International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

Early Detection of High-Risk Smartphone Addiction Through Behavioural Data

Kamalesh R[1], Kavin Joswa S[2], Govardhini S[3]

[1,2,3]*Department of Computer Science with Data Analytics, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India*

***Abstract*— Smartphone addiction has become a growing concern among students and young adults. This project presents a Smartphone Addiction Prediction System using Machine Learning techniques. The system analyzes behavioral factors such as daily usage time, sleep duration, screen unlock frequency, anxiety score, depression score, nomophobia score, and self-control score. A Random Forest Classifier is used to predict addiction levels categorized as No Addiction, Mild Addiction, or High Addiction. The system provides interactive visualizations and clear explanations for each prediction. Results show that the model achieves high accuracy and reliable performance in identifying addiction levels.**

***Keywords*—Addiction Prediction, Machine Learning, Random Forest, Behavioral Analysis, Smartphone Usage**

I. Introduction

Smartphones are widely used for communication, education, and entertainment. However, excessive usage can lead to smartphone addiction, affecting mental health and academic performance. Early detection of addiction patterns can help in preventive measures.This project develops a machine learning-based system to analyze behavioral data and predict smartphone addiction levels. The system provides not only predictions but also visual insights and explanation reports to improve understanding of user behavior.

II. System Architecture

The system is divided into multiple layers for efficient processing and prediction.

A. User Interface Layer

The web-based dashboard allows users to upload CSV datasets, view prediction results, analyze charts, and download reports. It provides an intuitive and user-friendly environment.

B. Input Processing Layer

This layer validates the uploaded CSV file, cleans column names, converts categorical values into numerical format, and handles missing or inconsistent data.

C. Feature Engineering Layer

Behavioral and psychological features are extracted, including usage time, sleep score, anxiety score, depression score, nomophobia score, and self-control score. Data normalization techniques are applied to improve model performance.

D. Machine Learning Layer

The Random Forest Classifier is used for prediction. The dataset is divided into training and testing sets. The model classifies users into No Addiction, Mild Addiction, or High Addiction categories.

E. Explanation Engine

The explanation module generates human-readable reasons for each prediction. It highlights major influencing factors such as high screen time, poor sleep duration, or low self-control.

F. Storage and Visualization Layer

Prediction results are stored in an SQLite database. The system provides interactive dashboards, including addiction distribution charts, usage analysis graphs, and demographic insights. Reports can be downloaded in CSV format.

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

III. Methodology

A. Dataset Description

The dataset consists of user behavioral and psychological attributes collected in CSV format. Each record represents one user and contains the following features:

- Daily Usage Time (hours)
- Screen Unlock Frequency
- Sleep Duration (hours)
- Anxiety Score
- Depression Score
- Nomophobia Score
- Self-Control Score
- Age
- Gender

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 03, March 2026)

The target variable is **Addiction Level**, classified into three categories:

- No Addiction
- Mild Addiction
- High Addiction

B. Data Preprocessing

Data preprocessing is an important step to ensure model accuracy and reliability.

The following preprocessing techniques were applied:

1. Removal of missing and inconsistent values.
2. Standardization of column names.
3. Conversion of categorical variables (e.g., Gender) into numerical format using encoding techniques.
4. Feature scaling using normalization to ensure uniform contribution of all features.

These steps reduce noise and improve classification performance.

C. Feature Engineering

Feature engineering was performed to enhance predictive capability. Behavioral indicators were carefully analyzed and transformed into numerical scores.

- Sleep Score derived from sleep duration.
- Psychological Risk Score computed from anxiety and depression values.
- Digital Dependency Score derived from usage hours and unlock frequency.

Normalization was applied to avoid bias toward higher-value features.

D. Machine Learning Model

The system uses the **Random Forest Classifier**, an ensemble learning method that builds multiple decision trees and combines their outputs for improved accuracy and reduced overfitting.

Model Training Process:

- Dataset split into 80% training and 20% testing data.
- Random Forest with multiple estimators used for classification.
- Gini Index used as splitting criterion.

Random Forest was selected due to:

- High accuracy
- Ability to handle non-linear relationships
- Robustness against overfitting
- Feature importance estimation capability

IV. Results And Performance

The performance of the proposed model was evaluated using standard classification metrics.

A. Evaluation Metrics

The following metrics were used:

- **Accuracy** – Overall correctness of predictions
- **Precision** – Correct positive predictions
- **Recall (Sensitivity)** – Ability to identify actual addiction cases
- **F1-Score** – Balance between precision and recall
- **ROC-AUC Score** – Model's discrimination ability

TABLE 1
Performance Evaluation of Random Forest Model

Metric	Score	Interpretation
Accuracy	87.5%	Correct predictions out of total cases
Precision	85.2%	Correctly predicted addiction cases
Recall	88.9%	Successfully identified actual addiction cases
F1-Score	0.87	Balanced precision and recall
ROC-AUC	0.91	Strong classification capability

C. Confusion Matrix Analysis

The confusion matrix shows that the model effectively distinguishes between No Addiction, Mild Addiction, and High Addiction categories. Misclassification is minimal, primarily occurring between Mild and High categories due to overlapping behavioral characteristics.

D. Feature Importance

Random Forest provides feature importance ranking. The most influential features identified were:

1. Daily Usage Time
2. Sleep Duration
3. Self-Control Score
4. Anxiety Score

This indicates that behavioral and psychological factors strongly impact addiction prediction.

E. Visualization Insights

The system generates interactive dashboards including:

- Addiction Distribution Pie Chart
- Average Usage Bar Graph
- Demographic Analysis Graph

These visualizations provide meaningful insights into user behavior patterns.

V. CONCLUSION

The proposed Smartphone Addiction Prediction System demonstrates strong predictive performance using behavioral and psychological attributes.

The Random Forest model achieved high accuracy and reliable classification results. The inclusion of feature importance analysis and explainable insights enhances system transparency.

Future enhancements may include real-time mobile monitoring, deep learning integration, and larger-scale dataset evaluation.

REFERENCES

[1] Kwon, M., Kim, D. J., Cho, H., and Yang, S., "The Smartphone Addiction Scale: Development and Validation of a Short Version for Adolescents," Computers in Human Behavior, vol. 31, pp. 374–378, 2013.

[2] Demirci, K., Akgönül, M., and Akpinar, A., "Relationship of Smartphone Use Severity with Sleep Quality, Depression, and Anxiety in University Students," Journal of Behavioral Addictions, vol. 4, no. 2, pp. 85–92, 2015.

[3] Breiman, L., "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[4] Forman, G., "An Extensive Empirical Study of Feature Selection Metrics for Text Classification," Journal of Machine Learning Research, vol. 3, pp. 1289–1305, 2003.

[5] Han, J., Kamber, M., and Pei, J., Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011.

[6] Pedregosa, F., et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[7] Kuhn, M., and Johnson, K., Applied Predictive Modeling, Springer, 2013.