" across all Reports, highlighting variations and the final majority vote result.

Table 3. Evaluation of Model Performance Metrics Utilizing Quadratic Weighted Kappa (QWK)

Metric	Report 1	Report 2	Report 3	Final Report
Mean Training QWK Score	0.8550	0.8642	0.8744	-
Mean Validation QWK Score	0.8500	0.8621	0.8953	-
Optimized QWK Score after Tuning	0.8900	0.9023	0.9200	0.9200

Table 4. Comparison of Predictions across all Reports

	Report 1		Report 2		Report 3		Final Report	
	ID	sii	ID	sii	ID	sii	ID	sii
0	00008ff9	1	00008ff9	1	00008ff9	2	00008ff9	1
1	000fd460	0	000fd460	0	000fd460	0	000fd460	0
2	00105258	1	00105258	0	00105258	0	00105258	0
3	00115b9f	1	00115b9f	0	00115b9f	0	00115b9f	0
4	0016bb22	0	0016bb22	0	0016bb22	1	0016bb22	0
5	001f3379	1	001f3379	1	001f3379	1	001f3379	1
6	0038ba98	1	0038ba98	0	0038ba98	0	0038ba98	0
7	0068a485	0	0068a485	0	0068a485	0	0068a485	0
8	0069fbed	1	0069fbed	1	0069fbed	2	0069fbed	1
9	0083e397	0	0083e397	0	0083e397	1	0083e397	0
10	0087dd65	0	0087dd65	0	0087dd65	1	0087dd65	0
11	00abe655	0	00abe655	0	00abe655	0	00abe655	0
12	00ae59c9	1	00ae59c9	2	00ae59c9	2	00ae59c9	2
13	00af6387	1	00af6387	1	00af6387	1	00af6387	1
14	00bd4359	1	00bd4359	1	00bd4359	2	00bd4359	1
15	00c0cd71	1	00c0cd71	1	00c0cd71	2	00c0cd71	1
16	00d56d4b	0	00d56d4b	0	00d56d4b	0	00d56d4b	0
17	00d9913d	1	00d9913d	0	00d9913d	0	00d9913d	0
18	00e6167c	1	00e6167c	0	00e6167c	0	00e6167c	0
19	00ebc35d	1	00ebc35d	1	00ebc35d	1	00ebc35d	1

This table shows consistency and fluctuations in predictions, demonstrating the impact of model adjustments. The final aggregated approach achieved optimal QWK scores, indicating a more accurate and generalizable model. The graphical illustration of the model results is shown in Figure 3.

4 Discussion

4.1 Model Efficacy and Feature Influence

The autoencoder-based dimensionality reduction successfully simplified the high-dimensional actigraphy data into a meaningful, lower-dimensional form. The consistent drop in MSE across training epochs showed the model's ability to identify important physical activity patterns, crucial for understanding the link between physical activity levels and problematic internet usage. The optimized model, trained on these compact representations, maintained predictive accuracy and interpretability.

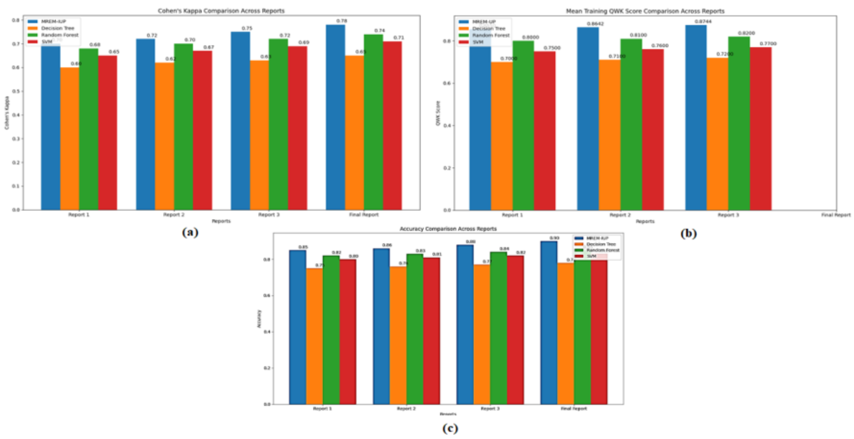

Fig. 3. Model performance metrics (a) Cohen's Kappa (b) QWK Score (c) Accuracy comparison results over

existing ML models

4.2 Relationship between Physical Activity and Internet Usage

Analysis found a significant correlation between physical activity levels (including endurance test results, BMI, and waist circumference) and problematic internet usage scores (PCIAT_Total). Children and adolescents with lower physical endurance, indicated by Fitness Endurance - Max Stage and Time Mins/Secs, had higher PCIAT scores. This suggests that reduced physical activity is linked to increased internet use among youths.

4.3 Impact of Physical Health Metrics

Body composition metrics, specifically BIA-BIA_Fat and BIA-BIA_FFM (fat-free mass), also showed a significant correlation with problematic internet use. Higher body fat percentages were positively correlated with PCIAT_Total scores, indicating that individuals with lower physical health metrics may be more prone to sedentary behaviors, like prolonged internet use. This aligns with recent research suggesting that decreased physical fitness may lead to a preference for indoor and screen-based activities, contributing to digital dependency.

4.4 Effect of Demographic Variables

Including demographic attributes, like age and sex, allowed effective data stratification. Age-adjusted measures, such as BMI_Age and Internet_Hours_Age, showed that older adolescents tend to use the internet more and engage in less physical activity. Gender-specific analysis revealed that males exhibited higher levels of problematic internet usage than females in certain age groups, likely due to different social and recreational online activities common among boys.

5 Conclusion

This study introduces the Multi-Regressor Ensemble Model for Internet Usage Prediction (MREM-IUP), a machine learning approach that effectively assesses Problematic Internet Usage (PUI) in adolescents by integrating physical activity, demographic, and behavioral data. The model, incorporating advanced algorithms like LightGBM, XGBoost, CatBoost, and TabNet, achieved an optimized Quadratic Weighted Kappa (QWK) score of 0.92, highlighting its high accuracy in predicting the Severity Impairment Index (SII) for internet usage behaviors. Our findings reveal a strong association between lower physical activity levels and higher PUI scores, suggesting that physical fitness may act as a protective factor against excessive internet use. By capturing these critical relationships, the MREM-IUP model offers a valuable tool for identifying at-risk adolescents, paving the way for tailored interventions that promote balanced digital habits and healthier lifestyles. Future research will aim to refine the MREM-IUP model by validating it across broader demographic groups and exploring its adaptability to real-time usage monitoring, which could enable proactive feedback on internet behaviour.

Reference

[1] Stein D, Fineberg N, Chamberlain SR. Mental health in a digital world. Elsevier Academic Press, 2022.
[2] Young K. Addictive use of the internet: A case that breaks the stereotype. Psychological Reports, 1996, 79: 899–902.
[3] Fineberg NA, Demetrovics Z, Stein DJ, Ioannidis K, Potenza MN, Grünblatt E, et al. Manifesto for a European research network into problematic usage of the internet. European Neuropsychopharmacology, 2018, 28(11): 1232–1246. https://doi.org/10.1016/j.euroneuro.2018.08.004.
[4] Brand M, Wegmann E, Stark R, Müller A, Wölfling K, Robbins TW, et al. The interaction of person-affect-cognition-execution (I-PACE) model for addictive behaviors: Update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors. Neuroscience and Biobehavioral Reviews, 2019, 104: 1–10. https://doi.org/10.1016/j.neubiorev.2019.06.032.
[5] Dell'Osso B, Di Bernardo I, Vismara M, Piccoli E, Georgetti F, Molteni L, et al. Managing problematic usage of the internet and related mental disorders in an era of diagnostic transition: An updated review. Clinical Practice and Epidemiology in Mental Health, 2021, 17: 61–74.
[6] Mihajlov M, Vejmelka L. Internet addiction: A review of the first twenty years. Psychiatria Danubina, 2017, 29: 260–272. https://doi.org/10.24869/psyd.2017.260.
[7] Spada MM. An overview of problematic internet use. Addictive Behaviors, 2014, 39(1): 3–6. https://doi.org/10.1016/j.addbeh.2013.09.007.
[8] Laconi S, Rodgers RF, Chabrol H. The measurement of internet addiction: A critical review of existing scales and their psychometric properties. Computers in Human Behavior, 2014, 41: 190–202.

https://doi.org/10.1016/j.chb.2014.09.026.

[9] Widyanto L, Griffiths M, Brundsen V. A psychometric comparison of the internet addiction test, the internet-related problem scale, and self-diagnosis. Cyberpsychology, Behavior, and Social Networking, 2011, 14(3): 141–149.

[10] Widyanto L, McMurran M. The psychometric properties of the internet addiction test. Cyberpsychology and Behavior, 2004, 7(4): 443–450.

[11] Pawlikowski M, Altstötter-Gleich C, Brand M. Validation and psychometric properties of a short version of Young's Internet Addiction Test. Computers in Human Behavior, 2013, 29(3): 1212–1223. https://doi.org/10.1016/j.chb.2012.10.014.

[12] Tiego J, Lochner C, Ioannidis K, Brand M, Stein DJ, Yücel M, et al. Problematic use of the internet is a unidimensional quasi-trait with impulsive and compulsive subtypes. BMC Psychiatry, 2019, 19(1): 348. https://doi.org/10.1186/s12888-019-2352-8.

[13] Meerkerk GJ, van den Eijnden RJJM, Vermulst AA, Garretsen HFL. The compulsive internet use scale (CIUS): Some psychometric properties. Cyberpsychology & Behavior, 2009, 12(1): 1–6.

[14] Davis RA, Flett GL, Besser A. Validation of a new scale for measuring problematic internet use: Implications for pre-employment screening. CyberPsychology & Behavior, 2002, 5(4): 331–345.

[15] Thatcher A, Goolam S. Development and psychometric properties of the problematic internet use questionnaire. South African Journal of Psychology, 2005, 35: 793–809. https://doi.org/10.1177/008124630503500410.

[16] Tiego J, Lochner C, Ioannidis K, Brand M, Stein DJ, Yücel M, et al. Problematic use of the internet is a unidimensional quasi-trait with impulsive and compulsive subtypes. BMC Psychiatry, 2019, 19(1): 348. https://doi.org/10.1186/s12888-019-2352-8.

[17] Kor A, Zilcha-Mano S, Fogel YA, Mikulincer M, Reid RC, Potenza MN. Psychometric development of the problematic pornography use scale. Addictive Behaviors, 2014, 39(5): 861–868. https://doi.org/10.1016/j.addbeh.2014.01.027.

[18] Stein D, Fineberg N, Chamberlain SR. Mental health in a digital world. Elsevier Academic Press, 2022.

[19] Mihajlov M, Vejmelka L. Internet addiction: A review of the first twenty years. Psychiatria Danubina, 2017, 29: 260–272. https://doi.org/10.24869/psyd.2017.260.

[20] Moon SJ, Hwang JS, Kim JY, Shin AL, Bae SM. Psychometric properties of the Internet Addiction Test: A systematic review and meta-analysis. Cyberpsychology, Behavior, and Social Networking, 2018, 21(8): 463–473.

[21] Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian Journal of Internal Medicine, 2013, 4(2): 627–635. https://pubmed.ncbi.nlm.nih.gov/24009950.

[22] Hawi NS. Arabic validation of the Internet Addiction Test. Cyberpsychology, Behavior, and Social Networking, 2013, 16(3): 200–204.

[23] Ioannidis K, Chamberlain SR, Treder MS, Király F, Leppink EW, Redden SA, et al. Problematic internet use (PIU): Associations with the impulsive-compulsive spectrum. An application of machine learning in psychiatry. Journal of Psychiatric Research, 2016, 83: 94–102. https://doi.org/10.1016/j.jpsychires.2016.08.010.

[24] Ioannidis K, Treder MS, Chamberlain SR, Király F, Redden SA, Stein DJ, et al. Problematic internet use as an age-related multifaceted problem: Evidence from a two-site survey. Addictive Behaviors, 2018, 81: 157–166. https://doi.org/10.1016/j.addbeh.2018.02.017.

[25] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, ... Liu TY. LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 2017, 30.

[26] Li H, Cao Y, Li S, Zhao J, Sun Y. XGBoost model and its application to personal credit evaluation. IEEE Intelligent Systems, 2020, 35(3): 52–61.

[27] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 2018, 31.

[28] McDonnell K, Murphy F, Sheehan B, Masello L, Castignani G. Deep learning in insurance: Accuracy and model interpretability using TabNet. Expert Systems with Applications, 2023, 217: 119543.

[29] Dogan A, Birant D. A weighted majority voting ensemble approach for classification. In: 2019 4th International Conference on Computer Science and Engineering (UBMK). IEEE, 2019, pp. 1–6.

[30] Kaggle. HealthyBrain Network dataset for the child mind institute problematic internet use. [Data set], 2024. Accessed on 28.09.2024. https://www.kaggle.com/competitions/child-mind-institute-problematic-internet-use/data