Emerging Trends in Drugs, Addictions, and Health 4 (2024) 100145

Contents lists available at ScienceDirect

Emerging Trends in Drugs, Addictions, and Health

journal homepage: www.elsevier.com/locate/etdah

# Predicting anxiety using Google and Youtube digital traces

Joshua Rochotte [a,*], Aniket Sanap [b], Vincent Silenzio [c], Vivek K. Singh [a]

[a] School of Communication & Information, Rutgers University, NJ, USA
[b] Department of Computer Science, Rutgers University, New Brunswick, NJ, USA
[c] School of Public Health, Rutgers University, New Brunswick, NJ, USA

A R T I C L E I N F O

*Keywords:*
Anxiety
Machine learning
Google
YouTube
Digital traces

A B S T R A C T

Anxiety is a widespread and serious mental health issue that has been exacerbated by the COVID-19 pandemic and other stressors. In this study, we explore how online behavior data from Google and YouTube can be used to infer anxiety levels in individuals. We collected and processed digital traces from nearly 100 participants over eight weeks and applied various machine learning techniques to extract features and build predictive models. We found that combining data from multiple media modalities can yield highly accurate predictive models for anxiety as self-reported by a clinical GAD-7 scale (AUC > 0.86). We also found that the semantic categories of online engagement can affect the predictive performance of the models. This study contributes to the field of computational social science and digital mental health and demonstrates the potential of using online behavior data to monitor psychological well-being and design interventions for anxiety.

## Introduction

Anxiety refers to a class of disorders that “share features of excessive fear and related behavioral disturbances concerned with the anticipation of a future threat”. It ranks as one of the most frequently occurring and costly conditions of all mental disorders (Bandelow, 2015). Anxiety disorders affect 17 % i.e., 1 in 6, of all individuals across their lifetime. Other research has shown that anxiety has direct consequences on public health, including increased risk of mortality (Meier et al., 2016), cardiovascular disease (Liblik et al., 2022), and depression (Elhai et al., 2017). Traditional anxiety detection methods include a visit to a medical practitioner to participate in surveys and detailed medical discussions in order to detect anxiety. However, individuals may face the issues of stigma, cost, or access, and there is often an under-availability of mental health care providers. Hence, there is a need to explore ways to triage anxiety levels that are easier for patients and also aid medical practitioners prioritizing those in most need.

Emerging research seeks to perform digital trace analysis (e.g., search behavior logs, YouTube consumption logs) to illuminate factors which impact an individuals’ health and well-being (Fantasia et al., 2023), including anxiety (Zaman et al., 2020). Given the amount of time individuals spend online, this development is unsurprising and quite understandable from a conceptual standpoint since many theories, such as the Uses and Gratification Theory (U&G) and even work on eudaimonic and hedonic media use, suggest a link between media consumption and well-being (Zhang et al., 2011; Gadino et al., 2023; Eden et al., 2020). U&G theory assumes users of a given source of media engage in goal-directed activities, using it in an effort to satisfy an internal need. The collective identity, needs, and personality of a given user will result in different motivations and behaviors (Katz et al., 1973). This means that each actor, in their own way, engages themselves fully in pursuits which will benefit them, like relieving anxiety, feelings of loneliness, or other information seeking needs. This suggests an interconnection between a user’s (goal directed) behavior and underlying psychological situation, including the level of anxiety. Multiple studies have uncovered both strong positive and strong negative associations between technology use and anxiety in different scenarios, which speaks to the opportunity that exists to leverage social media and social networking sites for modeling aspects of mental health (Seabrook et al., 2016; Przybylski, 2017). Many previous studies concerning smartphone use and mental health have focused around problematic smartphone use and the challenges of marrying technology and mental health. However, this study seeks to illuminate the possibilities for using these tools, which users are already engaging with, to better understand and provide future mental health recommendations for users (Elhai et al., 2017).

Some studies have looked at how machine learning models may use individual-level traces from Google and YouTube to predict anxiety (Zaman et al., 2020). From a theoretical perspective, it is important to

* Corresponding author.
*E-mail address:* josh.rochotte@rutgers.edu (J. Rochotte).

https://doi.org/10.1016/j.etdah.2024.100145
Received 31 October 2023; Received in revised form 15 February 2024; Accepted 25 February 2024
Available online 8 March 2024

look at Google and YouTube because these platforms collect various aspects of online behavior that may have an impact on anxiety in various ways. Though Google may be more frequently used for active pursuits, like information seeking (Case, 2016), YouTube is more frequently associated with passive media consumption via its platform affordances like recommended videos that auto-play and live video engagement. A recent analysis of YouTube finds that 70 % of all content being viewed is recommended to users, without any active search being required (Rodrigues, 2018). This motivates a hypothesis that the information coming from use patterns across different modalities might yield complementary information that may improve prediction accuracy over data coming from a single modality. However, this question remains under-explored in the context of anxiety prediction using web traces. A notable exception is Zaman et al. (2020), which uses Google, YouTube traces, and past anxiety scores to predict future anxiety scores. Hence, while the Zaman et al. (2020) approach requires past self-reported anxiety scores, the approach in this work aims to make predictions on anxiety levels directly from the digital web traces. To the best of our knowledge, this is the first attempt to combine data from Google and YouTube to directly infer anxiety scores.

To build such a predictive model, this work follows the recommendations from a recent review article by Jon Elhai (Elhai et al., 2019) that focuses on a closely related problem of studying the interconnections between anxiety and problematic smartphone use (SMU). Much of the existing literature around problematic SMU has regarded frequency and not pattern of use and relying on self-reports of data. Hence, this work goes beyond self-report to use online digital traces and considers the patterns of use across semantic categories (e.g., YouTube views across sports, music, news & politics etc.) rather than aggregated frequency to study the predictive interconnections between anxiety and online media use. We do this through the use of automated predictive models that are generated using individual traces of data obtained via YouTube and Google. Specifically, the research questions (RQ) for this study are:

*RQ1: Can online data (Google and YouTube traces) be used to build automatic prediction models for a clinically validated scale of anxiety (GAD-7)?*

*RQ2: Are there systematic differences in terms of Google and YouTube's predictive ability for anxiety?*

## Methods

### *Data collection and protection*

We collected two types of data: (1) individual-level digital traces from Google and YouTube and (2) a self-reported survey from 92 participants over 8 weeks in 2021. The digital trace data was downloaded by individual participants using Google Takeout[1] and shared with the researchers via a secure mechanism. Meanwhile, the survey information concerning the participants' health, wellness, and personal behavior were collected weekly through Qualtrics. We acknowledge the challenges associated with the use of personal data for predictions. On one hand, such data can be very useful for personalized healthcare. On the other hand, such data can be used to profile and target individuals. We posit that responsible use of personal data with guard rails has the potential to transform human health and scientific research. These studies of the associations between human wellbeing and digital traces should be undertaken in a transparent manner. The pros and cons should be shared with the participants, and eventual participation should always be an individual decision.

In this study, several steps were taken to support the above perspective. First, before taking part in the study, participants learned about its objectives and how data would be collected. They also knew they could leave the study anytime during the ten weeks. Only those who signed consent forms and agreed to the conditions participated in the study. The participants received monetary incentives for their time. Second, participants' data (e.g., names, addresses, and phone numbers) were anonymized using Google's Cloud Data Loss Prevention (DLP) API before the research team accessed it. Next, the data was kept in a safe and secure system. Finally, findings based on participants' data are only reported as aggregate trends or associations instead of individual results. The study received approval from the Rutgers University Institutional Review Board (IRB). We report on and use gender, in which participants self-selected as male, female or wish not to specify. One participant chose to not specify their gender, and no participants added additional self-descriptions. A clinical expert (co-author) was available to intervene and provide referrals in case any serious mental health episodes were witnessed.

### *Participants*

Participation in the study was open to English-speaking individuals, 18 or older, living in the United States. In addition, participation also required using Google and Google services to search, send/receive emails, and share locations in the three months preceding the study. Participants were recruited through online advertisements, social media, and university mailing lists. Also, recruitment efforts focused on a large public university in the Northeastern United States. As a result, the most common age group for participants (43.5 %) was between 18 and 21. They were also primarily White (39.1 %) or Asian (35.9 %), Female (68.5 %) and mostly single (81.5 %). Additional demographic characteristics of the population are summarized in Table 1. Each participant received a total compensation of $120 for their involvement in the study spread over three months. Individuals received $25, $35, and $60 respectively for completing the study until week 1, week 6, and week 10 respectively.

Participant selection criteria of English-speaking individuals ensured that they understood the instructions of the surveys and active use of Google services was to ensure that the digital traces had a good chance of capturing the participant's daily activities. The recruitment effort focused around the campus of a North American university, which can be considered a convenience sample. Such convenience sampling allows for feasibility analysis for new technical methodology proposed in the work, which must be replicated and evaluated in bigger, more diverse settings, if found useful in the considered context.

### *Anxiety measurement: GAD-7*

The Generalized Anxiety Disorder 7 (GAD-7) scale was used to measure participants' anxiety levels (Williams, 2014). The GAD-7 can also aid in the screening of other different anxiety disorders, but for this study's purpose, the analysis seeks to understand, broadly, anxiety severity. The scale contains seven statements, such as "Feeling nervous,

**Table 1**
Demographic Distribution of Participants.

| Measure | Item | Count | Percentage (%) |
|---|---|---|---|
| Gender | Female | 63 | 68.5 |
| | Male | 28 | 30.4 |
| | Did Not Specify | 1 | 1.1 |
| Age | 18 - 21 | 40 | 43.5 |
| | 22 - 24 | 23 | 25.0 |
| | 25 - 34 | 17 | 18.5 |
| | 35 - 44 | 4 | 4.3 |
| | 45 - 54 | 7 | 7.6 |
| | 65 or older | 1 | 1.1 |
| Primary Race / Ethnicity | White | 36 | 39.1 |
| | Asian American | 33 | 35.9 |
| | African American | 11 | 11.9 |
| | Hispanic/Latino | 10 | 10.9 |
| | Self-Described | 2 | 2.2 |

[1] https://takeout.google.com//.

anxious, or on edge", "Not being able to stop or control worrying" and asks participants to consider how often they have been bothered by the above statements over the last 2 weeks. The participants choose how much they identify with the statement by choosing one of four responses from:

- Not at all (+0 points)
- Several days (+1 point)
- More than half the days (+2 points)
- Nearly every day (+3 points)

Following the recommended scoring mechanism as per Williams (2014), the sum of the points recorded from every statement was used to get the total GAD-7 score of each participant. This score ranges from 0 to 21 with higher values denoting higher levels of anxiety. This study chose to use this scale since it has been validated and widely used in literature (Williams, 2014; Choudhary et al., 2022). We classified individuals with GAD-7 $< 5$ as not anxious and GAD-7 $\geq 5$ as anxious based on the guidelines in Williams (2014), which suggests 5 as the cut-off for mild level of anxiety. The GAD-7 survey was collected on a weekly basis in the study. The ratio of the two classes in our dataset was 51.6 % anxious and 48.4 % not anxious, with scores across participants ranging from 0 to 21, and averaging around 5.18 over the 10 weeks of data collection.

*Google and YouTube traces*

These features were computed based on Google search and YouTube video engagement data. Since our survey data were collected weekly, we also created temporal aggregate features in this data. This includes using the average value of the features from the previous two weeks to correspond to the two-week duration used for questions in the GAD-7 survey. This methodology allowed us to understand better whether the feature's value for the current week differed from the previous behavior for a specific user. These additional features are prefixed with prev k, followed by the name of the feature used to calculate the mean.

*Google Search Features*

- num google searches: Total number of Google searches during a week.
- num websites visited: Total number of websites visited (through Google search results) during a week. daily use count google: Total number of times Google search was used per week.
- flu terms google search: Total number of Google searches that included flu-related glossary[2] terms per week.
- url category: Categorization of the URLs visited by the user from num websites visited using an API.[3] Many of the categories had very sparse representation and here we focus on the four most frequent categories: Business and Finance, Shopping, Style & Fashion, Technology & Computing.
- prev k num google searches: Average number of Google searches conducted in the last two weeks.
- prev k num websites visited: Average websites visited through Google search results in the previous two weeks.

*YouTube features*

- num videos watched: Total number of videos watched during a week on YouTube.
- average num sessions per day: Total number of YouTube sessions in a week. Here, two videos belong in a session if they were watched within 60 min of each other.
- daily use count YouTube: Total number of times YouTube was used in any capacity during a week.
- yt category x: Categorization of YouTube urls watched by the participant using the YouTube API V3. Due to sparsity, we focus on ten categories: Sports, Music, Education, People & Blogs, News & Politics, Film & Animation, Comedy, Entertainment, Gaming, How to & Style.
- prev k num videos watched: Average number of YouTube videos that were watched during the last two weeks.

*Data preprocessing*

The dataset was curated such that each row corresponded to each participant, week pair. In other words, the data contained 920 rows (the number of participants × the number of weeks). The study included data for ten weeks. However, there was a ramping-up effect in week 1 (e.g., some people signed up on Monday, while others on Saturday). The exact process happened in reverse during the last week of the study. Some participants stopped sharing their data on Monday while others shared until Saturday. This would lead to inaccurate timescales for many of the participants, as some would only have a partial week as opposed to those who may have more of a full week on the study. Rather than look at participant time on study, this study used study week as the scale, keeping all participants engaged weekly on the same schedule, so as to reduce any effects which may arise from the days of week users were engaging. As a result, data from weeks 1 and 10 were removed, and the study focused on data from weeks 2 to 9 instead. We also note that week 6 of the study corresponded to Spring Break for the university. Given the significant number of participants who were students, this may have partially impacted the fluctuations in various measurements. The eight-week data were divided into a training and test set.

The data from the first six weeks were used for training, and the data from the last two weeks were used for testing. The missing values in the data were imputed using the median value. All participants were included in both the training and test set, and all responses collected from participants and their devices were anonymized, encrypted, and stored securely on an internal lab drive. Here, the underlying assumption was that we should be able to track anxiety levels passively over time once individuals sign up for this service and actively engage with it for a few weeks.

*Machine learning modeling*

Four different machine learning models were used in this study: Decision Tree, Support Vector Machine (SVM), Logistic Regression and XGBoost, each of which were implemented using Python libraries, such as sklearn and xgboost. The Logistic Regression model was run with all parameters set to the default configuration, and except where mentioned, the following models ran with default settings. The Decision Tree was modified to use a max depth of 7, the SVM used a C (regularization parameter) of 45 and probability set to True and the XGBoost model ran with a max depth of 3 and the number of estimators set to 100. For every model, the selected features mentioned previously were used in each model in order to predict anxiety, while dropping all infrequent categories from specific inclusion in the model. Finally, ROC-AUC (Receiver Operating Characteristic Area Under Curve), accuracy and F1-score, were used to evaluate the models once they were computed (Muller, 2016). Scores closer to 1.0 signified a better fitting or more accurate model. The parameter selection was based on optimization for ROC-AUC on the training data.

## Results

There was some variance in the total number of anxious participants per week as shown in Fig. 1**(a)**. The total number of anxious participants in a week ranged from 40 to 54 with a mean of 46.88 anxious

[2] https://acl.gov/sites/default/files/common/Words- To- Know- About-the-Coronavirus-in-Plain-Language.pdf.

[3] https://website-categorization.whoisxmlapi.com/api.

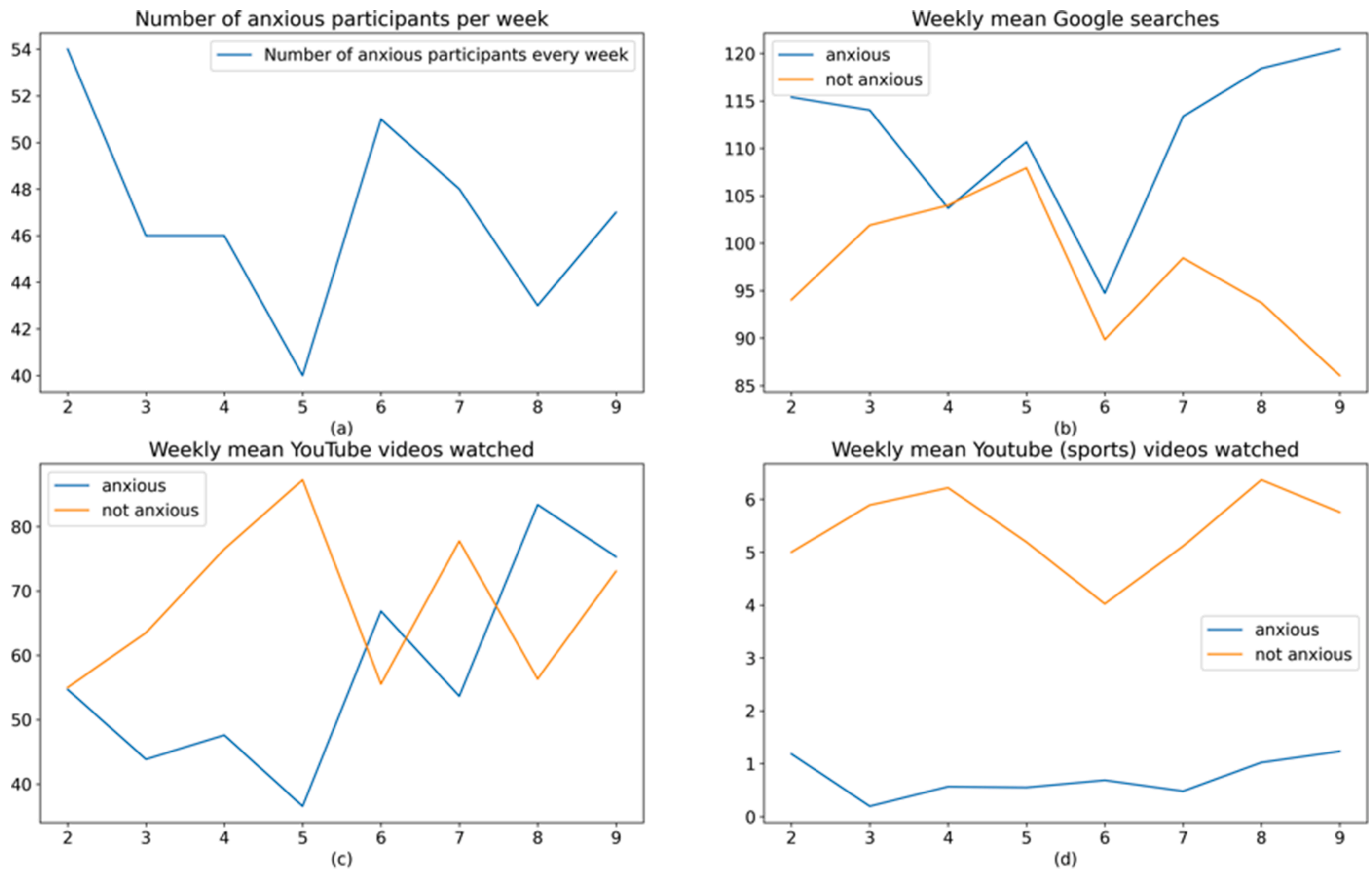


**Fig. 1.** (a) Number of anxious participants per week (b) Weekly average number of Google searches (c) Weekly average number of YouTube videos watched (D) Weekly average number of YouTube (sports) videos watched.

participants per week. The week with the lowest levels of anxiety corresponds with the week before university spring break. The weekly variations in the mean number of Google searches, YouTube videos watched, and videos watched in a specific category (sports) are shown in Fig. 1**(b), (c), (d),** respectively. We see differences between the anxious and not anxious group in terms of these features. While the average number of Google searches is higher for anxious group across almost all weeks (Fig. 1**b**), the trend is not as clear in terms of the number of YouTube videos watched (Fig. 1**c**). However, zooming into specific categories of YouTube videos (e.g., sports in Fig. 1**d**) demonstrates a much clearer demarcation between anxious and not anxious groups. To understand the behavioral patterns beyond the primary usage frequency we look at multiple categorical and temporal aggregate features and the mean, median and standard deviation for all the features are shown in Table 2.

Further, to understand the associations between different behavioral features and anxiety, we plot the change in means between the anxious and not anxious groups for every feature in Fig. 2. As the level of difference varied quite significantly over different features, we normalize them over a log scale for the ease of visualization and interpretation. This plot was created by taking the log of percentage change in means between the anxious and not anxious groups for every feature. Fig. 2 shows that on an aggregate basis (i.e., disregarding the semantic category of use), more Google searches were observed in the anxious group, and more YouTube videos were watched by the non-anxious group.

However, to create a richer picture one needs to analyze the semantic category of use. For Google searches, clicks on shopping related URLs were most common for the non-anxious group, while clicks on technology & computing related URLs were common for the anxious group. For YouTube videos, the anxious group was observed to watch news & politics, how-to, films & animation, and comedy videos. The non-anxious group was more likely to watch videos related to sports, music, and gaming. Further analysis of these patterns across categories to understand the root causes behind these differences would be an

**Table 2**
Descriptive statistics of Google and YouTube features.

| Feature Name | Mean | Median | Std |
|---|---|---|---|
| num_google_searches | 104.39 | 76.0 | 90.5 |
| num_websites_visited | 63.14 | 40.0 | 60.41 |
| num_videos_watched | 63.24 | 16.0 | 117.07 |
| average_num_sessions | 10.62 | 6.0 | 11.14 |
| num_comments | 0.03 | 0.0 | 0.22 |
| daily_use_count_google | 182.44 | 134.5 | 151.14 |
| daily_use_count_YouTube | 72.41 | 20.0 | 127.59 |
| flu_terms_google_search | 1.75 | 0.0 | 3.41 |
| url_category_Business & Finance | 9.35 | 6.0 | 10.1 |
| url_category_Shopping | 1.61 | 0.0 | 2.64 |
| url_category_Style & Fashion | 1.77 | 0.0 | 3.88 |
| url_category_Tech & Computing | 5.42 | 2.0 | 8.19 |
| yt_category_Film & Animation | 2.51 | 0.0 | 7.79 |
| yt_category_Music | 12.03 | 1.0 | 35.36 |
| yt_category_Sports | 3.07 | 0.0 | 13.69 |
| yt_category_Gaming | 3.03 | 0.0 | 11.08 |
| yt_category_People & Blogs | 9.51 | 2.0 | 22.61 |
| yt_category_Comedy | 4.53 | 0.0 | 10.31 |
| yt_category_Entertainment | 12.64 | 1.5 | 26.25 |
| yt_category_News & Politics | 2.0 | 0.0 | 7.59 |
| yt_category_Howto & Style | 2.22 | 0.0 | 4.97 |
| yt_category_Education | 2.61 | 0.0 | 6.0 |
| prev_k_num_videos_watched | 46.91 | 15.0 | 88.59 |
| prev_k_num_websites_visited | 56.16 | 36.0 | 54.42 |
| prev_k_num_google_searches | 90.53 | 67.0 | 78.59 |

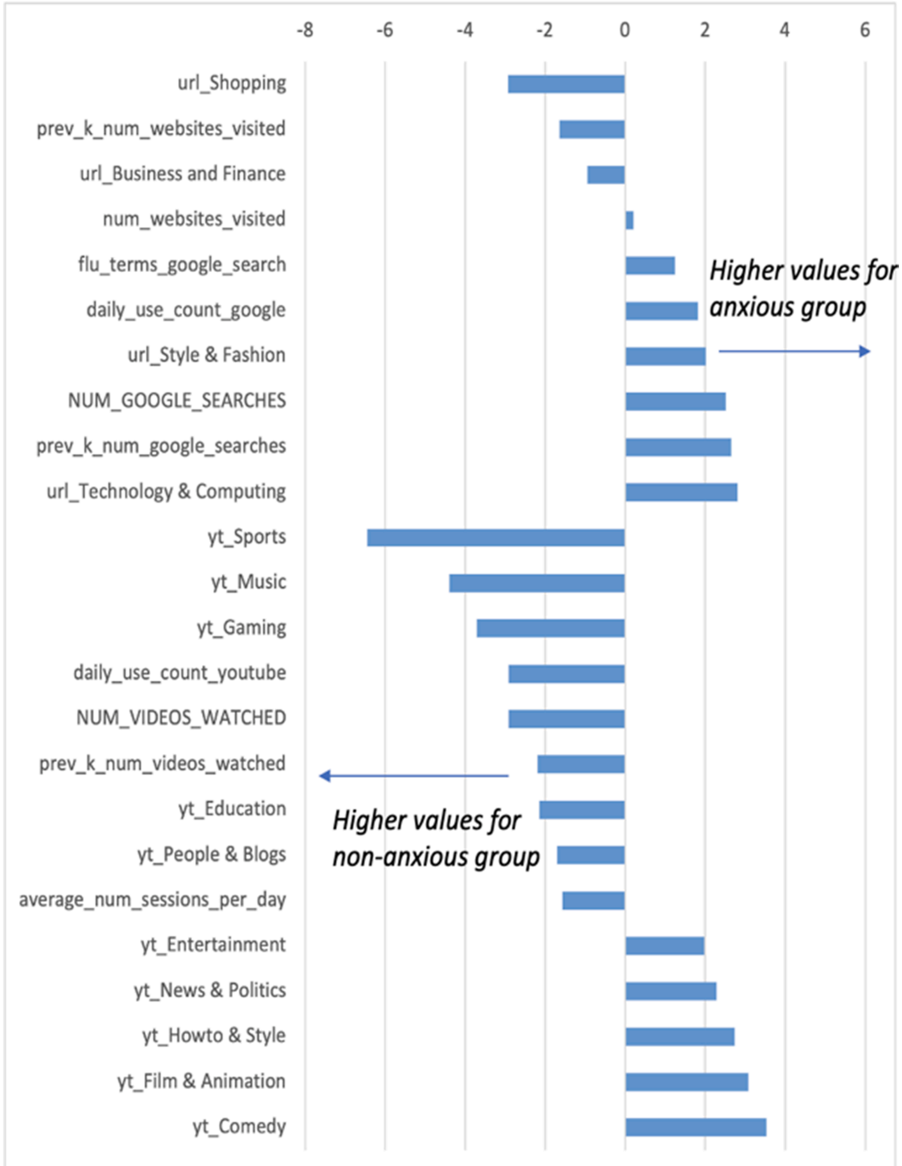


**Fig. 2.** Log of percent change in means between anxious and not anxious groups.

important avenue for future work. Nevertheless, the results support (Elhai et al., 2019) and indicate clear value in moving past aggregate use counts for the task of predicting mental health related outcomes such as anxiety.

The predictions for the test set using all the behavioral (Google and YouTube) features are shown in Table 3. We observe reasonable performance across all models used with the best performance coming from the XGBoost model (AUC = 0.8619). To understand the relative impact of various types of features, we zoom into the performance with XGBoost for different feature sets in Table 4. As a baseline, we also included a model based on demographic features Gender, Age, Race. The model built on demographic features yielded an AUC of 0.7137. The combined model that uses both Google and YouTube features yielded an AUC of 0.8619 and performs better than the baseline/component models.

**Table 3**
Performance of different machine learning models.

| Model | AUC | Accuracy | F1-score |
|---|---|---|---|
| Decision Tree | 0.7296 | 0.6684 | 0.7136 |
| SVM | 0.7884 | 0.6956 | 0.7171 |
| Logistic Regression | 0.7080 | 0.6576 | 0.6985 |
| XGBoost | 0.8619 | 0.7717 | 0.7765 |

**Table 4**
Performance of XGBoost with different feature sets.

| Feature Subset | AUC | Accuracy | F1-score |
|---|---|---|---|
| Only YouTube Features | 0.8033 | 0.6793 | 0.7093 |
| Only Google Features | 0.7850 | 0.7445 | 0.7486 |
| Only Demographic Features | 0.7137 | 0.6630 | 0.6593 |
| Google + YouTube Features | **0.8619** | **0.7717** | **0.7765** |

## Discussion

The first RQ of this work was “Can online data (Google and YouTube traces) be used to build automatic prediction models for a clinically validated scale of anxiety (GAD-7)? To answer this question, the study analyzed user-level digital trace data and multiple machine learning algorithms. Based on the results, we report that a combination of Google and YouTube data can approximate the scores that would have been obtained via a clinical self-reported anxiety scale (GAD-7) with a high level of accuracy (AUC > 0.86). These results were obtained using the XGBoost machine learning model. It should also be noted that the best performing XGBoost model was obtained when demographic characteristics were removed, which is a potential benefit for future modeling practices, where we need not consider these features.

The second RQ for this study was "Are there systematic differences in terms of Google and YouTube's predictive ability for anxiety"? The study uncovered that users' online behavior on distinct platforms as well as engagement with different categories of content (e.g., sports vs. comedy) had different associations and predictive power with anxiety. These findings mirror the theoretical implications of U&G, but also reflect similar findings from studies on mental gratifications where there are significant differences between anxious group and their counterparts (Gadino et al., 2023). In the current study, while the XGBoost machine learning model could predict anxiety using Google + YouTube trace data with 0.8619 AUC, the same model could infer anxiety with lower AUC using a single modality of data (e.g., only Google or only YouTube). This suggests certain complementarity in the information coming from the two modalities, allowing for higher predictive power based on combining modalities.

At a broad level, more Google searches were observed in the anxious group, and more YouTube videos were watched by the non-anxious group. This validates our hypothesis, given the information seeking needs of anxious individuals, their search terms, especially for covid and flu related terms would be higher, and that in general, the needs fulfillment of the non-anxious group would be much more diverse. Google searches for the anxious group trended on topics like technology and computing, which covered searches about digital services and style & fashion, which would have accounted for things like mask wearing. Furthermore, there were noticeable differences based on the semantic category of the content, and a significantly different volume of videos watched by the non-anxious group, which tended to be spread out across a variety of categories. For instance, while the anxious group was more likely to watch news & politics, how-to, films & animation, and comedy videos, the non-anxious group was more likely to watch videos related to sports, music, and gaming. While there isn't a hard line that can be drawn about the categories, the anxious group was more likely to engage in information seeking media by engaging in news and politics concerning themselves with current events, and also tended to watch content that has been traditionally considered pleasurable, like animated cartoons and comedy videos in an attempt to regulate their mood. This connects back with the Uses and Gratification Theory and eudaimonic use and suggests that different kinds of content (across media type as well as semantic categories) serve different purposes in an individual's life and would hence have different correlations with offline outcomes like anxiety (Katz et al., 1973; Eden et al., 2020).

The findings of this study have important implications for public health researchers and practitioners who are interested in understanding and addressing the issue of anxiety and its related health outcomes. By using online behavior data from Google and YouTube, we can potentially identify and monitor individuals who are at risk of developing or experiencing anxiety, and provide them with timely and personalized interventions (Zaman et al., 2020). For example, we can use the predictive models to screen for anxiety symptoms among online users, and offer them online resources or referrals to professional help. We can also use the semantic categories of online engagement to tailor the interventions to the users' preferences and needs, such as suggesting coping strategies or positive content that match their interests. Given the lack of mental health resources in large parts of the world, and associated stigma, building automated methods for initial triaging could have an important public health impact.

As it relates to the COVID-19 pandemic, Holmes et al. (2020) call for broader methods and research leading to public policy and health improvements, and the work undertaken here can begin to provide methods for better understanding users experiencing distress. With the growing risks of misinformation around these public health emergencies, it is becoming more apparent that we need effective methods to aid the most vulnerable populations. We can use the online behavior data from online media such as search engines and video sharing services to evaluate the effectiveness of the interventions, and track the changes in anxiety levels over time, as well as promote broader health improvements (Prestin, 2020; Wagner et al., 2021; Wolfers, 2021). Our study demonstrates the feasibility and utility of using online behavior data as a proxy for self-reported anxiety scales, and opens up new avenues for public health research and practice in the digital age.

The current study has a few limitations. Data for this study was collected during the COVID-19 pandemic, a period that had increased observation of social distancing and digital engagement. This significant external factor, while affecting people across the globe, may have had an effect on the populations existing rates of depression within the group, as well as their increased media use due to social isolation policies enacted. In addition, the participants involved in this study were those living in North America and primarily between 18 and 21 and White or Asian. Whether the trends observed will hold over different periods of time with different populations, remains to be tested. Further, anxiety was measured using self-reported survey data, which could be subject to reporting bias.

Additionally, while we did not attempt to differentiate between subtypes of anxiety disorders, the GAD-7 is validated to assess the severity of symptoms across four main disorders, including generalized anxiety disorder, panic disorder, social anxiety disorder, and post-traumatic stress disorder. While the GAD-7 is a clinically validated instrument, we also feel it important to note that what is considered "normal" or "anxious" behavior in one culture or for one individual might be different to another. If an algorithm is biased towards a particular cultural perspective, the underlying data may be misinterpreted and could lead to misclassification. We believe a more nuanced understanding of personal media interpretation, health search behavior and cultural context would make for stronger conclusions, something which a set of interviews may help with in future work. Lastly, the machine learning approach in this work focuses on predicting anxiety levels for the same individuals as those who are present in the training set. While this allows for generalization in settings where individuals provide self-reported data for a few weeks and then expect online data logs and machine learning to automatically approximate the scores, its generalization to different populations needs further evaluation.

Future work in this area should explore generalization of the model to diverse populations at different points of time. For instance, validating the proposed approach with clinically diagnosed patient populations would be a major step forward. It would also be important to consider different modalities of user data (e.g., Google, social media, phone, sensor logs) and combine them to understand the differential impact of various media use on individual anxiety. Lastly, the usability and adoption of such predictive mechanisms should be tested with wider populations.

## Conclusion

The study supports previous literature and shows that users' digital trace data across different sources of media could be used to identify and mitigate anxiety. The present study also validates theories, like the Uses and Gratification Theory, which suggests there exists different motivations which could encourage people to go online and engage with various forms of media. It also suggests an opportunity for the creation of low-cost automated approaches to understand population level shifts in anxiety and/or for the design of systems that will allow individuals to receive helpful tips when their anxiety levels may be increasing. Hence, this work contributes to the growing interest in using online digital traces and machine learning approaches for inferring offline health outcomes. With refinement and validation, the proposed approach can be used to support passive long-term monitoring for a large number of individuals, especially those who may not have access to traditional forms of mental health support.

## CRediT authorship contribution statement

**Joshua Rochotte:** Writing – original draft, Investigation, Formal analysis, Writing – review & editing. **Aniket Sanap:** Data curation, Visualization, Formal analysis. **Vincent Silenzio:** Writing – review & editing, Conceptualization, Funding acquisition. **Vivek K. Singh:** Writing – review & editing, Supervision, Conceptualization, Project administration, Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Funding

This research was partially funded by the Rutgers Center for COVID-19 Response and Pandemic Preparedness, and had no further involvement in study conduct or submissions of author works.

## References

Bandelow, B., Michaelis, S., 2015. Epidemiology of anxiety disorders in the 21st century. Dialogues Clin. Neurosci. 17 (3), 327–335. https://doi.org/10.31887/DCNS.2015.17.3/bbandelow.

Case, D.O., Given, L.M., 2016. Looking For information: A survey of Research On Information seeking, needs, and Behavior. Emerald Group Publishing.

Choudhary, S., Thomas, N., Alshamrani, S., Srinivasan, G., Ellenberger, J., Nawaz, U., Cohen, R., 2022. A machine learning approach for continuous mining of nonidentifiable smartphone data to create a novel digital biomarker detecting generalized anxiety disorder: prospective cohort study. JMIR Med. Inform. 10 (8), e38943. https://doi.org/10.2196/38943.

Eden, A.L., Johnson, B.K., Reinecke, L., Grady, S.M., 2020. Media for Coping During COVID-19 Social Distancing: stress, Anxiety, and Psychological Well-Being. Front. Psychol. 11. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.577639.

Elhai, J.D., Levine, J.C., Dvorak, R.D., Hall, B.J., 2017. Non-social features of smartphone use are most related to depression, anxiety and problematic smartphone use. Comput. Human Behav. 69, 75–82. https://doi.org/10.1016/j.chb.2016.12.023.

Elhai, J.D., Levine, J.C., Hall, B.J., 2019. The relationship between anxiety symptom severity and problematic smartphone use: a review of the literature and conceptual frameworks. J. Anxiety Disord. 62, 45–52. https://doi.org/10.1016/j.janxdis.2018.11.005.

Fantasia, A.T., Prybutok, G., Prybutok, V., 2023. The relationship between post-traumatic stress disorder and social media addiction: a qualitative study. Emerg. Trends Drugs, Addict., Health 3, 100056. https://doi.org/10.1016/j.etdah.2023.100056.

Gadino, N., Ellithorpe, M.E., Ulusoy, E., Wirz, D.S., Eden, A., 2023. Binge-watching to feel better: mental health gratifications sought and obtained through binge-watching. Psychol. Popul. Media. https://doi.org/10.1037/ppm0000485.

Holmes, E.A., O'Connor, R.C., Perry, V.H., Tracey, I., Wessely, S., Arseneault, L., Ballard, C., Christensen, H., Silver, R.C., Everall, I., Ford, T., John, A., Kabir, T., King, K., Madan, I., Michie, S., Przybylski, A.K., Shafran, R., Sweeney, A., Bullmore, E., 2020. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 7 (6), 547–560. https://doi.org/10.1016/S2215-0366(20)30168-1.

Katz, E., Blumler, J.G., Gurevitch, M., 1973. Uses and gratifications research. Public Opin. Q. 37 (4), 509–523. https://doi.org/10.1086/268109.

Liblik, K., Mulvagh, S.L., Hindmarch, C.C.T., Alavi, N., Johri, A.M., 2022. Depression and anxiety following acute myocardial infarction in women. Trends Cardiovasc. Med. 32 (6), 341–347. https://doi.org/10.1016/j.tcm.2021.07.005.

Meier, S.M., Mattheisen, M., Mors, O., Mortensen, P.B., Laursen, T.M., Penninx, B.W., 2016. Increased mortality among people with anxiety disorders: total population study. Br. J. Psychiatry 209 (3), 216–221. https://doi.org/10.1192/bjp.bp.115.171975.

Muller, A.C., Guido, S., 2016. Introduction to Machine Learning With Python: a Guide For Data Scientists. O'Reilly Media, Inc.

Prestin, A., Nabi, R., 2020. Media prescriptions: exploring the therapeutic effects of entertainment media on stress relief, illness symptoms, and goal attainment. J. Commun. 70 (2), 145–170. https://doi.org/10.1093/joc/jqaa001.

Przybylski, A.K., Weinstein, N., 2017. A large-scale test of the goldilocks hypothesis: quantifying the relations between digital-screen use and the mental well-being of adolescents. Psychol. Sci. 28 (2), 204–215. https://doi.org/10.1177/0956797616678438.

Rodrigues, A. 2018. YouTube's recommendations drive 70% of what we watch. https://qz.com/1178125/YouTubes-recommendations-drive-70-of-what-we-watch.

Seabrook, E.M., Kern, M.L., Rickard, N.S., 2016. Social networking sites, depression, and anxiety: a systematic review. JMIR Ment. Health 3 (4), e50. https://doi.org/10.2196/mental.5842.

Wagner, B.E., Folk, A.L., Hahn, S.L., Barr-Anderson, D.J., Larson, N., Neumark-Sztainer, D., 2021. Recreational screen time behaviors during the COVID-19 pandemic in the U.S.: a mixed-methods study among a diverse population-based sample of emerging adults. Int. J. Environ. Res. Public Health 18 (9), 9. https://doi.org/10.3390/ijerph18094613. Article.

Williams, N., 2014. The GAD-7 questionnaire. Occup. Med. (Chic Ill) 64 (3). https://doi.org/10.1093/occmed/kqt161, 224–224.

Wolfers, L.N., Schneider, F.M., 2021. Using media for coping: a scoping review. Commun. Res. 48 (8), 1210–1234. https://doi.org/10.1177/0093650220939778.

Zaman, A., Zhang, B., Silenzio, V., Hoque, E., Kautz, H., 2020. Individual-level Anxiety Detection and Prediction from Longitudinal Youtube and Google Search Engagement Logs (arXiv:2007.00613). arXiv. http://arxiv.org/abs/2007.00613.

Zhang, Y., Tang, L.S.-T., Leung, L., 2011. Gratifications, collective self-esteem, online emotional openness, and traitlike communication apprehension as predictors of facebook uses. Cyberpsychol., Behav., Soc. Netw. 14 (12), 733–739. https://doi.org/10.1089/cyber.2010.0042.