IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 6, Issue 11, April 2026

Impact Factor: 8.2

Predictive Model for Detecting Excessive Screen Time Using Machine Learning

A. Manivarsh, P. Manish, C. Gnandev

Department of CSE

Sreenidhi Institute of Science and Technology, Telangana, India

manivarsh33@gmail.com, manish.pabba21@gmail.com, chinthagnandev@gmail.com

Abstract: *The rapid increase in smartphone usage has raised significant concerns regarding excessive screen exposure and its impact on user well-being. This study presents a machine learning-driven web-based system for predicting daily screen-on time using user demographics and device-related parameters. The proposed system integrates multiple regression models, including Linear Regression, Random Forest, and Gradient Boosting, to estimate screen usage patterns with high accuracy. Among these, the Random Forest model demonstrates superior performance with an R^2 score of 0.934, outperforming other models in terms of mean squared error and mean absolute error. The system incorporates a complete pipeline consisting of data preprocessing, categorical encoding, model inference, and performance evaluation. Additionally, it employs explainable artificial intelligence techniques using SHAP (SHapley Additive Explanations) to provide feature-level interpretability for individual predictions, enhancing transparency and trust in the model. A risk assessment mechanism is further integrated to classify predicted screen time into low, moderate, and high-risk categories, accompanied by a normalized risk score and user-centric recommendations. The application is deployed through a web framework, enabling real-time user interaction, data visualization, and prediction generation. Visual analytics, including distribution plots and feature relationships, support better understanding of usage trends. The proposed system demonstrates high predictive capability and practical applicability in promoting healthier digital habits by offering interpretable insights and actionable feedback*

Keywords: Machine Learning, Screen Time Prediction, Random Forest, Explainable AI, SHAP, Regression Models, Risk Analysis, Data Visualization, User Behavior Analysis, Web Application

I. INTRODUCTION

The exponential increase in the use of smartphones has led to a substantial change in human behaviour on a daily basis, particularly among the younger population groups, therefore there has been an increase in concerns about excessive exposure to screens and its potential health consequences. Recent research stresses that prolonged interaction with digital devices is highly linked with the changes in behavior, low physical activities as well as emerging health risks like obesity and mental health disorders [1][2]. As mobile usage becomes prevalent, people are leaving a huge amount of behavioral data behind and it can be used to gain insight into how they are using the device and how to anticipate future screen-based behavior. Traditional statistical techniques can often not capture the complexity of such patterns and can therefore lead to the use of advanced machine learning techniques. The power of machine learning models in forecasting behavioral and health-related outcomes such as gadget addiction and mental health risks using multidimensional data sets has been shown in research [3][5]. Furthermore, the joint training of interpretable machine learning approaches can help to better understand the contribution of features, which can improve model transparency and reliability. In this regard, the prediction of screen-on time is of paramount importance regarding promotion of digital well-being and the effectiveness of early intervention approaches. By using machine learning methods, the

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

scenario opens up the chance of creating intelligent systems that can precisely predict the screen time of users and offer tangible insights, which could aid in creating healthier digital lifestyles and decision-making processes.

The use of machine learning in predictive analytics has proven to be a great success in a variety of areas such as healthcare, environmental monitoring, and behavioral analysis. Ensemble learning techniques such as Random Forest and Gradient Boosting have become popular mostly because of their robustness, high accuracy and ability to work on complex, non-linear relationships present in data [4][6]. These models are better than the traditional models by including multiple learners to decrease variance and bias, improving the predictive performance. Furthermore, the advancement of deep learning and AutoML frameworks has also driven the improvement in the capabilities of automatically selecting and optimizing models, which in turn makes predictive systems more efficient and scalable [8][9]. In parallel, explainable artificial intelligence (XAI) methods such as SHAP have become prominent because of their ability to provide interpretations of model predictions at the global level as well as local level. This is especially important in applications involving behavior and health, where the influence of individual features on the user is important for building trust and adoption. Prior researches have successfully applied explainable models to predict conditions such as myopia and chronic diseases, showing the importance of explainability of machine learning systems [4][6]. These developments collectively suggest that the combination of powerful predictive models and interpretability mechanisms can contribute to more reliable and user-centric solutions.

In addition to the question of accurate prediction, the practical deployment of machine learning systems also requires the integration of such systems with user-friendly platforms that enable real-time interaction as well as decision support. Web-based applications have become a viable medium to develop predictive analytics solutions that can share data, visualize findings and provide individualized feedback to users. Recent advances in digital phenotyping have shown the potential of passive data harvesting from smartphones into digital data [10] in order to predict behavioral and psychological outcomes by using real-time monitoring systems. Similarly, machine learning-based frameworks for gadget addiction prediction illustrate the importance of converting the insights from the predictive model into some recommendations for the user's [3]. The introduction of risk assessment mechanisms into these systems further improves their usefulness in classifying users into various risk levels and in giving tailored guidance on this. This approach is congruent with the wider goal of fostering digital wellbeing through informed awareness and behavioral change. Therefore, machine learning models, explainability methods, and interactive web interfaces, as a comprehensive solution for screen time prediction and analysis, are formed. Such systems not only drive increases in the accuracy of the prediction but also give the users interpretable information and meaningful recommendations, which adds to healthier and more balanced digital participation.

II. LTERATURE SURVEY

A. Machine Learning in Health and Behavior Prediction

Machine learning has emerged as a powerful tool for predicting health and behavioral outcomes by leveraging large-scale datasets and identifying hidden patterns. Several studies have demonstrated its effectiveness in predicting conditions such as obesity, mental health disorders, and behavioral risks. For instance, interpretable machine learning models have been used to assess childhood and adolescent obesity with high accuracy, highlighting the role of predictive analytics in preventive healthcare [1][2]. Similarly, machine learning approaches have been applied to analyze behavioral trends and identify risk factors associated with lifestyle habits. These models enable early detection and intervention, making them valuable in both clinical and non-clinical settings for improving overall well-being and decision-making processes [5].

B. Prediction of Digital Behavior and Gadget Usage

The rapid increase in smartphone and digital device usage has led researchers to focus on predicting user behavior and identifying patterns of excessive usage. Machine learning techniques have been widely adopted to analyze digital interactions and detect gadget addiction or abnormal usage trends. Studies have shown that predictive models can effectively estimate user engagement levels and classify individuals based on their digital habits [3]. Furthermore, digital phenotyping approaches utilize passive smartphone data to monitor user activity and predict psychological conditions, demonstrating the potential of continuous behavioral tracking [10]. These

DOI: 10.48175/IJARSCT-36620

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 11, April 2026 Impact Factor: 8.2

advancements highlight the importance of predictive systems in understanding digital behavior and promoting responsible technology usage among different user groups.

C. Role of Ensemble and Advanced Learning Models

Ensemble learning techniques, such as Random Forest and Gradient Boosting, have gained significant attention due to their superior predictive performance and robustness. These models combine multiple learning algorithms to reduce errors and improve generalization, making them highly effective for complex prediction tasks. Research has shown that ensemble models outperform traditional regression techniques in various domains, including health prediction and behavioral analysis [4][6]. Additionally, advancements in deep learning and hybrid models have further enhanced prediction capabilities by capturing intricate relationships within data. The adoption of such advanced models enables more accurate and reliable predictions, thereby improving the effectiveness of machine learning systems in real-world applications [2][9].

D. Explainable Artificial Intelligence in Prediction Systems

As machine learning models become more complex, the need for interpretability has become increasingly important. Explainable Artificial Intelligence (XAI) techniques, such as SHAP, provide insights into how individual features influence model predictions, thereby enhancing transparency and trust. Research studies have emphasized the importance of interpretability in applications involving human behavior and health, where understanding model decisions is critical [5]. Explainable models have been successfully applied in predicting medical conditions and behavioral outcomes, allowing users and practitioners to make informed decisions based on clear insights [4][6]. The integration of XAI ensures that predictive systems are not only accurate but also understandable and accountable.

E. Integration of Machine Learning with Real-Time Systems

The deployment of machine learning models in real-time applications has significantly improved their practical usability and impact. Web-based systems and interactive platforms enable users to input data, receive predictions, and visualize results instantly. Research has highlighted the effectiveness of integrating predictive models with user-centric interfaces to enhance accessibility and usability [8]. Additionally, real-time monitoring systems that utilize continuous data streams have been developed to track behavioral patterns and provide timely interventions [10]. These systems often incorporate risk assessment mechanisms to classify users and offer personalized recommendations. Such integration bridges the gap between theoretical models and real-world applications, making machine learning solutions more impactful and user-friendly [3].

III. PROPOSED METHODOLOGY

A. Data Acquisition and the Selection of Features

The proposed methodology starts with structured collection of data with user demographics and aspects related to a device. Some key input features are device model, operating system, battery consumption, number of installed applications, daily data usage, age and gender. These features are chosen on the basis of their direct or indirect influence on the behavior of screen use. The target variable is defined as the amount of time spent daily behind the screen measured in hours. The dataset is arranged into tabular format so that supervised learning is supported. Feature relevance is guaranteed by taking into account behavioral and usage-driven parameters for the model to learn meaningful relationship and generate relevant predictions for screen time estimation.

B. Processing and Encoding Data

Data preprocessing is done to obtain consistency, quality and compatibility to machine learning models. Categorical features such as device model, operating system and gender are converted using label encoding techniques to numerical representations. There are data cleaning processes such as white space, standardization, and possible inconsistencies of data for a categorical value. The data is then split into training and testing data sets using an 80:20 ratio in order to effectively evaluate datasets. All the processed features are then converted to numbers which are then used to detect compatibility with regression algorithms and provide a reliable input pipeline for further model training and prediction tasks.

C. Model Selection and Training

Multiple regression models are used to determine the best method of predicting screen time. The models are Linear Regression, Random Forest, and Gradient Boosting. Each model is processed with the training data set for learning the correlation between input features and screen-on time. Ensemble methods like Random Forest and Gradient Boosting are majorly focused upon because of

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

their capacity to catch the non-linear relationship and lower the error in the prediction. Model performance is assessed by using statistical measures, which enables one model to be compared to another. Best performing model is chosen depending upon the predictive accuracy and error minimization of the model to be deployed into the system.

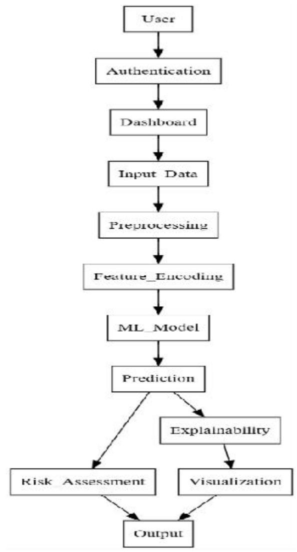

Fig 1: System Architecture

D. Prediction and Risk Assessment mechanism

The chosen model is used to make predictions based on input data supplied by the user using an interactive user interface. The predicted screen time is further analysed to determine the classification of user behavior. A risk assessment mechanism is used to convert the predicted value to normalized score between 0 and 100. Based on pre-set levels, the users are placed in the low, moderate or high-risk category. This classification can be used to obtain meaningful interpretation of the raw predictions as well as increasing usability. Additionally, personalized feedback is also generated, in the form of descriptive messages, which helps users to understand their digital habits and encourages them to use their screens in a healthier manner.

E. Interpretability of the System & Features

To bring more transparency and trust to the prediction system, explainable artificial intelligence techniques are implemented. SHAP (SHapley Additive Explanations) is used for analysis to determine the contribution of each feature towards the predicted output. This approach offers the interpretation (on a global but also local scale) by quantifying feature importance values. A visual representation is produced to focus on the impact of individual features relevant for the prediction of screen time. This allows users to learn what factors have the greatest influence their behavior when using a product. The incorporation of explainability guarantees that the system is meticulously not just correct but interpretable for the people supporting and able to make informed decisions while boosting the confidence of the users.

F. System Integration and Web Deployments

The whole system is applied in the form of a web-based application for ease of real-time user interaction and accessibility. The application has modules for the user authentication, data input, model inference, and results visualization. Backend processing is processed using light weight framework that connects trained model with the preprocessing step and the prediction pipeline. Components of visualization are provided to present data trends and feature relationships. The system architecture ensures free communication between the user inputs and prediction outputs. This deployment approach allows one to practically use the model for immediate prediction and insights within an intuitive and interactive interface.

Algorithm Used

Step 1: Firing up the system and set up the web application environment and the needed libraries and configurations.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 11, April 2026 Impact Factor: 8.2

Step 2: Load the previously trained machine learning model with corresponding label encoders into the memory for prediction purposes.
Step 3: Make a connection to the database and allow for user authentication (registration and login ability).
Step 4: Loading the dataset and seeing some basic information i.e. Number of records, attributes, sample preview.
Step 5: Perform data preprocessing i.e. splitting of input features and target variable (screen-on time).
Step 6: Encoding categorical variables With the help of the label encoding, all those categorical variables should be converted into the numerical format which can be easily processed by the model.
Step 7: Categorize the dataset into subsets for training and testing the model to assess the model's performance.
Step 8: Accept user input using the interface to convert the input to structured form for prediction.
Step 9: Pass processed input data into the trained model to derive predicted value of screen time and then calculate risk scores and classify
Step 10:Generate output results such as predicted value, risk level and feature importance visualization and display them to the user.

IV. RESULTS AND DISCUSSION

A. Overview of Model Performance

The designed system tests various regression models in order to identify the most appropriate methodology while detecting daily screen-on time. The performance comparison suggests that the ensemble-based approaches are having better performance than the case of the traditional/linear approaches because they were able to capture the complex and non-linear relationships in the dataset. Out of the tested models, Random Forest proves to have the best predictive power and accuracy with less error rates. The results show the importance of model selection when making reliable predictions. The system essentially helps to determine the best performing model in order to ensure that the final deployment is based on the final algorithm that performs best and is robust for real-world use scenarios.

B. Evaluation Metrics Analysis

The performance of the models is evaluated using standard performance evaluation techniques, such as R2 score, Mean Squared Error (MSE), and the Mean Absolute Error (MAE). R2 score gives an idea to how good the model has captured the variance in the dependent variable, MSE and MAE gives an idea of prediction errors. The values of R2 are higher, and the error measures are lower for better performance. The results of the analysis are that all models have achieved high R2 scores higher than 0.91, which means high predictive ability. However, variations in the values of error shows differences in terms of model efficiency; therefore, it is imperative to use multiple metrics to get a comprehensive evaluation on the performance of models.

Evaluation Metrics

Table 1: Model Evaluation Metrics

Model	R^2 Score	MSE	MAE
Linear Regression	0.919	0.725	0.666
Random Forest	0.934	0.593	0.609
Gradient Boosting	0.926	0.668	0.620

C. Comparative Analysis of Performance

A comparative analysis between the models shows a huge improvement in the performance of the algorithms that are ensemble techniques compared to the linear regression model. Random Forest gives highest R2 value and least error value thus making it most reliable model for prediction. Gradient Boosting also has a good performance but performs slightly less when compared to Random Forest as far as reduction in error is concerned. Linear Regression, though simple and efficient, doesn't perceive complex relations present in the data thus producing comparatively higher errors. This comparison proves that ensemble learning techniques are thus more suited in behavioral prediction tasks that deal with multiple interacting features and non-linear dependencies.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 11, April 2026
Impact Factor: 8.2

D. Accuracy and Reliability of Prediction

The accuracy of the predictions made by the system is verified through test data that was not seen, while also making sure that the model generalizes not only in the set of data it was trained on. The low Mean Absolute Error shows that the values predicted for screen time are close to the actual values, with a low level of deviation. This accuracy enables a degree of reliability when application to real-world scenarios where users expect consistent and meaningful predictions. The high R2 score further confirms that the model is good in capturing the underlying patterns in the data. These results show that the system can be trusted to give correct estimations of user screen time behaviour.

E. Risk Assessment and Classification and Result

The system goes beyond prediction in that its system incorporates a risk assessment mechanism which categorizes users in different 'levels' based on the predicted screen time. This classification is based on preset thresholds, which allows for the result to be interpreted in a significant way. The risk scoring mechanism pictures predictions in a normalized scale for easy understanding and comparing user behavior. The classification gives a translatable insight and the user can determine if the screen usage statistics are within acceptable limits. This extra layer makes the system more practical in its application by converting the raw predictions to become interpretable and useful for users.

Table 2: Criteria used for Risk Classification

Predicted Screen Time (hours/day)	**Risk Level**	**Score Range**
≤ 4	Low Risk	0 – 40
> 4 and ≤ 6	Moderate Risk	41 – 60
> 6	High Risk	61 – 100

F. Contribution to Feature and Interpretability

The combination of explainable artificial intelligence techniques allows to analyze in detail the contributions of the features to each prediction. Feature importance visualization shows the importance of individual attributes of the predicted screen time. This interpretability means the users understand the items which are driving their usage patterns, such as data consumption, number of applications or device characteristics. By giving clear explanations, the system helps to build user trust and facilitate informed decision-making. The ability to interpret predictions is especially important in behavioral analysis applications, where end-users want to understand how something is generated.

G. Discussion and Effectiveness of System

The overall results show that the proposed system combines well the accuracy for prediction with the interpretability and usability. The approach of ensemble learning for ensuring high performance and combining risk assessment with explainability lends themselves to high practical value. The system manages to advance the divide between complex machine learning models and user-central applications. It offers robust predictions, meaningful insights, and actionable recommendations, thus suitable for real-world deployment. The findings in our study suggest that machine learning-based approaches may play a significant role in the promotion of more healthy digital habits by helping the user to effectively monitor and manage screen usage.

V. CONCLUSION AND FUTURE ENHANCEMENT

The proposed system presents a good usage of machine learning techniques in the prediction of daily screen on time through user and device related features. By using regression models like Linear Regression, Random Forest, and Gradient Boosting Models the system has huge prediction accuracy, with Random Forest becoming the best performed model. A combination of preprocessing, feature encoding, and model evaluation ensures a robust and reliable prediction pipeline. Furthermore, the addition of explainable artificial intelligence techniques improves transparency by giving them insights on feature contribution. The inclusion of a risk assessment mechanism allows researchers to go from making raw prediction to meaningful classification, allowing the user to better understand and manage their digital behavior. The deployment of the system by a web-based interface further enhances its accessibility and usability to be used for real-time applications. Overall, the system succeeds in its combination of accuracy, interpretability, and user interaction towards the overall goal of the promotion of healthier digital habits and informed decision-making.

There are further improvements that can be made to the system to further enhance its scalability, accuracy and applicability in the real world. One of the potential improvements is through integration of real-time data collection from mobile devices or wearable sensors in order to allow continuous monitoring, dynamic prediction. The addition of deep learning models and hybrid models could help improve performance by capturing more complex facial behavior patterns. Additionally, enhancing the explainability aspect through more robust interpretability methods is able to give more insights about model decisions. From the system standpoint, one can add secure authentication mechanisms and cloud-based deployment for improving scalability and data security. The application can also be added to have personalized recommendations, adaptive feedback systems, and longitudinal tracking of behaviors. These changes would make the system a more comprehensive digital well-being platform, with the ability to support proactive intervention and long-term behavioral improvement.

REFERENCES

[1] Zheng, F., Wang, X., Xue, M., Wang, Q., Zhang, W., Zhang, Z., & Niu, W. (2026). Interpretable Machine Learning Models for Childhood and Adolescent Obesity Prediction According to Chinese and WHO Standards: A Two-Wave Cross-Sectional Study.

[2] Kumar, S., & Kumar, A. (2025). Deephealthnet: Ai-driven adolescent obesity prediction using machine learning. Science and Technology, 7(02).

[3]Dharnasi, P. (2026). Electronic gadget addiction prediction using machine learning. International Journal of Research Publications in Engineering, Technology and Management (IJRPETM), 9(2), 500-505.

[4]Sifat, I. K., Jisa, T. A., Roy, J. S., Sultana, N., Hasan, F., Mosharaf, M. P., & Kibria, M. K. (2025). Prediction of Myopia Among Undergraduate Students Using Ensemble Machine Learning Techniques. Health Science Reports, 8(5), e70874.

[5]Arya, T. (2025). Digital Behavior and Mental Health Prediction Through Explainable AI.

[6]Pechprasarn, S., Wetchasit, P., &Pongsuwan, S. (2025). Optimizing chronic kidney disease prediction: A machine learning approach with minimal diagnostic predictors. Journal of Current Science and Technology, 15(1), 76-76.

[7]McNutt, A. T., Li, Y., Meli, R., Aggarwal, R., & Koes, D. R. (2025). GNINA 1.3: the next increment in molecular docking with deep learning. Journal of Cheminformatics, 17(1), 28.

[8]Saqr, M., Tlili, A., & López-Pernas, S. (2025). Automating individualized machine learning and AI prediction using AutoML: the case of idiographic predictions. In Advanced Learning Analytics Methods: AI, Precision and Complexity (pp. 565-593). Cham: Springer Nature Switzerland.

[9]Huan, J., Fan, Y., Xu, X., Zhou, L., Zhang, H., Zhang, C., ... & Gu, S. (2025). Deep learning model based on coupled SWAT and interpretable methods for water quality prediction under the influence of non-point source pollution. Computers and Electronics in Agriculture, 231, 109985.

[10]MALIK, A. N., TARIQ, A. A., NADEEM, W., & SHAFIQ, H. (2025). DIGITAL PHENOTYPING USING PASSIVE SMARTPHONE DATA FOR EARLY PREDICTION OF PSYCHOTIC EPISODES IN CLINICALLY HIGH-RISK ADOLESCENTS: A LONGITUDINAL MACHINE LEARNING APPROACH. TPM–Testing, Psychometrics, Methodology in Applied Psychology, 32(S6 (2025): Posted 15 September), 2352-2358