修 士 論 文 概 要 書

Master's Thesis Summary

Date of submission: 07/25/2025 (MM/DD/YYYY)

専攻名 (専門分野) Department	Computer Science and Communications Engineering	氏 名 Name	Kein Yamada	指 導 教 員 Advisor	Hiroshi Watanabe 印 Seal
研究指導名 Research guidance	Research on Audiovisual Information Processing	学籍番号 Student ID number	CD 5123FG44-9		
研究題目 Title	Brain Rot Profiling: Understanding Social Media Addiction Through Large-Scale User Activity Logs and Short Videos				

1. Introduction

Social media platforms like TikTok, Instagram, and YouTube dominate modern digital life but have sparked concerns about **"brain rot,"** a term for compulsive scrolling and overstimulation. Such behaviors, common in the young generation, are linked to reduced focus, anxiety, and other mental health issues. Traditional research methods, like surveys or lab experiments, cannot fully capture large-scale, real-world engagement patterns. This thesis aims to identify measurable indicators of addiction and provide a foundation for ways to confront social media addiction by focusing on actual social media platform metadata. For this purpose, we utilize user interaction logs and short videos for analyzing trends and patterns.

2. Related Work

Studies on social media addiction have primarily focused on psychological and neuroscientific perspectives, linking excessive use to altered brain activity, attention deficits, and negative emotional outcomes. Common methods such as EEG [1], fMRI [2], and self-reported surveys [3] offer valuable insights but are limited by small sample sizes, subjective biases, and a lack of real-world scalability.

In contrast, computational research that uses large-scale user interaction data remains scarce. This thesis addresses this gap by not only modeling user interaction trends through social media log metadata but also examining video content via text-based methods such as ASR and video caption generation, aiming to capture a more holistic view of social media addiction.

3. Method

3.1. Large-scale User Activity Log

For this study, we make use of the KuaiSAR dataset [4], originally designed for recommendation system research, which provides detailed user interaction logs but lacks explicit labels for "brain rot." To address this, we define a hypothetical ground truth by identifying behavioral patterns common among addicted users, focusing on metrics such as the number of active days and average views per day. The dataset includes core features of user actions, such as follows, likes, clicks, forwards, playing time, timestamps, and search activity. To enhance analysis, we derive new features (e.g. watch efficiency and hour entropy) and evaluate categorical ratios of the content viewed to better capture user engagement trends.

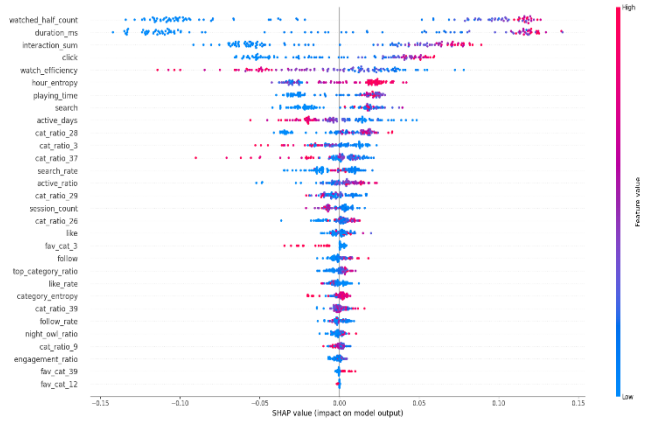

Fig. 1. SHAP result of the user activity features that display relevance to brain rot. Feature values define its impact based on the horizontal direction (left: non-brain rot, right: brain rot).

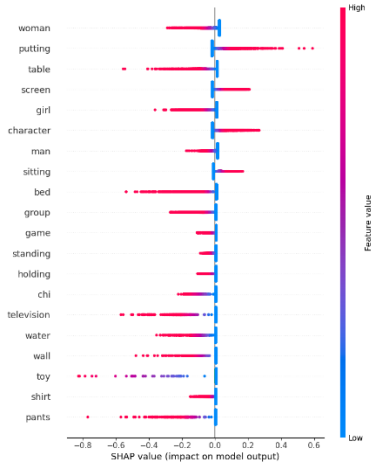

Fig. 2. SHAP result of TF-IDF indicators for BLIP-2 generated captions. Feature values define its impact based on the horizontal direction (left: non-addictive, right: addictive).

3.2. Social Media Short Videos

We utilize the short-video dataset by Shang et al. [5], originally designed for recommendation system research, which provides both video content and associated metadata. As with the user activity log, we establish a hypothetical ground truth for "brain rot" by evaluating users' active days and average views per day. While the dataset includes default ASR-generated text for each video, we supplement this by using BLIP-2 [6], an image captioning model, to produce three captions per video based on the first, middle, and last frames. This approach not only enriches the analysis but also demonstrates how captions can be generated when ASR text is unavailable. For this research, we select BLIP-2 over state-of-the-art

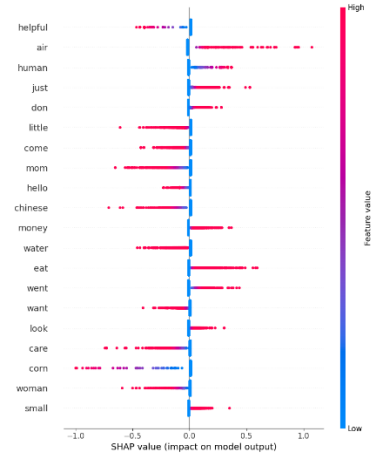

Fig. 3. SHAP result of TF-IDF indicators for provided ASR text. Feature values define its impact based on the horizontal direction (left: non-addictive, right: addictive).

(SOTA) video captioning models due to practical constraints, such as the dataset's size being over 1TB of video and the excessive time required for full video-based captioning.

4. Experiment

4.1. Evaluation for User Activity Log

For the user activity log analysis, we first downsample non-brain-rot users to match the number of users labeled as brain rot, ensuring balanced training data. We employ a Random Forest Classifier with 100 estimators and a maximum depth of 8 to capture non-linear patterns in user behaviors. To better understand feature contributions, we rely on SHAP (SHapley Additive exPlanations), which quantifies the impact of each feature on the model's predictions. Since our brain rot labels are hypothetical and not empirically verified by clinical experts, we focus more on SHAP-based interpretability rather than metrics like accuracy or precision, due to the true baseline performance being inherently unknown.

4.2. Evaluation for Short Videos

For the short video analysis, we label videos viewed by brain-rot users as addictive and those viewed by others as non-addictive. To ensure balance, we sample 15,000 videos from each group, removing any overlapping content that was watched by both user types. We apply a TF-IDF (Term Frequency–Inverse Document Frequency) approach to the ASR and BLIP-2 caption texts, using logistic regression combined with SHAP analysis to identify key terms contributing to the classification of addictive versus non-addictive content. As a complementary approach, we employ Latent Dirichlet Allocation (LDA) with 10 latent topics, assigning a dominant topic to each video's text data to uncover broader thematic patterns.

5. Results and Discussion

As shown in the SHAP analysis of Fig. 1, watched half count, duration_ms, and interaction_sum are the most influential indicators separating brain-rot users, with derived metrics like watch efficiency and hour entropy also contributing significantly. For the short videos,

TABLE I. LDA TOPICS OF TOP KEYWORDS FOR ASR AND CAPTIONS

Topic	Top Keywords (ASR)	Top Keywords (Captions)
0	one, look, like, good, big	holding, putting, person, hand, phone
1	first, game, also, new, two	girl, hat, boy, asian, pink
2	china, chinese, us, people, yang	car, street, city, walking, driving
3	yuan, money, buy, boss, video	game, screenshot, screen, character, showing
4	man, woman, mother, old, father	shirt, white, black, blue, red
5	eat, water, food, small, fish	sitting, bed, table, chair, baby
6	love, heart, like, life, song	standing, front, sexy, suit, tie
7	people, good, many, like, person	chinese, poster, chi, character, characters
8	go, come, let, going, want	group, standing, front, screen, people
9	hello, said, could, would, human	bowl, eating, cup, plate, preparing

TF-IDF feature importance highlights several key caption terms (Fig. 2) and ASR terms (Fig. 3) that correlate with addictive versus non-addictive content, though these signals are weaker compared to user activity logs. The LDA topic modeling results found in Table 1 reveal clear differences between ASR and caption text: while ASR topics emphasize abstract or conversational words (e.g., *look*, *good*, *money*, *people*), the captions focus on concrete visual elements (e.g., *holding*, *person*, *car*, *shirt*). This contrast reflects the inherent difference between spoken or transcribed video content and image-based descriptions.

6. Conclusion

In this thesis, we explore social media overuse, or "brain rot," by analyzing large-scale user activity logs and short video content. Using the KuaiSAR dataset, we derive behavioral features such as watch efficiency and hour entropy, with SHAP analysis revealing that metrics found in activity logs strongly indicate addictive behaviors. Complementary analysis of ASR text and BLIP-2 captions, using TF-IDF and LDA, showed differences between abstract conversational terms in ASR and concrete visual descriptions in captions, offering additional insights into content patterns. Overall, our data-driven approach demonstrates that user behavior features provide the clearest signal for detecting overuse, while content-level analysis adds valuable context for understanding how addictive content is consumed.

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] Z. Sun et al., "KuaiSAR: A unified search and recommendation dataset," in *Proc. 32nd ACM Int. Conf. Inf. Knowl. Manag. (CIKM)*, New York, NY, USA, 2023, pp. 5407–5411. doi:10.1145/3583780.3615123.

[5] 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

[6] 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.

Brain Rot Profiling: Understanding Social Media Addiction Through Large-Scale User Activity Logs and Short Videos

A Thesis Submitted to the Department of Computer Science and Communications Engineering, the Graduate School of Fundamental Science and Engineering of Waseda University in Partial Fulfillment of the Requirements for the Degree of Master of Engineering

Submission Date: July 21st, 2025

Kein Yamada

(5123FG44-9)

Advisor: Prof. Hiroshi Watanabe

Research guidance: Research on Audiovisual Information Processing

Acknowledgements

To my professor, I extend my deepest gratitude for the unwavering support and encouragement, particularly during the challenging moments of this research journey.

His willingness to guide me and provide constructive feedback has been invaluable in shaping my work.

Another heartfelt applause goes to him for accepting my bold decision to completely change my research direction, showing both trust and understanding.

Never once did I feel alone in this endeavor, as my friends, colleagues, and lab mates constantly offered strong advice and insightful words that kept me on track.

Kind words and wisdom from these individuals have greatly contributed to the clarity and depth of this thesis.

Years of perseverance would not have been possible without the foundation built by my parents.

Over the past 25 years, they have supported me through every challenge with unwavering faith and love.

Under their guidance and care, I have been able to grow not just academically but personally as well.

Having their encouragement has truly been the cornerstone of my success.

Without them, this path of life would not have been possible. Hence, I would like to thank everyone that I met along and wish them all the best.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Background

Social media platforms have become an integral part of modern life, shaping how individuals communicate, consume information, and spend their leisure time. With countless number of active users across platforms such as Instagram, TikTok, YouTube, and X (formerly Twitter), these services offer unparalleled connectivity, real-time interaction, and endless streams of entertainment [1]. While these benefits are undeniable, a growing body of research has raised concerns about the long-term psychological and behavioral effects of excessive social media use. In particular, the phenomenon of compulsive engagement, characterized by habitual scrolling [2], binge-watching short-form content [3], and a persistent need for digital stimulation, has become increasingly prevalent, especially among adolescents and young adults [4].

This pattern of usage, often referred to colloquially as "brain rot," reflects a broader public unease about the addictive design of modern social platforms. The term, once considered informal internet slang, gained significant cultural traction in recent years and was notably chosen as Oxford's Word of the Year in 2024 [5]. This selection reflects not only the ubiquity of the concept but also a growing societal recognition of the potential harm associated with unmoderated content consumption. A closely related behavior, "doom scrolling"—the act of continuously consuming negative or emotionally charged content—has similarly been shown to exacerbate psychological fatigue and diminish users' emotional resilience [6]. Both terms underscore a cultural shift in how we conceptualize digital overexposure, especially in environments driven by endless feeds and algorithmic engagement maximization.

From a psychological standpoint, excessive social media use has been linked to various adverse outcomes, including decreased attention span, disrupted sleep patterns, heightened anxiety, and symptoms of depression [7]. Neuro-cognitive studies suggest that the constant influx of novel

stimuli—likes, comments, and algorithmically curated videos—can rewire reward pathways in the brain, mimicking the effects observed in substance-based addictions [8, 9]. This is particularly concerning among adolescents, whose brains are still in critical stages of development and are more susceptible to reinforcement-driven behaviors. Moreover, the fear of missing out (FOMO), online social comparison, and performance pressures contribute to a deteriorating sense of well-being, amplifying the mental health burden among heavy users [10].

1.2 Research Objectives

The convergence of behavioral, social, and technological factors has prompted researchers across disciplines to investigate the underlying mechanisms of digital addiction. In the field of computer science, there is an urgent need to develop data-driven frameworks to quantify and classify usage patterns that may signal problematic engagement. By leveraging large-scale behavioral logs, machine learning models, and explainable AI techniques, it becomes possible to identify indicators of overuse and detect at-risk individuals. This thesis aims to contribute to this growing research effort by analyzing user interaction data from short-form video platforms, with the ultimate goal of developing interpretable models that can assist in the early detection of social media overexposure—colloquially described as "brain rot."

Detecting such addiction-like behavior remains a significant challenge, largely due to the complex, multifaceted nature of user interaction patterns and the diversity of content formats. Traditional approaches, which often rely on surveys, interviews, or psychological self-assessment questionnaires, have limitations in scalability, objectivity, and temporal resolution [11, 12, 13, 14]. While these methods have been instrumental in shaping foundational understanding, they fall short of capturing real-time, large-scale behavioral nuances necessary for modeling digital addiction. In contrast, the growing availability of user interaction logs such as view histories, like or forward activities, and time-series engagement data presents a powerful but underutilized opportunity to identify problematic usage behaviors through computational means.

Notably, existing research on social media addiction and brain rot is dominated by approaches rooted in medicine and neuroscience. Many studies employ electroencephalography (EEG), functional MRI, or controlled observation of brain signals to analyze the neurological underpinnings of addictive digital behavior [15, 16, 17]. These efforts, while scientifically valuable, are often constrained by small sample sizes and laboratory settings, limiting their applicability to the broader population. Furthermore, within computer science, relatively few studies have explored this topic in depth, despite its growing relevance to digital well-being and human-computer interaction. The lack of large-scale, data-driven studies represents a critical research gap—particularly when considering

that billions of users interact with algorithm-driven social platforms on a daily basis.

To date, very few studies have directly applied real-world social media usage data to study the phenomena of digital overuse or addiction. Most research either simulates user behavior or extrapolates findings from indirect user-reported measures. There is virtually no precedent for applying machine learning to comprehensive, high-volume social media interaction datasets with the specific aim of detecting addiction-like patterns at scale. Such datasets are essential not only for building generalizable models but also for ensuring that proposed detection methods can scale to real-world applications. Given the widespread impact of social media across age groups, regions, and demographics, research that embraces the computational potential of big behavioral data is urgently needed.

This thesis responds to that need by utilizing a large-scale dataset of user activities on short-form video platforms to uncover indicators of "brain rot" behavior. Through the integration of temporal features, interaction metadata, and machine learning interpretability tools, we aim to provide both a methodological foundation and empirical insight into how social media addiction manifests in measurable digital behavior. Our ultimate goal is to establish a data-driven framework that advances the understanding of social media overuse and offers viable pathways for detection and intervention in practical settings.

1.3 Thesis Outline

The outline of this thesis is as follows:

Chapter 1: This chapter provides an overview of social media addiction and the societal discourse surrounding "brain rot." It outlines the research objectives and presents the structure of the thesis.

Chapter 2: We review prior research on social media addiction and cognitive decline, highlighting both psychological studies and computational approaches.

Chapter 3: This chapter investigates behavioral traits associated with "brain rot" using large-scale user activity data. We introduce a framework for hypothetically labeling users exhibiting addictive behavior, propose relevant features, and construct derived features to enhance behavioral analysis.

Chapter 4: We explore methods for analyzing short-form video content to understand differences in user behavior. This includes the use of caption generation techniques to identify recurring keywords and themes potentially linked to addictive engagement patterns.

Chapter 5: Building on Chapter 3, we introduce an alternative approach to user labeling and reapply behavioral analysis techniques to evaluate consistency in detecting "brain rot" traits across

different assumptions.

Chapter 6: The thesis concludes with a summary of findings, emphasizing the potential of large-scale social media datasets for identifying behavioral markers of digital addiction and advancing data-driven approaches in this domain.

Chapter 2

Related Work

2.1 Analytic Approaches to Social Media Addiction

A growing body of research seeks to analyze and predict social media addiction through computational and psychological approaches. These efforts are primarily rooted in the use of survey or questionnaire-based data, through which participants self-report their usage patterns, emotional responses, and perceived levels of dependence. While such approaches have contributed foundational knowledge to the field, they often lack the scale, granularity, and objectivity necessary for modeling real-world behavior. Nevertheless, these studies provide valuable insight into the psychological dimensions of social media addiction and demonstrate the feasibility of applying statistical and machine learning techniques to subjective self-assessments.

Akter et al. [11] investigated the prediction of social media addiction during the COVID-19 pandemic by collecting responses from 504 participants in Bangladesh through structured questionnaires and interviews. Participants were categorized into several addiction levels based on self-reported behavioral patterns and psychological symptoms, such as restlessness and distraction. Utilizing machine learning classifiers such as logistic regression, decision trees, and support vector machines, the study achieved high classification accuracy, with logistic regression performing best at approximately 94%. However, the data used in the study lacks behavioral granularity, such as timestamped usage logs, types of consumed content, or frequency of engagement, thereby limiting its applicability to real-time detection or deployment on actual social media platforms.

In a similar study, Mardiah and Kusnawi [12] analyzed user behavior using a publicly available Kaggle dataset composed of 481 questionnaire responses. The study aimed to compare the effectiveness of linear regression and Random Forest algorithms in predicting a user's social media addiction level. Features were selected through recursive feature elimination, and model performance was evaluated using standard error metrics. Interestingly, linear regression slightly outperformed Ran-

dom Forest in terms of root mean square error (RMSE) and mean absolute error (MAE), suggesting that simpler models may suffice when working with limited and structured survey data. Despite this, the study's dataset was constrained by its small sample size and reliance on self-reported variables, offering limited insights into temporal dynamics or content-specific engagement patterns.

Expanding on this paradigm, Ehsan and Basit [13] developed a machine learning framework to classify users into addiction risk groups using behavioral and psychological indicators collected via questionnaires. Their study employed a Random Forest classifier and examined correlations between high social media use and symptoms such as irritability and mental fatigue. Feature importance analysis identified social comparison behavior and the need for self-validation as strong predictors of addiction. While their work emphasizes the link between social media use and mental health, it also underscores the dependence on psychological self-disclosure, which is not readily accessible or observable in actual platform usage logs.

Another notable contribution comes from Çiftci and Yıldız [14], who explored the relationship between social media addiction and subjective well-being, particularly happiness and life satisfaction. In their study of over 1,000 adults, they observed a negative correlation between addiction scores and well-being measures. A mediation analysis revealed that life satisfaction partially mediated the effect of addiction on happiness. Moreover, they trained several regression models to predict happiness levels, with elastic net regression emerging as the most effective. SHAP (Shapley Additive Explanations) [21] analysis confirmed that social media addiction and life satisfaction were among the most influential features in predicting user happiness. Though the study introduced interpretable modeling into addiction research, it still depended on user surveys and omitted direct behavioral metrics obtainable from platform data.

Taken together, these studies reflect a common trend in social media addiction research: a reliance on voluntary, small-scale survey or questionnaire-based datasets. While informative, these datasets often lack connection to the actual metadata and usage characteristics of short-form media content such as videos and reels. This disconnection limits the applicability of such research in real-world platforms, where identifying addictive usage patterns requires high-resolution behavioral traces, including session duration, engagement frequency, content type, and algorithmic exposure.

Furthermore, although a focus on psychological symptoms and self-perceptions provides important clinical and social insights, these dimensions are not practically observable by social media platforms themselves. Platform operators cannot directly access or infer psychological traits such as self-esteem, distractibility, or emotional fatigue without explicit user disclosure, making such features difficult to operationalize for real-time monitoring or intervention.

Given these limitations, there is a pressing need for a shift in research emphasis—from subjective, small-scale psychological analysis to large-scale, metadata-driven investigation of user behav-

ior. Incorporating interaction logs, video metadata, and content recommendation traces offers a path toward scalable, objective, and actionable models of social media addiction. By grounding analysis in actual user engagement patterns and content characteristics, future research can enable more accurate detection of addiction-like behavior and potentially support platform-level interventions that promote healthier digital consumption habits.

2.2 Medical Approaches of Cognitive Analysis

Researchers have increasingly applied neuroscience methods to uncover how excessive social media use affects the brain. Using resting-state EEG and network analysis, for example, Yin et al. [15] examined the brain network topology associated with "fear of missing out" (FoMO) on social media. They found that individuals with high FoMO exhibited a more scale-free brain network organization (higher Kappa and leaf fraction in EEG-based minimum spanning tree metrics) compared to low-FoMO peers. Moreover, the link between this altered network topology and FoMO was partly mediated by psychosocial factors – specifically, increased loneliness and more problematic SNS use among high-FoMO individuals. This suggests that feeling socially isolated and compulsively checking social networks may reinforce each other alongside measurable brain-network differences in heavy social media users. In a related EEG study, Sun et al. [16] compared Internet-addicted individuals to healthy controls and observed a disrupted functional connectivity pattern: the addicted group's brain networks had significantly lower clustering coefficients and shorter path lengths (notably in beta and gamma bands), indicative of a shift toward a more random, less optimally small-world organization. These EEG findings imply that problematic Internet/SNS use is accompanied by global topological changes in brain connectivity, potentially reflecting less efficient information integration in the brain's resting-state networks.

Complementing the EEG-based evidence of functional network alterations, other work has identified structural brain differences associated with social media overuse. Using MRI-based voxel-based morphometry, He et al. [17] showed that higher SNS addiction levels correlate with reduced grey matter volume in the bilateral amygdala, a key region of the brain's reward/impulsive system. This amygdala shrinkage mirrors patterns seen in substance and gambling addictions and may indicate a neuroplastic adaptation (pruning of neurons) that makes the reward system more "efficient" or sensitized. Interestingly, He et al. also found that, unlike classic drug addictions where chronic use is linked to reduced anterior cingulate cortex (ACC) volume (impaired self-control), SNS addicts did not show ACC volume loss; in fact, ACC volume was positively correlated with SNS addiction severity. In other words, the neural profile of social networking addiction exhibited a mix of addiction-like changes (in the amygdala) and distinct features (no structural deficit in the ACC). This points to both similarities and differences between technology-related addictions and traditional ad-

dictions in terms of brain anatomy, and it highlights that social media overuse may hyper-engage impulsive reward circuits while leaving regulatory control areas relatively intact or even heightened.

Despite these neuroscientific insights, there are clear limitations in translating them to real-time, large-scale applications. Most evidence comes from controlled lab studies with small samples and specialized equipment that are not feasible to deploy broadly. Even if the cost-effective EEG-based machine learning models are to be proposed for classifying Internet addiction and suggested they could be practical at scale, in reality an SNS platform cannot easily obtain users' brain data in the wild. Real-time monitoring of millions of users' EEG signals or brain scans is logistically and ethically prohibitive since users are not going to wear EEG devices during everyday social media use, and brain imaging (MRI) is even less accessible. Thus, an operator of a social media service cannot implement these neuro-assessment techniques on their own as a tool to detect or mitigate addiction among their user base. Any intervention based on brain metrics would require medical cooperation and individual user participation well outside the normal social media experience. In summary, neuroscientific approaches deepen our understanding of social media addiction's neural correlates, but their practical utility for real-time detection or large-scale management of user addiction remains very limited under current technology and infrastructure. The insights gained are valuable for the research and clinical context, yet social media companies must rely on other strategies to address problematic use in practice.

TABLE I. STRENGTHS AND WEAKNESSES OF EACH METHOD FOR EXAMINING SOCIAL MEDIA ADDICTION

Method	Accessibility	Credibility	Complexity	Cost Efficiency	Scalability
Survey or questionnaires	✓	△	✓	✓	×
EEG	△	✓	×	△	×
MRI	×	✓	×	×	×

Chapter 3

Brain Rot Analysis on Behavioral Insights from Large-Scale User Interactions

3.1 Recommendation Dataset: KuaiSAR

In recent years, large-scale datasets sourced from real-world social media platforms have become increasingly vital for advancing personalized recommendation systems. Among these, there are three datasets based on Kuaishou short video platform publicly available: KuaiRec [18], KuaiRand [19], and KuaiSAR [20]. All three datasets contain user interaction logs centered around short-form video consumption, capturing key behavioral signals from a large number of users. These resources have significantly contributed to the development and evaluation of recommendation systems tailored for short video content—an area that has gained prominence due to the explosive growth of platforms like TikTok, Instagram Reels, and Kuaishou itself.

While the primary intention behind these datasets is to facilitate research in recommendation systems, they also hold substantial value for behavioral and user modeling tasks. In particular, their rich, time-resolved interaction records make them suitable for exploring patterns of excessive or compulsive engagement—traits commonly associated with what is colloquially referred to as "brain rot." The granular nature of the data allows researchers not only to study user preferences and ranking algorithms but also to quantify engagement intensity, identify anomalous usage patterns, and potentially classify at-risk individuals. This dual utility—supporting both recommendation performance and behavioral analysis—positions these datasets as ideal candidates for research on social media overuse and digital addiction.

Among the three datasets, KuaiSAR (Short video Active Recommendation dataset) was selected for this thesis due to its comprehensive coverage of user interactions and the inclusion of detailed user- and item-level features. Compared to KuaiRec and KuaiRand, KuaiSAR provides a broader

interaction scope and more robust metadata, making it particularly suitable for large-scale behavioral analysis. Specifically, KuaiSAR is composed of four core components: rec_inter.csv, src_inter.csv, user_features.csv, and item_features.csv.

The rec_inter.csv file contains 14,605,716 recommendation actions, representing passive content exposure and active user engagement with recommended short videos. These interactions form the central focus of our analysis, as they most closely mirror the habitual and sometimes compulsive behavior seen in short-form video consumption. Each row in this file corresponds to a single recommendation event and includes the following fields: user_id, item_id, playing_time, duration_ms, time, click, forward, like, follow, search, and search_item_related. These attributes collectively offer a high-resolution view of user engagement, enabling detailed analysis of behavioral patterns such as passive scrolling, binge-watching, and repetitive interactions—all of which are indicative of addictive tendencies in the context of "brain rot."

In contrast, the src_inter.csv file records 5,059,169 search actions, where users actively searched for content. While this information is valuable for understanding intent-based behavior, our research prioritizes scroll-based passive consumption over active retrieval. Thus, we focus solely on rec_inter.csv, as it better aligns with the involuntary and habitual traits often discussed in the context of digital overuse.

The user_features.csv file includes demographic and behavioral metadata for 25,877 users. One particularly important feature is the activity level, which is measured as the ratio of days a user was active over the entire logging period. This measure provides a quantitative basis for distinguishing between casual users and potentially addicted users, especially when analyzed alongside engagement signals such as session frequency, duration, and interaction density. These user-level attributes are essential for defining profiles of high-risk individuals within a large, heterogeneous user base.

Additionally, the item_features.csv file describes content characteristics for 6,890,707 unique items (i.e., short videos or images). These features include categorical labels such as genre or topic, which may offer further insight into the types of content that correlate with high or repetitive engagement. While our primary focus remains on user behavior, content-level analysis can complement user modeling by identifying whether certain themes—e.g., humor, drama, or visually stimulating clips—are more frequently consumed by users exhibiting overuse patterns.

By employing KuaiSAR's rich dataset structure, we aim to uncover trends and indicators of potential addiction-like behavior. The sheer scale of the data, combined with its detailed feature set, enables a realistic and scalable framework for studying "brain rot" in real-world user populations. Unlike prior studies that relied on survey-based assessments or small experimental samples, our work utilizes millions of genuine user interactions to infer behavioral patterns, increasing both the generalizability and applicability of our findings. This makes KuaiSAR a foundational component of

our research, allowing us to investigate the intersection of recommendation systems, user behavior, and digital well-being through a data-driven lens.

3.2 Proposed Method

3.2.1 Labelling Users with ”Brain Rot”

A fundamental challenge in this study lies in the absence of explicit diagnostic labels for social media addiction—or “brain rot”—within existing large-scale behavioral datasets. The KuaiSAR dataset, like other recommendation-based benchmarks, is designed to support research in recommender systems and does not include ground truth labels that indicate whether a user is addicted or cognitively affected by excessive platform use. Consequently, it becomes necessary to manually define and assign labels that serve as proxies for the concept of brain rot.

However, this process is inherently constrained by the lack of medical or psychological assessments in the dataset. There exists no authoritative signal that definitively determines which users are experiencing behavioral addiction. Without expert diagnosis or clinical ground truth, we adopt a hypothetical labeling scheme guided by behavioral patterns observed in prior literature on social media addiction. In particular, we use an existing attribute in the KuaiSAR dataset known as rec_active_level, which categorizes users into four levels (0 to 3) based on the number of days they were active on the platform. Higher levels correspond to a greater number of active days, reflecting more frequent engagement with the platform’s content.

Given that excessive daily interaction is a hallmark of social media overuse, the rec_active_level metric presents a logical candidate for approximating user addiction. Individuals with high daily platform access are more likely to exhibit compulsive scrolling behavior, reduced self-regulation, and the type of persistent consumption often described colloquially as “brain rot.” Based on this rationale, we define our hypothetical ground truth as follows: users labeled with rec_active_level 3 are considered to exhibit brain rot-like behavior, while those in levels 0 through 2 are treated as non-affected users for the purpose of binary classification.

The distribution of users across these activity levels provides a reasonably balanced label ratio. Specifically, 6,362 users fall into level 3, while levels 2, 1, and 0 contain 8,030, 7,226, and 4,259 users, respectively. Although this definition is not medically validated, it enables a scalable and behaviorally grounded approach to label generation, which is critical for training and observing machine learning models aimed to classify and analyze traits relating to activities of problematic social media use.

3.2.2 Engineering New Activity Log Features

The KuaiSAR dataset offers a robust foundation for modeling user behavior on short-form video platforms by capturing a wide range of user interactions. Among the most fundamental activity log features included are: forward, like, follow, search, click, duration_ms, and playing_time. These features reflect core user behaviors that collectively describe how content is consumed, interacted with, and potentially engaged with at a deeper level.

Each of these actions is intuitively defined. The forward action indicates when a user shares or forwards a video, typically signaling a high level of interest or a desire to distribute content further. The like action reflects explicit positive feedback, which is commonly associated with content preference or user satisfaction. The follow action indicates a decision to subscribe to or continue tracking the content from a particular creator, suggesting a deeper or more long-term form of engagement. The search action captures the act of manually inputting a query to seek specific content, which can be interpreted as goal-directed or curiosity-driven behavior. The click action measures when a user initiates a video playback, serving as a base event for subsequent interactions.

The last two features—duration_ms and playing_time—warrant further clarification. The duration_ms attribute represents the total length of the video in milliseconds, providing a fixed reference for content length. Meanwhile, playing_time records how long the user actually watched the video, also in milliseconds. The ratio between playing_time and duration_ms thus becomes an informative indicator of how attentively users are consuming content, which could relate to behavioral tendencies like skipping or binge-watching.

TABLE II. BOOL ACTION LIKELIHOOD AND NUMERIC ACTION MEAN COMPARED BETWEEN ADDICTED AND NON-ADDICTED USERS

Actions	Brain Rot	non-Brain Rot
forward (bool)	0.2123%	0.3466%
like (bool)	2.4071%	3.4128%
follow (bool)	0.3654%	0.4357%
search (bool)	0.2971%	0.4038%
click (bool)	46.8938%	50.9669%
duration_ms (ms)	94711.0808	90627.8481
playing_time (ms)	24147.1827	25399.6694

To better understand how users labeled with and without brain rot differ, we first conducted a simple descriptive analysis by calculating the average values of these features for each group. Table II shows the results of this comparison. At first glance, there appear to be observable differences in interaction patterns between Brain Rot and non-Brain Rot users. Notably, non-addicted users have slightly higher engagement in most boolean actions, such as forwarding (0.3466% vs. 0.2123%), liking (3.4128% vs. 2.4071%), following (0.4357% vs. 0.3654%), and searching (0.4038% vs. 0.2971%). Likewise, the mean click rate is higher for non-Brain Rot users (50.97%) compared to

Brain Rot users (46.89%). Interestingly, in terms of video consumption, Brain Rot users exhibit slightly lower average playing_time (24,147 ms) and higher duration_ms (94,711 ms) than their counterparts, who show 25,399 ms and 90,627 ms respectively.

Although these statistics offer an initial lens into behavioral distinctions, the overall magnitude of difference is relatively small. The boolean action differences lie mostly below a 1% margin, and even numerical features such as playing_time or duration_ms show limited divergence in means. These weak signals suggest that while default action features are useful for establishing a baseline understanding of user behavior, they may not be sufficient to form a strong basis for detecting or modeling brain rot-like tendencies.

To address this limitation, we engineered a set of composite behavioral features derived from the fundamental ones. These new metrics are designed to better capture nuanced aspects of user engagement, temporal activity, and consumption patterns. The engineered features include:

Interaction Rates: These include like_rate, follow_rate, and search_rate, each defined as the proportion of a specific action to total click count. For example, like_rate = (like count) / (click count). These rates provide normalized indicators of how often specific behaviors occur relative to general activity, helping control for individual differences in overall usage volume.

$$\text{InteractionRate}_u^{(a)} = \frac{\sum_{i=1}^{N_u} a_{u,i}}{\sum_{i=1}^{N_u} \text{click}_{u,i}}, \quad a \in \{\text{like}, \text{follow}, \text{search}\}. \tag{3.1}$$

Watch Efficiency: Defined as the ratio of playing_time to duration_ms, this feature quantifies how thoroughly users watch content. Users with low watch efficiency might be rapidly skipping through content, while high-efficiency users may be more engrossed.

$$\text{WatchEfficiency}_u = \frac{\sum_{i=1}^{N_u} \text{playing\_time}_{u,i}}{\sum_{i=1}^{N_u} \text{duration\_ms}_{u,i}}. \tag{3.2}$$

Engagement Ratio: This composite metric sums the like and follow counts and normalizes them by the click count. It reflects how interactive a user is per content viewed, offering a richer signal of affective or social engagement.

$$\text{EngagementRatio}_u = \frac{\sum_{i=1}^{N_u} (\text{like}_{u,i} + \text{follow}_{u,i})}{\sum_{i=1}^{N_u} \text{click}_{u,i}}. \tag{3.3}$$

Session Count: To model temporal clustering of activity, we define a session as a group of clicks separated by less than 30 minutes. Session count thus reflects the number of distinct viewing bursts and can distinguish between habitual short bursts versus prolonged single sessions.

$$\text{SessionCount}_u = \left|\left\{ s \in \mathscr{S}_u \mid \text{gap}(s_j, s_{j-1}) \geq 30 \text{ min} \right\}\right|. \tag{3.4}$$

Night Owl Ratio: This metric quantifies the proportion of user interactions that occur between midnight and 5:00 AM. Elevated values may suggest irregular usage habits or possible signs of dependency, especially if combined with high frequency.

$$\text{NightOwlRatio}_u = \frac{\sum_{i=1}^{N_u} \mathbb{1}_{\{00:00 \leq t_i < 05:00\}}}{N_u}. \tag{3.5}$$

Top Category Ratio: By examining the most frequently watched content category and calculating its viewing frequency over all interactions, this metric assesses content diversity. A high top category ratio may suggest obsessive consumption of particular content genres.

$$\text{TopCategoryRatio}_u = \frac{\max\limits_{c \in \mathscr{C}} \left(\sum_{i=1}^{N_u} \mathbb{1}_{\{\text{cat}_{u,i} = c\}} \right)}{N_u}. \tag{3.6}$$

User Category Ratio: This equation defines the ratio $r_{u,c}$ of interactions that user u has with content category c. The numerator $n_{u,c}$ represents the number of interactions in category c, normalized by the total interactions across all categories $\mathscr{C}$.

$$r_{u,c} = \frac{n_{u,c}}{\sum_{c' \in \mathscr{C}} n_{u,c'}}. \tag{3.7}$$

Category Entropy: This equation defines the category entropy $H_u^{(cat)}$ for a user u, which measures the diversity of content categories the user engages with. The term $p_u(c)$ represents the probability (or proportion) of interactions from user u within category c, and $\mathscr{C}$ denotes the set of all categories. A small constant 10^{-9} is added inside the logarithm to prevent numerical instability when $p_u(c) = 0$. Higher entropy values indicate a more diverse viewing pattern, while lower values suggest a strong preference for specific categories.

$$H_u^{(cat)} = -\sum_{c \in \mathscr{C}} p_u(c) \log_2 \left(p_u(c) + 10^{-9} \right). \tag{3.8}$$

Hour Entropy: This equation defines the hour entropy $H_u^{(hour)}$ for a user u, which quantifies the diversity of the user's activity across the 24 hours of a day. The term $p_u(h)$ represents the probability of user u being active during hour h. A small constant 10^{-9} is added inside the logarithm to avoid numerical issues when $p_u(h) = 0$. A higher value of $H_u^{(hour)}$ indicates that the user's activity is spread

across many hours, while a lower value suggests activity concentrated within specific hours.

$$H_u^{(hour)} = -\sum_{h=0}^{23} p_u(h) \log_2 \left(p_u(h) + 10^{-9} \right). \tag{3.9}$$

Watched Half Count: This equation defines the *watched_half_count* for a user u, which represents the total number of videos that the user watched at least halfway. For each video i in the set of N_u videos, the indicator function $\mathbb{1}(\cdot)$ counts 1 if the ratio of playing time to video duration is at least 0.5.

$$\text{watched\_half\_count}_u = \sum_{i=1}^{N_u} \mathbb{1}\left(\frac{\text{playing\_time}_{u,i}}{\text{duration\_ms}_{u,i} + 1} \geq \frac{1}{2} \right). \tag{3.10}$$

Active Ratio: Calculated as the total number of boolean actions divided by total playing_time, this feature captures action density during consumption. A user with a high active ratio is constantly interacting with the platform, which could signal compulsive behavior.

$$\text{ActiveRatio}_u = \frac{\sum_{i=1}^{N_u} (\text{like}_{u,i} + \text{follow}_{u,i} + \text{search}_{u,i} + \text{click}_{u,i} + \text{forward}_{u,i})}{\sum_{i=1}^{N_u} \text{playing\_time}_{u,i}}. \tag{3.11}$$

These engineered features provide a more expressive representation of user behavior and allow for finer-grained classification and interpretation. They also facilitate the modeling of higher-level behavioral traits—such as impulsiveness, habitual engagement, or goal-directed exploration—that may not be captured by the basic features alone. By augmenting the dataset with these additional metrics, we aim to enhance the sensitivity and interpretability of downstream machine learning models designed to detect addictive usage patterns and assess the severity of social media overexposure.

3.3 Experiment

To validate our approach to classifying users with brain rot tendencies, we conduct a comprehensive machine learning experiment using features derived from both raw activity logs and engineered behavioral metrics. The objective of this experiment is to assess the feasibility of detecting potentially addicted users using interpretable models grounded in interaction behavior, temporal activity, and content consumption patterns.

We begin by loading two primary datasets: user_features_wBR.csv, which contains user IDs alongside their assigned brain rot level (as defined in the labeling process), and feature_merged.csv, which contains granular user interaction data including timestamps and categorical labels. Before

feature construction, we parse timestamps into datetime objects and extract the date and hour to support temporal feature generation. From this parsing, we achieve features such as active_days, late_night_actions, and avg_actions_per_day to represent user-level temporal engagement.

As mentioned in Subsection 3.2.2, we engineer a suite of custom behavioral metrics from the logs and apply them for the analysis on the brain rot user prediction model. For instance, hour_entropy measures the randomness of viewing hours. Category-based features such as top_category_ratio and category_entropy capture content preferences and diversity of interest. We also encode user affinity to the top five most frequent content categories using one-hot encoding (fav_cat_* and cat_ratio_* features). Together, these features offer a multi-faceted representation of user behavior across temporal, categorical, and engagement dimensions.

Once all features are computed and merged, we remove users with missing data and assign binary labels—1 for users with brain rot level 3, and 0 otherwise. To ensure class balance, we apply down-sampling on the majority class, resulting in a balanced dataset for training. The final feature set includes base activity metrics along with the addition of engineered features.

We choose a Random Forest Classifier as the primary model due to its robustness, interpretability, and compatibility with SHAP explainability tools. The model is configured with 100 estimators and a maximum depth of 8. While we apply 5-fold cross-validation with the F1 score as a consistency check, our primary goal is not to optimize predictive accuracy. Instead, the experiment focuses on uncovering which behavioral features are most indicative of brain rot-like tendencies.

For this purpose, we split the dataset into 70% training and 30% testing using stratified sampling. Although a standard classification pipeline is followed, the emphasis is not on evaluation metrics such as precision or recall. This is because the labels are based on a hypothetical ground truth rather than clinically verified diagnoses. Rather than interpreting the model's predictive accuracy, we use the trained classifier to analyze feature importance and identify which user behaviors are most strongly associated with the hypothesized condition.

To understand which features contribute most significantly to the model's decisions, we apply SHAP. We compute SHAP values on the test set and visualize the results using a beeswarm plot. This analysis confirms the influence of features such as watch_efficiency, night_owl_ratio, and category-specific interactions, and provides transparency into how different behaviors are weighted in classifying potential brain rot users.

This experimental pipeline demonstrates that user behavioral logs—when properly engineered and interpreted—can offer measurable signals indicative of problematic usage. While limitations remain regarding ground truth reliability, the methodology presents a scalable and interpretable framework for behavioral health assessment in digital media environments.

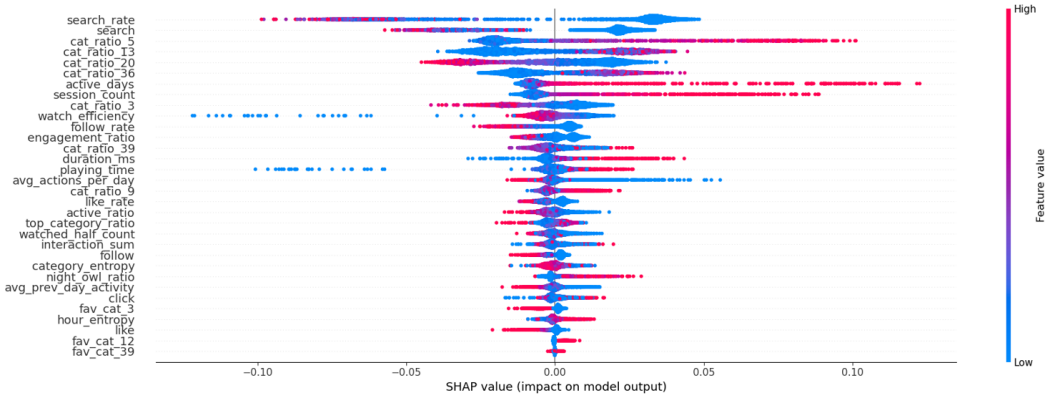

Figure 3.1. Application of SHAP on a random forest algorithm defining the features that contribute to the classification.

3.4 Results and Discussion

To gain insight into which behavioral patterns most strongly influence the model's predictions, we visualize the SHAP value distribution across all features using a beeswarm plot, as shown in Fig. 3.1. This plot illustrates both the importance and directional impact of each feature on the classification outcome. Each point represents an individual user, colored according to the corresponding feature value (red for high, blue for low), and positioned horizontally based on its SHAP value—i.e., its effect on increasing or decreasing the likelihood of being labeled as brain rot.

The most influential feature is search_rate, where low feature values strongly contribute to a brain rot prediction. This suggests that users who are less likely to perform search actions rather click or scroll through the videos, exhibiting behavior aligned with excessive platform exploration, possibly reflecting compulsive seeking of stimulation. Similarly, search, the raw count of search actions, reinforces this finding.

Several category-related features also appear prominently in the top ranks, including cat_ratio_5, cat_ratio_13, cat_ratio_20, cat_ratio_36, respectively representing topics "Style," "San Nong," "pixiv," and "Real Estate Home Furnishings." These indicate the proportion of videos consumed from particular top-level content categories. A high concentration in certain content categories may reflect narrow consumption habits or obsessive interest in specific themes—common characteristics in behavioral addiction literature. The SHAP values suggest that strong preference for these dominant categories contribute to defining the brain rot and non-brain rot users.

Other significant features include active_days and session_count, both of which reflect high-frequency platform usage and habitual re-engagement. watch_efficiency, defined as the ratio of playing_time to duration_ms, also plays a key role, where lower efficiency (i.e., skipping behavior) is associated with brain rot classification. This aligns with the observation that heavy users may

engage in fragmented or inattentive content consumption.

Engagement-related features such as follow_rate, engagement_ratio, and like_rate appear in the middle of the importance spectrum. While they provide moderate contributions, their directionality is mixed, suggesting that interaction volume alone is not a strong predictor unless contextualized with behavioral patterns like time of use or content preference.

Toward the bottom of the ranking are features such as click, hour_entropy, and some of the one-hot encoded favorite category indicators (e.g., fav_cat_3, fav_cat_12, respectively categories for ”game” and ”food”). Their low SHAP magnitudes indicate minimal influence in the final decision-making process of the classifier, despite their theoretical relevance.

Overall, the SHAP analysis confirms that high search activity, content fixation, and habitual usage patterns are among the most informative behavioral signals for identifying brain rot tendencies. It also demonstrates that while many features contribute marginally, only a select few drive the majority of model predictions. These insights validate our engineered features and support their use in modeling problematic platform use—even under hypothetical ground truth conditions.

Chapter 4

Investigation on Short Video Feature Trends Through Text Generation

4.1 Short-video Dataset and Video Captioning

4.1.1 Option for Short Videos

The analysis of user behavior and content trends in short-video platforms requires not only activity logs but also access to the video content itself. In our previous work, we rely on the KuaiSAR dataset, which provides extensive user interaction data and metadata for items such as video IDs, categories, and popularity metrics. However, KuaiSAR does not include the short-video files, which limits our ability to analyze content-level characteristics. To investigate visual traits and patterns in short-form videos, we require a dataset that combines both user behavior logs and the actual video content.

Such datasets are scarce due to privacy concerns, copyright restrictions, and the large storage requirements of short-video files. After surveying available resources, the dataset "A Large-scale Dataset with Behavior, Attributes, and Content of Mobile Short-video Platform" by Shang et al. emerges as the most viable option within the publicly available datasets [22]. This dataset contains not only detailed user activity data and video attributes but also includes 153,561 short-video files, making it a valuable resource for content-based analysis. The dataset offers a wide variety of features such as user-item interaction logs, video metadata, and categorical labels, which together provide a comprehensive view of user preferences and platform dynamics.

Our goal is to utilize this dataset to uncover content-related traits that might correlate with addictive or repetitive user engagement. However, to translate the visual information in these videos into a form that can be analyzed alongside behavioral logs, we require a mechanism for generating

textual descriptions of video content. We approach this task through video captioning techniques. At this stage, we briefly note that our focus is on extracting representative descriptions of videos using automated caption generation; the details of this process will be explored in depth in the subsequent section.

4.1.2 Video Captioning Approaches

Video captioning is an active research area that has achieved remarkable progress in recent years, with several state-of-the-art (SOTA) models dominating benchmark evaluations. The most prominent ones are mPLUG-2, MaMMUT, and VideoCoCa.

- mPLUG-2 is a multimodal foundation model that integrates image, text, and video tasks using a modular transformer-based architecture with modality-specific encoders [23]. It is pre-trained on large-scale image-text and video-text data and evaluated on datasets such as MSR-VTT, MS COCO, Flickr30k, and Kinetics-400, achieving state-of-the-art results in video captioning and QA.
- MaMMUT uses a simple vision-language design with a single vision encoder and a text decoder, trained jointly for contrastive image-text alignment and generative captioning [24]. It extends seamlessly to video tasks with spatio-temporal tokens and performs strongly on MSR-VTT, MSVD, and VQAv2 benchmarks.
- VideoCoCa adapts Google's CoCa model to video by processing sampled frames through the original image-text transformer architecture with minimal changes [25]. It achieves strong results on MSR-VTT, ActivityNet Captions, VATEX, and YouCook2, excelling in video captioning, QA, and retrieval tasks.

While these models are powerful, they come with substantial computational costs. Running these video captioning pipelines on large datasets requires both high-end GPUs and significant inference time. The short-video dataset we use, which contains 153,561 videos with a combined size of approximately 3 TB, poses a major challenge in this regard. Even if the dataset were split into smaller subsets—such as 1/3 of the total—it would still amount to nearly 1 TB of data, making the direct use of heavy video captioning models computationally impractical in our environment.

To mitigate these computational demands, we opt for a lighter yet effective approach: utilizing an image captioning model instead of full-scale video captioning. Specifically, we adopt BLIP-2 (Bootstrapping Language-Image Pre-training) [26], which is a cutting-edge vision-language model capable of generating accurate textual descriptions from single video frames. By sampling key frames from each video and applying BLIP-2, we can approximate video captions while avoiding

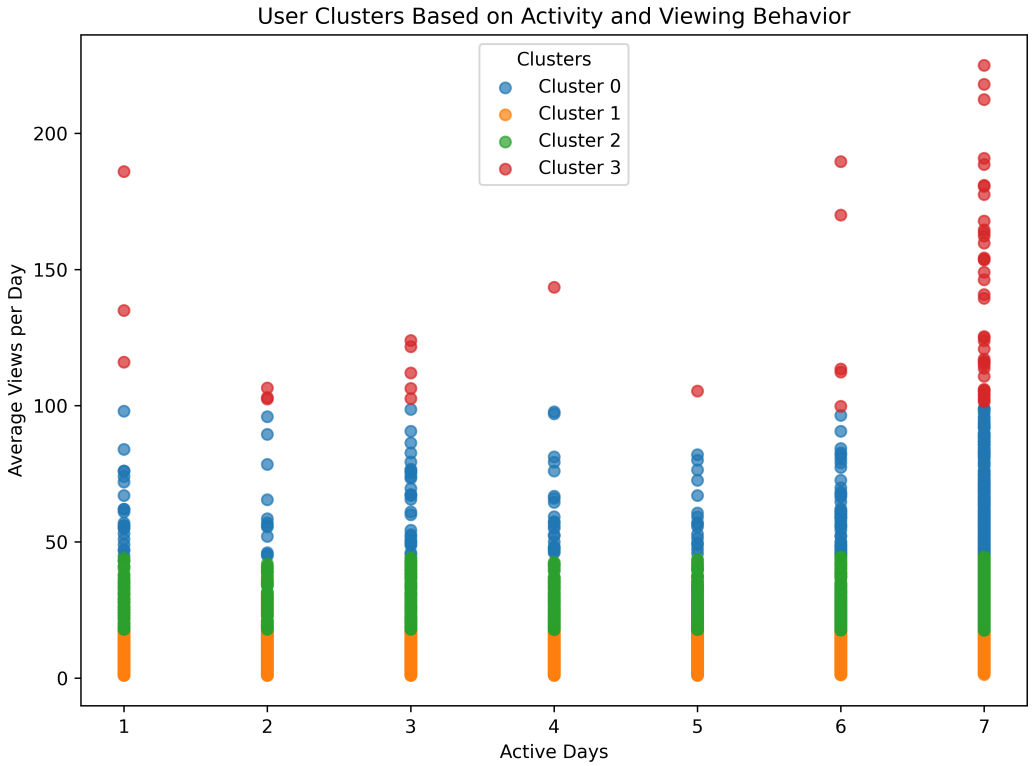

Figure 4.1. Graph of cluster classification output of social media users from short-Video dataset [22].

the time-intensive process of frame-by-frame temporal modeling required by SOTA video caption models. This trade-off allows us to handle the scale of our dataset efficiently while still extracting meaningful textual representations of video content for subsequent analysis.

This strategy offers a practical balance between computational feasibility and content interpretability. In the following sections, we detail how we implement this image-based captioning pipeline and how the generated captions are integrated into our broader investigation of short-video trends.

4.2 Proposed Method

To investigate content-related indicators of addictive tendencies, we first establish labels indicating whether each video is associated with “brain rot” behavior. Similar to our previous work with user-level activity data, this short-video dataset—being primarily designed for recommendation system research—does not include any labels indicating addictive or compulsive consumption. Therefore, we construct a hypothetical ground truth by analyzing user activity patterns and clustering users based on their engagement metrics.

Since the dataset by Shang et al. does not provide an active_level attribute per user, we derive behavioral clusters using the number of interactions, specifically the average views per day. In addition, we also consider the active days of each user. As shown in Fig. 4.1, applying a clustering algorithm, we form four distinct user groups that represent different levels of platform activity. We focus on two representative clusters for comparison:

- Cluster 3: The most active cluster, which includes 57 users who exhibit the highest number of interactions and activity days,
- Cluster 1: The least active cluster, which consists of 4,761 users with minimal platform engagement.

We label the users in the most active cluster as "brain rot" users and those in the least active cluster as "non brain rot" users. The "brain rot" group has collectively watched 34,316 videos, while the "non brain rot" group has watched 65,832 videos. For the purpose of binary classification, we assign the videos viewed by the "brain rot" cluster to the addictive class, while the videos from the "non brain rot" cluster are assigned to the non-addictive class.

However, certain videos are consumed by both user groups, which introduces potential noise in our labeling. To address this, we identify overlapping videos—those watched by both addicted and non-addicted users—and exclude them from the dataset. As a result, the set of videos labeled as addictive is reduced from 34,316 items to approximately 15,000 unique videos. This filtering step ensures that the content associated with the "brain rot" group is distinct, allowing for a cleaner comparison against the non-addictive content.

Once the video subsets are finalized, we utilize BLIP-2, a lightweight yet state-of-the-art image captioning model, to generate textual descriptions of video content. Since BLIP-2 operates on single frames rather than entire video sequences, we adopt a key-frame sampling strategy, extracting the first, middle, and last frame from each video. For each selected frame, we prompt BLIP-2 with an instruction to "describe the current scene," resulting in three captions per video. These captions collectively provide a concise textual summary of the video's content.

To prepare the caption data for analysis, we apply text filtering and processing to reduce excessive repetition and remove artifacts such as redundant phrases. This refinement step ensures that the generated textual descriptions are both clean and semantically rich, which is essential for subsequent feature extraction and trend analysis.

4.3 Experiments

As mentioned earlier in the Proposed Method section, we prepare a balanced dataset consisting of 15,000 videos labeled as addictive and 15,000 videos labeled as non-addictive. These videos are a subset of the short-video dataset by Shang et al., which we download directly from their hosted server. The downloading process is computationally intensive; retrieving all 30,000 videos required for the experiment takes approximately one full day, consuming around 700–800 GB of storage. If we are to download the complete set of 153,561 videos, it requires roughly 5–7 days and multiple terabytes of disk space.

For video captioning, we utilize the BLIP-2 model (as described in the Proposed Method), running inference on the first, middle, and last frame of each video to generate three descriptive captions per video. This process is performed on a NVIDIA RTX 2080Ti GPU with 8 GB of VRAM, which provides sufficient computational resources to handle batch processing of frames at scale. BLIP-2's relatively lightweight architecture enables the captioning of 30,000 videos within a reasonable timeframe, avoiding the excessive overhead associated with heavy state-of-the-art video captioning models.

Once the captions are generated, we analyze them using a TF-IDF (Term Frequency–Inverse Document Frequency) approach to quantify the importance of words across the addictive and non-addictive video groups. As part of this analysis, we train a logistic regression model to classify videos based on their captions. However, as stated previously, accuracy, precision, and other classification metrics are not central to our study, since we are working with hypothetical ground truth labels rather than clinically validated or objectively confirmed data. The purpose of this model is not performance benchmarking but rather identifying linguistic features that are more prominent in addictive versus non-addictive content.

For training, we split the data into 70% training and 30% testing sets, ensuring that captions from the same video are not leaked across both splits. The text features are extracted using TF-IDF vectorization, where we configure TfidfVectorizer with lowercasing enabled, English stop-word removal, and a maximum of 1,000 features to capture the most significant terms across the caption dataset. After vectorization, we train a logistic regression classifier with a maximum of 1,000 iterations (max_iter=1000) and set class_weight="balanced" to compensate for any potential class imbalance between addictive and non-addictive videos. The trained model is then evaluated using the test set, and while metrics such as accuracy, precision, and recall are computed, they are not central to our analysis since we work with a hypothetical ground truth rather than validated labels.

Additionally, the script includes SHAP value analysis to interpret feature importance. After fitting the logistic regression model, SHAP values are computed on the test set to rank TF-IDF

features by their influence on the model's predictions. The top features (i.e., words with the highest mean absolute SHAP values) are then visualized in a summary plot for further analysis

To complement the TF-IDF analysis, we apply Latent Dirichlet Allocation (LDA) to extract high-level thematic structures from the video captions, providing an unsupervised perspective on recurring narrative and visual elements in both addictive and non-addictive content. The captions are preprocessed by converting text to lowercase, tokenizing, and removing English stopwords and non-alphabetic tokens. We construct a dictionary and bag-of-words corpus, filtering out terms that appear in fewer than five captions or in more than 50% of captions, thereby eliminating extremely rare and overly generic words that add little semantic value. The LDA model is trained with 10 latent topics, using a fixed random seed for reproducibility and 10 passes to ensure stable convergence. For each caption, we assign a dominant topic and analyze the distribution of these topics across the two classes, supported by visualizations of topic frequencies and normalized proportions.

The choice of 10 topics reflects a balance between thematic granularity and interpretability, as preliminary experiments with fewer topics such as 6 topics had difficulties in capturing the diversity of themes present in the data, while a larger number of topics led to overly fragmented or redundant themes. Similarly, the no_below=5 and no_above=0.5 thresholds were selected to mitigate noise from rare words and remove high-frequency terms that lack discriminative power. We also compute the difference in topic prevalence between the two classes, ranking topics by absolute skew to highlight those that are overrepresented in addictive content. This analysis offers a complementary lens to TF-IDF, enabling a deeper understanding of the thematic patterns that may correlate with compulsive user engagement.

Since the dataset also has the texts of Auto Speech Recognition (ASR) for each videos, we also test these files as input for TF-IDF analysis and LDA topic analysis.

4.4 Results and Discussion

4.4.1 TF-IDF

As shown in Fig. 4.2, the SHAP results for the TF-IDF analysis of captions reveal both notable strengths and clear limitations. On the positive side, several high-ranking words provide meaningful distinctions between addictive and non-addictive videos. These terms often reflect visually descriptive elements or recurring thematic content that may characterize short videos with high engagement. The top words identified by SHAP exhibit strong influence scores, indicating that the TF-IDF-based logistic regression model captures some degree of discriminative signal from the textual data. However, a key weakness is that many of the top features are generic or overly broad terms, which might

not strongly correlate with addictive content patterns. This suggests that while caption-based TF-IDF features offer interpretability, their predictive power is constrained by the lack of temporal or semantic context beyond individual keywords.

In contrast, the SHAP results for the ASR-generated transcripts present a slightly different behavior, though similar strengths and weaknesses persist. Looking at Fig. 4.3, the top-ranked words extracted from speech recognition output tend to reflect spoken cues, dialogue snippets, or narrative elements present in the videos. While some of these words seem relevant for distinguishing addictive content, particularly when tied to conversational or repetitive speech patterns, others appear to be noise or filler terms with limited semantic value. This introduces additional challenges, as ASR transcripts are often prone to errors, especially when dealing with noisy audio or background music commonly found in short videos. Nevertheless, the ASR-based features still provide complementary insights that are not always captured by the visual captions, adding a layer of linguistic nuance to the analysis.

The nature of these SHAP results stems largely from the characteristics of TF-IDF and the nature of the data. Since TF-IDF emphasizes word frequency without capturing context or word order, both caption and ASR analyses tend to highlight words that frequently appear in one class (addictive vs. non-addictive) while neglecting higher-level semantics. Moreover, BLIP-2 captions are generated by an image captioning model that focuses on object and scene descriptions, while ASR captures spoken language, which may include storytelling or commands. The combination of limited linguistic depth and variability in video content leads to a sparse, keyword-driven feature space where certain words stand out, but the overall discriminative signal remains weak.

When comparing the SHAP graphs for captions and ASR, we observe minimal structural differences. Both exhibit the characteristic “T-shape” distribution of SHAP values, where most features cluster near zero, with a few influential words creating extended horizontal spreads. The primary difference lies in the specific set of words that dominate each graph, reflecting the modality of origin—visual descriptions for captions versus spoken language for ASR. Beyond these variations in word content, the overall patterns and distribution of SHAP values remain similar, underscoring the shared limitations of keyword-based feature extraction in both approaches.

Overall, these findings highlight both the utility and the constraints of relying on TF-IDF-based textual features for modeling addictive short-form video content. While SHAP analysis provides interpretable evidence of which words influence the classification, the results underscore the necessity of incorporating richer contextual and multimodal features to improve predictive performance. The limitations observed, particularly the overemphasis on generic or context-independent words suggest that future work could benefit from advanced language representations such as word embeddings or transformer-based encoders, which capture semantic relationships and sequential dependencies be-

yond isolated terms. Moreover, expanding the analysis with alternative prompts, such as generating descriptions of video dynamics, whether the content is visually flashy, fast-paced, or intensive in movement may reveal behavioral or stylistic patterns that correlate more strongly with addictive content. Integrating these enhanced textual features with temporal activity patterns and visual cues could lead to a more robust and holistic understanding of the signals underlying addictive content.

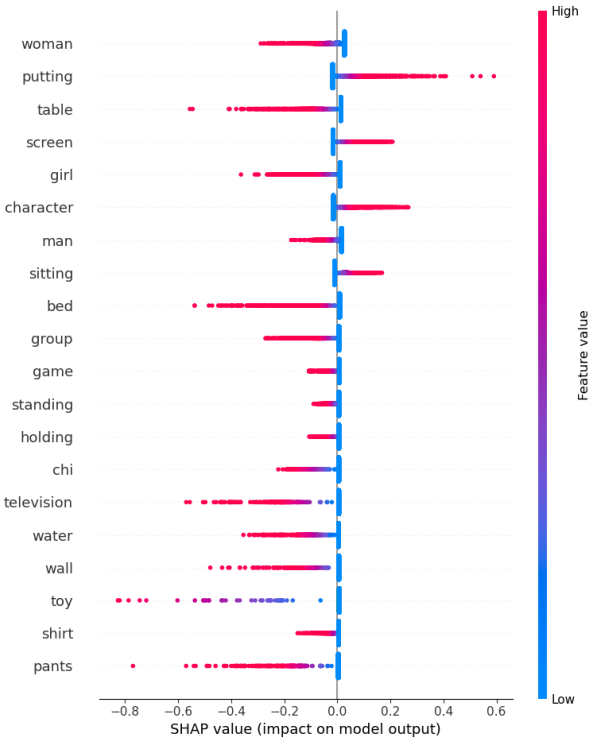

Figure 4.2. SHAP result of TF-IDF analysis on BLIP-2-driven video caption generation.

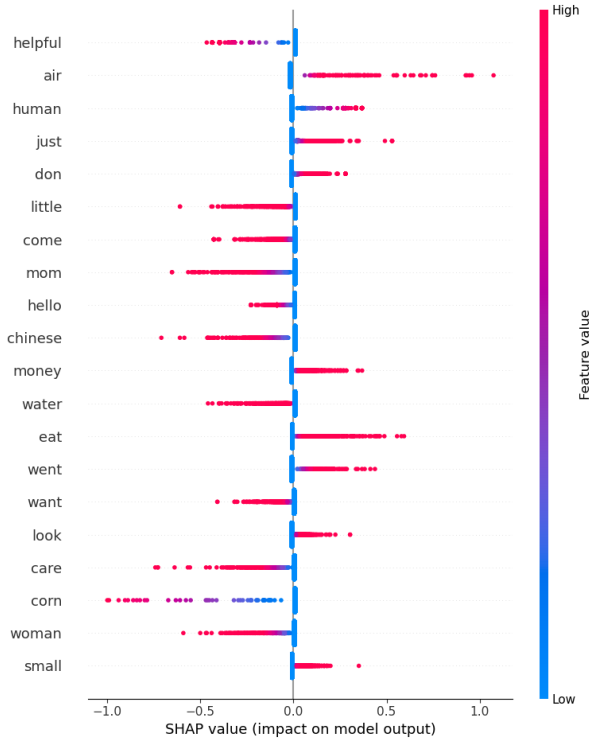

Figure 4.3. SHAP result of TF-IDF analysis on provided ASR text.

4.4.2 Latent Dirichlet Allocation (LDA)

We apply LDA to uncover dominant topics present in both the BLIP-2 captions and the ASR-generated transcripts. Table III summarizes the top keywords associated with each topic, while Fig. 4.4 and Fig. 4.5 illustrate the topic distributions for non-addicted and addicted classes. These results highlight recurring themes in both visual descriptions and spoken content, providing a higher-level perspective on content trends.

TABLE III. LDA TOPICS WITH TOP KEYWORDS OF ASR AND CAPTIONS

Topic	Top Keywords (ASR)	Top Keywords (Captions)
0	one, look, like, good, big	holding, putting, person, hand, phone
1	first, game, also, new, two	girl, hat, boy, asian, pink
2	china, chinese, us, people, yang	car, street, city, walking, driving
3	yuan, money, buy, boss, video	game, screenshot, screen, character, showing
4	man, woman, mother, old, father	shirt, white, black, blue, red
5	eat, water, food, small, fish	sitting, bed, table, chair, baby
6	love, heart, like, life, song	standing, front, sexy, suit, tie
7	people, good, many, like, person	chinese, poster, chi, character, characters
8	go, come, let, going, want	group, standing, front, screen, people
9	hello, said, could, would, human	bowl, eating, cup, plate, preparing

For the ASR topics, Topic 0 is associated with descriptions of size or appearance, including words such as “one,” “look,” and “good.” Topic 1 focuses on games and comparisons, with terms

like “first,” “game,” and “new.” Topic 2 is related to countries and people, featuring references to “china,” “chinese,” and “us.” Topic 3 emphasizes money and commerce, while Topic 4 highlights family and gender roles with terms like “mother” and “father.” Topics 5 and 6 correspond to food and eating and love or emotions, respectively. Topic 7 represents general characteristics of people, while Topic 8 focuses on actions and movements (e.g., “go,” “come,” “want”). Finally, Topic 9 captures conversations or greetings, including words like “hello,” “said,” and “would.”

For the caption topics, we observe a slightly different thematic structure. Topic 0 centers around handling objects such as “holding,” “hand,” and “phone.” Topic 1 is focused on people and appearance, including references like “girl,” “boy,” and “asian.” Topic 2 highlights vehicles and urban scenes, with terms such as “car,” “street,” and “walking.” Topic 3 is strongly linked to games and screenshots, while Topic 4 focuses on clothing and color themes. Topics 5 and 6 refer to indoor scenes or family settings (e.g., “sitting,” “bed,” “baby”) and standing or posing scenes (e.g., “front,” “suit,” “tie”). Topic 7 stands out with words indicating Chinese text or visual media, while Topic 8 focuses on group activities or TV/stage performances. Topic 9 captures food and kitchen scenes, with words like “bowl,” “eating,” and “preparing.”

Analyzing the distributions shown in Fig. 4.4 and Fig. 4.5, we see that the dominant topics are fairly evenly spread between addictive and non-addictive content. For captions, topics such as 3 (games) and 7 (Chinese text) show slightly higher frequency in addictive content, while ASR results highlight topics 3 (money) and 9 (conversation) as more frequent in addictive videos. This suggests that addictive content may lean toward videos depicting social interaction, active instructions, or games, while non-addictive content includes similar themes but with marginally different proportions.

The general shape of the topic distributions remains consistent between captions and ASR, showing no significant structural difference beyond the specific keywords associated with each topic. This consistency reinforces the observation that both modalities, visual captions and audio transcripts, capture similar underlying content themes. However, ASR adds a conversational dimension (e.g., greetings, commands) that is largely absent in captions, while captions tend to emphasize objects, attire, and visual attributes.

In summary, the LDA-based topic modeling of captions and ASR transcripts provides valuable insights into the recurring themes and narrative elements present across both addictive and non-addictive videos. While the overall topic distributions remain relatively balanced between the two classes, subtle and slight differences such as the prominence of social interactions, action-oriented content, games and money in addictive videos indicate potential behavioral and contextual cues that could inform content characterization. The complementary nature of captions and ASR transcripts further suggests that combining visual and conversational topics may yield a richer understanding

of video dynamics. Future work could explore more descriptive prompts or scene-level annotations, such as identifying whether a video is fast-paced, flashy, or emotionally charged, to capture stylistic patterns beyond static keywords. Such refinements could enhance the interpretability and discriminative power of topic-based features when analyzing addictive video content.

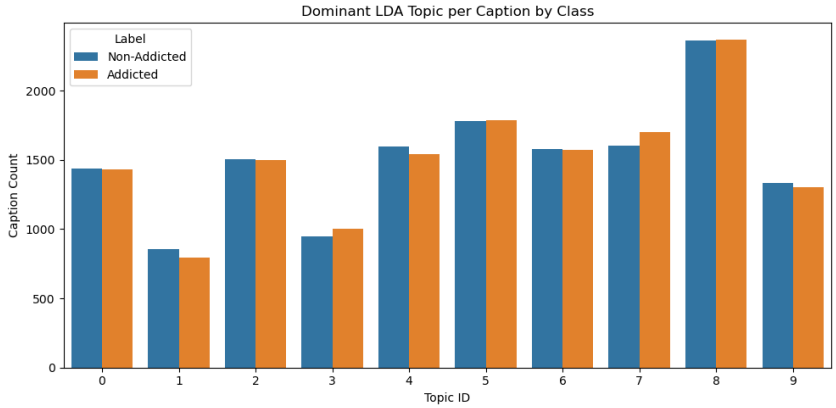

Figure 4.4. A graph showing the dominant LDA topic observed for each BLIP-2-generated caption. Please refer to Table III for the information about each Topic ID.

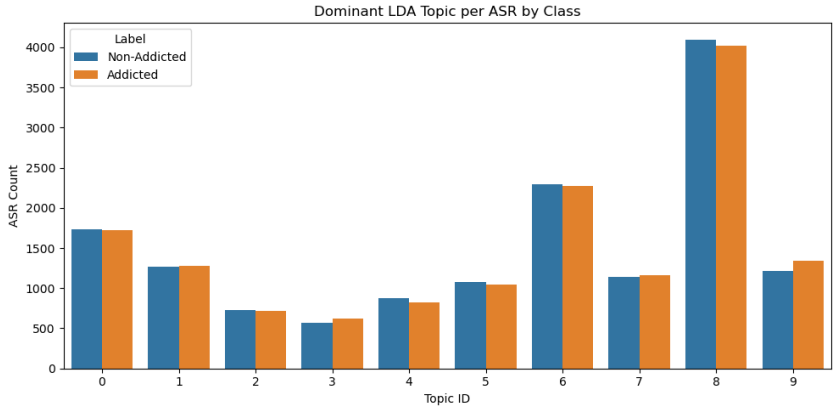

Figure 4.5. A graph showing the dominant LDA topic observed for each ASR provided from the dataset. Please refer to Table III for the information about each Topic ID.

Chapter 5

Revisiting Activity Log Analysis for Brain Rot Classification

5.1 Configuring a Different Hypothetical Label for Addiction

In Chapter 3, the classification of users in the KuaiSAR dataset into "brain rot" and "non-brain rot" categories was primarily based on the recommendation-mode activity level, which relied on the total number of active days recorded for each user. While this labeling strategy provided an initial baseline for identifying highly engaged users, it assumes that the number of days a user accesses Kuaishou directly reflects their level of addictive behavior. However, this assumption may not always hold true. For example, a user might log into the platform on a large number of days but consume only a small number of videos per session—potentially fewer than three digits per day—indicating a less intense engagement level despite the high number of active days.

To address this limitation, we propose a revised labeling approach that takes into account both average views per day and active days, in a manner consistent with the feature adjustments discussed in Chapter 4. By combining these two metrics, we aim to capture not just the frequency of platform access but also the intensity of content consumption during each active session. This approach provides a more nuanced characterization of user behavior, as it distinguishes between users who frequently check the platform for short periods and those who engage in extended binge-watching sessions, which are more indicative of brain rot tendencies.

To implement this revised labeling strategy, we partition the users into four clusters based on their joint distribution of average views per day and total active days. This clustering allows us to identify distinct user engagement patterns, ranging from sporadic viewers with low daily consumption to highly active users with consistently large volumes of viewed content. Similar to the original framework, we continue to define the most active cluster—those users exhibiting both high daily

views and frequent activity—as representing the "brain rot" group. The remaining three clusters, which reflect lower levels of engagement and consumption intensity, are classified as non-brain rot.

This redefinition of labels provides a more realistic approximation of addictive behavior, as it accounts for both platform access frequency and content consumption volume. It also aligns better with the behavioral dynamics observed in real-world social media usage, where addiction is typically characterized by not just how often a user checks the platform but also by how much content is consumed during each session. This refined labeling method serves as the foundation for the subsequent activity log analyses, allowing for a more accurate exploration of user behaviors and patterns that may signal problematic or compulsive usage.

5.2 Experiments

To validate the revised labeling approach introduced in this chapter, we conduct a machine learning experiment that mirrors the methodology outlined in Chapter 3, with only minor adjustments to accommodate the new label configuration. The objective remains to evaluate the potential of behavioral and temporal features in distinguishing users classified under the updated brain rot label.

We utilize the same two primary datasets: user_features_wBR2.csv, which contains user IDs alongside the new binary brain rot labels based on average daily views and active days, and feature_merged.csv, which provides detailed user interaction logs. As before, timestamps are parsed into date and hour components to generate temporal features such as active_days, late_night_actions, and avg_actions_per_day. We also retain the engineered behavioral metrics from Chapter 3, including hour_entropy, category_entropy, and category affinity ratios.

The labeling adjustment alters the class distribution slightly, but we maintain a balanced dataset by down-sampling the majority class. The most active cluster of users, as defined by the joint distribution of average daily views and active days, is assigned a label of 1 (brain rot), while all other clusters are labeled as 0 (non-brain rot).

We again employ a Random Forest Classifier with 100 estimators and a maximum depth of 8, as this configuration offers both robust performance and compatibility with SHAP-based feature explainability. A 70/30 train-test split with stratified sampling is used, ensuring proportional representation of both classes. As in Chapter 3, we prioritize interpretability over raw predictive performance, using SHAP to identify the features that most strongly influence the classification.

5.3 Results and Discussion

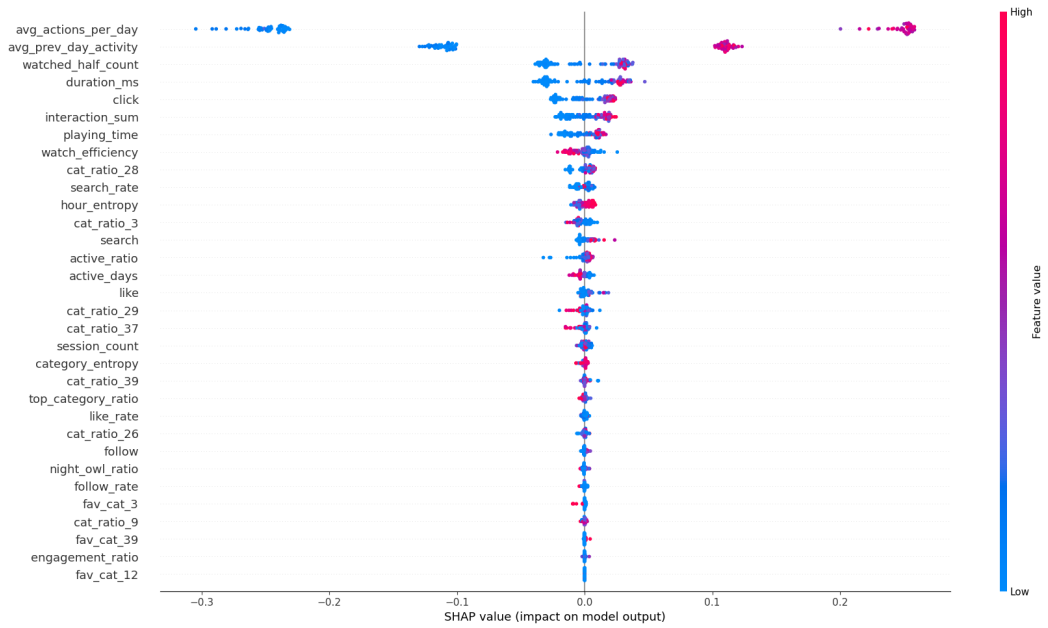

Figure 5.1. SHAP result of the user activity features with new brain rot labels (including watched half count and duration_ms).

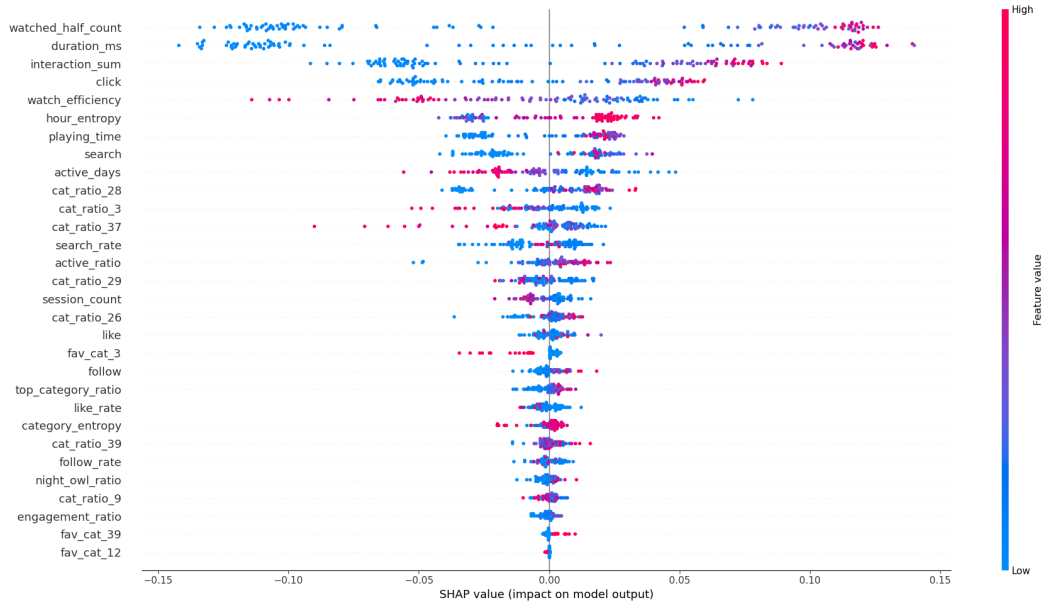

Figure 5.2. SHAP result of the user activity features with new brain rot labels (excluding watched half count and duration_ms).

Fig. 5.2 illustrates the SHAP value distribution of features derived from user activity logs under the newly configured brain rot labels. Compared to Fig. 3.1, which represents the earlier labeling method based solely on active days, we observe notable shifts in the relative importance of

behavioral and category-based features. In particular, click, watched_half_count, duration_ms, and interaction_sum emerge as dominant predictors in Fig. 5.2, suggesting that the revised labels place stronger emphasis on engagement intensity and content consumption depth rather than merely the frequency of access.

The cat_ratio_XX features, which reflect the proportion of videos from specific content categories viewed by a user, also play a significant role. For example, cat_ratio_28 corresponds to real-time information content, while cat_ratio_3 represents game-related videos. Similarly, cat_ratio_37, cat_ratio_29, and cat_ratio_26 correspond to strange people and strange phenomena, science, and photography, respectively. The relatively high and low feature values of these features in Fig. 5.2 suggest that users labeled as brain rot tend have the tendency to primarily watch some contents while avoiding others. Such findings can be an indicator for repetitive or binge-like viewing patterns.

When comparing Fig. 3.1 and Fig. 5.2, one clear difference is the elevated role of engagement-based features (e.g., watched_half_count and interaction_sum) in 5.2, whereas Fig. 3.1 showed a stronger focus on search-related and session-based metrics. For instance, in Fig. 3.1, search_rate and various cat_ratio features dominated the top positions, implying that the earlier model captured more exploratory behavior rather than sustained engagement for classifying the users. The current analysis suggests that once labels account for average daily views alongside active days, the model prioritizes depth of interaction (e.g., total watch duration, partial video completion) as stronger indicators of brain rot behavior.

Additionally, watch_efficiency remain consistently important across both figures, indicating that temporal viewing patterns and the ability to sustain attention on videos are robust behavioral markers regardless of labeling strategy. However, Fig. 5.2 demonstrates a more balanced contribution between temporal features and content category features, which may imply that the revised labeling better captures a holistic picture of user behavior. Furthermore, hour_entropy appears to be a stronger indicator for the new labels, noting that users with brain rot relentlessly watches contents on any hours.

In conclusion, the comparison between Fig. 3.1 and Fig. 5.2 highlights a shift in emphasis from exploratory metrics (e.g., search behaviors) to features representing viewing depth, interaction consistency, and focused content preferences. While features such as search_rate and session_count lost some influence under the new labeling, indicators like watched_half_count, click, and hour_entropy gained prominence, reflecting the importance of intensive and constant viewing sessions as a core element of brain rot classification. This adjustment in feature importance validates the decision to incorporate average daily views into the labeling process, as it leads to a feature landscape more aligned with binge-like user behavior patterns.

Chapter 6

Conclusion

6.1 Summary

This thesis explored the potential of identifying social media addiction, referred to as brain rot, through the analysis of large-scale activity logs and video content features. The findings demonstrate that extracting activity log features relevant to brain rot tendencies was highly effective. By adopting engineered behavioral metrics and explainable machine learning models, such as the Random Forest classifier with SHAP-based interpretation, we successfully uncovered patterns that distinguish between addicted and non-addicted users. The SHAP plots and values revealed clear behavioral contrasts between the two groups, highlighting key indicators such as watch duration, interaction depth, and content category preferences.

A key refinement of the study was the relabeling of hypothetical ground truth based on both average daily views and active days, rather than solely on active days. This adjustment emphasized the behavioral differences between user groups, validating the relevance of more intensity-driven metrics. The relabeling and subsequent reanalysis proved meaningful, as the revised SHAP feature rankings better captured signals of binge-like consumption, making the retest of user activity logs both worthwhile and insightful.

In contrast, the video caption analysis utilizing TF-IDF and SHAP for feature interpretation proved less effective than the activity log approach. While the predictive power of caption-based features was relatively weak, some patterns emerged through discriminative TF-IDF terms, which pointed to recurring themes or descriptive words associated with addictive videos. The SHAP values, although subtle, suggested that certain keywords had consistent, albeit limited, influence on classification.

To complement this, the LDA topic evaluation provided a broader view of thematic trends across

captions and ASR-generated transcripts. While the differences between addicted and non-addicted content were not drastic, a handful of topics—such as food preparation, group activities, and fast-paced or visually dynamic content—appeared more frequent among addictive videos. At the same time, the analysis underscored that both groups shared a considerable overlap in the types of topics viewed, suggesting that addiction may arise not only from specific content themes but also from viewing patterns and intensity.

Lastly, the comparative analysis of ASR transcripts and BLIP-2 captions highlighted the complementary strengths and weaknesses of both modalities. Captions provided clear visual descriptions but lacked conversational context, while ASR transcripts captured dialogue and narrative cues at the cost of noise and transcription errors. Both approaches exhibited similar limitations in terms of keyword-driven feature extraction, reinforcing the notion that textual features alone may not sufficiently capture the complexity of addictive video content.

6.2 Future Work

While this thesis demonstrates promising results in identifying brain rot tendencies through user activity logs and textual analysis, several avenues remain for further exploration and refinement. One immediate improvement lies in the labeling strategy. The current approach, though effective, relies on hypothetical ground truth based on engagement metrics such as average views and active days. Future research could integrate additional behavioral signals—such as session duration, content revisit rates, or user feedback—to construct more robust and realistic labels. Collaborating with psychological studies or incorporating survey-based assessments could also provide a partial ground truth to validate machine learning outputs.

Another key direction is the enhancement of textual feature modeling. While TF-IDF provided a baseline for caption and ASR analysis, its bag-of-words nature fails to capture deeper semantics or context. Transformer-based models such as BERT or CLIP-text encoders could be explored to represent captions and transcripts with richer contextual embeddings. Additionally, prompt engineering for image or video captioning models could focus on generating descriptive cues about visual style, pacing, or emotional tone (e.g., whether a video is “fast-paced,” “flashy,” or “intense”), potentially offering more direct signals of addictive content.

Expanding into multimodal analysis also presents a valuable opportunity. By combining user activity logs with video-based features—such as motion intensity, editing patterns, or visual dynamics—future studies could better understand how both user behavior and content characteristics jointly influence addiction tendencies. The inclusion of audio cues, sentiment analysis of speech, or engagement signals like comment activity could further strengthen the holistic view of user interac-

tion.

Lastly, future work could investigate temporal evolution of brain rot behavior by modeling how user patterns change over time. Sequential models, such as recurrent neural networks or transformers tailored for time-series data, could capture the progression of viewing habits and identify early warning indicators of excessive engagement. By pursuing these enhancements, future research can build on this thesis to achieve a deeper, more accurate understanding of social media addiction and its underlying mechanisms.

List of Publications

International Conferences

1. **Kein Yamada** and Hiroshi Watanabe, ”Brain Rot Detection in Users: Analyzing Behavioral Insights from Large-Scale User Interactions,” 2025 IEEE 14th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 2025 (paper accepted)

2. Takahiro Shindo, **Kein Yamada**, Taiju Watanabe and Hiroshi Watanabe, ”Image Coding For Machines With Edge Information Learning Using Segment Anything,” 2024 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 2024, pp. 3702-3708.

3. Takahiro Shindo, **Kein Yamada**, Taiju Watanabe and Hiroshi Watanabe, “Object Detection Method for Drone Videos Using Optical Flow,” IIEEJ International Conference on Image Electronics and Visual Computing (IEVC2024), Tainan City, Taiwan, 2024, LBP-05.

4. **Kein Yamada**, Takahiro Shindo and Hiroshi Watanabe, ”Data Augmentation with 3D-rendered Models for Livestock Recognition Using Drone Footage,” IIEEJ International Conference on Image Electronics and Visual Computing (IEVC2024), Tainan City, Taiwan, 2024, LBP-29.

5. Takahiro Shindo, Taiju Watanabe, **Kein Yamada** and Hiroshi Watanabe, ”Image Coding for Machines with Object Region Learning,” 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2024, pp. 1040-1041.

6. Takahiro Shindo, Taiju Watanabe, **Kein Yamada** and Hiroshi Watanabe, ”VVC Extension Scheme for Object Detection Using Contrast Reduction,” 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE), Nara, Japan, 2023, pp. 1097-1098.

7. Taiju Watanabe, Takahiro Shindo, **Kein Yamada** and Hiroshi Watanabe, ”Future Object Detection Using Frame Prediction,” 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE), Nara, Japan, 2023, pp. 944-945.

8. **Kein Yamada**, Takahiro Shindo, Taiju Watanabe and Hiroshi Watanabe, "Accuracy Consistency of Object Detection With Contrast Reduction by Pixel Value Limitation," 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE), Nara, Japan, 2023, pp. 1101-1104.

9. Takahiro Shindo, Taiju Watanabe, **Kein Yamada** and Hiroshi Watanabe, "Accuracy Improvement of Object Detection in VVC Coded Video Using YOLO-v7 Features," 2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, 2023, pp. 247-251.

Bibliography

[1] DataReportal, "Digital 2024: Global overview report," *DataReportal*, Apr. 2024. [Online]. Available: https://datareportal.com/reports/digital-2024-global-overview-report [Accessed: Jul. 17, 2025]

[2] J. Park and Y. Jung, "Unveiling the dynamics of binge-scrolling: A comprehensive analysis of short-form video consumption using a stimulus-organism-response model," *Telemat. Inform.*, vol. 95, 2024. doi: 10.1016/j.tele.2024.102200

[3] U. Ilyas, R. Maqsood, H. Ahsan, A. Rauf, and S. Afzal, "Binge-Watching as Behavioral Addiction: A Systematic Review", *APR*, vol. 2, no. 2, pp. 39-65, Dec. 2023.

[4] H. Al-Samarraie, K.-A. Bello, A. I. Alzahrani, A. P. Smith, and C. Emele, "Young users' social media addiction: Causes, consequences and preventions," *Inf. Technol. People*, vol. 35, no. 7, pp. 2314–2343, 2022. doi: 10.1108/ITP-11-2020-0753

[5] Oxford University Press, "Oxford Word of the Year 2024: Brain rot," *Oxford Languages*, Dec. 2024. [Online]. Available: https://corp.oup.com/word-of-the-year/ [Accessed: Jul. 17, 2025]

[6] S. Rajeshwari and S. Meenakshi, "The age of doom scrolling – Social media's attractive addiction," *J. Educ. Health Promot.*, vol. 12, Art. no. 21, Jan. 2023. doi: 10.4103/jehp.jehp_838_22

[7] D. J. Yu, Y. K. Wing, T. M. H. Li, and N. Y. Chan, "The impact of social media use on sleep and mental health in youth: A scoping review," *Curr. Psychiatry Rep.*, vol. 26, no. 3, pp. 104–119, 2024. doi: 10.1007/s11920-024-01481-9

[8] D. De, M. El Jamal, E. Aydemir, and A. Khera, "Social media algorithms and teen addiction: Neurophysiological impact and ethical considerations," *Cureus*, vol. 17, no. 1, Art. no. e77145, 2025. doi: 10.7759/cureus.77145

[9] A. R. Dores, M. Peixoto, C. Fernandes, A. Marques, and F. Barbosa, "The effects of social feedback through the 'like' feature on brain activity: A systematic review," *Healthcare (Basel)*, vol. 13, no. 1, Art. no. 89, 2025. doi: 10.3390/healthcare13010089

[10] M. Gupta and A. Sharma, “Fear of missing out: A brief overview of origin, theoretical underpinnings and relationship with mental health,” *World J. Clin. Cases*, vol. 9, no. 19, pp. 4881–4889, 2021. doi: 10.12998/wjcc.v9.i19.4881

[11] M. Akter, K. F. Ritu, M. T. Habib, M. S. Rahman and F. Ahmed, “A Machine Learning Approach To Predict Social Media Addiction During COVID-19 Pandemic,” in *Proc. Int. Conf. Appl. Artif. Intell. Comput. (ICAAIC)*, Salem, India, 2022, pp. 401-405, doi: 10.1109/ICAAIC53929.2022.9793193.

[12] M. Mardiah and K. Kusnawi, “Analysis of Social Media Addiction: A Comparison of the Performance of Linear Regression and Random Forest Algorithms in Predicting User Behaviour,” in *Proc. 12th Int. Conf. Inf. Commun. Technol. (ICoICT)*, Bandung, Indonesia, 2024, pp. 411-418, doi: 10.1109/ICoICT61617.2024.10698019.

[13] 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.

[14] N. Çiftci and M. Yıldız, “The relationship between social media addiction, happiness, and life satisfaction in adults: Analysis with machine learning approach,” *Int. J. Ment. Health Addiction*, vol. 21, pp. 3500–3516, 2023. doi: 10.1007/s11469-023-01118-7

[15] 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.

[16] Y. Sun, H. Wang, and S. Bo, “Altered topological connectivity of internet addiction in resting-state EEG through network analysis,” *Addict. Behav.*, vol. 95, pp. 49–57, 2019.

[17] 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/srep45064

[18] C. Gao *et al.*, “KuaiRec: A fully-observed dataset and insights for evaluating recommender systems,” in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manag. (CIKM)*, Atlanta, GA, USA, 2022, pp. 540–550. doi: 10.1145/3511808.3557220.

[19] C. Gao *et al.*, “KuaiRand: An unbiased sequential recommendation dataset with randomly exposed videos,” in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manag. (CIKM)*, New York, NY, USA, 2022, pp. 3953–3957. doi: 10.1145/3511808.3557624.

[20] Z. Sun *et al.*, “KuaiSAR: A unified search and recommendation dataset,” in *Proc. 32nd ACM Int. Conf. Inf. Knowl. Manag. (CIKM)*, New York, NY, USA, 2023, pp. 5407–5411. doi: 10.1145/3583780.3615123.

[21] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NeurIPS*, Red Hook, NY, USA, 2017, pp. 4768–4777.

[22] 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

[23] H. Xu *et al*., “MPLUG-2: A modularized multi-modal foundation model across text, image and video,” in *Proc. 40th Int. Conf. Mach. Learn. (ICML)*, vol. 202, 2023, Art. no. 1614, pp. 38728–38748.

[24] W. Kuo *et al*., “MaMMUT: A simple vision-encoder text-decoder architecture for multimodal tasks,” *Trans. Mach. Learn. Res., 2023*. [Online]. Available: https://arxiv.org/abs/2303.16839

[25] S. Yan *et al*., “VideoCoCa: Video-text modeling with zero-shot transfer from contrastive captioners,” *arXiv preprint arXiv:2212.04979*, 2022.

[26] 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.