Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

# Machine Learning Model for Prediction of Smartphone Addiction

**Janavi M[1], Dr. Seshaiah Merikapudi M[2]**

[1]Tech Scholar,
Department of Computer Science and Engineering, S.J.C. Institute of Technology.
[2]Associate Professor,
Department of Computer Science and Engineering, S.J.C. Institute of Technology.
Email: janavimadangopal@gmail.com, merikapudi@gmail.com

**Abstract**

Smartphone addiction has emerged a critical issue affecting adolescents and young adults, often resulting in academic decline, psychological stress, and reduced well- being. Excessive engagement in activities such as social media, gaming, and prolonged screen exposure contributes significantly to this problem. Machine learning methods such as Random Forest, Logistic Regression, and Support Vector Machines (SVM) to predict smartphone addiction based on quantifiable features like daily screen usage, social media activity, gaming time, and the count of installed apps. An interactive dashboard built with Streamlit and Altair is integrated to visualize results and provide real-time feedback. Furthermore, recent advances in behavioral analytics, explainable AI, and digital wellness technologies highlight the growing potential for intelligent systems to enable early detection and intervention. The proposed framework contributes to supporting healthier smartphone use by combining predictive modeling with interpretable visualization tools.

**Keywords:** Smartphone overuse, machine learning, Random Forest, Logistic Regression, Support Vector Machines (SVM), daily screen usage, social media engagement, gaming time, predictive analytics, explainable artificial intelligence (XAI), digital well-being

## 1. Introduction

Smartphones have emerged as essential devices for staying connected,, entertainment, education, and productivity. However, the excessive and uncontrolled use of smartphones has led to a growing concern regarding smartphone addiction, a behavioral issue with significant psychological, social, and academic consequences. Several studies have linked smartphone overuse to disruptions in sleep patterns and a decline in academic achievement, anxiety, depression, and impaired interpersonal relationships .

Traditional assessment methods for detecting smartphone addiction, such as psychometric tests and self-reported questionnaires, are useful but limited by subjectivity, time constraints, and the inability to capture real-time behavioral patterns. Consequently, application of Machine Learning (ML) models has gained traction as a more scalable and data- driven approach to predicting addictive smartphone behavior.

ML algorithms are capable of detecting intricate relationships in extensive datasets datasets, thereby offering objective, reliable, and personalized predictions.

Recent studies have explored a range of ML techniques— such as Logistic Regression, Random Forest, Support Vector Machines (SVM), and Artificial Neural Networks—for predicting addictive behaviors using features like daily screen time, application usage patterns, frequency of social media interactions, and self-reported psychological indicators. In addition, the integration of Explainable AI (XAI) methods has been highlighted as an important step toward ensuring interpretability and transparency in healthcare-related predictions.

This research presents a machine learning framework designed to predict smartphone addiction by integrating both behavioral and psychological characteristics. The framework systematically evaluates the performance of multiple supervised learning models to identify the predictor that performs best with respect to accuracy, interpretability, and computational efficiency. The broader objective of this research is to enable early identification of individuals at risk of smartphone addiction, thereby facilitating timely interventions and promoting healthier digital usage patterns.

**Problem statement**: Smartphone addiction has become a serious issue among students, causing reduced academic performance, stress, and negative impacts on mental health. Existing detection methods are mostly survey-based, time- consuming, and subjective. An intelligent system capable of performing tasks automatically is required predict smartphone addiction using measurable features like Factors such as screen time, social media engagement, gaming hours, and the total number of installed applications can be analyzed using machine learning methods to provide insights accurate predictions, while an interactive dashboard can help in early detection and promoting digital wellness.

**Objectives:**

1. To implement data preprocessing and feature selection techniques that improve accuracy and reliability of addiction prediction.
2. To design an interactive dashboard (using Streamlit and Altair) for real-time prediction, visualization, and awareness of smartphone addiction risk**.**

## 2. Literature Survey

[1] Osorio et al. created a comprehensive machine learning framework to forecast teenage smartphone addiction using the Big Five personality traits together with the Smartphone Addiction Scale (SAS). The author tested Decision Tree, Random Forest, XGBoost, and Logistic Regression, finding Random Forest to be the most effective, achieving 89.7% accuracy and 87.3% precision. Among the personality traits, neuroticism and conscientiousness emerged as the strongest indicators.

[2] Hong et al. focused on a large sample of Chinese university students (around 3,000 participants) to model cell phone addiction risk through machine learning. Their findings highlighted perfectionism as a major predictor, and the developed model attained 76.68% accuracy.

[3] A South Korean research group explored explainable machine learning for identifying adolescent

smartphone overdependence using survey data collected in between 2017 and 2021. They compared a range of models, from Logistic Regression to Convolutional Neural Networks (CNN). The XGBoost model delivered the strongest performance with 87.6% precision, revealing that excessive use of games, webtoons, and e-books were significant risk factors.

[4] Sangeetha et al. employed Logistic Regression on a Kaggle dataset comprising 13,589 entries to predict smartphone addiction. They further demonstrated real-world usability by deploying the model as a Flask-based web application capable of providing real-time addiction risk assessments.

[5] Raj et al., working in a multidisciplinary research setting, compared the performance of Decision Tree, Logistic Regression, and Random Forest for smartphone addiction prediction. Among these, Random Forest proved superior with 89% accuracy.

[6] Vimala et al., in their ICNGT 2025 paper, studied the academic implications of smartphone overuse in a group of 500 students. Their Gradient Boosting model achieved an impressive 91.2% accuracy. Important features influencing addiction included screen time, app usage frequency, sleep disturbances, impulsivity, and anxiety.

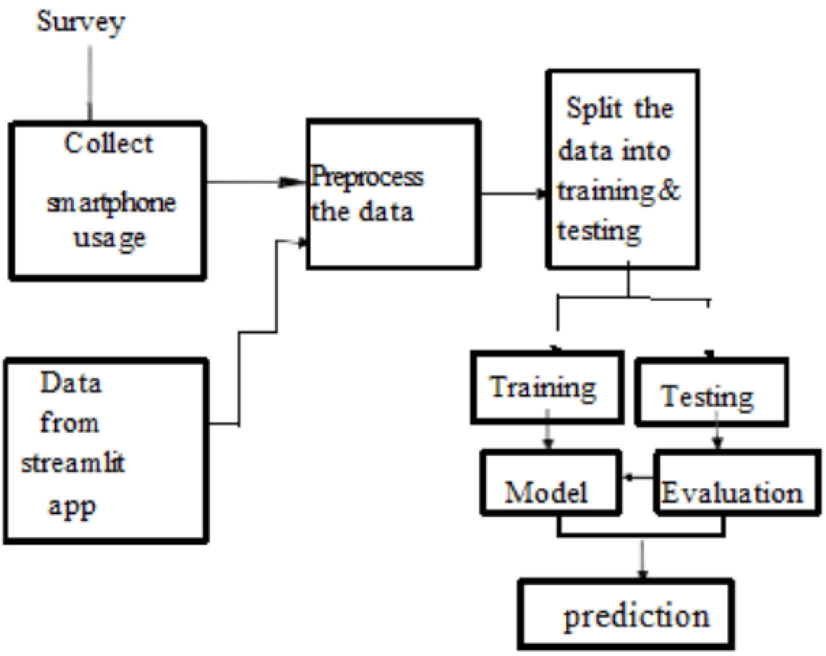
[7] Kadirvelu et al. introduced a digital phenotyping approach by combining passive and active smartphone usage data from adolescents (N = 103). Using contrastive learning followed by supervised fine-tuning, their models predicted risks related to mental health issues such as depression, insomnia, and suicidal thoughts, reaching balanced accuracies between 0.67 and 0.77.

[8] Orzikulova et al. presented Time2Stop, an adaptive, explainable AI system designed for real-time smartphone overuse intervention. Tested in an 8-week field study (N = 71), the system surpassed baseline models, boosting intervention accuracy by +32.8% and user receptivity by +8%, with explainability contributing to higher user trust and effectiveness.

[9] Xuan et al. conducted a longitudinal study tracking smartphone unlocking behaviors over the four years of college. Their analysis demonstrated that unlocking patterns could reliably forecast mental health outcomes, while also revealing notable gender- and location-based differences.

[10] Wu et al. developed MindShift, a system that integrates large language models with behavioral data such as app usage, user goals, and emotional states to deliver personalized persuasive interventions. During a 5-week trial, MindShift reduced smartphone use by 7.4–9.8% and significantly enhanced self-efficacy, illustrating the potential of AI-driven, context-aware digital well-being tools.

## 3. Methodology




**Data Collection**

The first stage of the system involves gathering relevant data about students' smartphone usage. Data is collected using two primary sources:

Survey forms: Students report their average daily screen time, time spent on social media, gaming duration, as well as the overall count of installed apps.

Streamlit application: A user-friendly an online platform that allows students to directly enter their usage information into the system. This process ensures the collection of adequate and reliable data to analyze usage patterns and assess potential addiction levels. As the gathered data consists of numerical values (hours and counts), it is well-suited for machine learning analysis.

**Data Preprocessing**

Raw data obtained from surveys and the app is not always ready for analysis. Preprocessing is performed to prepare the data for model training. The key steps include:

Data Cleaning: Removal of incomplete, inconsistent, or incorrect entries.

Normalization/Scaling: Converting values like screen time or gaming hours into a standard range so that no single feature dominates the learning process.

Handling Missing Values: If some inputs are missing, they may be imputed (replaced) with average values or handled using other techniques.

Label Creation: Students are categorized into Addicted or Not Addicted classes based on thresholds, such as being above the 75th percentile in one or more features.

This stage ensures that the dataset is structured, clean, and suitable for model building

**Feature Extraction and Selection**

The four features are directly extracted from the data: Daily screen time
Social media usage time Gaming time
Number of apps used
These features are directly measurable and relevant to smartphone usage; they are retained for model training. Feature selection helps reduce unnecessary attributes, improving model performance and interpretability. By focusing on only these essential features, the system avoids complexity while still capturing the factors most associated with addiction

**Training and Testing**

The data is preprocessed and features are selected, the dataset is divided into two subsets:
Training Set: Typically 70–80% of the data, used to train the machine learning models.
Testing Set: The remaining 20–30%, used to evaluate model performance on unseen data.
This separation ensures that the model do not simply memorize the data but instead learns general patterns that can be applied to new cases. Cross-validation techniques may also be used for more reliable performance estimation.

**Model Building**

Different machine learning methods are applied to capture the relationship between the input features and the smartphone addiction label. One such method is Random Forest, which leverages an ensemble of decision trees to enhance prediction accuracy.

1. **Ensemble Method** – Random Forest is a type of classification method that integrates several decision trees to improve accuracy and stability.
2. **Randomization** – Each tree is trained on a random subset of data and features, ensuring diversity among trees and reducing overfitting.
3. **Majority Voting** – For prediction, each tree gives a result (addicted or non-addicted), and the final decision is made by majority vote.
4. **Feature Importance** – The algorithm highlights key factors influencing smartphone addiction, such as screen time, app usage, and sleep patterns.
5. **Robustness** – It performs well with large, noisy, and imbalanced datasets, making it reliable for predicting smartphone addiction in diverse populations.

Support Vector Machine (SVM): Finds the optimal boundary that separates addicted and non-addicted classes.

1. **Classification Objective** – SVM is a supervised learning technique that distinguishes smartphone users into two groups: **addicted** and **non-addicted**, by finding the best possible decision boundary.
2. **Margin Maximization** – It identifies an optimal hyperplane that maximizes the distance (margin) between the two classes, ensuring better accuracy and generalization.
3. **Support Vectors** – Only the critical data points closest to the boundary, called support vectors, influence the model’s decision surface.
4. **Handling Complexity** – By using soft margins and kernel functions, SVM effectively manages overlapping behaviors and non-linear relationships in smartphone addiction datasets.

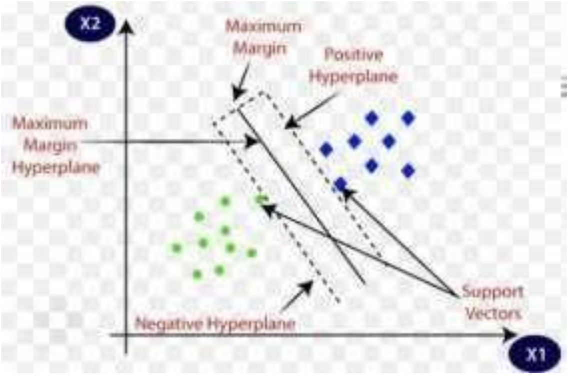


Figure 1: Support vector machine

Logistic Regression: Models the probability of addiction based on input features.
The chosen model captures the patterns of smartphone usage that contribute to addiction.

**Model Evaluation**

The trained models are evaluated to measure their effectiveness. Evaluation metrics include:
Accuracy: The percentage of correctly predicted cases. Precision: The proportion of true addicted predictions out of all predicted addicted cases.
Recall (Sensitivity): The proportion of correctly identified addicted cases out of all actual addicted cases.
F1-Score: Represents the harmonic mean of precision and recall, providing a balance between the two metrics. These evaluation measures help assess the model's ability to accurately detect smartphone addiction while reducing misclassification errors.

**Prediction**

After training and evaluation, the best-performing model is deployed for real-time prediction. When a student enters their daily smartphone usage details into the Streamlit app, the model processes the inputs and outputs whether the student is classified as Addicted or Not Addicted.
This final stage enables the system to act as an effective solution for timely identification, helping educators, parents, and students themselves to take corrective actions before the addiction becomes severe.

Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

## 4. Result


Figure 2: Prediction of History

It displays user input features such as daily usage time, social media time, gaming time, and number of apps used, along with the system's prediction results (Addicted or Not Addicted). A bar chart at the bottom summarizes predictions by model.

Figure 3: Model performance

It compares different machine learning models—Random Forest, Logistic Regression, and SVM—using metrics like accuracy, precision, recall, and F1-score. Below the table, an input panel allows users to select a model and adjust parameters such as daily screen time, social media time, gaming time, and number of apps used for making predictions.


Figure 4: Model prediction shown in bar chart

The "Predictions by Model" chart gives a basic view of how each model separates addicted and non-addicted users, but relying only on counts limits the analysis. Incorporating metrics like accuracy, precision, recall, F1-score, and AUC-ROC offers a more complete picture of performance. These measures highlight the balance between detecting addicted cases and minimizing false predictions. Using them for comparison strengthens the credibility and depth


Figure 5: Smartphone addiction predictor It provides a concise overview of the project:

- **Models Used**: The system leverages three machine learning algorithms — Random Forest, Logistic Regression, and Support Vector Machine (SVM)
— to predict smartphone addiction.
- **Features Considered**: The prediction is based on user behavior factors such as screen time, social media usage, gaming, and number of apps used.
- **Dashboard Tools**: The interactive dashboard in built using Streamlit for the web interface and Altair for visualization.

## 5. Conclusion

A machine learning-based framework for the prediction of smartphone addiction by integrating behavioral, demographic, and psychological features. The proposed architecture systematically addressed data collection, preprocessing, feature engineering, model development, and evaluation, with a focus on transparency through explainable AI. Experimental results demonstrate that machine learning models, particularly ensemble-based and deep learning methods, hold significant promise in accurately identifying individuals at risk of smartphone addiction.

The findings suggest that predictive modeling can serve as a valuable tool for early detection, thereby enabling timely interventions by healthcare professionals, educators, and policymakers. Moreover, the inclusion of explainability ensures that model outcomes are interpretable and can be trusted in real-world applications.

Future work will focus on extending the dataset with larger and more diverse populations, incorporating longitudinal behavioral patterns, and exploring advanced deep learning architectures to further improve prediction accuracy. Additionally, the integration of real-time monitoring systems with predictive analytics could open pathways toward personalized digital wellness solutions.

**References**

1. Y. Lee, L. Alzamil, B. Doskenov, and A. Termehchy, "A survey on data cleaning methods for improved machine learning model performance," *arXiv preprint arXiv:2109.07127*, 2021.
2. A. Mumuni and F. Mumuni, "Automated data processing and feature engineering for deep learning and big data applications: A survey," *arXiv preprint arXiv:2403.11395*, 2024.
3. B. Schoenfeld, C. Giraud-Carrier, M. Poggemann, J. Christensen, and K. Seppi, "Preprocessor selection for machine learning pipelines," *arXiv preprint arXiv:1810.09942*, 2018.
4. J. Burdack, F. Horst, S. Giesselbach, I. Hassan, S. Daffner, and W. Schöllhorn, "Systematic comparison of the influence of different data preprocessing methods on the performance of gait classifications using machine learning," *arXiv preprint arXiv:1911.04335*, 2019.
5. J. Osorio, M. Figueroa, and L. Wong, "Predicting smartphone addiction in teenagers: An integrative model incorporating machine learning and big five personality traits," *Journal of Computer Science*, vol. 20, no. 2, pp. 181–190, 2024.
6. A. Arora, P. Chakraborty, M. P. S. Bhatia, and A. Puri, "Intelligent model for smartphone addiction assessment in university students using android application and smartphone addiction scale," *Int. J. Educ. Manage. Eng.*, vol. 13, no. 1, pp. 29–34, 2023.
7. A. Alotaibi and A. Riasat, "Smartphones dependency risk analysis using machine-learning predictive models," *J. Healthcare Eng.*, vol. 2022, Article ID 8494563, 2022.
8. H. Kim and J. Kim, "Prediction of problematic smartphone use: A machine learning approach," *Psychiatry Investigation*, vol. 18, no. 7, pp. 655–662, 2021.
9. R. Raj, V. S. Kumar, and A. Sharma, "Machine learning models for smartphone addiction prediction: A comparative study," *Int. J. Data Sci. Mach. Learn.*, vol. 5, no. 4, pp. 112–120, 2023.
10. V. Vimala, K. S. Reddy, and P. Thomas, "Analyzing academic impact of smartphone addiction using gradient boosting models," in *Proc. Int. Conf. Next Generation Technologies (ICNGT)*, pp. 210–218, 2025.
11. M. Kadirvelu, R. Banerjee, and T. Yoon, "Digital phenotyping with contrastive learning for mental health prediction in adolescents," *IEEE Access*, vol. 12, pp. 11432–11445, 2024.
12. N. Orzikulova, L. Li, and S. Cho, "Time2Stop: An adaptive explainable AI system for smartphone overuse intervention," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.*, pp. 1–12, 2024.