Original Paper

Utilizing Machine Learning on Internet Search Activity to Support the Diagnostic Process and Relapse Detection in Young Individuals With Early Psychosis: Feasibility Study

Michael Leo Birnbaum[1,2,3*], MD; Prathamesh "Param" Kulkarni[4*], PhD; Anna Van Meter[1,2,3], PhD; Victor Chen[5], BSc; Asra F Rizvi[1,2], MA; Elizabeth Arenare[1,2], BA; Munmun De Choudhury[5], PhD; John M Kane[1,2,3], MD

[1]The Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY, United States

[2]The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, United States

[3]Hofstra Northwell School of Medicine, Hempstead, NY, United States

[4]Cornell Tech, Cornell University, New York City, NY, United States

[5]Georgia Institute of Technology, Atlanta, GA, United States

[*]these authors contributed equally

Corresponding Author:
Michael Leo Birnbaum, MD
The Zucker Hillside Hospital
Northwell Health
75-59 263rd Street
Glen Oaks, NY, 11004
United States
Phone: 1 718 470 8305
Email: Mbirnbaum@northwell.edu

Abstract

Background: Psychiatry is nearly entirely reliant on patient self-reporting, and there are few objective and reliable tests or sources of collateral information available to help diagnostic and assessment procedures. Technology offers opportunities to collect objective digital data to complement patient experience and facilitate more informed treatment decisions.

Objective: We aimed to develop computational algorithms based on internet search activity designed to support diagnostic procedures and relapse identification in individuals with schizophrenia spectrum disorders.

Methods: We extracted 32,733 time-stamped search queries across 42 participants with schizophrenia spectrum disorders and 74 healthy volunteers between the ages of 15 and 35 (mean 24.4 years, 44.0% male), and built machine-learning diagnostic and relapse classifiers utilizing the timing, frequency, and content of online search activity.

Results: Classifiers predicted a diagnosis of schizophrenia spectrum disorders with an area under the curve value of 0.74 and predicted a psychotic relapse in individuals with schizophrenia spectrum disorders with an area under the curve of 0.71. Compared with healthy participants, those with schizophrenia spectrum disorders made fewer searches and their searches consisted of fewer words. Prior to a relapse hospitalization, participants with schizophrenia spectrum disorders were more likely to use words related to hearing, perception, and anger, and were less likely to use words related to health.

Conclusions: Online search activity holds promise for gathering objective and easily accessed indicators of psychiatric symptoms. Utilizing search activity as collateral behavioral health information would represent a major advancement in efforts to capitalize on objective digital data to improve mental health monitoring.

KEYWORDS

Introduction

Schizophrenia can be associated with significant impairment [1]. Although early intervention services have demonstrated the potential to improve outcomes [2], several challenges persist, limiting the established benefits of effective care. These include lengthy delays to early and accurate diagnostic ascertainment [3,4], as well as high rates of relapse, particularly during the early course of illness [5]. Under-recognized or misdiagnosed symptoms contribute to poorer outcomes such as social isolation, unemployment, and comorbid depression, anxiety, and substance abuse [6]. Furthermore, each new relapse can be associated with costly emergency room visits, psychiatric hospitalizations, family burden, medical complications, and suicide [7].

These challenges are compounded by the fact that psychiatry is still nearly entirely reliant on patient self-report. In contrast to all other areas of medicine, there are no reliable tests, biomarkers, or objective sources of collateral information available to inform diagnostic procedures or to assess mental health status [8-10]. Clinicians must therefore rely on subjective information, collected through patient and family interviews, to support diagnoses and make treatment recommendations. Technology offers the opportunity to collect objective digital data to complement self-reports and facilitate more informed treatment decisions [11-13]. Online search activity is a source of objective data with great potential.

Google search is one of the most popular websites worldwide, managing over 3 billion searches daily across over 600 million daily visitors [14]. Moreover, searching online has become a primary resource for youth seeking mental health–related information [15-20]. This is particularly true for stigmatized illnesses such as schizophrenia as the internet provides an easy and anonymous setting to gather information about symptoms and treatment options [21]. Importantly, online search engines store search activity as time-stamped digital records, offering a reliable source of objective, easily accessed, and detailed collateral information about an individual over an extended period of time.

Prior work has highlighted opportunities to utilize large-scale anonymized search logs to detect signals associated with the emergence and progression of medical illnesses [22]. For example, search activity, including content and patterns of use, has been used to identify individuals with lung cancer, Parkinson disease, and pancreatic cancer with high degrees of accuracy up to a year in advance of the diagnosis [23-25]. The success of these algorithms may lead to the development of a new generation of digital tools designed to assist in the screening and early identification of individuals at risk for medical conditions. Similar methods have been employed successfully in psychiatry using digital data extracted from social media sites [26-33]. However, few studies to date have explored the use of computational approaches to detect search patterns associated with psychiatric disorders [34]. Furthermore, while promising, online activity research thus far has been limited by the fact that it has been conducted primarily utilizing data extracted from anonymous individuals online who self-disclose having a particular diagnosis [35], and has yet to be carried out in real-world clinical settings using participant-contributed search data with clinically validated diagnoses.

Toward the goal of improving early diagnostic accuracy and relapse detection, we sought to conduct one of the first ecologically valid investigations into the relationship between online search activity and behavioral health. Specifically, we aimed to develop computational algorithms designed to accurately identify individuals with schizophrenia spectrum disorders (SSD) and to predict psychotic relapse based on internet search activity. We hypothesized that significant differences in the timing, content, and pattern of online search activity would differentiate participants with SSD from healthy volunteers, and that changes in these features would accurately predict a psychotic relapse in individuals with SSD.

Methods

Participants and Data Collection

Participants between the ages of 15 and 35 years were recruited from Northwell Health's inpatient and outpatient psychiatry departments. Individuals with SSD were recruited primarily from the Early Treatment Program, Northwell Health's specialized early psychosis intervention clinic (N=37). Additional participants diagnosed with SSD (N=7) were recruited from a collaborating institution located in East Lansing, Michigan. Recruitment occurred between March 2016 and December 2018. The study was approved by the Institutional Review Board (IRB) of Northwell Health (the coordinating institution) as well as by the local IRB at the participating site. Written informed consent was obtained for adult participants and legal guardians of participants under 18 years of age. Assent was obtained for participating minors. Healthy volunteers were approached and recruited from an existing database of eligible individuals who had already been screened for prior research projects at Zucker Hillside Hospital and had agreed to be recontacted for additional research opportunities (N=58). Additional healthy volunteers (N=21) were recruited from a southeastern university via an online student community research recruitment site. Healthy status was determined either by the Structured Clinical Interview for DSM Disorders [36] conducted within the past 2 years or the Psychiatric Diagnostic Screening Questionnaire [37]. If clinically significant psychiatric symptoms were identified during the screening process, participants were excluded.

Participants requested their search archive (known as "takeout") through a simple process supported by Google. Participation involved a single visit during which all historical search activity was downloaded and collected. Each archive included a time-stamped record of search terms and browser history. Using hospitalization dates pulled from participants' medical records, each participant's search data was segmented into 4-week periods immediately before and after each hospitalization. A 4-week period was selected as it represents an interval of time long enough to identify symptomatic changes [38,39] and also to contain sufficient online data required to train an algorithm [33,40]. For healthy participants (who did not have a hospitalization date), we randomly selected 4 weeks' worth of search data to serve as a control.

Diagnostic Classifier

A diagnostic classifier was built utilizing 4 weeks' worth of search data immediately preceding the first psychiatric hospitalization. Data prior to the first hospitalization were selected to reduce the potential confounding influence of receiving a psychiatric diagnosis, being hospitalized, and receiving psychiatric interventions (such as therapy or prescriptions for psychiatric medications) on search activity. Concurrently, we built the diagnostic classifier using data closest to the time when the diagnosis is typically made (at the point of initial hospitalization) [41] to enhance the classifier's potential clinical utility as a diagnostic support tool. A 4-week period before hospitalization was selected as it represents a period of time when psychotic symptoms would likely be most prominent. To match the data extraction period for both groups, we randomly selected 4 weeks' worth of search data from each healthy participant to serve as a comparison group. This strategy also reduced possible effects of seasonality on search behavior. Participants diagnosed with SSD who did not have any search data in the 4-week period before their first hospitalization were excluded from this classifier. For healthy volunteers, if no search data existed in the randomly selected 4-week period, that participant was excluded.

Relapse Classifier

A relapse classifier was built by segmenting the search data into 4-week periods of "relative health" and "relative illness." Periods of relative illness were defined as the 4 weeks immediately preceding each relapse hospitalization, as it represents a period of time prior to hospitalization during which psychiatric symptoms are typically the most prominent. When less than 1 month existed between two consecutive hospitalizations, these data were not included in the classification model. Healthy periods were defined as the 4-week period immediately following discharge from a relapse hospitalization, as this represents a period of time when symptoms are typically better managed and less pronounced. If less than 2 months' worth of search data existed between consecutive hospitalizations, these data were not included in the classifier, as we did not expect this period to represent a true period of relative health. Search data prior to the first hospitalization were not included in the relapse classifier. In total, 38 participants were included in the relapse classifier consisting of 51 periods of relative health and 42 periods of relative illness.

Defining Features

We defined features of search content and search behavior using linguistic and temporal parameters. For linguistic features, we used linguistic inquiry and word count (LIWC) [42]. LIWC is a language analytic tool designed to capture and count the frequency of 51 different word categories, with established psychometric properties, including emotions, mood, cognition, thinking styles, and social concerns. A rich body of literature has identified associations between the use of LIWC categories and psychological health and illness [42,43]. We concatenated the Google search streams for the selected periods before passing them to LIWC as the input text for computing features. For the search behavioral features, we constructed histograms of length and frequency of queries using 1-hour bins as well as 4-day bins. This was done to explore search features that might accompany changes in circadian patterns associated with SSD. The 1-hour bin histogram helped to model finer changes in the length and frequency of search behaviors throughout the day, whereas the 4-day bin histogram was used to model broader changes in search behaviors. The 1-hour bin histograms were computed by creating 24 bins corresponding to each hour of the day and aggregating (through summation) each participant's data across the 28 days. We chose hourly bins as this approach has been successfully implemented in prior research [27,44-46] exploring fluctuations in mood.

In addition, we included the total number of queries and the average query length for the 4-week period. We also included the standard deviation of the 4-day bin histograms (length and frequency) to represent the variation in search behaviors. Finally, we included directional changes in search behavior by computing first- and second-order statistics on the derivative of the 4-day histograms. All LIWC features were normalized based on the number of words in all searches concatenated for each participant, whereas the other features were normalized by subtracting the mean and dividing by the standard deviation. This process controls for any discrepancies in the feature values (ie, differences in the number of searches). A summary of all feature types along with the dimension of each feature is shown in Table 1.

XSL•FO

RenderX

Table 1. Feature categories along with the dimensionality of each feature type.

Feature type	Dimensions
24-hour histogram of length of queries with 1-h bin	24
24-hour histogram of frequency of queries with 1-h bin	24
32-day histogram of length of queries with 4-day bin	8
32-day histogram of frequency of queries with 4-day bin	8
SD of 4-day frequency of queries bins	1
SD of 4-day length of queries bins	1
Average of the derivative of 4-day frequency of queries bins	1
Average of the derivative of 4-day length of queries bins	1
SD of the derivative of 4-day frequency of queries bins	1
SD of the derivative of 4 day length of queries bins	1
Linguistic inquiry and word count	51
Total number of queries in 1 month	1
Average query length in 1 month	1

Classifier Analyses

For both the diagnostic classifier and relapse prediction, we tested three classifiers: random forest (RF) [47], support vector machine (SVM) [48], and gradient boosting (GB) [49]. We used the standard python-based scikit-learn [50] library for evaluating classification performance. We performed hyperparameter tuning using a held-out validation dataset, which resulted in selection of optimal hyperparameters for the classifiers. For example, for SVM, we selected the radial basis function kernel over the standard linear kernel. Each classifier was validated using a 5-fold crossvalidation technique to avoid overfitting. To prevent bias in selection of healthy volunteer data, we tried 10 different iterations of randomly selected 4-week periods and found that the results were consistent. We calculated the average F1 score, average accuracy, and average area under the receiver operating characteristic curve (AUC) across 5 folds for each classifier. Since both diagnostic and relapse classifiers were trained on unbalanced datasets, we chose to evaluate the classifiers based on the AUC since it is a parameter that is agnostic to class imbalance [51].

Feature Importance

A total of 123 features were used for each classifier. We used the permutation feature importance [52] method to compute the rank-ordered feature importance for each classifier. Under this method, feature importance is defined by the difference in the model's score when the feature is randomly shuffled. Feature importance is proportional to the drop in the model score when the feature is shuffled. We used the AUC value as the model score. The feature importance was calculated on the validation set in 5-fold crossvalidation and the average score was computed across the 5 folds. We used this method as it is model-agnostic and enabled comparison of three different classifier models in an unbiased manner.

Results

A total of 123 search archives (44 individuals diagnosed with SSD and 79 healthy volunteers) were available for analysis, and 116 (42 individuals with SSD an 74 healthy volunteers) met the inclusion criteria. Of these, 38 participants with SSD were available for the relapse classifier. An overview of the final dataset is shown in Table 2.

With respect to the diagnostic classifier (Table 3), the RF was selected for further feature analysis given its superior AUC compared to that of the other models. Figure 1 shows the receiver operating characteristic curves of the RF diagnostic classifier for each of the 5 folds. To explore consistency, this process was repeated 10 times with differing randomly selected 4-week periods of healthy volunteer data. Classifier performance remained consistent. Table 4 shows the quantity of search data provided per group for the diagnostic classifier.

Table 2. Participant demographics (N=116).

Characteristic	Value
Age (years), mean (SD)	24.38 (5.18)
Sex, n (%)	
Male	51 (44.0)
Female	65 (56.0)
Race, n (%)	
Asian	18 (15.5)
African American	32 (27.6)
Caucasian	60 (51.7)
Mixed/Other	6 (5.2)
Hispanic, n (%)	11 (9.5)
Diagnosis, n (%)	
Schizophrenia	16 (13.7)
Schizophreniform	13 (11.2)
Schizoaffective	2 (1.8)
Unspecified SSD[a]	11 (9.5)
Healthy volunteers, n (%)	74 (63.8)

[a]SSD: schizophrenia spectrum disorders.

Table 3. Diagnostic classifier results.

Classifier type	Mean F1	Precision (HV[a])	Precision (SSD[b])	Recall (HV)	Recall (SSD)	Mean Accuracy	Mean (SD) AUC[c]
Support vector machine	0.49	0.73	0.51	0.73	0.5	0.65	0.66 (0.09)
Random forest	0.54	0.75	0.72	0.86	0.48	0.73	0.74 (0.06)
Gradient boost	0.47	0.71	0.53	0.77	0.44	0.65	0.68 (0.09)

[a]HV: healthy volunteers.

[b]SSD: schizophrenia spectrum disorders.

[c]AUC: area under the receiver operating characteristic curve.

Figure 1. Receiver operating characteristic curves of the random forest diagnostic classifier for each of the 5 folds. AUC: area under the curve.
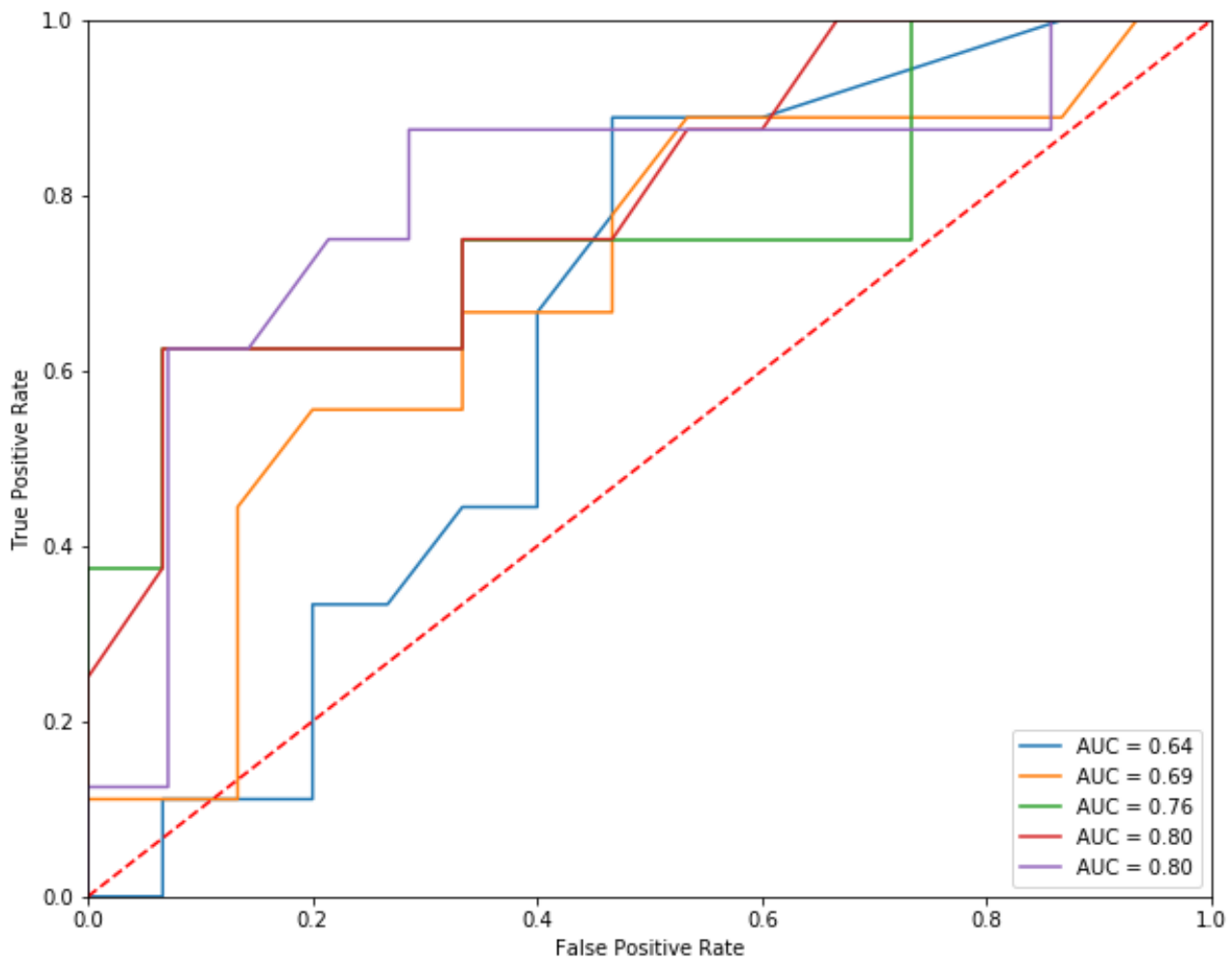
Table 4. Quantity of search data provided per group for the diagnostic classifier.

Metric	Healthy volunteers	Participants with SSD[a]
Total average queries (SD)	332.93 (298.1)	192.76 (214.19)
Weekly average queries (SD)	80.37 (71.92)	48.19 (52.91)

[a]SSD: schizophrenia spectrum disorders.

For the relapse classifier (Table 5), the SVM and GB models had the same AUC, and therefore both were considered for feature analysis. Further analysis of the feature importance of the SVM and GB relapse model revealed differing features. Herein, we report the SVM model as the identified features included search terms/themes that were deemed to be clinically interpretable and demonstrated some consistency with previous findings [33]; see Multimedia Appendix 1 for a comparison of the important features highlighted by both models. Figure 2 shows the receiver operating characteristics of the SVM relapse classifier for each of the 5 folds. The average F1 score for the SVM model was 0.36 and the average accuracy was 0.63. Table 6 shows the quantity of search data provided per group for the relapse classifier.

For each of the selected models, we calculated the top 20 features using the permutation feature selection method. The features sorted in decreasing order of feature importance for diagnostic and relapse classifies are shown in Table 7 and Table 8, respectively. For the two classifiers, both linguistic and behavioral features accounted for the top 20 features, indicating that both categories of features were important drivers of the classification result. Top features pertaining to the diagnostic classifier included a reduced search length between 12 am and 12 pm, lower overall number/frequency of search queries, as well as differences in the use of search terms/words from the "inhibition," "positive affect," and "anxiety" categories. Top features pertaining to the relapse classifier included differences in the use of search terms/words from the "sexual," "health," "hear," "anger," "sadness," and "perception" LIWC categories, as well as reductions in search length and search frequency prior to a relapse hospitalization.

Table 5. Relapse classifier results.

Classifier type	Mean F1	Precision (HV[a])	Precision (SSD[b])	Recall (HV)	Recall (SSD)	Mean Accuracy	Mean (SD) AUC[c]
Support vector machine	0.36	0.61	0.77	0.92	0.26	0.63	0.71 (0.16)
Random forest	0.53	0.61	0.61	0.69	0.48	0.61	0.69 (0.09)
Gradient boost	0.57	0.66	0.63	0.75	0.53	0.65	0.71 (0.10)

[a]HV: healthy volunteers.

[b]SSD: schizophrenia spectrum disorders.

[c]AUC: area under the receiver operating characteristic curve.

Figure 2. Receiver operating characteristic curves of the support vector machine relapse classifier for each of the 5 folds. AUC: area under the curve.
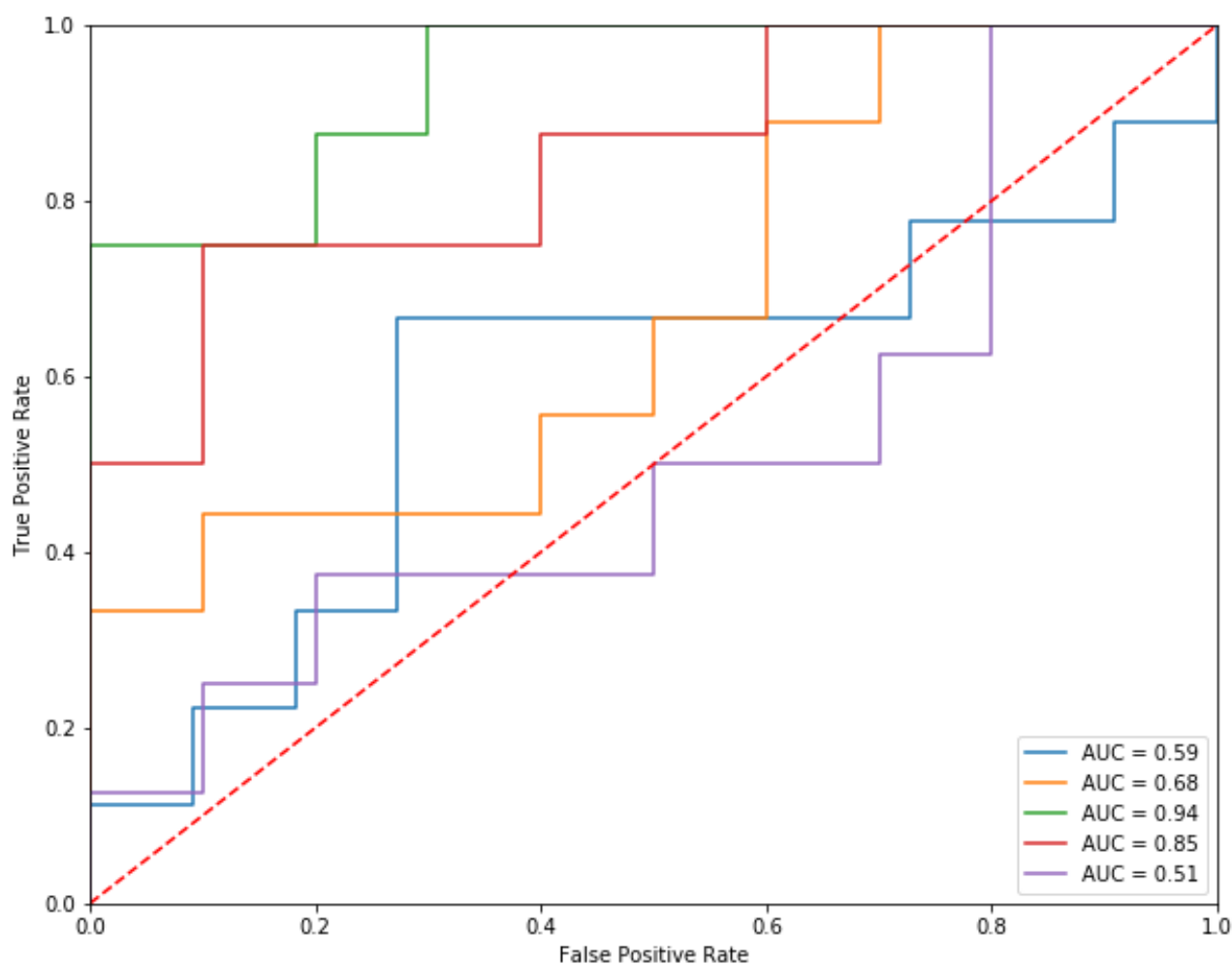
Table 6. Quantity of search data provided per group for the relapse classifier.

Metric	Periods of relative health	Periods of relative illness
Total average queries (SD)	96.80 (98.77)	168.29 (250.18)
Weekly average queries (SD)	24.2 (11.17)	42.07 (39.9)

Table 7. Feature importance of diagnostic classifiers sorted by decreasing order of importance.

Diagnostic classifier features	Average feature importance (random forest)
Reduced search lengths between 8-9 am in participants with SSD[a] compared to HV[b]	0.0315
Reduced search lengths between 6-7 am in participants with SSD compared to HV	0.0255
Length of queries from 23-20 days prior to first hospitalization is lower in participants with SSD compared to HV	0.0178
Reduced usage of "relative" LIWC[c] features in participants with SSD compared to HV	0.0112
Variance in frequency of search lengths is lower in participants with SSD	0.0111
Reduced search lengths between 11am to 12 pm in participants with SSD compared to HV	0.0091
Reduced usage of "inhibition" LIWC features in participants with SSD compared to HV	0.0078
Reduced search lengths between 4 and 5 am in participants with SSD compared to HV	0.0073
Reduced usage of "quantifier" LIWC features in participants with SSD compared to HV	0.0072
Reduced search lengths between 1 and 2 am in participants with SSD compared to HV	0.0071
Reduced usage of "positive affect" LIWC features in participants with SSD compared to HV	0.0071
Reduced search lengths between 12 am and 1 am in participants with SSD compared to HV	0.0070
Reduced usage of "anxiety" LIWC features in participants with SSD compared to HV	0.0064
Lower overall number of queries in participants with SSD compared to HV	0.0062
Reduced usage of "preposition" LIWC features in participants with SSD compared to HV	0.0061
Reduced usage of "inclusive" LIWC features in participants with SSD compared to HV	0.0059
Frequency of search 19-16 days prior to first hospitalization is lower in participants with SSD compared to HV	0.0059
Reduced usage of "insight" LIWC features in participants with SSD compared to HV	0.0057
Number of queries between 2 and 3 am is lower in participants with SSD compared to HV	0.0056
Number of queries between 11 pm and 12 am is lower in participants with SSD compared to HV	0.0051

[a]SSD: schizophrenia spectrum disorders.

[b]HV: healthy volunteers.

[c]LIWC: linguistic inquiry and word count.

Table 8. Feature importance of relapse classifiers sorted by decreasing order of importance.

Relapse classifier features	Average feature importance (support vector machine)
Reduced length of queries during relapse periods	0.0688
Increased usage of "sexual" LIWC[a] features during relapse periods	0.0523
Reduced length of queries 3-0 days prior to relapse hospitalization	0.0506
Reduced frequency of search activity during relapse periods	0.0263
Reduced usage of "health" LIWC features during relapse periods	0.0245
Increased usage of "hear" LIWC features during relapse periods	0.0224
Increased usage of "bio" LIWC features during relapse periods	0.0223
Increased searches in the 4 days before relapse hospitalization	0.0209
Reduced length of queries in the 7-4 days prior to relapse hospitalization	0.0196
Reduced frequency of searches 23-20 days prior to relapse hospitalization	0.0194
Increased usage of "percept" LIWC features during relapse periods	0.0186
Increased length of queries in the 31-28 days prior to relapse hospitalization	0.0162
Increased usage of "inclusive" LIWC features during relapse periods	0.0143
Denser searches during relapse periods	0.0140
Increased usage of "anger" LIWC features during relapse periods	0.0131
Reduced frequency of searches 19-16 days prior to relapse hospitalizations	0.0125
Reduced length of queries 11-8 days prior to relapse hospitalization	0.0105
Reduced usage of "sadness" LIWC features during relapse periods	0.0105
Increased usage of "indefinite pronoun" LIWC features during relapse periods	0.0104
Reduced frequency of searches 15-12 days prior to relapse hospitalization	0.0097

[a]LIWC: linguistic inquiry and word count.

Discussion

Principal Findings

We aimed to explore the feasibility of using collateral online search activity to support the diagnostic process and relapse detection in individuals with SSD. Our results indicate that important differences exist in the timing, frequency, and content of search activity in individuals with SSD compared to healthy volunteers. Furthermore, linguistic and behavioral shifts were identified in the month preceding a relapse hospitalization in individuals with SSD. This study demonstrates the promise of online search activity to potentially serve as collateral information informing diagnostic procedures as well as relapse identification strategies. Much like physicians routinely use medical imaging and blood tests to obtain objective and reliable clinically meaningful patient data, our results support the prospect of incorporating real-time machine learning–based extraction and analysis of online activity into psychiatric assessment.

Features Relevant to the Diagnostic Classifier

Combining linguistic and behavioral features, the RF classifier distinguished individuals with SSD from healthy volunteers with an AUC of 0.74, suggesting that the integration of Google data with clinical information at the time of first hospitalization could potentially serve to improve the accuracy and reliability of clinical diagnoses [52]. Compared to healthy participants,

those with SSD made fewer searches and their searches consisted of fewer words. Reduced search activity may represent declining interests and engagement with the environment [53-55]; as positive and negative symptoms of schizophrenia escalate, individuals with SSD may become less invested in their environment and increasingly internally preoccupied. Alternatively, reduced search activity could be related to cognitive deficits that are commonly associated with schizophrenia [56]. Given that cognitive changes may be subtle early in the course of illness [57], having an objective way by which to identify cognitive markers of SSD could contribute valuable information to the diagnostic process and inform treatment recommendations. Future research will need to explore precisely when changes first manifest online as well as their clinical significance. Online search data typically exist from the origin of an individual's Google account, and the present results suggest that search data could prove to be particularly useful in charting the trajectory of an individual's illness, as well as in contributing useful information about the timing of symptomatic changes.

Compared to healthy participants, those with SSD were significantly less likely to search for content related to "positive affect" (eg, "happy," "good"), and less likely to search for content related to "anxiety" (eg, "nervous," "tense"). These findings are consistent with the experience of low mood, apathy, and reduced emotional expression often associated with SSD [58,59]. These symptoms often predate the positive symptoms

such as hallucinations and delusions, and therefore an objective method to identify them could help to overcome limitations of patient self-report to inform early intervention. Participants with SSD were also less likely to search using words from the relative (motion, space, time), inhibition (block, constrain), and inclusive (with, include) categories, and were less likely to use quantifiers (few, many, much) and prepositions (on, to, from). Determining the clinical significance of these differences requires additional research; however, they appear to be related to the complexity of the search query. Individuals with SSD often experience concrete thinking [60] in addition to the cognitive limitations noted above, and may therefore use less complex language when searching for information online.

Features Relevant to the Relapse Classifier

Relapse periods could be distinguished from healthy periods with an AUC of 0.71. During a relapse period, participants with SSD were more likely to use words from the hear (heard, listen, sound), bio (eat, blood, pain), perception (see, touch, listen), and anger categories. They were less likely to use words related to health. These changes could be consistent with increasing delusions, hallucinations, and irritability during a psychotic relapse [61-63]. Previous work has identified changes in language use on social media that occur alongside escalating psychotic symptoms [33]. Thus, future research should aim to identify the point in illness progression at which linguistic shifts emerge online so as to make the best clinical use of this information.

Compared with periods of relative health, search length became shorter and the frequency of search activity decreased closer to the date of the relapse hospitalization. This could be indicative of a further decline in cognition function [56], or perhaps due to the presence of distracting internal stimuli. Fewer searches may also represent disengagement from one's environment and reduced desire to ask questions and seek answers. This would be consistent with the avolition and negative symptoms commonly experienced by individuals with first-episode psychosis [59]. Additional research will be required to determine the precise clinical correlates.

Limitations

The first limitation is that our approach was limited by our characterization of monthly periods of relative health and relative illness. The illness trajectory for individuals with SSD does not neatly fall into distinct segments of "health" and "illness," and symptoms instead fluctuate over time. In addition, discharge from hospital does not necessarily mean full resolution of symptoms; therefore, we might have underestimated the potential differences between periods of illness and health. Furthermore, the inpatient hospitalization dates were obtained via medical records, and it is possible that some hospitalizations were missing from the record and therefore not included in our analyses. Related is the fact that the specific symptoms that define an exacerbation for each individual with SSD are often unique, and the impact of symptom heterogeneity on searches should also be explored in future work. To address these limitations and to improve the ability to find associations between online activity and psychotic symptoms, future studies need to monitor participants prospectively and utilize symptom

rating scales to more accurately assess symptom changes and severity as well as to determine the specificity and sensitivity of our findings in comparison to other diagnostic groups. Additionally, future research will need to consider the potential influence of various life events, including search patterns associated with work and school.

Second, some individuals with SSD are diagnosed well before or long after the first psychiatric hospitalization, and therefore the generalizability of our diagnostic classifier is currently unknown. Ongoing efforts focused on understanding search behavior throughout the entire course of illness development, progression, and care should explore potential differences in those who are diagnosed at various time points.

Third, some participants were more active online than others, providing varying degrees of extractable data. An important question for future research will be how much search data is necessary to make reliable clinical predictions.

Fourth, the archives used for our analyses were collected retrospectively. Although retrospective collection eliminates the possibility of altering behavior as a result of being monitored, to achieve the goal of making real-time predictions, identifying clinically meaningful patterns in search data prospectively will be necessary.

Fifth, the eligibility criterion for age was between 15 and 35 years to reflect the inclusion criteria of the Early Treatment Program; however, adolescents may engage with search engines in a distinct manner compared to young adults, and age will need to be considered in future initiatives.

Sixth, we used nonlinear kernels in our classification models to accommodate nonlinear feature dependencies in the models. Although this is recommended for improved classification performance, it can also limit the interpretability of the features based on linear permutation methods. The feature permutation does not test for nonlinear permutations in features.

Seventh, for the purpose of this feasibility study, we considered both classification accuracy as well as feature interpretability in selecting our models. Further research with additional data from more participants is required to test the scalability of the selected classifiers and features as well as their generalizability to other online search engines beyond Google.

Finally, Google takeout only extracts search data collected while an individual is signed in to their account. Some participants may have searched for information while signed out, and these data would not have been captured in their archives.

Conclusion and Prospects

Although search data alone are not sufficient to make a diagnosis or to predict a relapse, the integration of these data with information collected through traditional clinical means could be useful. Previous work has demonstrated that many people search for information online long before seeking help in person [9-11], and this study highlights the existence of a diagnostic signal in daily search patterns. Online services could one day facilitate the transition from information-seeking to help-seeking, hasten the diagnostic process, and help to reduce the burden of untreated psychosis. This approach could also be

XSL•FO

RenderX

beneficial for relapse identification, enabling earlier intervention. Prior initiatives have explored the utility of smartphone sensor data (ie, geolocation, physical activity, phone usage, and speech), wearables, and social media activity to predict symptom fluctuations [44,64-66]. Our results demonstrate that user-generated search activity represents another potentially critical source of digital data contributing to the diagnostic process and relapse identification. Future work combining digital data from multiple sources will likely result in the most effective clinical tools. However, how to effectively and ethically incorporate personalized patterns of activity into clinical workflow are critical questions of inquiry. Interdisciplinary teams of researchers, clinicians, and patients must continue to work together on exploring challenges in ethics, privacy, consent, clinical responsibility, and data ownership. As our analyses become increasingly sophisticated and our ability to predict health information improves, stakeholders must develop standards to protect the confidentiality and the rights of this sensitive population, and ensure that the enabled technologies are used in the service of positive outcomes for the patients.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

Top 20 features for the support vector machine (SVM) and gradient boost (GB) relapse classifiers in order of importance.
[DOCX File , 13 KB-Multimedia Appendix 1]

References

1. Kane JM, Correll CU. Past and present progress in the pharmacologic treatment of schizophrenia. J Clin Psychiatry 2010 Sep;71(9):1115-1124. [doi: 10.4088/JCP.10r06264yel] [Medline: 20923620]

2. Correll CU, Galling B, Pawar A, Krivko A, Bonetto C, Ruggeri M, et al. Comparison of Early Intervention Services vs Treatment as Usual for Early-Phase Psychosis: A Systematic Review, Meta-analysis, and Meta-regression. JAMA Psychiatry 2018 Jun 01;75(6):555-565 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.0623] [Medline: 29800949]

3. Murru A, Carpiniello B. Duration of untreated illness as a key to early intervention in schizophrenia: A review. Neurosci Lett 2018 Mar 16;669:59-67 [FREE Full text] [doi: 10.1016/j.neulet.2016.10.003] [Medline: 27717830]

4. Addington J, Heinssen RK, Robinson DG, Schooler NR, Marcy P, Brunette MF, et al. Duration of Untreated Psychosis in Community Treatment Settings in the United States. Psychiatr Serv 2015 Jul;66(7):753-756. [doi: 10.1176/appi.ps.201400124] [Medline: 25588418]

5. Robinson D, Woerner MG, Alvir JMJ, Bilder R, Goldman R, Geisler S, et al. Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. Arch Gen Psychiatry 1999 Mar 01;56(3):241-247. [doi: 10.1001/archpsyc.56.3.241] [Medline: 10078501]

6. Díaz-Caneja CM, Pina-Camacho L, Rodríguez-Quiroga A, Fraguas D, Parellada M, Arango C. Predictors of outcome in early-onset psychosis: a systematic review. NPJ Schizophr 2015 Mar 4;1(1):14005. [doi: 10.1038/npjschz.2014.5] [Medline: 27336027]

7. Ascher-Svanum H, Zhu B, Faries DE, Salkever D, Slade EP, Peng X, et al. The cost of relapse and the predictors of relapse in the treatment of schizophrenia. BMC Psychiatry 2010 Jan 7;10(1):2 [FREE Full text] [doi: 10.1186/1471-244x-10-2]

8. Lozupone M, Seripa D, Stella E, La Montagna M, Solfrizzi V, Quaranta N, et al. Innovative biomarkers in psychiatric disorders: a major clinical challenge in psychiatry. Expert Rev Proteomics 2017 Sep;14(9):809-824. [doi: 10.1080/14789450.2017.1375857] [Medline: 28870126]

9. Lozupone M, La Montagna M, D'Urso F, Daniele A, Greco A, Seripa D, et al. The Role of Biomarkers in Psychiatry. Adv Exp Med Biol 2019;1118:135-162. [doi: 10.1007/978-3-030-05542-4_7] [Medline: 30747421]

10. Kalia M, Costa E Silva J. Biomarkers of psychiatric diseases: current status and future prospects. Metabolism 2015 Mar;64(3 Suppl 1):S11-S15 [FREE Full text] [doi: 10.1016/j.metabol.2014.10.026] [Medline: 25467847]

11. Bzdok D, Meyer-Lindenberg A. Machine Learning for Precision Psychiatry: Opportunities and Challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018 Mar;3(3):223-230 [FREE Full text] [doi: 10.1016/j.bpsc.2017.11.007] [Medline: 29486863]

12. Tai AMY, Albuquerque A, Carmona NE, Subramanieapillai M, Cha DS, Sheko M, et al. Machine learning and big data: Implications for disease modeling and therapeutic discovery in psychiatry. Artif Intell Med 2019 Aug;99:101704. [doi: 10.1016/j.artmed.2019.101704] [Medline: 31606109]

13. Hsin H, Fromer M, Peterson B, Walter C, Fleck M, Campbell A, et al. Transforming Psychiatry into Data-Driven Medicine with Digital Measurement Tools. NPJ Digit Med 2018;1:37 [FREE Full text] [doi: 10.1038/s41746-018-0046-0] [Medline: 31304319]

14. Biswas S. Digital Indians: Ben Gomes. BBC News. 2013 Sep. URL: http://www.bbc.com/news/technology-23866614 [accessed 2019-01-01]

15. Burns JM, Davenport TA, Durkin LA, Luscombe GM, Hickie IB. The internet as a setting for mental health service utilisation by young people. Med J Aust 2010 Jun 07;192(S11):S22-S26. [Medline: 20528703]

16. Birnbaum ML, Rizvi AF, Faber K, Addington J, Correll CU, Gerber C, et al. Digital Trajectories to Care in First-Episode Psychosis. Psychiatr Serv 2018 Dec 01;69(12):1259-1263 [FREE Full text] [doi: 10.1176/appi.ps.201800180] [Medline: 30256181]

17. Birnbaum ML, Rizvi AF, Confino J, Correll CU, Kane JM. Role of social media and the Internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry 2017 Aug;11(4):290-295. [doi: 10.1111/eip.12237] [Medline: 25808317]

18. Van Meter AR, Birnbaum ML, Rizvi A, Kane JM. Online help-seeking prior to diagnosis: Can web-based resources reduce the duration of untreated mood disorders in young people? J Affect Disord 2019 Jun 01;252:130-134 [FREE Full text] [doi: 10.1016/j.jad.2019.04.019] [Medline: 30981056]

19. Schrank B, Sibitz I, Unger A, Amering M. How patients with schizophrenia use the internet: qualitative study. J Med Internet Res 2010 Dec 19;12(5):e70 [FREE Full text] [doi: 10.2196/jmir.1550] [Medline: 21169176]

20. Powell J, Clarke A. Investigating internet use by mental health service users: interview study. Stud Health Technol Inform 2007;129(Pt 2):1112-1116. [Medline: 17911888]

21. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827 [FREE Full text] [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]

22. Paparrizos J, White R, Horvitz E. Detecting Devastating Diseases in Search Logs. 2018 Aug Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; San Francisco p. 559-568. [doi: 10.1145/2939672.2939722]

23. Paparrizos J, White RW, Horvitz E. Screening for Pancreatic Adenocarcinoma Using Signals From Web Search Logs: Feasibility Study and Results. J Oncol Pract 2016 Aug;12(8):737-744. [doi: 10.1200/JOP.2015.010504] [Medline: 27271506]

24. White RW, Horvitz E. Evaluation of the Feasibility of Screening Patients for Early Signs of Lung Carcinoma in Web Search Logs. JAMA Oncol 2017 Mar 01;3(3):398-401. [doi: 10.1001/jamaoncol.2016.4911] [Medline: 27832243]

25. White RW, Doraiswamy PM, Horvitz E. Detecting neurodegenerative disorders from web search signals. NPJ Digit Med 2018;1:8 [FREE Full text] [doi: 10.1038/s41746-018-0016-6] [Medline: 31304293]

26. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoţiuc-Pietro D, et al. Facebook language predicts depression in medical records. Proc Natl Acad Sci USA 2018 Dec 30;115(44):11203-11208 [FREE Full text] [doi: 10.1073/pnas.1802331115] [Medline: 30322910]

27. De CM, Gamon M. Predicting depression via social media. 2013 Presented at: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media; June 2013; Boston, MA p. 128-137. [doi: 10.1109/ISADS.2017.41]

28. Reece AG, Reagan AJ, Lix KLM, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Oct 11;7(1):13006. [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]

29. De CM, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2016 May Presented at: Special Interest Group on Computer Human Interaction; May 2016; San Jose p. 2098-2110 URL: http://europepmc.org/abstract/MED/29082385 [doi: 10.1145/2858036.2858207]

30. Coppersmith G, Leary R, Crutchley P, Fine A. Natural Language Processing of Social Media as Screening for Suicide Risk. Biomed Inform Insights 2018;10:1178222618792860 [FREE Full text] [doi: 10.1177/1178222618792860] [Medline: 30158822]

31. De CM, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. 2012 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 2012; Austin. [doi: 10.1145/2470654.2466447]

32. Birnbaum ML, Ernala SK, Rizvi AF, De CM, Kane JM. A Collaborative Approach to Identifying Social Media Markers of Schizophrenia by Employing Machine Learning and Clinical Appraisals. J Med Internet Res 2017 Aug 14;19(8):e289 [FREE Full text] [doi: 10.2196/jmir.7956] [Medline: 28807891]

33. Birnbaum ML, Ernala SK, Rizvi AF, Arenare E, Van Meter AR, De Choudhury M, et al. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. NPJ Schizophr 2019 Oct 07;5(1):17 [FREE Full text] [doi: 10.1038/s41537-019-0085-9] [Medline: 31591400]

34. Zaman A, Acharyya R, Kautz H. Detecting Low Self-Esteem in Youths from Web Search Data. 2019 Presented at: World Wide Web Conference; 2019; San Francisco p. 2270-2280. [doi: 10.1145/3308558.3313557]

35. Ernala S, Birnbaum M, Candan K. Methodological gaps in predicting mental health states from social media: triangulating diagnostic signals. 2019 May Presented at: Conference of Human-Computer Interaction; May 2019; Glasgow. [doi: 10.1145/3290605.3300364]

36. First MB, Spitzer R, Gibbon M, Willams JB. Structured clinical interview for DSM-IV axis I disorders, clinician version (SCID-CV). Washington, DC: American Psychiatric Press; 1996.

37. Zimmerman M, Mattia JI. A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. Arch Gen Psychiatry 2001 Aug 01;58(8):787-794. [doi: 10.1001/archpsyc.58.8.787] [Medline: 11483146]

38. Birchwood M, Smith J, Macmillan F, Hogg B, Prasad R, Harvey C, et al. Predicting relapse in schizophrenia: the development and implementation of an early signs monitoring system using patients and families as observers, a preliminary investigation. Psychol Med 1989 Aug;19(3):649-656. [doi: 10.1017/s0033291700024247] [Medline: 2798634]

39. Henmi Y. Prodromal symptoms of relapse in schizophrenic outpatients: retrospective and prospective study. Jpn J Psychiatry Neurol 1993 Dec;47(4):753-775. [doi: 10.1111/j.1440-1819.1993.tb01824.x] [Medline: 7911165]

40. Buck B, Scherer E, Brian R, Wang R, Wang W, Campbell A, et al. Relationships between smartphone social behavior and relapse in schizophrenia: A preliminary report. Schizophr Res 2019 Jun;208:167-172. [doi: 10.1016/j.schres.2019.03.014] [Medline: 30940400]

41. Anderson KK, Fuhrer R, Malla AK. The pathways to mental health care of first-episode psychosis patients: a systematic review. Psychol Med 2010 Mar 18;40(10):1585-1597. [doi: 10.1017/s0033291710000371]

42. Chung CK, Pennebaker JW. Linguistic inquiry and word count (LIWC): pronounced "Luke,"... and other useful facts. In: McCarthy PM, Boonthum-Denecke C, editors. Applied Natural Language Processing: Identification, Investigation and Resolution -. Hershey, PA: IGI Global; 2012:206-229.

43. Tausczik YR, Pennebaker JW. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. J Lang Soc Psychol 2009 Dec 08;29(1):24-54. [doi: 10.1177/0261927X09351676]

44. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881. [doi: 10.1126/science.1202775] [Medline: 21960633]

45. Sadilek A, Homan C, Lasecki W. Modeling fine-grained dynamics of mood at scale. 2013 Presented at: Web Search and Data Mining; 2013; Rome.

46. Dzogang F, Lightman S, Cristianini N. Circadian mood variations in Twitter content. Brain Neurosci Adv 2017 Jan 01;1:2398212817744501. [doi: 10.1177/2398212817744501] [Medline: 29270466]

47. Flaxman AD, Vahdatpour A, Green S, James SL, Murray CJ, Population H. Random forests for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards. Popul Health Metr 2011 Aug 04;9:29 [FREE Full text] [doi: 10.1186/1478-7954-9-29] [Medline: 21816105]

48. Wu TF, Lin CJ, Weng RC. Probability estimates for multi-class classification by pairwise coupling. J Machine Learn Res 2004:975-1005 [FREE Full text]

49. Friedman J. Greedy Function Approximation: A Gradient Boosting Machine. Ann Statist 2001 Oct;29(5):1189-1232. [doi: 10.1214/aos/1013203451]

50. Varoquaux G, Buitinck L, Louppe G, Grisel O, Pedregosa F, Mueller A. Scikit-learn. GetMobile: Mobile Comp and Comm 2015 Jun;19(1):29-33. [doi: 10.1145/2786984.2786995]

51. Raeder T, Forman G, Chawla N. Learning from imbalanced datavaluation matters. In: Holmes DE, Lakhmi JC, editors. Data mining: Foundations and intelligent paradigms. Berlin, Heidelberg: Springer; 2012.

52. Zander E, Wyder L, Holtforth MG, Schnyder U, Hepp U, Stulz N. Validity of routine clinical diagnoses in acute psychiatric inpatients. Psychiatry Res 2018 Jan;259:482-487 [FREE Full text] [doi: 10.1016/j.psychres.2017.11.004] [Medline: 29154169]

53. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th edition. Washington DC: APA Publishing; 2013.

54. Lieberman JA, Perkins D, Belger A, Chakos M, Jarskog F, Boteva K, et al. The early stages of schizophrenia: speculations on pathogenesis, pathophysiology, and therapeutic approaches. Biol Psychiatry 2001 Dec 01;50(11):884-897. [doi: 10.1016/s0006-3223(01)01303-8] [Medline: 11743943]

55. Soares-Weiser K, Maayan N, Bergman H, Davenport C, Kirkham AJ, Grabowski S, et al. First rank symptoms for schizophrenia. Cochrane Database Syst Rev 2015 Jan 25;1:CD010653 [FREE Full text] [doi: 10.1002/14651858.CD010653.pub2] [Medline: 25879096]

56. Sheffield JM, Karcher NR, Barch DM. Cognitive Deficits in Psychotic Disorders: A Lifespan Perspective. Neuropsychol Rev 2018 Dec;28(4):509-533 [FREE Full text] [doi: 10.1007/s11065-018-9388-2] [Medline: 30343458]

57. Bora E, Yalincetin B, Akdede BB, Alptekin K. Duration of untreated psychosis and neurocognition in first-episode psychosis: A meta-analysis. Schizophr Res 2018 Mar;193:3-10. [doi: 10.1016/j.schres.2017.06.021] [Medline: 28634088]

58. Coentre R, Talina MC, Góis C, Figueira ML. Depressive symptoms and suicidal behavior after first-episode psychosis: A comprehensive systematic review. Psychiatry Res 2017 Jul;253:240-248. [doi: 10.1016/j.psychres.2017.04.010] [Medline: 28395229]

59. Gee B, Hodgekins J, Fowler D, Marshall M, Everard L, Lester H, et al. The course of negative symptom in first episode psychosis and the relationship with social recovery. Schizophr Res 2016 Jul;174(1-3):165-171. [doi: 10.1016/j.schres.2016.04.017] [Medline: 27131912]

60. Lysaker PH, Leonhardt BL, Pijnenborg M, van Donkersgoed R, de Jong S, Dimaggio G. Metacognition in schizophrenia spectrum disorders: methods of assessment and associations with neurocognition, symptoms, cognitive style and function. Isr J Psychiatry Relat Sci 2014;51(1):54-62. [Medline: 24858635]

61. Gleeson JF, Rawlings D, Jackson HJ, McGorry PD. Early warning signs of relapse following a first episode of psychosis. Schizophr Res 2005 Dec 01;80(1):107-111. [doi: 10.1016/j.schres.2005.07.019] [Medline: 16125373]

XSL•FO

RenderX

62. Alvarez-Jimenez M, Priede A, Hetrick SE, Bendall S, Killackey E, Parker AG, et al. Risk factors for relapse following treatment for first episode psychosis: a systematic review and meta-analysis of longitudinal studies. Schizophr Res 2012 Aug;139(1-3):116-128. [doi: 10.1016/j.schres.2012.05.007] [Medline: 22658527]

63. Herz MI, Melville C. Relapse in schizophrenia. Am J Psychiatry 1980 Jul;137(7):801-805. [doi: 10.1176/ajp.137.7.801] [Medline: 6104444]

64. Eisner E, Bucci S, Berry N, Emsley R, Barrowclough C, Drake RJ. Feasibility of using a smartphone app to assess early signs, basic symptoms and psychotic symptoms over six months: A preliminary report. Schizophr Res 2019 Jun;208:105-113 [FREE Full text] [doi: 10.1016/j.schres.2019.04.003] [Medline: 30979665]

65. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]

66. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting Mood Disturbance Severity with Mobile Phone Keystroke Metadata: A BiAffect Digital Phenotyping Study. J Med Internet Res 2018 Jul 20;20(7):e241 [FREE Full text] [doi: 10.2196/jmir.9775] [Medline: 30030209]

Abbreviations

AUC: area under the receiver operating characteristic curve.
GB: gradient boost
IRB: Institutional Review Board
LIWC: linguistic inquiry and word count
RF: random forest
SSD: schizophrenia spectrum disorders
SVM: support vector machine

XSL•FO

RenderX