# Graphical Approach of Identifying Patterns of Social Media Addiction

Kein YAMADA[†] and Hiroshi WATANABE[†]
† Waseda University

**Abstract** Social media has profoundly impacted human society, to the extent that social media addiction has emerged as a critical concern. We aim to identify short videos that are likely to be viewed by addicted users through caption analysis.

## 1. Introduction

Social media platforms such as TikTok, Instagram, and YouTube dominate today's digital landscape but have raised growing concerns about "brain rot," a term referring to compulsive scrolling and overstimulation of consuming too much online medias. These behaviors, particularly prevalent in younger generations, have been associated with decreased attention span, heightened anxiety, and other mental health challenges. Traditional research approaches, including surveys and controlled lab experiments, fall short of capturing large-scale, real-world engagement dynamics. Moreover, the limited use of actual short-form videos in studying social media addiction hampers a deeper understanding of the "brain rot" phenomenon. This thesis seeks to identify quantifiable indicators of addiction and lay the groundwork for strategies to address social media addiction by leveraging real platform metadata. To achieve this, we analyze user-generated short videos to uncover underlying trends and patterns.

## 2. Related Work

Research on social media addiction has largely centered on psychological and neuroscientific perspectives, linking excessive usage to changes in brain activity, attention impairments, and adverse emotional effects. Traditional approaches, including EEG [1], fMRI [2], and self-reported surveys [3], provide important insights but are constrained by small sample sizes, subjective bias, and limited applicability to real-world scenarios. In contrast, computational studies applying large-scale user interaction data remain relatively rare. We attempt to bridge that gap by analyzing video content through text-based techniques such as automatic speech recognition (ASR) and video caption generation, achieving a more beneficial and comprehensive understanding of social media addiction.

## 3. Proposed Method

We employ the short-video dataset introduced by Shang et al. [4], originally developed for recommendation system research, which provides both video content and detailed metadata. As the dataset lacks explicit "brain rot" labels, we construct a hypothetical ground truth by assessing user behavior through metrics such as active days and average daily views derived from activity logs. Using 4-cluster k-means clustering, we identify a group with the highest activity and another with the lowest, which we designate as addicted and non-addicted users, respectively. Videos viewed exclusively by addicted users are classified as "addictive," while those viewed solely by non-addicted users are classified as "non-addictive."

For short video analysis, although the dataset includes ASR-generated transcripts, we enhance the textual representation by applying BLIP-2 [5], an image caption model, to generate three captions per video from the first, middle, and last frames. This strategy not only enriches the content analysis but also serves as an alternative approach when ASR text is unavailable.

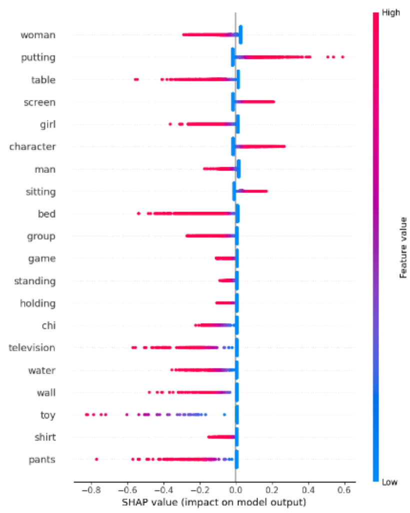


Fig. 1. SHAP result of TF-IDF indicators for BLIP-2 generated captions. Feature values define its impact based on the horizontal direction (left: non-addictive, right: addictive).

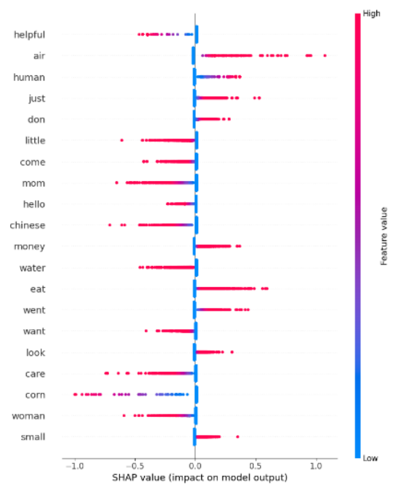


Fig. 2. SHAP result of TF-IDF indicators for provided ASR text. Feature values define its impact based on the horizontal direction (left: non-addictive, right: addictive).

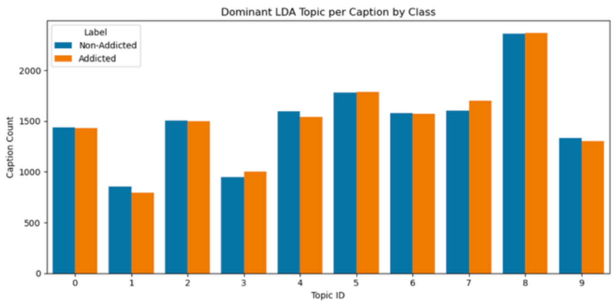


Fig. 3. Graph result of dominant LDA topic observed in BLIP-2-generated captions.

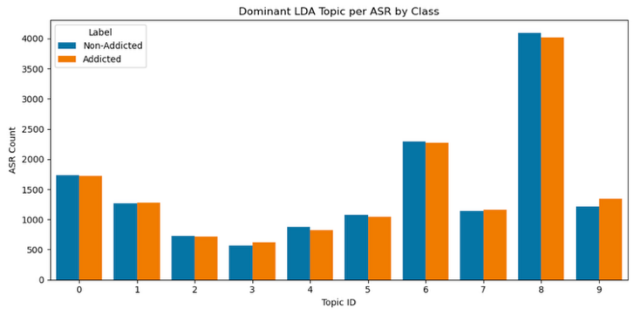


Fig. 4. Graph result of dominant LDA topic observed in provided ASR texts.

## 4. Evaluation

For the short video analysis, videos viewed by brain-rot users are labeled as addictive, while those viewed by non-brain-rot users are labeled as non-addictive. To maintain class balance, we randomly sample 15,000 videos from each category, excluding any overlapping content watched by both groups. We then apply the TF-IDF (Term Frequency–Inverse Document Frequency) technique to both ASR transcripts and BLIP-2 captions, training a logistic regression model and leveraging SHAP analysis to highlight the most influential terms distinguishing addictive from non-addictive content. Additionally, we utilize Latent Dirichlet Allocation (LDA) with 10 latent topics, assigning a dominant topic to each video's text to reveal broader thematic trends.

## 5. Results and Discussion

Fig. 1 and Fig. 2 present the SHAP analysis of TF-IDF features for distinguishing between non-addictive (left of the x-axis) and addictive (right of the x-axis) videos. The SHAP analysis reveals distinct thematic patterns between addictive and non-addictive video indicators. The SHAP analysis of BLIP-2 captions shows that words such as *"putting," "screen," "character,"* and *"sitting"* are strongly associated with addictive videos, indicating content with active scenes or visually engaging actions. In contrast, terms like *"woman," "table," "girl," "man," "bed," "group," "television,"* and *"water"* lean toward non-addictive videos, suggesting more static, descriptive, or object-oriented content. For ASR text, words such as *"air," "just," "don," "money," "eat," "went,"* and *"look"* are linked to addictive videos, reflecting conversational cues, action-oriented contexts, or material references that may capture viewer attention. Conversely, terms like *"helpful," "little," "come," "mom," "chinese," "water,"* and *"woman"* are associated with non-addictive videos, pointing to neutral, familial, or informational themes.

Moving to the analysis with LDA topic, For ASR topics, the themes range from size/appearance (Topic 0), games (Topic 1), and countries/people (Topic 2) to money (Topic 3), family roles (Topic 4), food (Topic 5), emotions (Topic 6), general traits of people (Topic 7), actions (Topic 8), and conversations or greetings (Topic 9). For caption topics, the themes include object handling (Topic 0), people/appearance (Topic 1), vehicles/urban scenes (Topic 2), games/screenshots (Topic 3), clothing (Topic 4), indoor or family settings (Topic 5), posing (Topic 6), Chinese text/media (Topic 7), group activities (Topic 8), and food/kitchen scenes (Topic 9).

As shown in Fig. 3 and Fig. 4, video captions for addicted and non-addicted labels exhibit largely similar distributions. Although certain topics, such as Topic 7 in Fig. 3 and Topic 9 in Fig. 4, display slightly higher peaks, the overall results suggest minimal differences in the themes and topics viewed by the two user groups.

## 6. Conclusion

We present an approach for analyzing short videos viewed by users with and without "brain rot." Although the findings are not highly conclusive, they offer valuable insights and suggest future directions for analysis—such as examining action intensity or rapid frame changes (flashes) rather than focusing solely on the topics or themes preferred by users.

## References

[1] Y. Yin, X. Cai, M. Ouyang, S. Li, X. Li, and P. Wang, "FoMO and the brain: Loneliness and problematic social networking site use mediate the association between the topology of the resting-state EEG brain network and fear of missing out," *Comput. Human Behav.*, vol. 141, Art. no. 107624, 2023.

[2] Q. He, O. Turel, and A. Bechara, "Brain anatomy alterations associated with social networking site (SNS) addiction," *Sci. Rep.*, vol. 7, Art. no. 45064, 2017. doi: 10.1038/srep45064Z.

[3] T. Ehsan and J. Basit, "Machine learning for detecting social media addiction patterns: Analyzing user behavior and mental health data," *Int. J. Inf. Sci. Technol. (IJIST)*, vol. 6, no. 4, pp. 1789-1807, Oct. 2024.

[4] Y. Shang, C. Gao, N. Li, and Y. Li, "A large-scale dataset with behavior, attributes, and content of mobile short-video platform," in *Companion Proc. ACM Web Conf. (WWW)*, New York, NY, USA, 2025, pp. 793–796. doi: 10.1145/3701716.3715296

[5] J. Li, D. Li, S. Savarese, and S. Hoi, "BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models," in *Proc. 40th Int. Conf. Mach. Learn. (ICML)*, vol. 202, 2023, Art. no. 814, pp. 19730–19742.

† 早稲田大学 基幹理工学研究科 情報理工情報通信専攻
〒169-0072 東京都新宿区大久保 3-14-9-401
TEL. 03-5286-2509 E-mail: stslm738.ymd@toki.waseda.jp