Predicting and Analyzing Mental Health Severity from Digital and Lifestyle Habits

Deborah Shilpi Raghani, Hadassah Olive Chinnam, Joshitha Dandlamudi, Pravanith Reddy Kankanala, Raju Goud Burra

SAINT LOUIS UNIVERSITY — DEPARTMENT OF HEALTH AND CLINICAL OUTCOMES RESEARCH

Background

- With the increasing dependency on digital devices, screen time has become a key concern for mental well-being.[1]
- Research highlights a strong correlation between depression symptoms and excessive use of digital devices and social media.[2]
- Lifestyle and demographic variables like age, gender, and physical activity levels are found to further influence how screen time impacts mental well-being.[2]
- Adolescents are found to be more vulnerable with screen usage leading to loneliness and reduced well-being.[3]

Goals of study:

- To Predict mental health status using machine learning models based on screen time and behavioral variables
- To Identify which screen-based behaviors and lifestyle factors are strongest predictors of poor mental health outcomes

Methods

Data Source:

- Impact of Screen Time on Mental Health Dataset from Kaggle (Khushi Kyad, 2024)[4].
- Approximately 2,000 records of individuals aged 13–64, based on a recent cross-sectional survey (2024–2025).

Data Preparation:

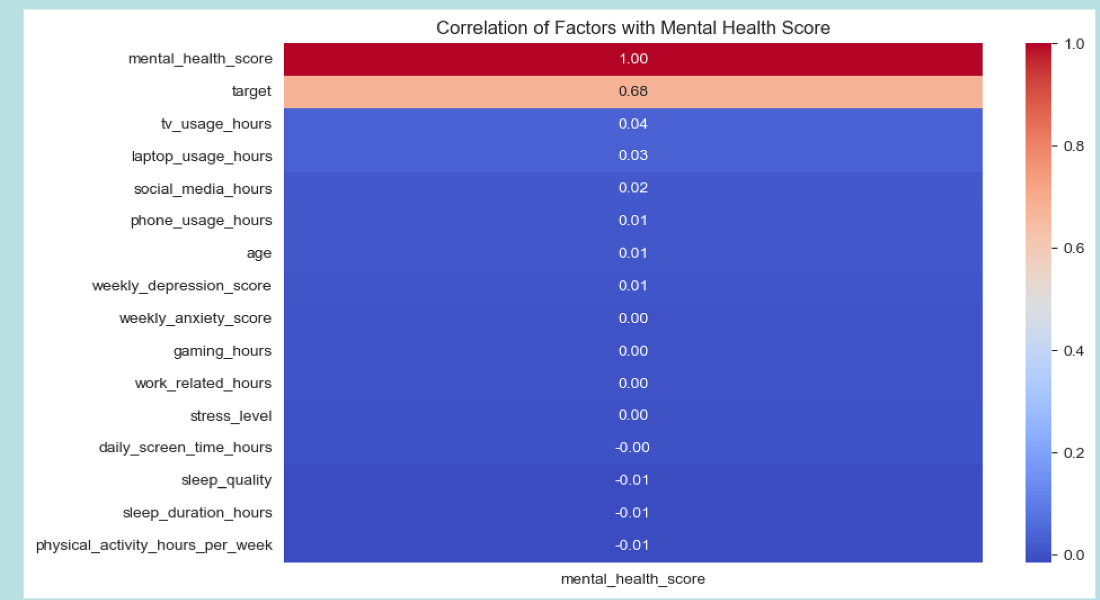
- Dropped irrelevant variables to prepare the dataset for modeling.
- Created binary target variable (mh_level) by binning continuous mental_health_score into "normal" (20-69) and "severe" (70-80).
- Converted categorical variables into numerical format using one-hot encoding.
- Split dataset into training (70%) and testing (30%) sets using stratified sampling to preserve class distribution.
- Applied SMOTE to training data to balance class distribution for model training.
- Standardized continuous features using StandardScaler for scale-sensitive models like SVM and KNN.

Machine Learning Methods:

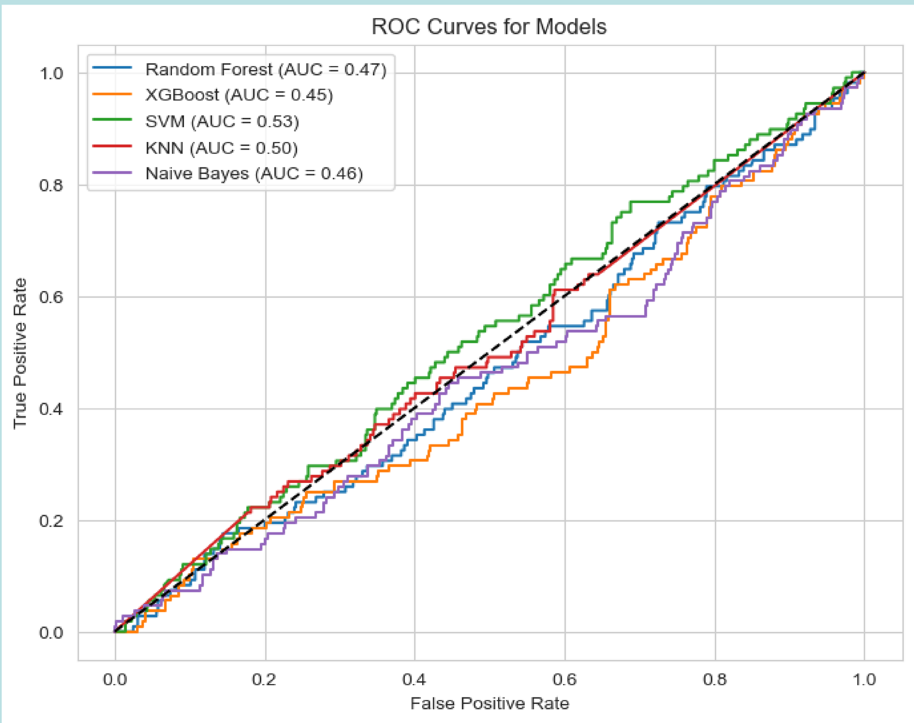
- Built and evaluated six supervised models: Logistic Regression, Random Forest, XGBoost, SVM, KNN, and Naive Bayes.
- Optimized hyperparameters with GridSearchCV using 3-fold cross-validation.

Evaluation Metrics:

- Assessed model performance using Accuracy, Precision, Recall, F1-score, and AUC-ROC.
- Focused primarily on maximizing recall for the "severe" class.



The chart reveals that the target variable has the strongest positive correlation (0.68) with the mental_health_score, which is expected since the target was created directly from this score.
All other variables, such as tv_usage_hours (0.04) and daily_screen_time_hours (-0.00), show correlation values very close to zero.
This indicates that none of these lifestyle or demographic factors have a meaningful linear relationship with mental health scores in this dataset, confirming that there isn't a single, simple predictor for mental health in this data.



As seen in the plot, all five model curves hug the dashed line, indicating that none of them are effective at distinguishing between the two classes.
The SVM model performs just slightly better than random with an AUC of 0.53, while KNN is exactly at random chance with an AUC of 0.50.
The other models—Random Forest (0.47), XGBoost (0.45), and Naive Bayes (0.46)—all perform worse than a random guess, which is a very poor result.
This visualization confirms that, based on the current features, none of these models can reliably separate the 'Normal' and 'Severe' groups.

Results

Model	Accuracy	AUC-ROC	Recall (Severe)	Precision (Severe)
KNN	0.55	0.50	0.43	0.18
Naive Bayes	0.58	0.46	0.30	0.16
SVM	0.73	0.53	0.17	0.20
XG Boost	0.75	0.45	0.11	0.19
Random Forest	0.76	0.47	0.07	0.17

Among the models, Random Forest achieved the highest overall accuracy at 76.3%, closely followed by XGBoost at 75.3%. However, this high accuracy did not translate into effective detection of the severe class, as both models exhibited very low recall for these cases. SVM, although slightly lower in accuracy (72.6%), demonstrated the best AUC-ROC score at 0.526, suggesting marginally better discrimination between the two classes across different thresholds. Interestingly, the KNN model, which had the lowest accuracy (55%), yielded the highest recall for the severe group at 43%, highlighting its potential utility in contexts where sensitivity is prioritized over precision. Naive Bayes offered more balanced results, with moderate recall and a lower false negative rate than the tree-based models, although it remained suboptimal overall.

Conclusion

This study set out to understand whether machine learning could help identify individuals at high risk of severe mental health issues based on their digital behavior and lifestyle habits. We tested multiple models including Random Forest, XGBoost, SVM, KNN, and Logistic Regression to see how well they could predict mental health severity. While models like Random Forest and XGBoost performed strongly in terms of overall accuracy, they had difficulty detecting the high-risk group, which was a much smaller portion of the dataset. This imbalance made it clear that accuracy alone is not enough when dealing with real-world health concerns where early identification is critical.
Among the models, KNN stood out for its ability to detect more severe cases, though it came with more false positives. SVM offered a more balanced performance, and Logistic Regression improved significantly once we addressed the data imbalance. Using CTGAN, a generative AI technique, we created synthetic samples that helped improve recall and fairness. Notably, Random Forest's accuracy increased to over 82%, and Logistic Regression became far more sensitive to severe cases — a major step forward. These results show that combining machine learning with thoughtful data balancing can significantly improve how we identify individuals who may be struggling, especially in underrepresented or overlooked groups.

References: 1. Devi KA, Singh SK. The hazards of excessive screen time: Impacts on physical health, mental health, and overall well-being. *Journal of Education and Health Promotion*. 2023;12(1). doi:10.4103/jehp.jehp_447_23 ; 2. Nagata JM, Al-Shoaibi AAA, Leong AW, et al. Screen time and mental health: a prospective analysis of the Adolescent Brain Cognitive Development (ABCD) Study. *BMC Public Health*. 2024;24(1). doi:10.1186/s12889-024-20102-x ; 3. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *International Journal of Adolescence and Youth*. 2019;25(1):79-93. doi:10.1080/02673843.2019.1590851; 4. Impact of screen time on mental health. Kaggle. https://www.kaggle.com/datasets/khushikyad001/impact-of-screen-time-on-mental-health. Published April 20, 2025.