Original Research Article

An EEG-Based Approach for Identifying Biomarkers of Internet Addiction Disorder

Khushee Goel

Scripps Ranch High School, 10410 Falcon Way, San Diego, CA 92131, United States

ABSTRACT

As smartphones became readily accessible to teenagers in 2010, longstanding trends for anxiety and depression skyrocketed. While statistics indicate that social media addiction is to blame, others argue that social media is more good than bad. This bears the question of how to distinguish between the productive and harmful use of social media. Employing EEG (electroencephalography) to measure brain activity, this research aims to identify biomarkers for internet addiction disorder (IAD). This study employed an open source Kaggle dataset of behavioral addiction disorders, with feature selection conducted through the Random Forest algorithm. This process led to the identification of 15 EEG biomarkers for addiction disorder, and then Odds Ratio analysis was used to identify the most significant ones. T-tests were conducted, beta coefficients were extracted, and logistic regression was used to further validate the findings. The discovered biomarkers now fill the void for quantitative measures of monitoring brain health upon social media consumption.

Keywords: EEG; Internet Addiction Disorder; Biomarkers; Machine Learning; Computational Neuroscience

INTRODUCTION

Research links heavy smartphone and social media use to negative mental health outcomes in adolescents, finding correlation between higher screen time and increased depression (1). In 2024, the prevalence of Internet Addiction Disorder (IAD) increased to an astounding 35% of the American population (2). Such climbing numbers are correlated to the over-consumption and misuse of social media. Today, over 67% of the world's population actively uses social media, translating to over 5 billion people constantly interacting through such platforms (3). This therefore raises alarm over the widespread threat IAD poses for worldwide populations.

Scientists have begun using fMRI (functional magnetic resonance imaging) data to analyze human brain activity during social media usage (4). However, results are derived from lab setups, where a fMRI scanner is used. This makes it expensive and inaccessible for the ordinary user to regularly acquire quantitative insight into the chemical balance–or disbalance–within their brain. More accessible are EEG (electroencephalography) devices. Offering high spatial temporality at a fraction of the cost, these devices are more accessible and can be used to measure electrophysiological activity in the human brain.

This study is based on the hypothesis that, if

Corresponding author: Khushee Goel, E-mail: goelkhushee@gmail.com.

Accepted October 15, 2025
https://doi.org/10.70251/HYJR2348.35842848

social media consumption causes electrophysiological disbalance, then brain waves measured through EEG (electroencephalography) readings must also exhibit a notable change, and such change could be used as a biomarker for the early and accessible diagnosis of IAD.

METHODS AND MATERIALS

This study leveraged a three-fold methodology workflow (Figure 1). The first stage included data preprocessing, during which the open source data was analyzed to assess applicability and was filtered to extract relevant subsets. The second stage involved feature selection, biomarker extraction, and an analysis of the results' validity. Finally, the extracted biomarkers were used to develop a risk prediction model for IAD.

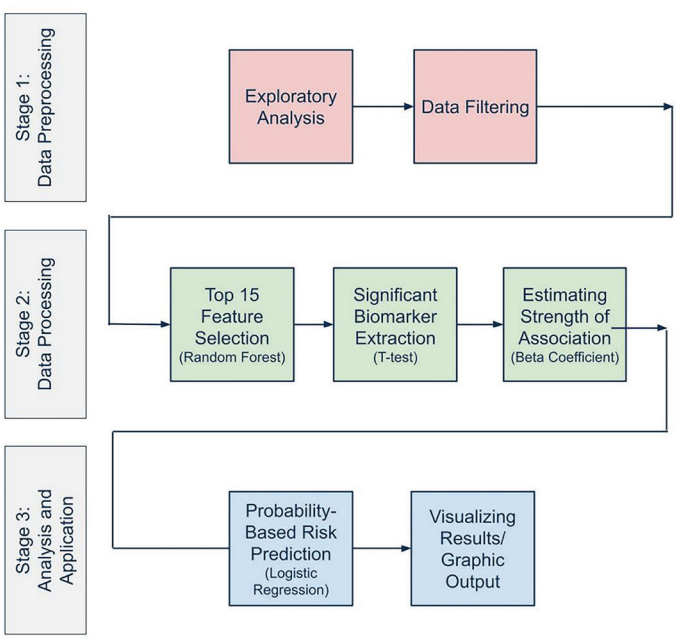

Figure 1. Three-Fold Methodology Workflow.

Data

The dataset for this research was acquired through Kaggle, an online platform for open source data sharing (5). It consisted of 945 participants, with a variety of psychiatric disorders including depression, anxiety, schizophrenia, eating disorders, and addictive behavior. According to the inclusion and exclusion criteria, all participants were between ages 18 to 70, and they had no medical history of neurological disorders, brain injury, tic disorder, attention deficit hyperactivity, neurocognitive disorders, or neurodevelopmental disorders (characterized as IQ<80) (6). Participants were recruited from the Seoul Metropolitan Government-Seoul National University (SMG-SNU) Boramae Medical Center in Seoul, South Korea, and were diagnosed based on MSN-IV or DSM-5 criteria. Diagnoses were then clinically confirmed against electrical medical records and psychological assessments. Among the subset of participants diagnosed with Addiction Disorders, each was further categorized as having either Alcohol Addiction disorder or Behavioral Addiction disorder. In this dataset, specific activities like Internet Gaming Disorder were categorized under Behavioral Addiction Disorder, and thus analysis of this data could be extended to similar conditions like internet and social media addiction (6).

For this study, 2 groups were extracted: healthy controls and participants with behavioral addiction disorder. The final analytic sample (n=188) included both control (n=95) and conditioned (n=93) participants. Demographic features like age, sex, IQ, and education level were obtained for the individuals. Additionally, the individuals were given a 20-10 system EEG device and were recorded for 5 minutes in the eyes closed state. Power and coherence values were then extracted from each electrode, resulting in over 1,000 channels of EEG data per participant. This data was collected in a retrospective fashion, where participants were clinically diagnosed with a psychiatric disorder prior to EEG recording. Therefore, the data reflects long-term brain activity patterns associated with the disorder, rather than brain activity tied to a specific episode of addiction.

Power channels refer to the frequency bands of the signal acquired from a single electrode. This signal was deconstructed into alpha, beta, high beta, theta, gamma and delta frequency subsets, allowing for more elaborate data analysis. Coherence measures, which reflect the degree of synchronicity across regions of the brain, were provided in accordance with the deconstructed power values.

Data Processing

To complete data pre-processing, python libraries including SciKit Learn, Numpy, and Pandas were leveraged. The use of these libraries helped visualize the spread of data, a step that was especially key to getting a broad level view of the demographic variation among the study's participants (Table 1). This data highlights the number of participants in both the control and

Table 1. Baseline Demographic Data

	Control	Behavioral Addiction Disorder	Total	p-value
N	95	93	188	
Women (%)	36.84	4.30	20.74	P < 0.001
Age (Mean)	25.72	25.09	25.41	P = 0.490
IQ (Mean)	116.24	104.38	110.47	P < 0.001
Education in Years (Mean)	14.91	13.16	14.04	P <0.001

conditioned groups, as well as the percent of females in the population. For features such as age, IQ, and education level, the average value among the respective control and conditioned populations was recorded.

Machine Learning

To filter through the data and extract the 15 most important features, multiple decision tree algorithms were evaluated. The Random Forest algorithm stood out because of its aggregated decision tree approach. Rather than using a single decision tree, through which bias can amply and skew the results, the Random Forest algorithm relies on multiple smaller decision trees, and then aggregates the results to output feature importance. Using this ensemble learning method, the 1,000+ power and coherence channels were ranked in order of diagnostic importance, and the top 15 most important biomarkers for behavioral addiction disorder were recorded (Figure 2).

To further analyze the statistical significance of these features, logistic regression models were leveraged and the odds ratios of each channel were computed. During this step, confounders were reintroduced to account for the demographic variation (Table 2).

RESULTS

Descriptives

The data reflected more males than females in both the conditioned and controlled groups. Additionally, significant variation was observed within the spread of ages, with conditioned parties being, on average, of younger age than controls. To account for this variation, demographic features were treated as covariates in this

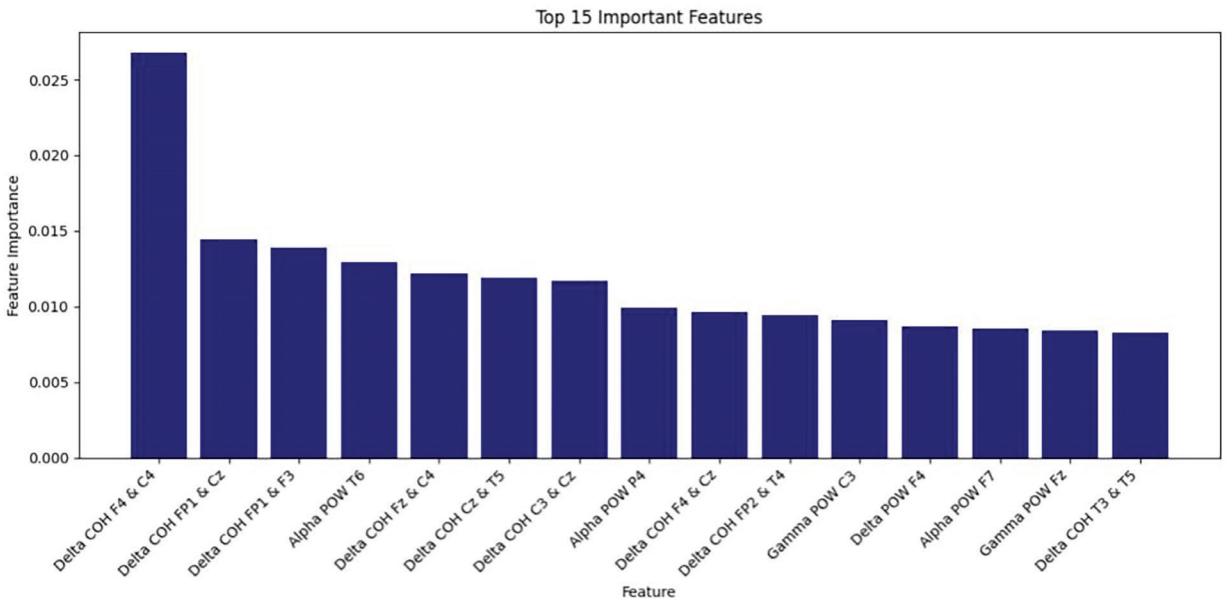

Figure 2. Feature Importance for the Top 15 Predicted Biomarkers.

Table 2. Odds Ratios and Statistic Significant After Adjusting for Confounders

Feature	Adjusted Odds Ratio	Lower 95% CI	Upper 95% CI	p-value
Delta COH F4 & C4	1.00	0.94	1.06	0.92
Delta COH FP1 & Cz	1.00	0.95	1.05	1.00
Delta COH FP1 & F3	0.99	0.96	1.03	0.61
Alpha POW T6	1.03	0.99	1.08	0.13
Delta COH Fz & C4	1.00	0.94	1.06	0.99
Delta COH Cz & T5	0.94	0.91	0.97	0.00
Delta COH C3 & Cz	0.97	0.93	1.02	0.23
Alpha POW P4	0.96	0.92	1.01	0.09
Delta COH F4& Cz	1.00	0.95	1.06	0.96
Delta COH FP2 & T4	1.04	1.00	1.08	0.04
Gamma POW C3	0.93	0.72	1.20	0.58
Delta POW F4	1.04	0.98	1.12	0.19
Alpha POW F7	0.99	0.95	1.03	0.63
Gamma POW Fz	0.99	0.83	1.18	0.90
Delta COH T3 & T5	1.04	1.01	1.07	0.01

study. Rather than including these variables in the main data processing pipeline, they were adjusted for after the fact.

Feature Selection

Using Random Forest to extract the top 15 most important features revealed that delta and alpha waves in the prefrontal cortex regions were most important in predicting between healthy controls and conditioned patients. Notably, 11 of 15 biomarkers were coherence values, while only 4 were power values for their respective channels.

T-Tests and Strength of Association

Adjusting the odds ratio to account for confounding variables resulted in only 3 of the 15 top features being significantly associated with Behavioral Addiction Disorder. These 3 features included Delta Coherence Cz-T5, Delta Coherence FP2-T4, and Delta Coherence T3-T5, and the confidence interval of each increased after adjustment (Figure 3).

Among the 3 features that remained significant after adjusting for confounders, increased activity in the Delta Coherence FP2-T4 and Delta Coherence T3-T5 channels were associated with greater likelihood

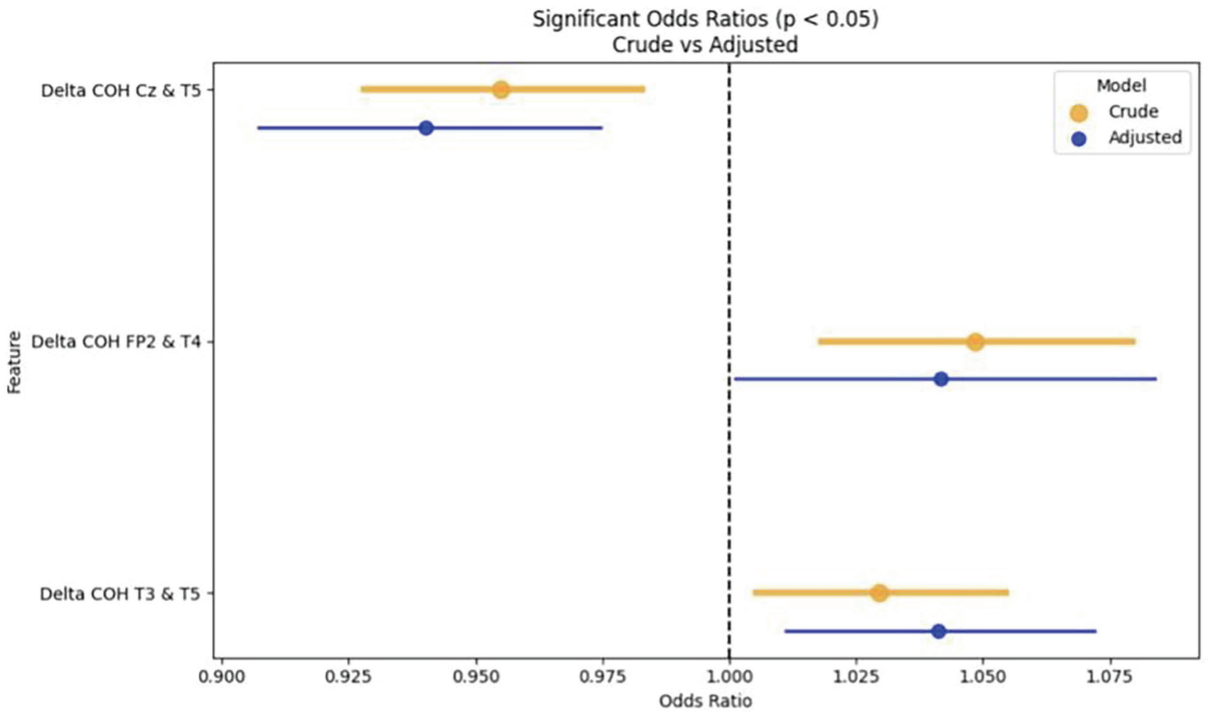

Figure 3. Crude and Adjusted Odds Ratios for Significant Biomarkers. The crude odds ratio is solely based on the prediction through logistic regression. The adjusted odds ratio accounts for confounding variables.

of IAD. Additionally, decreased activity in the Delta Coherence Cz-T5 channel was associated with a lower likelihood of IAD onset. Specific odds ratios for these channels indicate deviations of less than 0.1 from 1, revealing a strength of association of less than 10% between EEG activity in the respective channels and IAD diagnosis (Table 2). These findings prove the biomarkers are statistically significant, but their use in clinically diagnosing IAD would be incremental on their own. When applied in context of the risk measurement algorithm (discussed in section Discussion/Exploring Applications), these correlation measures are aggregated and scaled to result in a clinically significant diagnostic result.

DISCUSSION

Exploring Applications

To explore the application of the identified biomarkers, significant EEG channels were mapped to the commercially available Muse headset (7). This headset is an EEG recording device with 4 channels along the frontal and parietal regions of the brain (Figure 4). While correlating significant channels with electrodes on the Muse headset, both the delta wave coherence between the T3 and T5 electrodes and the Cz and T5 electrodes correlated to the TP9 region. Additionally, delta wave coherence between the FP2 and T4 electrodes was correlated to coherence between the AF8 and TP10 electrodes on the Muse headset, resulting in two EEG biomarkers of interest.

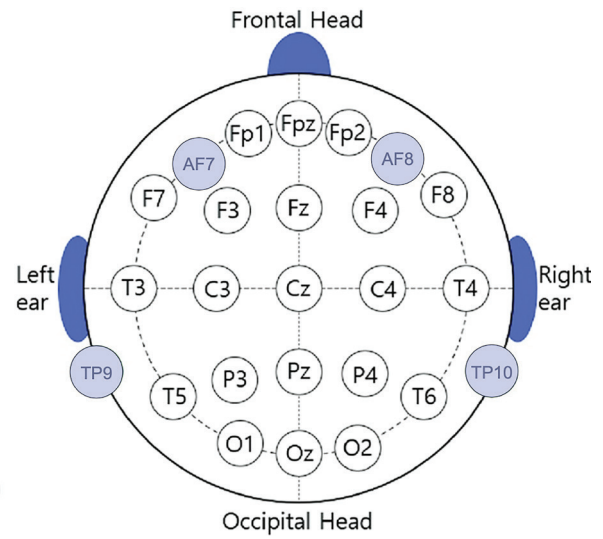

Figure 4. The highlighted electrodes indicate the available electrodes on the Muse EEG Headset.

To measure the risk of internet addiction disorder based on data acquired from the identified channels, a linear equation was constructed. The coefficient and intercept value for each channel was derived from the results of Logistic Regression, which was run during the machine learning process. The sigmoid function was then used to scale this risk output, and this value revealed the individual channel risks. When all the channels were computed and averaged, a cumulative risk measurement was outputted.

The results showed that an increase in TP9 electrode activity and a decrease in TP10 electrode activity is associated with higher risk of IAD onset. Through the development of this risk measurement algorithm, a potential application of this research effort was explored. This process has laid the framework for integration of the 20-10 EEG system and Muse EEG headset, and is proof that the Muse EEG headset can serve as an accessible and insightful option for regular mental health monitoring.

Contextualizing Findings

Using a dataset of participants (average age 25), it was found that decreased delta coherence activity in the Cz & T5 electrodes and increased delta coherence activity in the FP2 & T4 and T3 & T5 electrodes serve as indicators of IAD. Notably, the identified biomarkers pertained to the temporal and prefrontal cortex of the brain, and all three biomarkers were of the delta frequency subset (Figure 5).

Prior research finds that the prefrontal, frontal, and temporal regions of the brain are associated with emotions and decision making (8). These regions are most heavily impaired when addiction develops, and have been linked with the condition of substance abuse related addiction (9). Additional research has identified that delta and gamma waves are linked to the specific condition of Internet Addiction Disorder, and can be used to differentiate between different forms of addiction (10). Delta waves are prominent when the brain is in the resting or restorative development process, during which new neural connections are forming. As addiction develops, the neural connection associating the addictive behavior to the “dopamine high” is repeatedly reinforced and synaptic plasticity decreases as alternative neural connections weaken. This process is marked by delta wave activity, which explains why the majority of the biomarkers identified

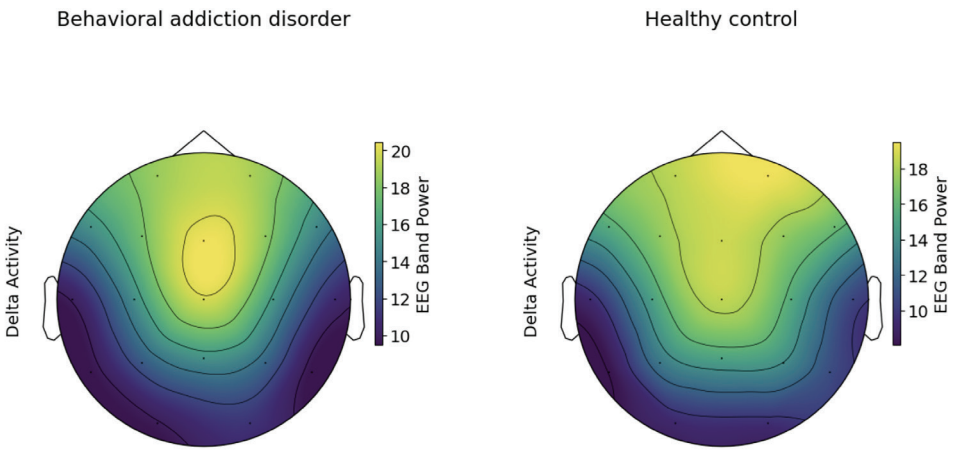

Figure 5. Delta Activity Heat Maps.

were from the delta frequency subset. Other research reports have proven that females are more prone to conditions like IAD, proving that demographic features do in fact impact addiction risk (11). This validates the importance of considering demographic variables as confounding variables. Considering these findings, it becomes further important to address the rising issue of IAD. Much of today's youth engages with digital media during their adolescence, commonly defined as the years between ages 10 to 25. During these years, the brain–and frontal cortex particularly–is constantly forming new neural connections and is undergoing development (12). Early patterns of addiction thwart this development and induce a stronger desire to experience "dopamine highs". This creates a cycle of addiction, one especially dangerous due to social media's unlimited quantity. In fact, research comparing social media to opioids finds that social media is supposedly worse due to the endless amount of "pleasure" it offers (13). Therefore, early identification of IAD is crucial for mitigating its progression and preventing the development of severe addictions in both youth and adults.

Strengths and Limitations

This research effort greatly benefited from access to the elaborate open source dataset titled "EEG Psychiatric Disorders Dataset". This allowed for the analysis of data that otherwise requires a complete lab setup to acquire. Additionally, the dataset primarily included participants near the age of 20, making it well suited for providing insight into the EEG activity of young minds with similar levels of emotional and cognitive development to teenagers. With a total of 945 participants to select from, the research process benefitted from the flexibility to choose specifically those with Behavioral Addiction Disorder and thus the results were more specific to the focus area. Finally, this research effort also benefited from the use of machine learning algorithms. Using such computational tools allowed for the extraction of biomarkers with high confidence and statistical significance.

Limitations still remain present. Subjects could not be recorded during active social media use due to EEG sensitivity to external noise when the eyes are open and moving. Therefore, the analysis had to be generalized to behavioral addiction patterns that are exhibited in the eyes-closed state. Also given that the dataset did not provide EEG data specific to Internet Addiction Disorder, this study pivoted to Behavioral Addiction Disorder–a broader form of addiction under which IAD is a subtype.

In light of these limitations, the resulting biomarkers can be used to identify if a subject is free of IAD. However, if a high risk of addiction is produced, further clinical diagnosis is required to determine the exact subtype.

CONCLUSION

The primary objective of this study was to identify biomarkers and explore the potential application of these biomarkers in identifying individuals at risk for IAD. Through the process of biomarker identification and the development of a risk measurement algorithm, these goals were accomplished. This study presents a complete data processing pipeline, which can serve as a framework for biomarker discovery. The proposed risk measurement algorithm offers promise for both

healthcare professionals and social media users in assessing IAD risk. Future directions include the development of a mobile application and broader adaptation of the algorithmic framework for use in other mental health contexts.

ACKNOWLEDGEMENTS

The author acknowledges the guidance of Kevin Gonzalez at the University of California, San Diego, throughout this research. The mentors at the San Diego Supercomputer Center (14) are also thanked for introducing the author to the field of computational neuroscience via the Research Experience for High School Students (REHS) program

CONFLICT OF INTERESTS

The author declares no conflicts of interest related to this work.

REFERENCES

1. Jean M. Twenge (n.d.). Increases in depression, self-harm, and suicide among U.S. adolescents after 2012 and links to technology use: Possible mechanisms. *Psychiatric research and clinical practice.* https://pubmed.ncbi.nlm.nih.gov/36101887/ (Accessed on 2025, October 10).
2. *Technology addiction statistics 2024*. The Center for Internet & Technology Addiction. (2025, April 22). https://virtual-addiction.com/technology-addiction-statistics-2024/ (Accessed on 2025, October 10).
3. Statista. Number of internet and social media users worldwide as of October 2024. *Statista*. https://www.statista.com/statistics/617136/digital-population-worldwide/. (Accessed on 2024, November 5).
4. Wang Y, Yin Y, Sun Y, Zhou Y, *et al*. Decreased Prefrontal Lobe Interhemispheric Functional Connectivity in Adolescents with Internet Gaming Disorder: A Primary Study Using Resting-State fMRI. *PLOS ONE*. 2015; 10 (3): e0118733. https://doi.org/10.1371/journal.pone.0118733 (Accessed on 2025, October 10).
5. Shashwat Tiwari. EEG Psychiatric Disorders Dataset. *Kaggle.com*. 2021, www.kaggle.com/datasets/shashwatwork/eeg-psychiatric-disorders-dataset (Accessed on 2025, October 10).
6. Park SM, Jeong B, Oh DY, Choi C-H, *et al*. *Identification of major psychiatric disorders from resting-state electroencephalography using a machine learning approach*. Frontiers. 2025a, October 10. https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2021.707581/full
7. Krigolson OE, Hammerstrom MR, Abimbola W, Trska R, *et al*. Using Muse: Rapid Mobile Assessment of Brain Performance. *Frontiers in Neuroscience*. 2021; 15. https://doi.org/10.3389/fnins.2021.634147 (Accessed on 2025, October 10).
8. Manuel AL, Roquet D, Landin-Romero R, Kumfor F, *et al*. Interactions between decision-making and emotion in behavioral-variant frontotemporal dementia and Alzheimer's disease. *Social Cognitive and Affective Neuroscience*. 2020. https://doi.org/10.1093/scan/nsaa085 (Accessed on 2025, October 10).
9. US Department of Health and Human Services. (2016, November). *THE NEUROBIOLOGY OF SUBSTANCE USE, MISUSE, AND ADDICTION*. Nih.gov; US Department of Health and Human Services. https://www.ncbi.nlm.nih.gov/books/NBK424849/. (Accessed on 2024, November 5)
10. Jung-Seok Choi, *et al*. Resting-State Beta and Gamma Activity in Internet Addiction. *International Journal of Psychophysiology*. 2013; 89 (3): 328–333. www.sciencedirect.com/science/article/pii/S0167876013001785?via%3Dihub, https://doi.org/10.1016/j.ijpsycho.2013.06.007. (Accessed on 2024, October 7)
11. Twenge JM, Haidt J, Lozano J & Cummins KM. Specification curve analysis shows that social media use is linked to poor mental health, especially among girls. *Acta Psychologica*. 2022; 224 (224): 103512. https://doi.org/10.1016/j.actpsy.2022.103512 (Accessed on 2025, October 10).
12. Arain M, Haque M, Johal L, Mathur P, *et al*. Maturation of the Adolescent Brain. *Neuropsychiatric Disease and Treatment*. 2013; 9 (9): 449–461. https://doi.org/10.2147/ndt.s39776 (Accessed on 2025, October 10).
13. Goldman B. *Addictive potential of social media, explained*. News Center. 2021, October 29. https://med.stanford.edu/news/insights/2021/10/addictive-potential-of-social-media-explained.html (Accessed on 2024, November 5)
14. *San Diego Supercomputer Center*. (n.d.). Www.sdsc.edu. https://www.sdsc.edu (Accessed on 2025, October 10).