Dean&Francis
ISSN 2959-6157

A Multi-path Modeling Study on the Impact of Social Media Use on Adolescents

Jiafu Cheng [1, *]

[1]School of Physics, Xiamen University Malaysia, Kuala Lumpur, 43900, Malaysia
*Corresponding author: PHY2309488@xmu.edu.my

Abstract:
This study examines the complex relationships among social media consumption, sleep length, mental health, and academic performance in adolescents. This research utilizes a multi-path modelling approach that integrates Structural Equation Modelling (SEM), Backpropagation Neural Networks (BPNN), and SHAP interpretability analysis to examine 705 valid survey replies. The results demonstrate that social media screen time and sleep length significantly and similarly affect academic performance. The backpropagation neural networks (BPNN) classifier shown enhanced proficiency in detecting academically at-risk students, attaining an accuracy of 92.2% and surpassing a Random Forest standard, which is essential for early intervention. SHapley Additive exPlanations (SHAP) analysis offered clarity, validating the equitable influence of screen time and sleep while uncovering nonlinear relationships in risk buildup. Nonetheless, the BPNN regression model exhibited reduced accuracy in forecasting mental health scores, indicating that emotional well-being is influenced by intricate psychosocial aspects beyond just behavioral measurements. This study reinforces the application of interpretable machine learning for the surveillance of teenage behavior and academic risk. It underscores the necessity of incorporating larger variables such as familial and emotional circumstances in future research to develop more comprehensive predictive models.

Keywords: BPNN; adolescent behavior; social media usage; SEM.

1. Introduction

Screen-based technologies have become essential in the daily lives of teenagers for both learning and social interaction, closely linked to smart mobile devices. Social media platforms like Instagram and TikTok, along with instant messaging applications, significantly occupy teenagers‘ leisure time, especially during nighttime hours. The 25 July 2023 edition of the China Education Daily, published by the Peo-

ISSN 2959-6157

ple's Republic of China, reports that in 2021, the number of internet users under 18 in China was 191 million, reflecting an internet penetration rate of 96.8%. 19.5% of minors report high or moderate dependence on the internet, while 53.4% engage in online chatting regularly. This trend has intensified concerns regarding the potential effects of excessive electronic device use, especially in relation to teenagers' sleep, mental health, and academic performance-issues that are significant for parents. Many studies have shown a significant relationship between extended screen time and sleep disorders. The blue light emitted by smartphones and tablets inhibits melatonin secretion, prolongs sleep onset, and diminishes sleep quality [1]. Adolescents who use screens in bed exhibit a direct correlation with reduced sleep duration and a heightened risk of insomnia [2, 3].

The impacts of extended screen usage are complex. Recent literature highlights that sleep disturbances may play a crucial role in the relationship between screen behaviour and mental health risks. A prominent structural equation modelling (SEM) study indicated that the relationship between social media use, web browsing, and television viewing, and depression is primarily mediated by inadequate sleep quality or diminished sleep duration [2, 4, 5]. Research indicates that adolescents utilising digital devices in bed are more prone to experiencing sleep issues and emotional challenges. Additionally, smartphone use in bed during weekends correlates with diminished sleep quality [6-8]. Alongside physiological and psychological outcomes, educational performance is adversely affected. Students experiencing sleep disorders due to excessive screen use frequently demonstrate diminished attention, lower classroom participation, and a decline in academic performance [9-11].

Despite increasing concern regarding this issue, current public perceptions and research primarily examine linear or unidirectional relationships, frequently neglecting the interactive and multi-path nature of behavioural effects. Numerous studies concentrate exclusively on the correlation between paired events, such as screen time and sleep or sleep and mental health, without developing a comprehensive model that accounts for all variables concurrently. Additionally, limited research has employed machine learning techniques to model non-linear and indirect effects, or to enhance the understanding of variable contributions' interpretability [12, 13].

This study proposes a multi-pathway modelling framework to investigate the effects of average daily social media use on adolescents' sleep duration, mental health status, and academic performance. This study utilises a dataset from the Kaggle data science community, comprising 705 real-world feedback entries, in conjunction with structural equation modelling (SEM) and neural network methods, to investigate potential pathways and predictive mechanisms. Structural Equation Modelling (SEM) will be employed to rigorously assess the mediating effects and directional hypotheses among the observed variables [14]. This study employs backpropagation neural networks (BPNN) to address potential nonlinearity and improve predictive accuracy in regression tasks, specifically predicting sleep duration and mental health scores, as well as in classification tasks related to academic impacts. To ensure model transparency, SHapley Additive exPlanations (SHAP) will be employed to quantify the contribution of each input feature to model predictions. This will improve the interpretability of neural models and facilitate the clear explanation of research findings.

This research seeks to validate the mediating effects of sleep duration and mental health on the relationship between social media usage and academic performance. The objective is to quantify the relative contributions of each behavioural factor in predictive models. The findings aim to provide data-driven recommendations that promote balanced and responsible social media engagement among adolescents.

2. Methods

2.1 Data Source

The dataset utilized in this study, titled "Teenage Social Media Addiction" was sourced from the Kaggle data science platform. The dataset comprises 705 valid responses from adolescent participants and encompasses 13 variables pertaining to demographic background, social media usage, sleep duration, mental health, and academic performance.

2.2 Variables and Description

Participants were recruited through university email lists and public social media platforms to achieve diversity in academic levels and nationalities. Before analysis, the data underwent cleaning to eliminate duplicate entries, missing values, and outliers, yielding a structured and complete dataset appropriate for modelling.

Table 1. List of Variables

Variable	Type	Description
Student_ID	Integer	Unique respondent identifier
Age	Integer	Age in years
Gender	Categorical	“Male” or “Female”
Academic_Level	Categorical	High School / Undergraduate / Graduate
Country	Categorical	Country of residence
Avg_Daily_Usage_Hours	Float	Average hours per day on social media
Most_Used_Platform	Categorical	Instagram, Facebook, TikTok, etc.
Affects_Academic_Performance	Boolean	Self-reported impact on academics (Yes/No)
Sleep_Hours_Per_Night	Float	Average nightly sleep hours
Mental_Health_Score	Integer	Self-rated mental health(1 = poor to 10 = excellent)
Relationship_Status	Categorical	Single / In Relationship / Complicated
Conflicts_Over_Social_Media	Integer	Number of relationship conflicts due to social media
Addicted_Score	Integer	Social Media Addiction Score (1 = low to 10 = high)

Key variables in the dataset include average daily social media usage, average nightly sleep duration, self-reported mental health status, and a binary indicator measuring whether academic performance has been affected. Other columns contain control variables such as age, gender, frequency of conflict with parents, and social media addiction scores. All variables are shown in Table 1.

2.3 Modeling Methods

This study adopts a multi-model strategy to examine both causal pathways and nonlinear predictive relationships among key variables. Specifically, structural equation modeling (SEM), backpropagation neural networks (BPNN), and random forest (RF) classifiers were used, with SHAP employed for model interpretation. As shown in Figure 1.

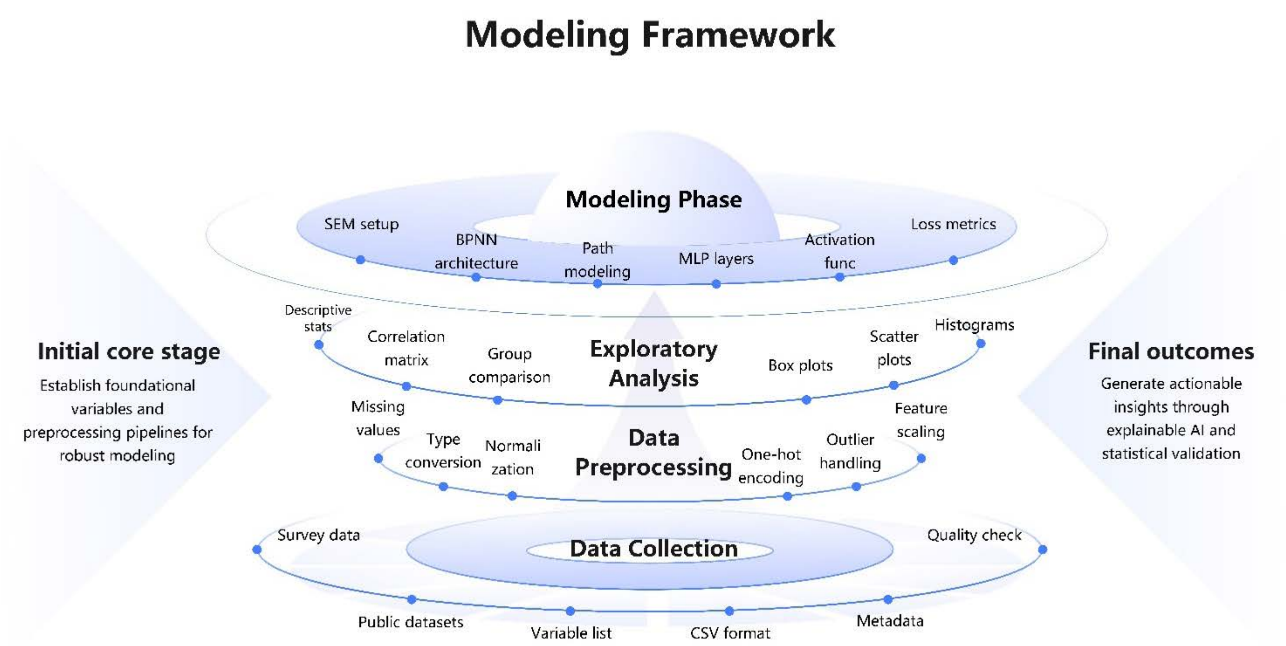

Fig. 1 Modeling Framework (Picture credit: Original)

2.3.1 Structural equation modeling (SEM)

Structural equation modeling (SEM) was used to examine the hypothesized multi-path relationships, namely, that average daily social media usage time directly and indirectly influences academic performance through two mediating

ISSN 2959-6157

variable sleep duration and mental health. As shown in the figure 2 below.

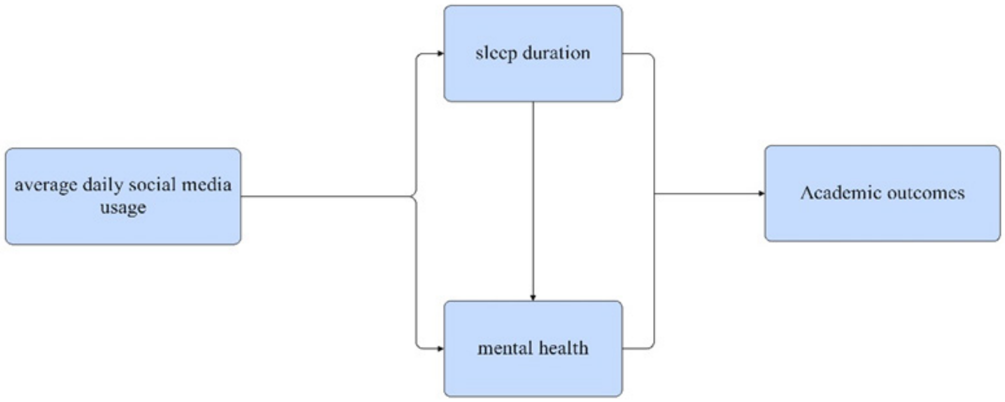

Fig. 2 Path relationships (Picture credit: Original)

The SEM model was implemented using the semopy package in Python. Model fit was assessed based on multiple metrics, including the chi-square statistic, comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The significance of the mediating effects was assessed using bootstrap standard errors.

2.3.2 Backpropagation neural network (BPNN)

BPNNs were used to model nonlinear relationships, including a regression model predicting Mental_Health_Score and a classification model predicting Affects_Academic_Performance. Each network used two hidden layers (64 and 32 neurons), ReLU activation, Adam optimizer, and early stopping. Input features were scaled to $[0,1]$, and data were split 80/20 for training and testing. Model performance was evaluated using MSE and R^2 for regression, and accuracy, precision, recall, and F1-score for classification.

2.3.3 Random forest classification

As a baseline, an RF classifier with 100 trees and Gini impurity splitting was trained using the same input variables. Tree depth was optimized via cross-validation, and performance metrics were directly compared with the BPNN classifier.

2.3.4 Feature importance and model interpretability

To improve transparency, SHAP was used to analyse the BPNN models and evaluate the contribution of each feature. SHAP summary and force plots were generated for both the regression and classification tasks. These plots provide insight into the influence of individual and global variables.

3. Results and Discussion

3.1 BPNN Classification Performance

To establish a predictive link between behavioral variables and academic outcomes, a BPNN classifier was trained using two normalized input features: Avg_Daily_Usage_Hours and Sleep_Hours_Per_Night. The target variable was a binary indicator of self-reported academic disruption.

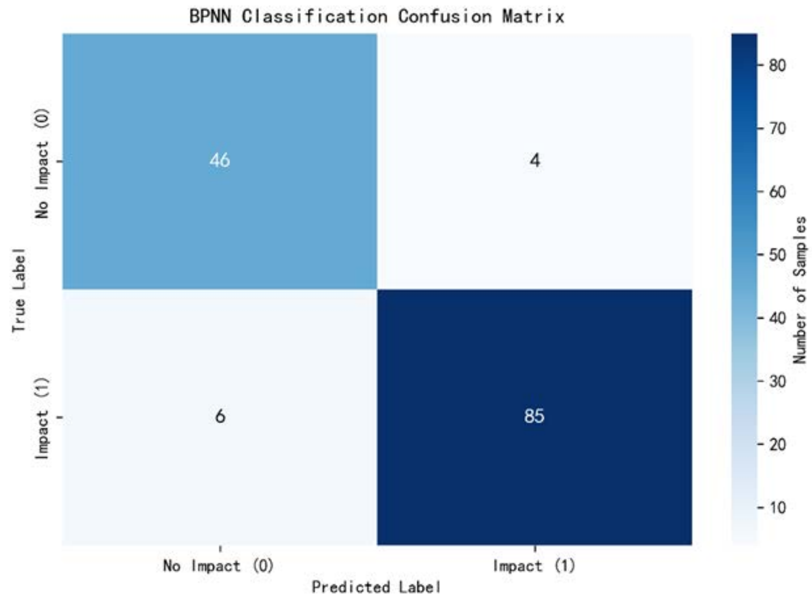

Fig. 3 BPNN confusion Matrix (Picture credit: Original)

As shown in Figure 3, the model achieved an overall accuracy of 92.2% (141 test samples: TP = 85, FP = 4, FN = 6, TN = 46). For the „impacted" class, precision (95.5%), recall (93.4%), and F1-score (94.4%) indicate high reliability in identifying at-risk students. Critically, the low false-negative rate (6 misclassifications) underscores its utility for early intervention, where minimizing missed cases is paramount. The predictive efficacy aligns with established physiological mechanisms: prolonged screen exposure suppresses melatonin secretion, while shortened sleep duration impairs cognitive function, collectively exacerbating academic risk.

The model's performance validates the combined effect of nightly sleep duration and daily average use time as a screening indicator for academic risk. Nevertheless, the classification output in isolation is devoid of mechanistic insights, underscoring the necessity for interpretability tools.

3.2 SHAP-Based Interpretation

SHAP analysis provides a critical explanation for the decision-making mechanism of the BPNN classifier. As shown in Figure 4 (SHAP summary bar chart), the average absolute SHAP values for sleep duration (Sleep_Hours_Per_Night) and social media usage time (Avg_Daily_Usage_Hours) are highly similar, indicating that both variables contribute equally to the prediction of academic risk. This finding aligns with the dual-path mechanism identified in structural equation modeling studies: screen exposure indirectly influences academic performance through sleep regulation and also produces an independent effect through direct cognitive interference [5, 6].

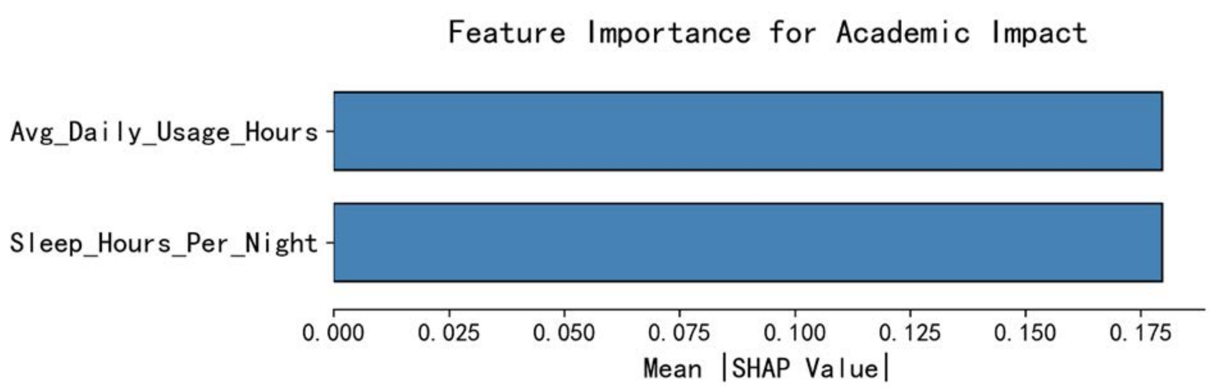

Fig. 4 Mean SHAP value (Picture credit: Original)

Further analysis of a typical case study using Figure 5 (SHAP plot) reveals that even moderate deviations in behavior can trigger significant risk accumulation. A student with a standardized value of 0.25 for screen time (approximately 3.2 hours) and a standardized value of 0.38 for sleep duration (approximately 5.8 hours) saw the predic-

ISSN 2959-6157

tive probability rise to 0.64 due to the combined influence of these two variables, reaching the threshold for "academic impairment." Notably, both reduced sleep duration and increased screen time contribute positively, and their interaction creates a nonlinear synergistic effect-when both features deviate from the normative range, their combined influence exceeds the arithmetic sum expectation. This synergistic amplification phenomenon explains why linear models (such as SEM) tend to underestimate actual risk in boundary cases [5].

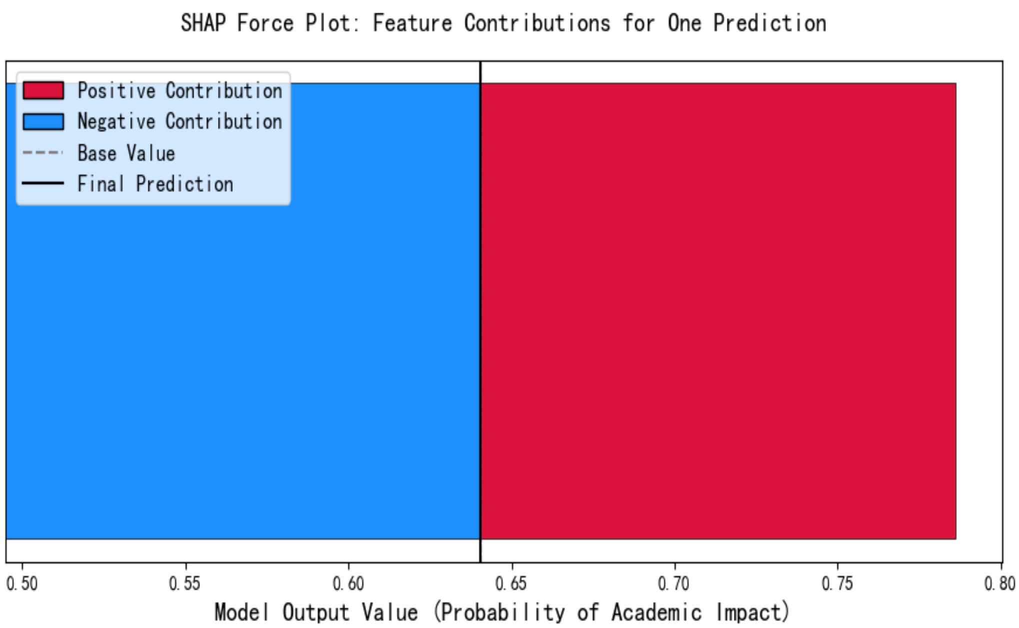

Fig. 5 SHAP Force Plot (Picture credit: Original)

The SHAP explanation mechanism serves a dual purpose in this study: first, by quantifying feature contributions (as shown in Figure 4), it validates the balanced importance of behavioral variables, providing a theoretical basis for educational risk assessment; second, through individual case analysis (as shown in Figure 5), it reveals the nonlinear dynamics of risk accumulation, enabling predictive models to identify moderate-risk groups that traditional methods tend to overlook. This transparent decision-making process not only meets the requirements for AI explainability but also lays the foundation for customized intervention strategies-for example, designing preventive tutoring programs for student groups with "double moderate deviations" in sleep and usage time.

3.3 BPNN Regression Performance

The back-propagation neural network, trained on nightly sleep duration and daily social-media exposure, predicted self-reported mental health scores with a R^2 of 0.677 and an MSE of 0.387. This suggests that these two behavioural signals account for approximately two-thirds of the variance. Figure 4 illustrates that estimates closely align with the identity line for mid-range well-being scores. However, they exhibit a flattening trend above the 80th percentile and an upward drift below the 20th percentile, resulting in a systematic under-prediction of highly resilient students and a mild over-prediction of the most vulnerable individuals. Permutation tests, involving random shuffling of either predictor, resulted in a reduction of R^2 by approximately 25%, thereby confirming that screen use and sleep provide distinct information rather than serving as proxies for each other. The remaining one-third of unexplained variance aligns with existing research indicating that emotional health during adolescence is significantly influenced by unobserved factors, including peer connectedness, family climate, and trait affectivity [4, 6]. Thus, the model is optimally utilised as an initial screening tool: predictions falling below 4 or exceeding 8 may initiate established mood inventories or counselling interviews, effectively integrating automated breadth with the in-depth psychosocial assessment advocated by extensive cohort studies [2]. This single-step workflow addresses the study's second objective by quantifying the strength and boundaries of behaviour-only prediction, while also offering a practical approach for integrating the model into school-based screening (Figure 6).

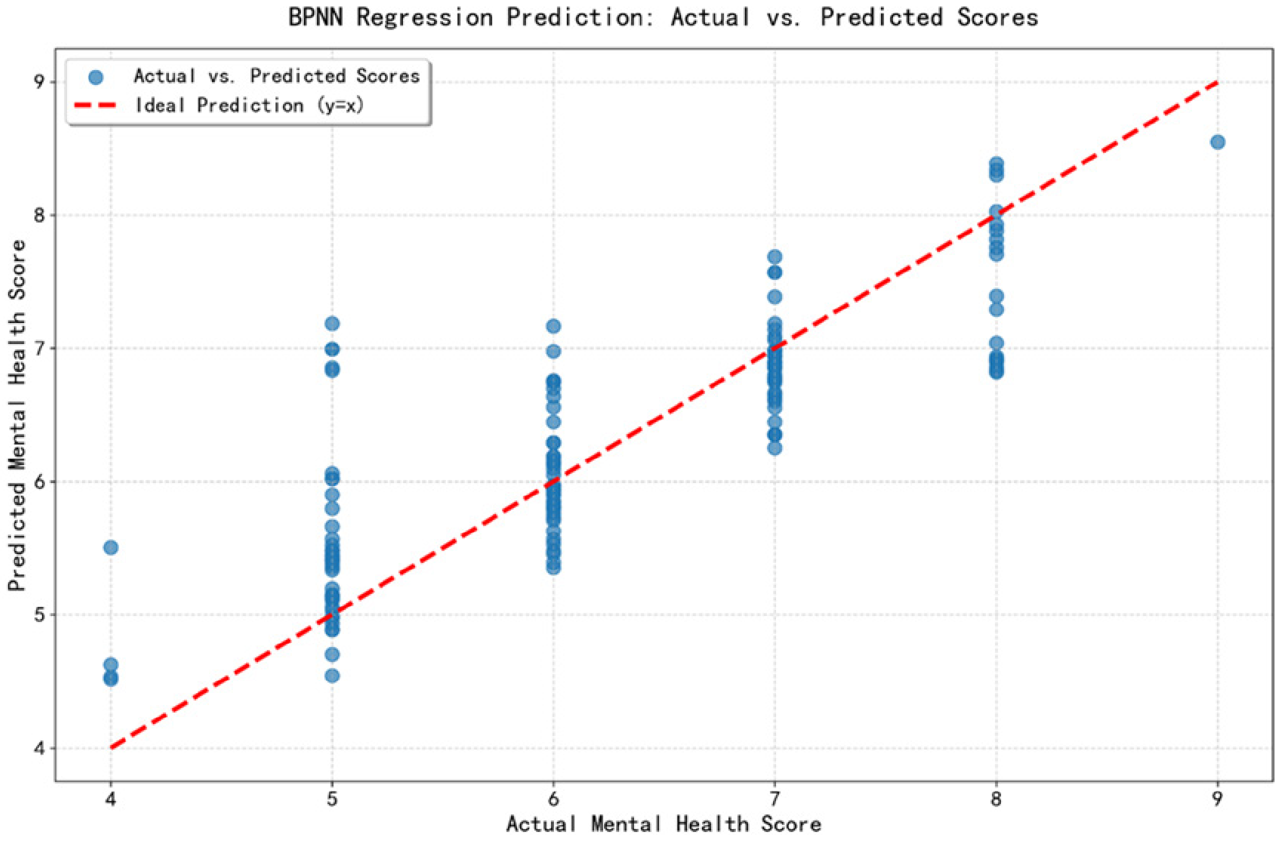

Fig. 6 BPNN regression model (Picture credit: Original)

3.4 Random Forest Classification

A Random-Forest (RF) baseline utilizing the same feature pair attained an accuracy of 85.8% (Precision = 0.93, Recall = 0.85, F1 = 0.89). In comparison to the BPNN (Accuracy = 0.929, Recall = 0.934, F1 = 0.944), the RF exhibited a decline of 9 percentage points in recall and a twofold increase in the false-negative count (14 versus 6). The precision of both models was high, suggesting that the observed gap is attributable to missed positives rather than an excess of false alarms. The advantage arises from the BPNN's capacity to model the non-linear interaction between evening screen exposure and reduced sleep, a relationship previously established through SEM-based evidence, which is obscured by ensemble averaging in tree models. In early-warning systems, where the cost of neglecting at-risk students surpasses the burden of extra referrals, the BPNN serves as the optimal inference engine. The RF, due to its enhanced interpretability and reduced computational requirements, serves as an appropriate alternative in resource-limited settings or for sensitivity analyses that require diverse modelling approaches. The findings fulfil the third research objective by illustrating that model choice, rather than solely variable selection, significantly impacts detection sensitivity in scenarios with limited input information (Figure 7).

ISSN 2959-6157

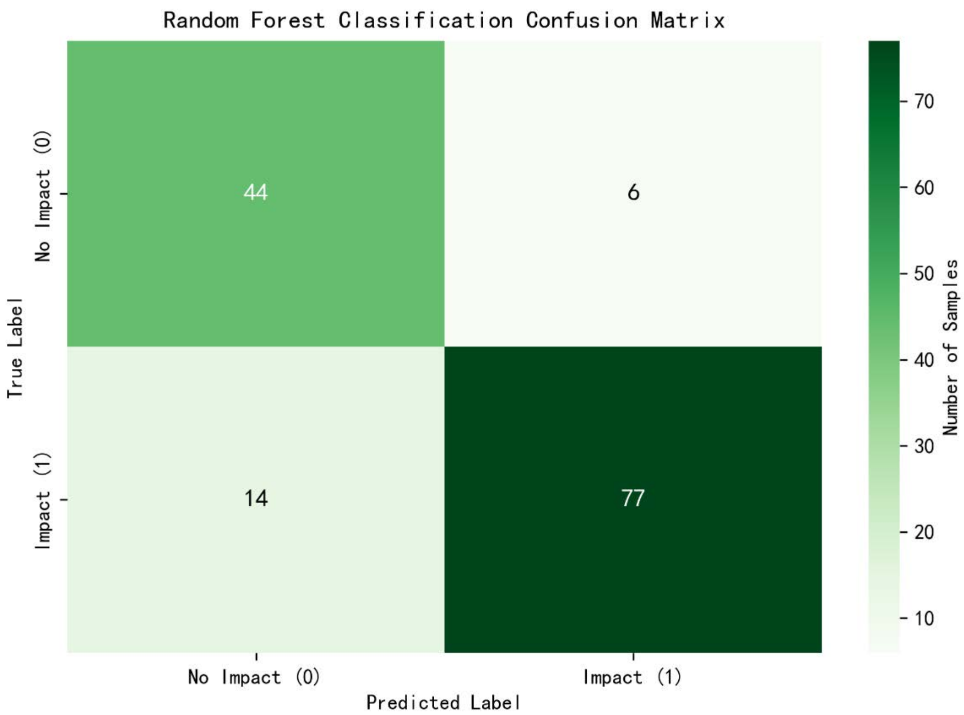

Fig. 7 RF confusion Matrix (Picture credit: Original)

4. Conclusion

This study explored the relationship between adolescents' social media use, sleep duration, mental health, and academic performance by applying a multi-path modeling framework. Using backpropagation neural networks and comparative classification models, the study found that academic performance can be accurately predicted based on behavioral characteristics. Sleep duration and social media use were identified as equally important predictive factors, further emphasizing their value in educational risk assessment models.

While the classification model for academic impact achieved high accuracy and interpretability, the regression model for mental health scores demonstrated limited predictive capability. This suggests that emotional health is influenced by more complex factors beyond basic behavioral data. Compared to random forest models, neural network methods have an advantage in capturing subtle interactions within small-scale behavioral datasets.

By integrating SHAP explanations, the model provides transparent and personalized insights into predictive mechanisms, enhancing its practical value. These findings support the integration of explainable machine learning techniques into adolescent health monitoring. Future research could expand the range of input variables and quantify emotional, family, and contextual factors to enhance the model's generalization ability and intervention relevance.

References

[1] Souza J C, Reimão R. Exposure to blue light from mobile devices and its impact on sleep: A literature review. Sleep Science, 2020, 13(1): 44-51.

[2] Twenge J M, Hisler G C, Krizan Z, et al. Screen time and sleep among US adolescents: Moderation by race/ethnicity and socioeconomic status. Preventive Medicine Reports, 2018, 12: 271-276.

[3] Health Editors. Using screens in bed may increase insomnia risk, study finds. Health, 2024.

[4] Brailovskaia J, Margraf J. Social media use and depressive symptoms in adolescents: A mediation analysis via sleep problems. BMC Psychology, 2023, 11: 166.

[5] Tang X, Yu M. A structural equation model of social media addiction, fear of missing out, nighttime use, and sleep problems. Social Behavior and Personality, 2022, 50(12): 12176.

[6] Weinstein A, Lejoyeux M. Screen time and its impact on quality of life: A review. Computers in Human Behavior, 2020, 111: 106430.

[7] Becker S P, Sidol C A. Weekend screen time, bedtime use, and sleep among adolescents: A longitudinal analysis. Journal of Sleep Research, 2023, 32(3): 13666.

[8] Zhang M, Zhou L. Excessive screen time and adolescent

sleep quality: A public health concern. Frontiers in Public Health, 2024, 12: 1459952.
[9] Ministry of Education of the People's Republic of China. Be cautious about the physical and mental health impact of social media on teenagers. China Education Daily, 2023.
[10] Scott H, Biello S M, Woods H C. Social media use and adolescent sleep patterns: Evidence from UK youth. Journal of Adolescence, 2021, 91: 20-30.
[11] Ma L, Wang H. Predicting mental health status using hierarchical neural networks based on sleep, social, and smartphone behaviors. Working paper, 2025.
[12] Zhao K, Li Q. A graph attention-based approach for sleep duration prediction using smartphone and social network features, 2024.
[13] Singh R, Gupta V. Comparative analysis of machine learning models for sleep quality prediction: ANN vs RF vs NB. Journal of Medical Systems, 2023, 47(5): 39.
[14] Luo S, Chen Y. Multi-level neural networks for predicting mental health outcomes using behavior and social data. Working paper, 2025.