CS 6474/CS 4803 Social Computing: Health and Well-Being

Munmun De Choudhury

munmund@gatech.edu

Week 7 | February 23, 2026

Social Media Use in the United States: Implications for Health Communication

Wen-ying Sylvia Chou[1,2], PhD, MPH; Yvonne M Hunt[2], PhD, MPH; Ellen Burke Beckjord[3], PhD, MPH; Richard P Moser[4], PhD; Bradford W Hesse[2], PhD

[1]Cancer Prevention Fellowship Program, National Cancer Institute, Bethesda, MD, USA
[2]Health Communication and Informatics Research Branch, National Cancer Institute, Bethesda, MD, USA
[3]RAND Corporation, Pittsburgh, PA, USA
[4]Behavioral Research Program, National Cancer Institute, Bethesda, MD, USA

Identify the sociodemographic and health-related factors associated with current adult social media users in the United States.

Table 1. Weighted sample characteristics: proportion of non-Internet and Internet users in each category

Characteristic	Non-Internet Users (N = 2566, 31.46%)	Internet Users (N = 5078, 68.54%)
Age	$P < .001$	
18-24	15.2%	84.8%
25-34	23.0%	77.0%
35-44	21.7%	78.3%
45-54	28.9%	71.1%
55-64	33.0%	67.0%
65+	66.4%	33.6%
Gender	$P = .003$	
Male	33.6%	66.4%
Female	29.5%	70.6%
Education	$P < .001$	
High school or less	50.5%	49.5%
Some college	17.1%	82.9%
College graduate	9.0%	91.0%
Race/ethnicity	$P < .001$	
Non-Hispanic white	25.0%	75.2%
Hispanic	50.7%	49.3%
Black/African American	43.3%	56.8%
Other[a]	25.8%	74.2%
General health	$P < .001$	
Excellent, very good, or good	26.7%	73.3%
Fair or poor	51.4%	48.6%
Psychological distress	$P < .001$	
Yes	43.4%	56.6%
No	28.4%	71.6%
Cancer experience	$P < .001$	
No personal experience with cancer	36.7%	63.3%
Had family with cancer	26.3%	73.7%
Cancer survivor	43.1%	56.9%
Have regular health care provider	$P < .001$	
Yes	29.1%	70.9%
No	35.7%	64.3%

[a] Other includes American Indian, Asian American, Pacific Islander, Native Hawaiian, Alaskan Native, and multiple races mentioned.

Original Article

Who Gives a Tweet: Assessing Patients' Interest in the Use of Social Media for Health Care

Jennifer Fisher DNP, APRN-BC, Margaret Clayton PhD, APRN-BC

✕ VIEW METRICS

✉ Address correspondence to Dr. Jennifer Fisher, University of Utah, College of Nursing, 10 South 2000 East, Salt Lake City, UT, 84112; iamjenni@gmail.com

Article

Empowering patients through social media: The benefits and challenges

Mowafa Househ
College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, National Guard Health Affairs, Saudi Arabia

Elizabeth Borycki and Andre Kushniruk
School of Health Information Science, University of Victoria, Canada

Abstract
This article explores the range of social media platforms used by patients and examines the benefits and challenges of using these tools from a patient perspective. A literature review was performed to investigate the use of social media technology by patients. The MEDLINE database was searched using the terms "social

How Social Media Will Change Public Health

Mark Dredze, *Johns Hopkins University*

Social media such as Twitter have created platforms for people to broadcast information, thoughts, and feelings about their daily lives. Since Twitter messages (called *tweets*) often reflect in-the-moment updates, they're filled with useful observations and information about the larger world. Researchers have examined a range of applications based on tweets, ranging from political polling[1] to earthquake monitoring,[2] that have demonstrated Twitter's ability to deliver fast, cheap, and reliable tools for monitoring real-world events.

These successes have drawn interest from the public-health community, whose goal is to study the health of a population and develop policies that improve health outcomes. Traditionally, this requires expensive, time-consuming monitoring mechanisms, primarily surveys and data collection from clinical encounters. Even high-priority projects, such as the US Centers for Disease Control and Prevention's (CDC's) FluView program that tracks the weekly US influenza rate, are still slow because they require clinical data aggregation. Twitter and other social media could reduce cost and provide real-time statistics about public health.

Recent work in machine learning and natural language processing has studied the health content of tweets and demonstrated the potential for extracting useful public-health information from their aggregation. This article examines the types of health topics discussed on Twitter, and how tweets can both augment existing public-health capabilities and enable new ones. I also discuss key challenges that researchers must address to deliver high-quality tools to the public-health community.

Discovering Health Topics on Twitter

Twitter's size and breadth make it difficult to determine exactly which types of public-health work it can support. Initial work in my research group[3,4] explored health-related tweets and topics on Twitter through the development of new computational models. Because many public-health activities are disease-oriented, we developed a model that discovered diseases (ailments) from raw tweets for guided exploration, rather than relying on predefined illnesses. We used supervised learning to filter tweets and find health-related messages, yielding 1.6 million English health tweets from March 2009 to October 2010.

To explore these tweets, we developed the Ailment Topic Aspect Model (ATAM), a probabilistic graphical model for uncovering ailments.[3] ATAM assumes that each message discusses a single ailment, manifested through the message's words, and associates three types of words (general disease words, symptoms, and treatments) with ailments. For example, the message "fever + headache = flu, home sick with Tylenol" discusses influenza, where "fever" and "headache" are symptoms, "Tylenol" a treatment, and "flu" a general word associated with the ailment.

Human annotators labeled 15 ailments discovered by ATAM, including headaches, influenza, insomnia, obesity, dental problems, and seasonal allergies. Examining the words, symptoms, and treatments most associated with each ailment, and the groups of messages that discuss each ailment, can support a variety of public-health initiatives.
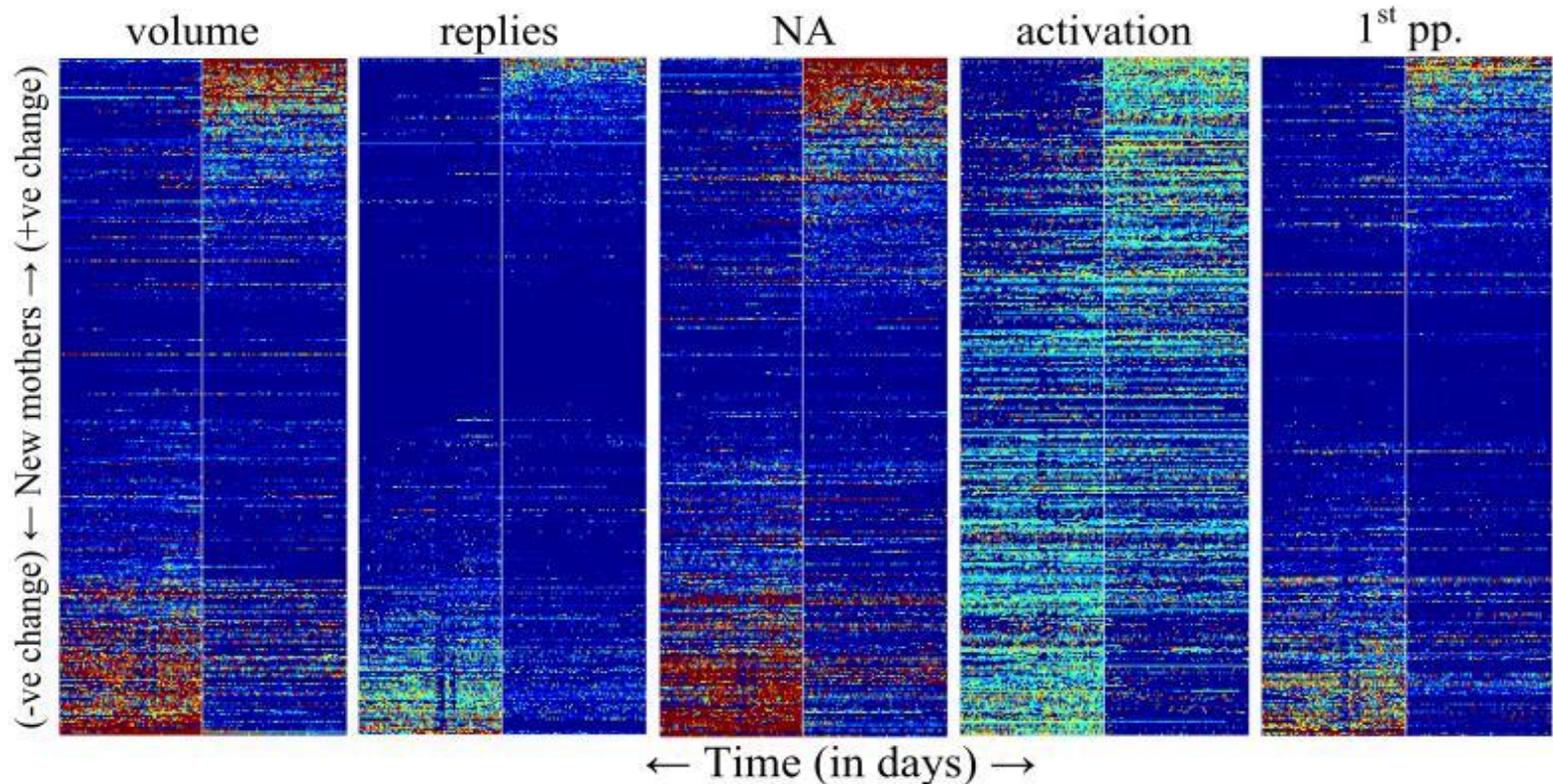
Augmenting Existing Public-Health Capabilities

A core capability of public-health programs, biosurveillance monitors a population for adverse health events, which include expected seasonal events, such as influenza or environmental allergies, disease outbreaks, such as the H1N1 virus, and other health threats, such as food poisoning or a biochemical contaminant. Surveillance is the

81

Figure 1. The rate of Twitter messages about seasonal allergies for June 2010. Messages were automatically coded using a machine-learning method and geo-located based on user-provided location. Overall shading indicates significant allergy messages, showing the heart of allergy season. States in the Northeast and Midwest are particularly active. Dashed states had insufficient data.

Figure 2. A word cloud visualization showing the words most associated with the ailment "insomnia" as discovered by a machine-learning model that examined 1.6 million tweets related to health. Larger fonts indicate more related terms, blue indicates general terms, red highlights symptoms, and green represents treatments. General words such as "hours," "awake," and "tired" characterize insomnia messages, with symptoms such as "nightmares" and "yawning" and treatments of "Benadryl" and "sleeping pills."

Social Media Derived Behavioral and Affective Markers Predict Postpartum Changes

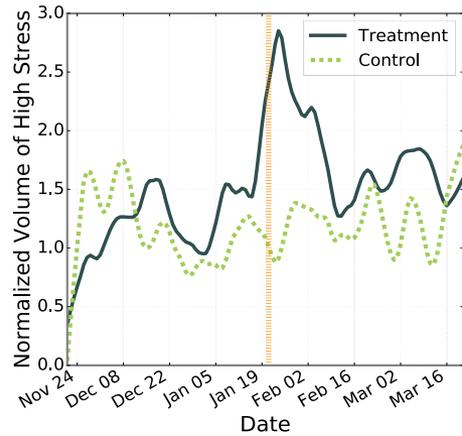# Social Media Derived Behavioral and Affective Markers Predict Postpartum Changes

376 users (new mothers); 40,426 posts between March 2011 and July 2012

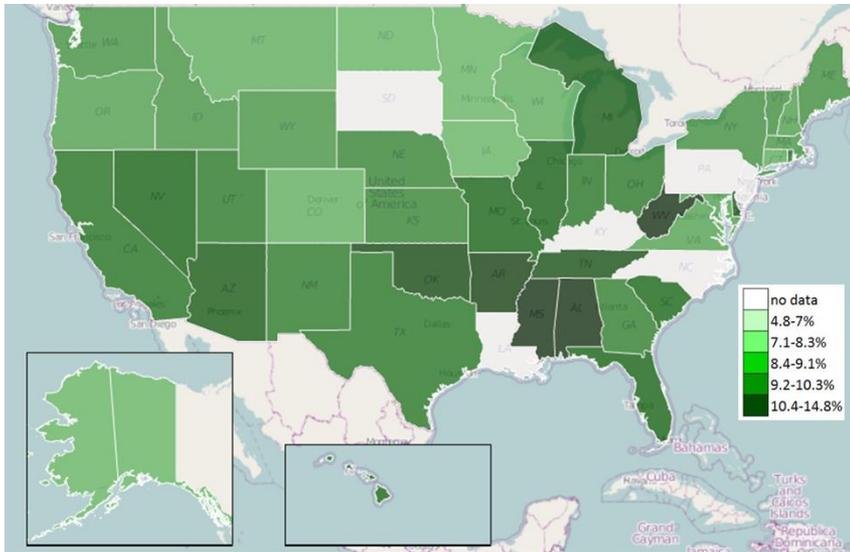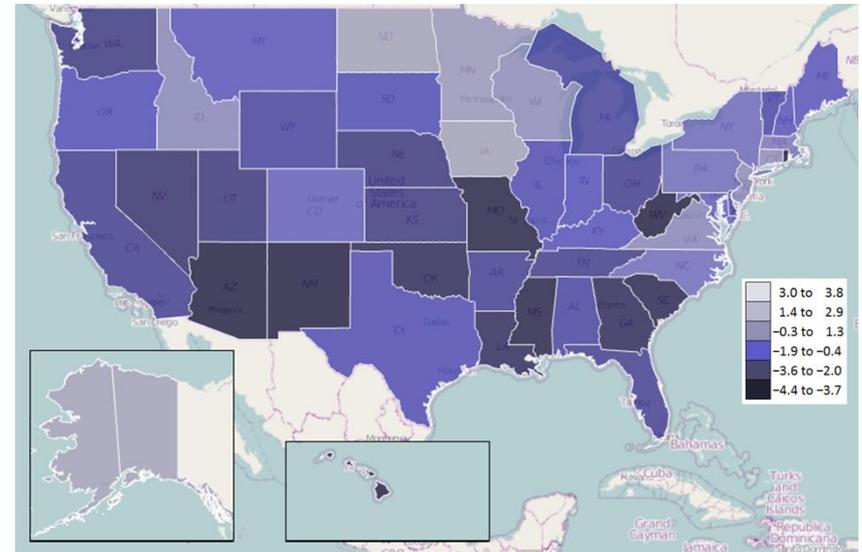# Measuring Levels of Acute Stress in College Campuses with Social Media
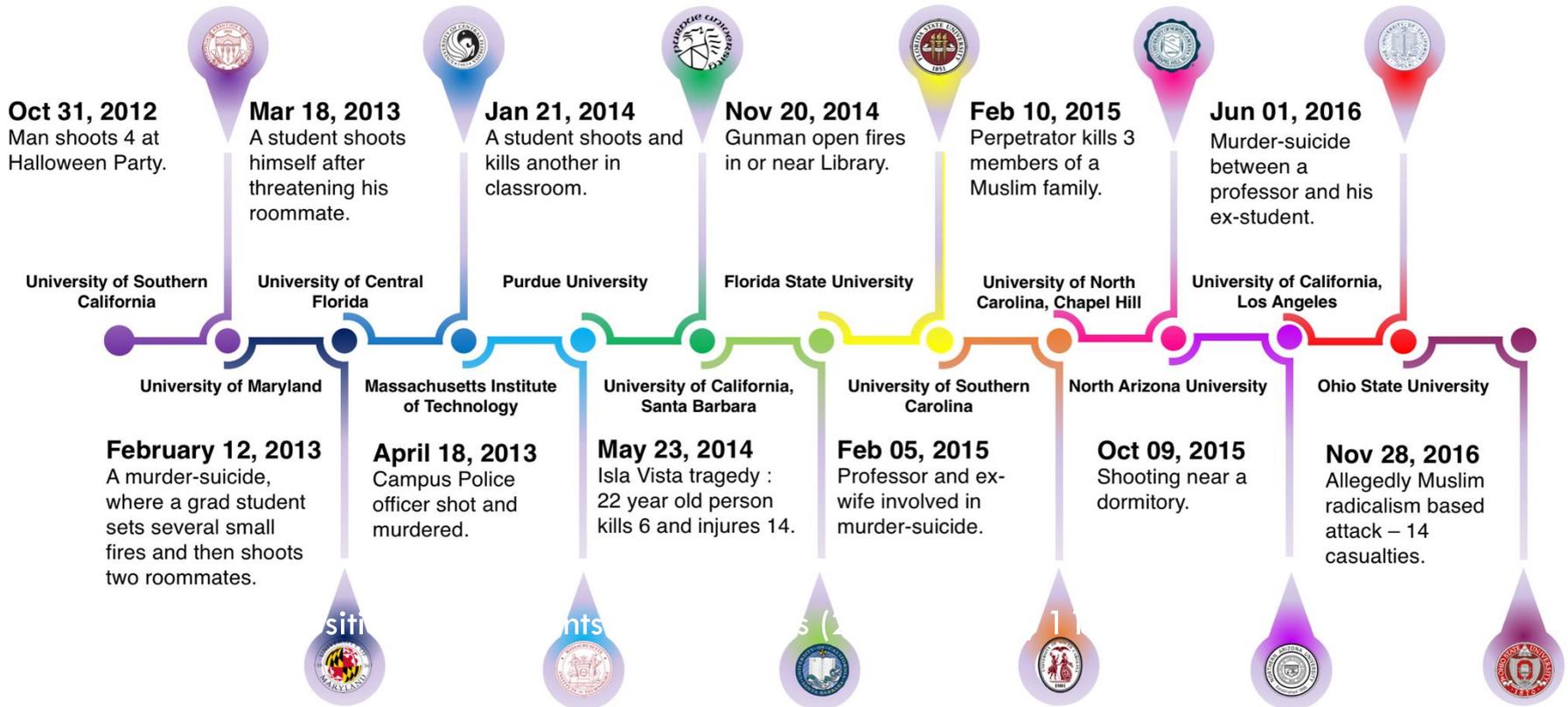
Temporal and Linguistic Patterns of Stress
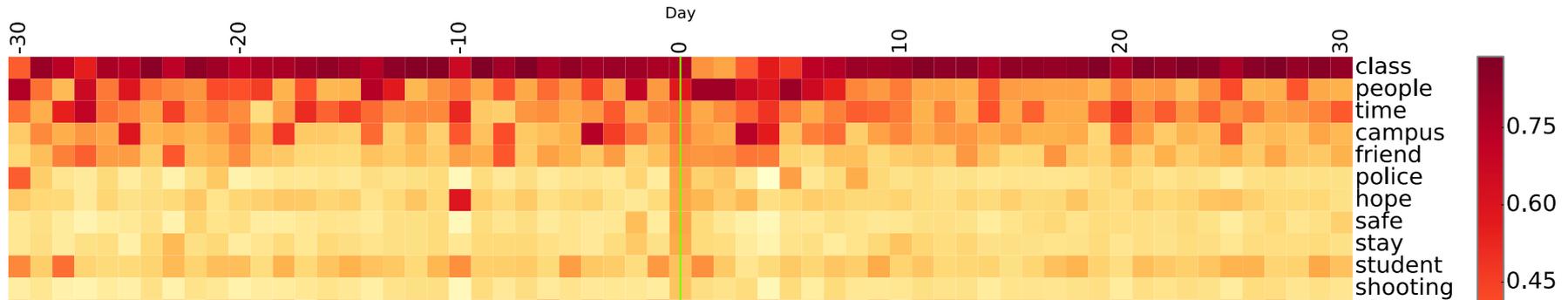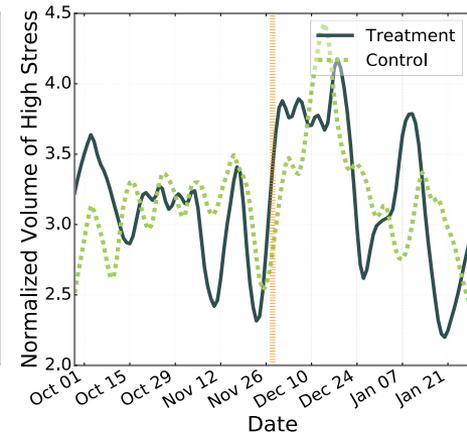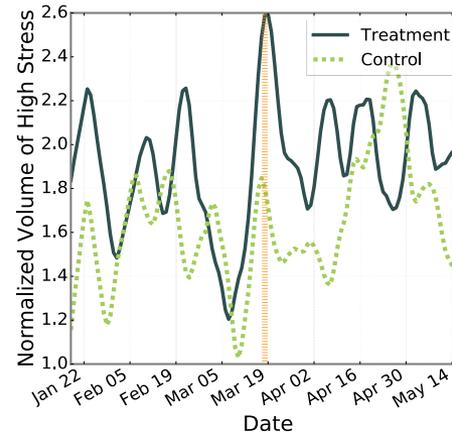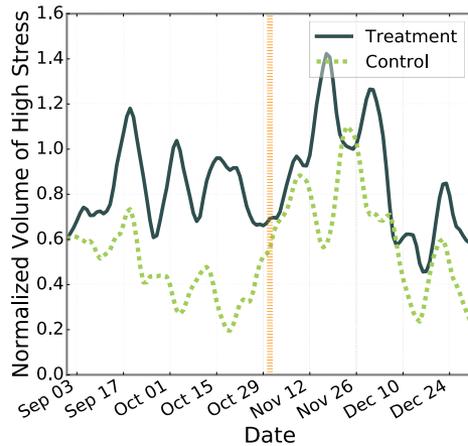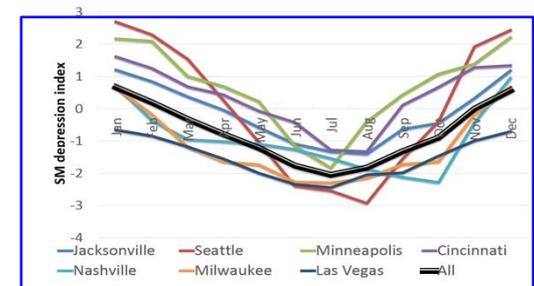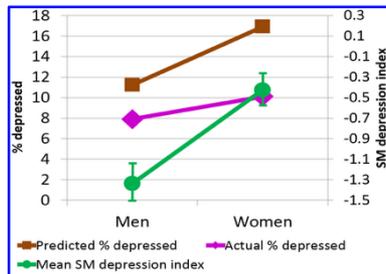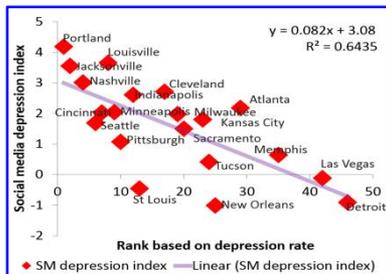
Social media depression index

actual (BRFSS data)

predicted (SMDI)

no data
4.8-7%
7.1-8.3%
8.4-9.1%
9.2-10.3%
10.4-14.8%

3.0 to 3.8
1.4 to 2.9
-0.3 to 1.3
-1.9 to -0.4
-3.6 to -2.0
-4.4 to -3.7

Socio-demographic, spatio-temporal patterns of prevalence of depression
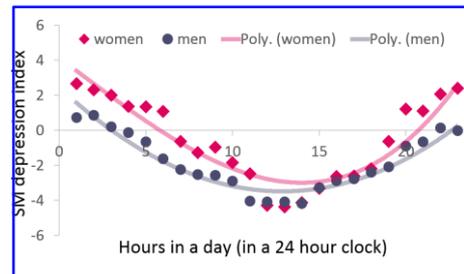
Multi-Task Learning for Mental Health using Social Media Text

Adrian Benton
Johns Hopkins University
adrian@cs.jhu.edu

Margaret Mitchell
Microsoft Research*
mmitchell@google.com

Abstract

We introduce initial groundwork for estimating suicide risk and mental health in a deep learning framework. By modeling multiple conditions, the system learns to make predictions about suicide risk and mental health at a low false positive rate. Conditions are modeled as tasks in a multi-task learning (MTL) framework, with gender prediction as an additional auxiliary task. We demonstrate the effectiveness of multi-task learning by comparison to a well-tuned single-task baseline with the same number of parameters. Our best MTL model predicts potential suicide attempt, as well as the presence of atypical mental health, with AUC > 0.8. We also find additional large improvements using multi-task learning on mental health tasks with limited training data.

1 Introduction

Suicide is one of the leading worldwide, and over 90% of by suicide experience mental However, detecting the risk as monitoring the effects of conditions, is challenging. rely on both self-reports and during short sessions with a often unclear when suicide is Consequently, conditions le suicides are often not adequa

*Now at Google Research.
[1]https://www.nami.org/Learn-More/Mental-Health-Conditions/R Conditions/Suicide#sthash.dMAhr
[2]Communication with clinicians shop (Hollingshead, 2016).

Discovering Shifts to S... from Mental Health Cont...

Munmun De Choudhury
Georgia Tech
Atlanta GA 30332
munmund@gatech.edu

Emre Kici...
Microsoft R...
Redmond WA
emre.k@micro...

Glen Coppersmith
Qntfy.io
Crownsville MD, 21032
glen@qntfy.io

ABSTRACT
History of mental illness is a major factor behind suicide risk and ideation. However research efforts toward characterizing and forecasting this risk is limited due to the paucity of information regarding suicidal ideation, exacerbated by the stigma of mental illness. This paper fills gaps in the literature by developing a statistical methodology to infer which individuals could undergo transitions from mental health discourse to suicidal ideation. We utilize semi-anonymous support communities on Reddit as unobtrusive data sources to infer the likelihood of these shifts. We develop language and interactional measures for this purpose, as well as a propensity score

---

Detecting depression and mental illness on social media: an integrative review

Sharath Chandra Guntuku[1], David B Yaden[1], Margaret L Kern[2], Lyle H Ungar[1] and Johannes C Eichstaedt[1]

Although rates of diagnosing mental illness have improved over the past few decades, many cases remain undetected. Symptoms associated with mental illness are observable on Twitter, Facebook, and web forums, and automated methods are increasingly able to detect depression and other mental illnesses. In this paper, recent studies that aimed to predict mental illness using social media are reviewed. Mentally ill users have been identified using screening surveys, their public sharing of a diagnosis on Twitter, or by their membership in an online forum, and they were distinguishable from control users by patterns in their language and online activity. Automated detection methods may help to identify depressed or otherwise at-risk individuals through the large-scale passive monitoring of social media, and in the future may complement existing screening procedures.

Automated analysis of social media potentially provides methods for early detection. If an automated process could detect elevated depression scores in a user, that individual could be targeted for a more thorough assessment, and provided with further resources, support, and treatment. Studies to date have either examined how the use of social media sites correlates with mental illness in users [3] or attempted to detect mental illness through analysis of the content created by users. This review focuses on the latter: studies aimed at predicting mental illness using social media. We first consider methods used to predict depression, and then consider four approaches that have been used in the literature. We compare the different approaches, provide direction for future studies, and consider ethical issues.

Prediction methods

Automated analysis of social media is accomplished by building predictive models, which use 'features,' or variables that have been extracted from social media data. For example, commonly used features include users' language encoded as frequencies of each word, time of posts, and other variables (see Figure 2). Features are

Addresses
[1] University of Pennsylvania, Philadelphia, PA, United States
[2] The University of Melbourne, Melbourne, Australia

Corresponding author: Eichstaedt, Johannes C (Johannes.penn@gmail.com)

Birnbaum et al

...entifying Social Media Markers of ... Machine Learning and Clinical

..., MS; Asra F Rizvi[1,2], MA; Munmun De Choudhury[4*], PhD;
...nd States

...cess in differentiating individuals who ... media have included expert input to evaluate ... more accurate identification ... l linguistic analysis of shared content is

...es of schizophrenia, was appraised for ... a classifier aiming to distinguish users ... expert appraisals on new, unseen Twitter

... including greater use of interpersonal ... on biological processes (P<.001). The ... rom control users with a mean accuracy ...ssifier's precision, recall, and accuracy

...ertise from multiple fields to combine ...online. These collaborations are crucial

...nalysis; Twitter

J Med Internet Res 2017 | vol. 19 | iss. 8 | e289 | p. 1

Facebook language predicts depression in medical records

Johannes C. Eichstaedt[a,1,2], Robert J. Smith[b,1], Raina M. Merchant[b,c], Lyle H. Ungar[a,b], Patrick Crutchley[a,b], Daniel Preoţiuc-Pietro[a], David A. Asch[b,d], and H. Andrew Schwartz[e]

[a]Positive Psychology Center, University of Pennsylvania, Philadelphia, PA 19104; [b]Penn Medicine Center for Digital Health, University of Pennsylvania, Philadelphia, PA 19104; [c]Department of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; [d]The Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center, Philadelphia, PA 19104; and [e]Computer Science Department, Stony Brook University, Stony Brook, NY 11794

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved September 11, 2018 (received for review February 26, 2018)

Depression, the most prevalent mental illness, is underdiagnosed and undertreated, highlighting the need to extend the scope of current screening methods. Here, we use language from Facebook posts of consenting individuals to predict depression in electronic medical records. We assessed the history of Facebook statuses posted by 683 patients visiting a large urban academic emergency department, 114 of whom had a diagnosis of depression in their medical records. Using only the language preceding their first documentation of a diagnosis of depression, we could identify depressed patients with fair accuracy [area under the curve (AUC) = 0.69], approximately matching the accuracy of screening surveys benchmarked against medical records. Restricting Facebook data to the 6 months immediately preceding the first documented diagnosis of depression yielded a higher prediction accuracy (AUC = 0.72) for those users who had sufficient Facebook data. Significant prediction of future depression status was possible as far as 3 months before its first documentation. We found that language predictors of depression include emotional (sadness), interpersonal (loneliness, hostility), and cognitive (preoccupation with the self, rumination) processes. Unobtrusive depression assessment through social media of consenting individuals may become feasible as a scalable complement to existing screening and monitoring procedures.

big data | depression | social media | Facebook | screening

Results

Prediction of Depression. To predict the first diagnosis of depression in the medical record, we built a prediction model by using the textual content of the Facebook posts, post length, frequency of posting, temporal posting patterns, and demographics (*Materials and Methods*). We then evaluated the performance of this model by comparing the probability of depression estimated by our algorithm against the actual presence or absence of depression for each patient in the medical record (using 10-fold cross-validation to avoid overfitting). Varying the threshold of this probability for diagnosis

the diagnosis of depression, which prior research has shown is feasible with moderate accuracy (15). Of the patients enrolled in the study, 114 had a diagnosis of depression in their medical records. For these patients, we determined the date at which the first documentation of a diagnosis of depression was recorded in the EMR of the hospital system. We analyzed the Facebook data generated by each user before this date. We sought to simulate a realistic screening scenario, and so, for each of these 114 users, we identified 5 random control patients without a diagnosis of depression in EMR, examining only the Facebook data they created before the corresponding depressed patient's first date of a recorded diagnosis of depression. This allowed us to compare depressed and control patients' data across the same time span and to model the prevalence of depression in the larger population (~16.7%).

Significance

Depression is disabling and treatable, but underdiagnosed. In this study, we show that the content shared by consenting users on Facebook can predict a future occurrence of depression in their medical records. Language predictive of depression includes references to typical symptoms, including sadness, loneliness, hostility, rumination, and increased self-reference. This study suggests that an analysis of social media data could be used to screen consenting individuals for depression. Further, social media content may point clinicians to specific symptoms of depression.

Author contributions: J.C.E., R.M.M., L.H.U., and H.A.S. designed research; J.C.E., P.C., D.P.-P., and H.A.S. performed research; J.C.E. and H.A.S. contributed new reagents/analytic tools; J.C.E., P.C., D.P.-P., and H.A.S. analyzed data; and J.C.E., R.M.M., L.H.U., D.A.A., and H.A.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]J.C.E. and R.J.S. contributed equally to this work.
[2]To whom correspondence should be addressed. Email: johannes.eph@gmail.com.
This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1802331115/-/DCSupplemental.
Published online October 15, 2018.

Each year, 7–26% of the US population experiences depression (1, 2), of whom only 13–49% receive minimally adequate treatment (3). By 2030, unipolar depressive disorders are predicted to be the leading cause of disability in high-income countries (4). The US Preventive Services Task Force recommends screening adults for depression in circumstances in which accurate diagnosis, treatment, and follow-up can be offered (5). These high rates of underdiagnosis and undertreatment suggest that existing procedures for screening and identifying depressed patients are inadequate. Novel methods are needed to identify and treat patients with depression.

By using Facebook language data from a sample of consenting patients who presented to a single emergency department, we built a method to predict the first documentation of a diagnosis of depression in the electronic medical record (EMR). Previous research has demonstrated the feasibility of using Twitter (6, 7) and Facebook language and activity data to predict depression (8), postpartum depression (9), suicidality (10), and post-traumatic stress disorder (11), relying on self-report of diagnoses on Twitter (12, 13) or the participants' responses to screening surveys (6, 7, 9) to establish participants' mental health status. In contrast to this prior work relying on self-report, we established a depression diagnosis by using medical codes from an EMR.

As described by Padrez et al. (14), patients in a single urban academic emergency department (ED) were asked to share access to their medical records and the statuses from their Facebook timelines. We used depression-related International Classification of Diseases (ICD) codes in patients' medical records as a proxy for

---

Natural Language Processing of Social Media as Screening for Suicide Risk

Glen Coppersmith, Ryan Leary, Patrick Crutchley and Alex Fine
Qntfy, Boston, MA, USA.

ABSTRACT: Suicide is among the 10 most common causes of death, as assessed by the World Health Organization. For every death by suicide, an estimated 138 people's lives are meaningfully affected, and almost any other statistic around suicide deaths is equally alarming. The pervasiveness of social media—and the near-ubiquity of mobile devices used to access social media networks—offers new types of data for understanding the behavior of those who attempt to take their own lives and suggests new possibilities for preventive intervention. We demonstrate the feasibility of using social media data to detect those at risk for suicide. Specifically, we use natural language processing and machine learning (specifically deep learning) techniques to detect quantifiable signals around suicide attempts, and describe designs for an automated system for estimating suicide risk, usable by those without specialized mental health training (eg, a primary care doctor). We also discuss the ethical use of such technology and examine privacy implications. Currently, this technology is only used for intervention for individuals who have "opted in" for a triage and intervention, but the technology enables scalable screening for suicide risk, potentially identifying many people who are at risk preventively and prior to any engagement with a health care system. This raises a significant cultural question about the trade-off between privacy and prevention—we have potent ally life-saving technology that is currently reaching only a fraction of the possible people at risk because of respect for their privacy. Is the current trade-off between privacy and prevention the right one?

KEYWORDS: Suicide, suicide prevention, social media, data science, natural language processing

Introduction

An estimated 16 million suicide attempts occur each year. Of these, approximately 800000 people will die from those attempts.[1] Suicide deaths have increased by 24% in the past 20 years, making suicide one of the top 10 causes of death in the United States,[2] a pattern that seems to be constant across geographic region within the country.[3] Not only is the magnitude of the problem large and worsening, there has been little progress made over the past 50 years in understanding suicide and improving outcomes in at-risk individuals.[4] The stubbornness of the problem reflects its complexity, and the densely interwoven causal factors underlying it. Here we focus on one piece of the puzzle: how can we identify those who are at risk of taking (or attempting to take) their own life, and how can this screening be used to foster effective interventions?

Assessing an individual's risk for suicidal behavior is difficult. Experienced and talented clinicians frequently struggle to correctly interpret signals in their patients' behavior that are indicative of suicide risk. Setting aside the profound difficulties associated with understanding an individual's personal history and its relationship to their capacity and motivations for self-harm, there are at least 2 practical reasons that assessing suicide risk is difficult: (1) the latency between the onset of acute risk for suicide and the suicide attempt itself may be too small for interventions requiring contact with a health professional, and (2) most existing methods for detecting high risk of suicide require that individuals disclose their wish to harm themselves to a health professional. In this article, we explore the possibility that *digital life data*—that is, the interactions that a person

has with digital devices, though the daily course of their life—collected passively but with consent might at least partially address each of these difficulties.

Individuals can be at risk for suicide at different temporal intervals relative to suicide attempts. For instance, the kind of social isolation that is frequently associated with suicide can gradually accumulate over the course of a person's life or may become acute in a very short period of time after a traumatic life event such as the loss of a loved one.

Moreover, once an individual is engaged with a health care professional, standard methods of suicide intervention require both that the clinician administer a standardized risk assessment (often in the form of a questionnaire) and that the patients disclose their intention to harm themselves. Each of these presents its own challenges. First, administering a suicide screening tool may place an unreasonable burden on the health care provider. The standard for suicide screening within the health care system is Beck's Scale for Suicide Ideation, a 5- or 19-item questionnaire examining the patient's active and passive desire for suicide, and any specific plans they might have.[5] Many patients who are at risk for suicide only interact with primary care physicians (PCPs) or emergency departments (EDs) rather than with psychiatric specialties. Such health care providers may lack the time or the training to administer a specific questionnaire for suicide risk. Indeed, enabling PCPs and EDs to better screen for suicide risk has been posited as a method for reducing the suicide rate.[6] Second, patients cannot always be relied upon to disclose suicidal thoughts in the clinical setting.[4] These factors have the

# Predicting Depression via Social Media

Summary

- Can social media activities and connectedness predict risk to major depressive disorder?

- Recruitment of a sample of Twitter users through a survey methodology over Amazon's Mechanical Turk
  - ~40% provided access to Twitter data
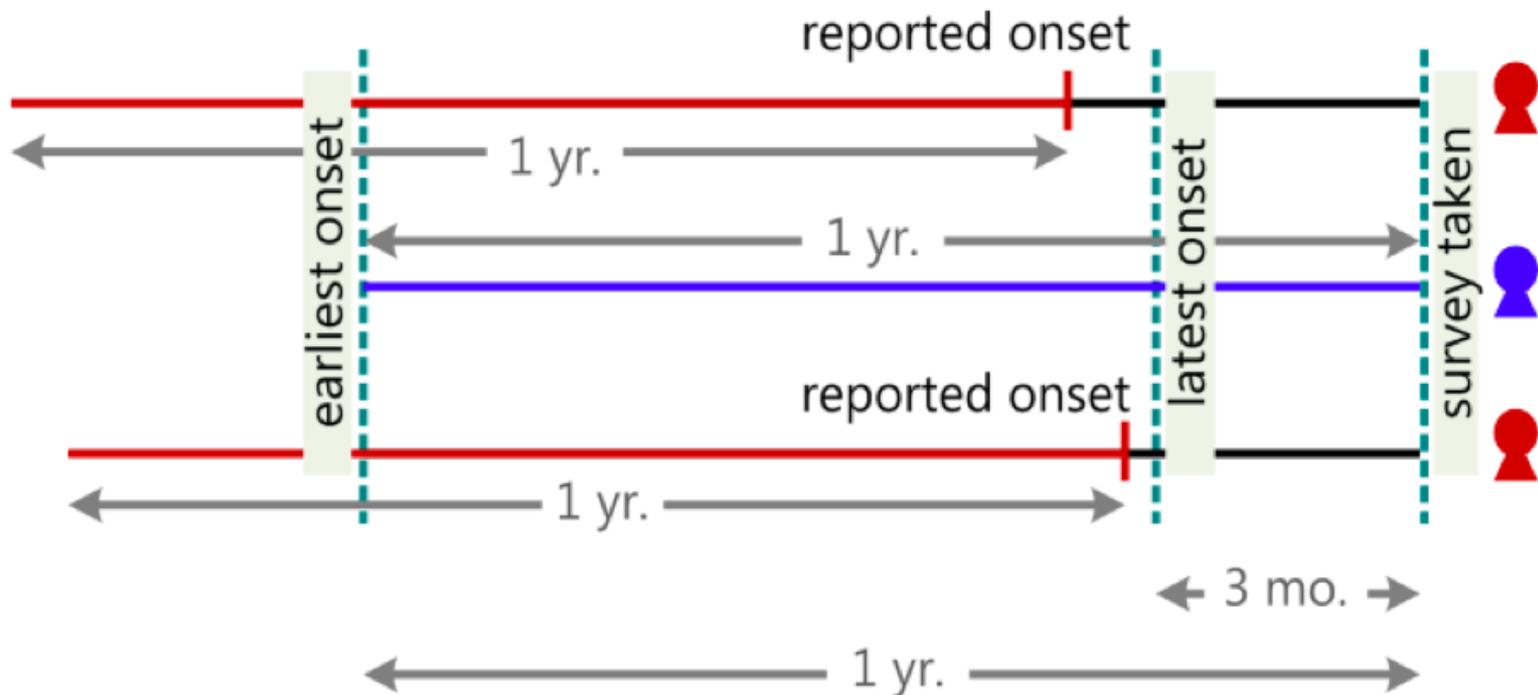
Summary

- Social engagement

- "Insomnia index" – mean z-score of an individual's volume of Twitter activity per hour

- Ego-centric social graph – nodal properties (*inlinks, outlinks*); dyadic properties (*reciprocity, interpersonal exchange*); neighborhood properties (*density, clustering coefficient, two-hop neighborhood, embeddedness, number of ego components*)

- Language
  - Depression lexicon – top uni- and bigrams compiled from Yahoo! Answers category on mental health
  - Linguistic style

Summary
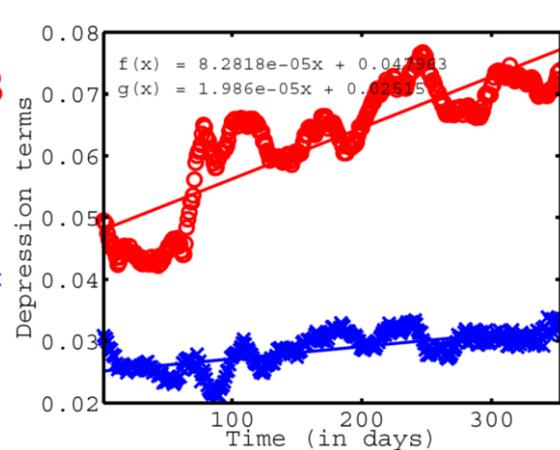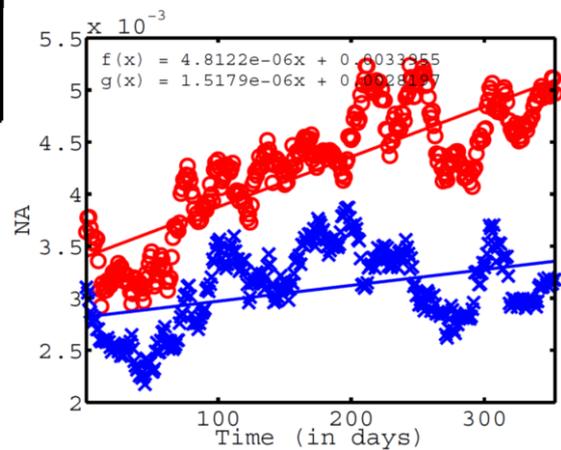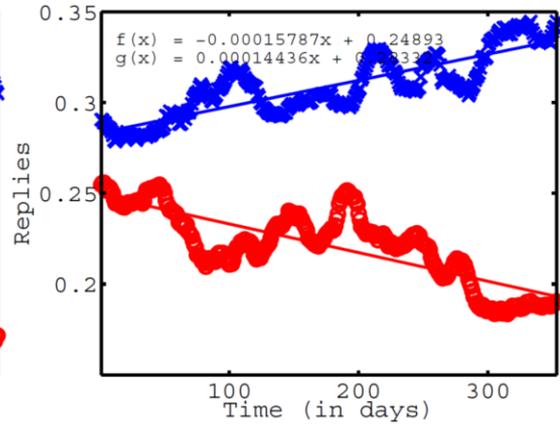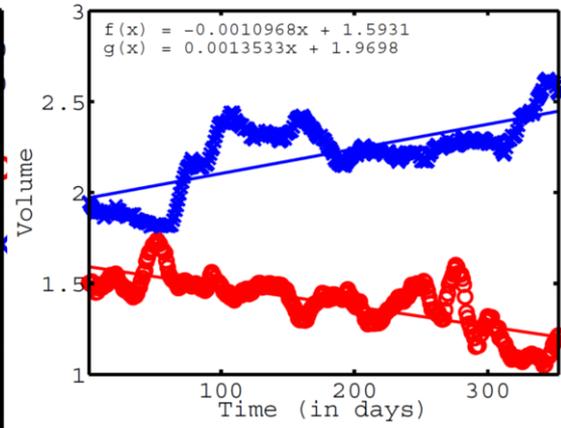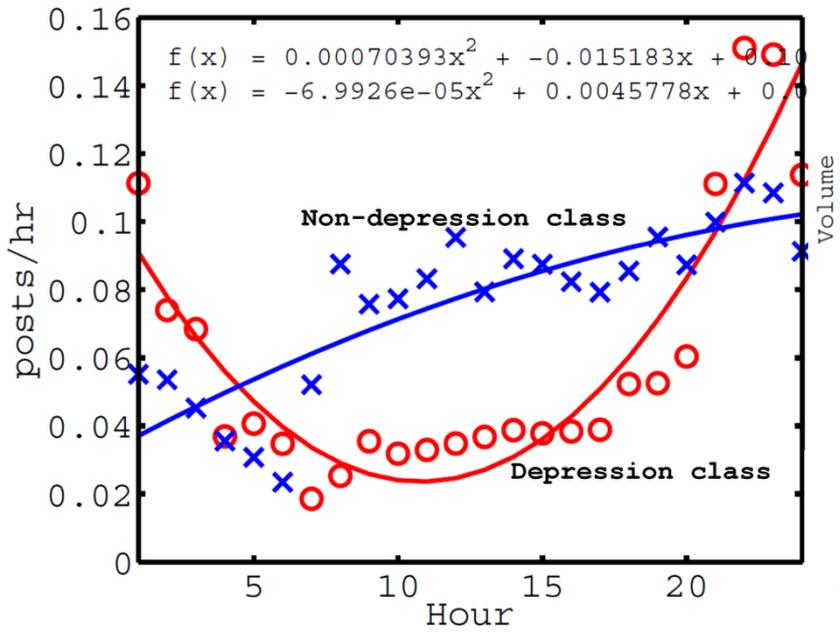
Summary

Egonetwork measures	Depres. class	Non-depres. class
#followers/inlinks	26.9 (σ=78.3)	45.32 (σ=90.74)
#followees/outlinks	19.2 (σ=52.4)	40.06 (σ=63.25)
Reciprocity	0.77 (σ=0.09)	1.364 (σ=0.186)
Prestige ratio	0.98 (σ=0.13)	0.613 (σ=0.277)
Graph density	0.01 (σ=0.03)	0.019 (σ=0.051)
Clustering coefficient	0.02 (σ=0.05)	0.011 (σ=0.072)
2-hop neighborhood	104 (σ=82.42)	198.4 (σ=110.3)
Embeddedness	0.38 (σ=0.14)	0.226 (σ=0.192)
#ego components	15.3 (σ=3.25)	7.851 (σ=6.294)

Discussion Point I

In this paper, the ground truth was obtained from Amazon mechanical turk workers. Anything unique about this population that may have affected the findings? What would be alternative ways of recruiting people or gathering high quality ground truth?

Discussion Point II

Depression is not an online condition, but one that spans both the online and the offline life. The paper does not take offline attributes into their models.

Is there a way to that into account? What would be the most significant offline attributes to consider?

Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance

Mauricio Santillana[1,2,3]*, André T. Nguyen[1], Mark Dredze[4], Michael J. Paul[5], Elaine O. Nsoesie[6,7], John S. Brownstein[2,3]

1 Harvard School of Engineering and Applied Sciences, Cambridge, Massachusetts, United States of America, 2 Boston Children's Hospital Informatics Program, Boston, Massachusetts, United States of America, 3 Harvard Medical School, Boston, Massachusetts, United States of America, 4 Department of Computer Science, Johns Hopkins University, Baltimore, Maryland, United States of America, 5 Department of Information Science, University of Colorado, Boulder, Colorado, United States of America, 6 Department of Global Health, University of Washington, Seattle, Washington, United States of America, 7 Institute for Health Metrics and Evaluation, Seattle, Washington, United States of America

* msantill@fas.harvard.edu

Abstract

We present a machine learning-based methodology capable of providing real-time ("nowcast") and forecast estimates of influenza activity in the US by leveraging data from multiple data sources including: Google searches, Twitter microblogs, nearly real-time hospital visit records, and data from a participatory surveillance system. Our main contribution consists of combining multiple influenza-like illnesses (ILI) activity estimates, generated indepen-

PDF | (f) More ▽ | " Cite | © Permissions

Original Investigation | Health Informatics   🔓

December 23, 2020

Development of a Machine Learning Model Using Multiple, Heterogeneous Data Sources to Estimate Weekly US Suicide Fatalities

Daejin Choi, PhD[1]; Steven A. Sumner, MD[2]; Kristin M. Holland, PhD[3]; John Draper, PhD[4]; Sean Murphy, PhD[4]; Daniel A. Bowen, MPH[3]; Marissa Zwald, PhD[3]; Jing Wang, MD[3]; Royal Law, PhD[5]; Jordan Taylor, BS[6]; Chaitanya Konjeti, BS[6]; Munmun De Choudhury, PhD[6]

≫ Author Affiliations | Article Information

Key Points

Question Can real-time streams of secondary information related to suicide be used to accurately estimate suicide fatalities in the US in real time?

Discussion Point III

But are models trained on aggregated group-level differences useful at the individual level?
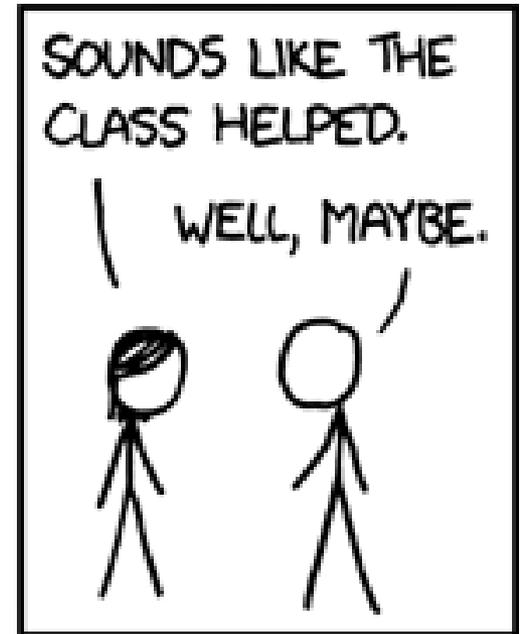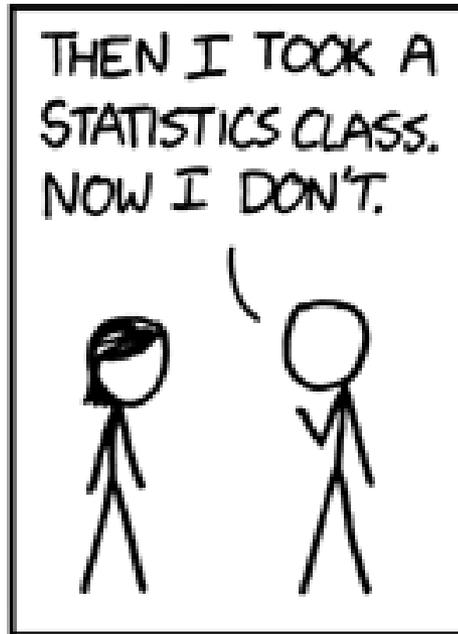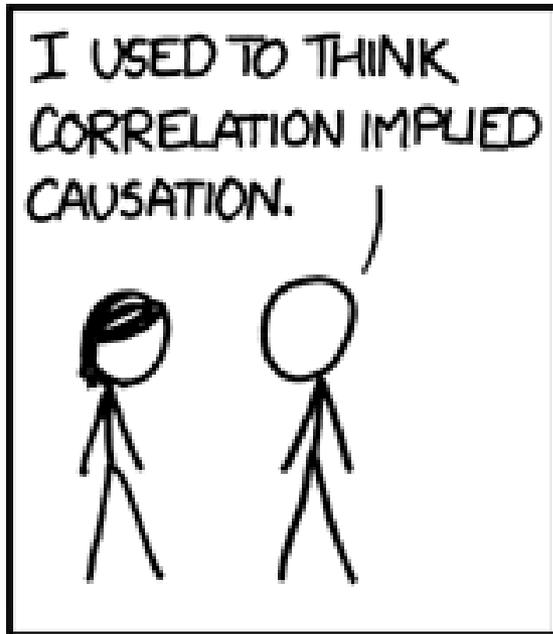
Correlation and causation
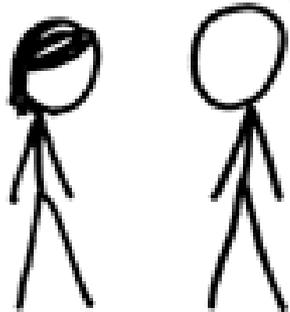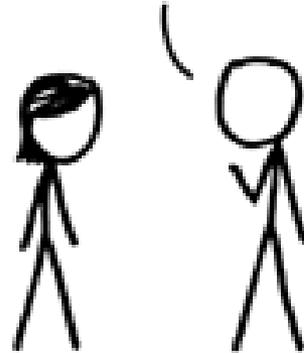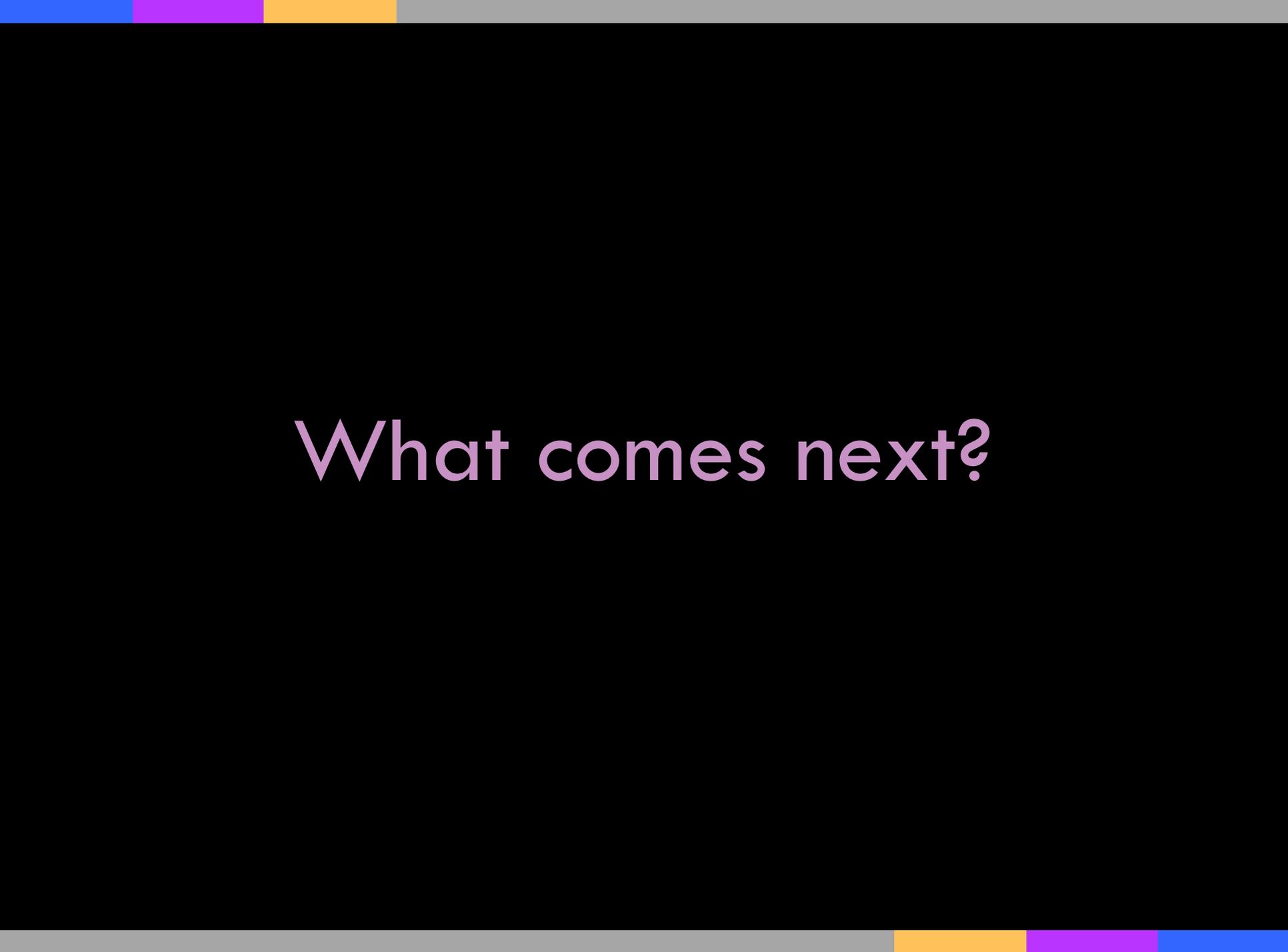
What comes next?

What comes next?

Social Media + Machine Learning for clinical interventions
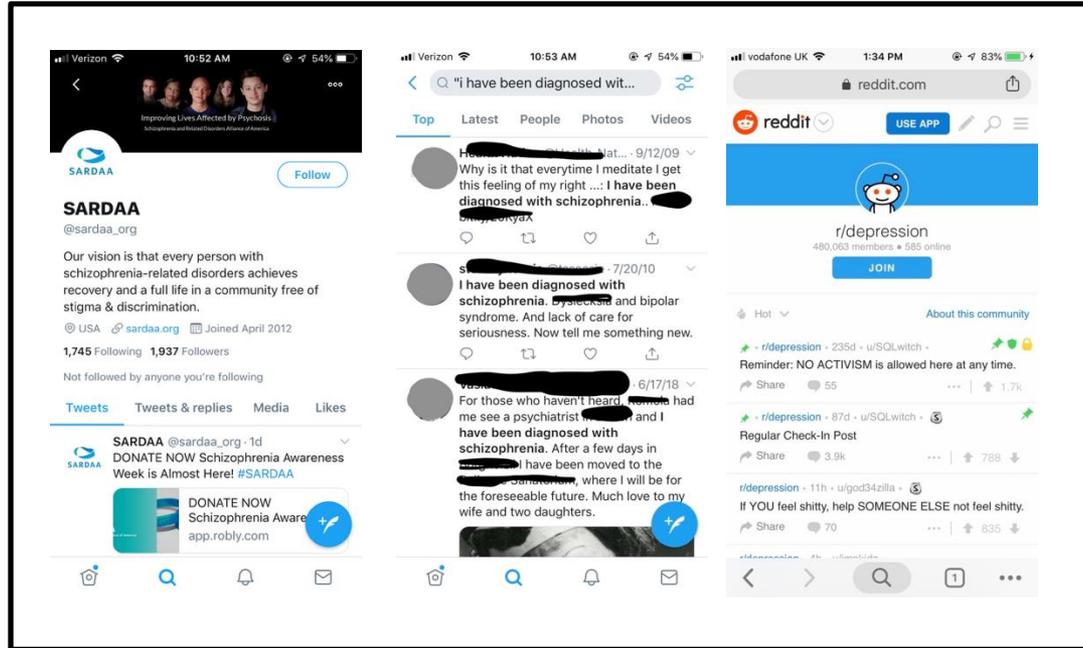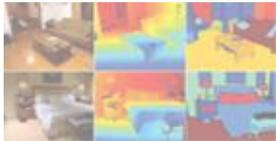
Efficacy ?

Validity ?

SOCIAL MEDIA + MACHINE LEARNING

Ground truth label:
Readily available

Proxy ground truth labels

Ground truth label:
clinical assessment

Construct Validity: Do the proxy diagnostic signals **objectively and accurately measure what they claim to measure** (clinical mental illness diagnosis)

Theoretical/Clinical grounding: Is what is being measured by

the proxy diagnostic signals

valid in itself?

Proxy data sets: diagnostic signals for schizophrenia on Twitter

Affiliation Data	N = 861
Self-reports Data	N = 412
Appraised Data	N = 153

Matched Control Data

N = 640

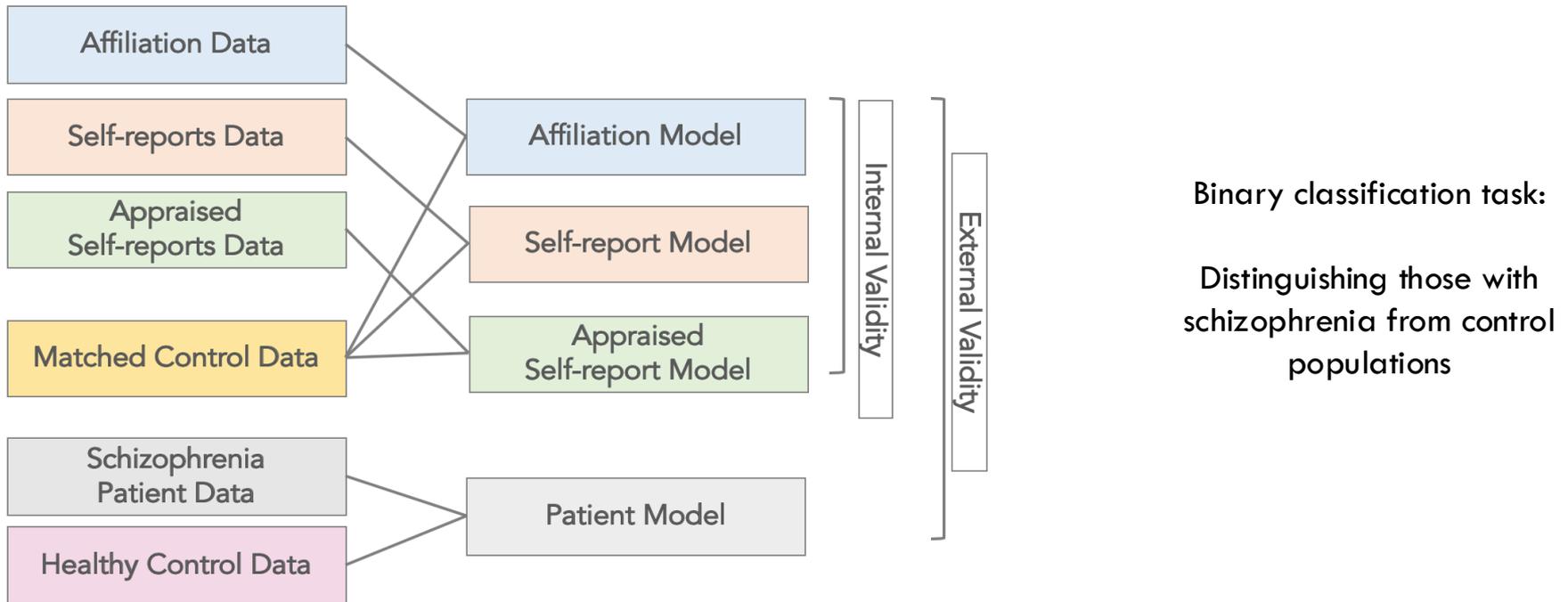# Patient's social media data

Schizophrenia
Patient Data

N = 88

Healthy Control Data

N = 55

Northwell
Health℠

Methodology: Triangulation

Binary classification task:

Distinguishing those with schizophrenia from control populations

Efficacy

High internal validity
Very low external validity

	Cross Validation	Testing on patient data
Affiliation Model	**0.89**	**0.21**
Self-report Model	**0.72**	**0.48**
Appraised Model	**0.80**	**0.55**
Patient Model	0.72	0.76

Issues with Construct Validity

Affiliation	β	Appraised	β	Patient	β
i'm	-0.825	*NegAffect*	0.063	*cog mech*	-0.003
stigma	0.665	negation	0.074	*present*	-0.002
mhchat	0.696	*present*	0.40	*body*	-0.002
body	0.729	help	0.401	*verbs*	-0.002
bipolar	0.774	thought	0.41	*social*	-0.002
work	0.919	i'm	0.44	*aux verbs*	-0.002
self	0.961	die	0.45	help	0.0002
social	1.109	alone	0.45	feeling	0.001
care	1.111	hard	0.457	i'm	0.002
depression	1.116	cry	0.50	gonna	0.002
suicide	1.133	*body*	0.52	angel	0.002
thanks	1.445	feeling	0.523	burning	0.002
illness	1.447	*verbs*	0.58	pray	0.003
help	1.632	sorry	0.662	lifetime	0.005
mental health	1.866	gonna	0.63	attack	0.006

Main Takeaway

If the broader research agenda is to use social media data to inform clinical decision-making, such as early diagnosis, treatment or patient-provider interventions, **(social media) data collection and machine learning model development should happen in context**

Facebook language predicts depression in medical records

Johannes C. Eichstaedt[a,1,2], Robert J. Smith[b,1], Raina M. Merchant[b,c], Lyle H. Ungar[a,b], Patrick Crutchley[a,b], Daniel Preoţiuc-Pietro[a], David A. Asch[b,d], and H. Andrew Schwartz[e]

[a]Positive Psychology Center, University of Pennsylvania, Philadelphia, PA 19104; [b]Penn Medicine Center for Digital Health, University of Pennsylvania, Philadelphia, PA 19104; [c]Department of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, PA 19104; [d]The Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center, Philadelphia, PA 19104; and [e]Computer Science Department, Stony Brook University, Stony Brook, NY 11794

Depression, the most prevalent mental illness, is underdiagnosed and undertreated, highlighting the need to extend the scope of current screening methods. Here, we use language from Facebook posts of consenting individuals to predict depression recorded in electronic medical records. We accessed the history of Facebook statuses posted by 683 patients visiting a large urban academic emergency department, 114 of whom had a diagnosis of depression in their medical records. Using only the language preceding their first documentation of a diagnosis of depression, we could identify depressed patients with fair accuracy [area under the curve (AUC) = 0.69], approximately matching the accuracy of screening surveys benchmarked against medical records. Restricting Facebook data to only the 6 months immediately preceding the first documented diagnosis of depression yielded a higher prediction accuracy (AUC = 0.72) for those users who had sufficient Facebook data. Significant prediction of future depression status was possible as far as 3 months before its first documentation. We found that language predictors of depression include emotional (sadness), interpersonal (loneliness, hostility), and cognitive (preoccupation with the self, rumination) processes. Unobtrusive depression assessment through social media of consenting individuals may become feasible as a scalable complement to existing screening and monitoring procedures.

big data | depression | social media | Facebook | screening

Each year, 7–26% of the US population experiences depression (1, 2), of whom only 13–49% receive minimally adequate treatment (3). By 2030, unipolar depressive disorders are predicted to be the leading cause of disability in high-income countries (4). The US Preventive Services Task Force recommends screening adults for depression in circumstances in which accurate diagnosis, treatment, and follow-up can be offered (5). These high rates of underdiagnosis and undertreatment suggest that existing procedures for screening and identifying depressed patients are inadequate. Novel methods are needed to identify and treat patients with depression.

By using Facebook language data from a sample of consenting patients who presented to a single emergency department, we built a method to predict the first documentation of a diagnosis of depression in the electronic medical record (EMR). Previous research has demonstrated the feasibility of using Twitter (6, 7) and Facebook language and activity data to predict depression (8), postpartum depression (9), suicidality (10), and posttraumatic stress disorder (11), relying on self-report of diagnoses on Twitter (12, 13) or the participants' responses to screening surveys (6, 7, 9) to establish participants' mental health status. In contrast to this prior work relying on self-report, we established a depression diagnosis by using medical codes from an EMR.

As described by Padrez et al. (14), patients in a single urban academic emergency department (ED) were asked to share access to their medical records and the statuses from their Facebook timelines. We used depression-related International Classification of Diseases (ICD) codes in patients' medical records as a proxy for

the diagnosis of depression, which prior research has shown is feasible with moderate accuracy (15). Of the patients enrolled in the study, 114 had a diagnosis of depression in their medical records. For these patients, we determined the date at which the first documentation of a diagnosis of depression was recorded in the EMR of the hospital system. We analyzed the Facebook data generated by each user before this date. We sought to simulate a realistic screening scenario, and so, for each of these 114 patients, we identified 5 random control patients without a diagnosis of depression in the EMR, examining only the Facebook data they created before the corresponding depressed patient's first date of a recorded diagnosis of depression. This allowed us to compare depressed and control patients' data across the same time span and to model the prevalence of depression in the larger population (~16.7%).

Results

Prediction of Depression. To predict the future diagnosis of depression in the medical record, we built a prediction model by using the textual content of the Facebook posts, post length, frequency of posting, temporal posting patterns, and demographics (*Materials and Methods*). We then evaluated the performance of this model by comparing the probability of depression estimated by our algorithm against the actual presence or absence of depression for each patient in the medical record (using 10-fold cross-validation to avoid overfitting). Varying the threshold of this probability for diagnosis

Class Exercise

Describe a design idea where we can use social media based depression (or other mental health condition like schizophrenia) predictors to help people. How would it negotiate privacy and ethical issues?

Improving "Blanket" Interventions
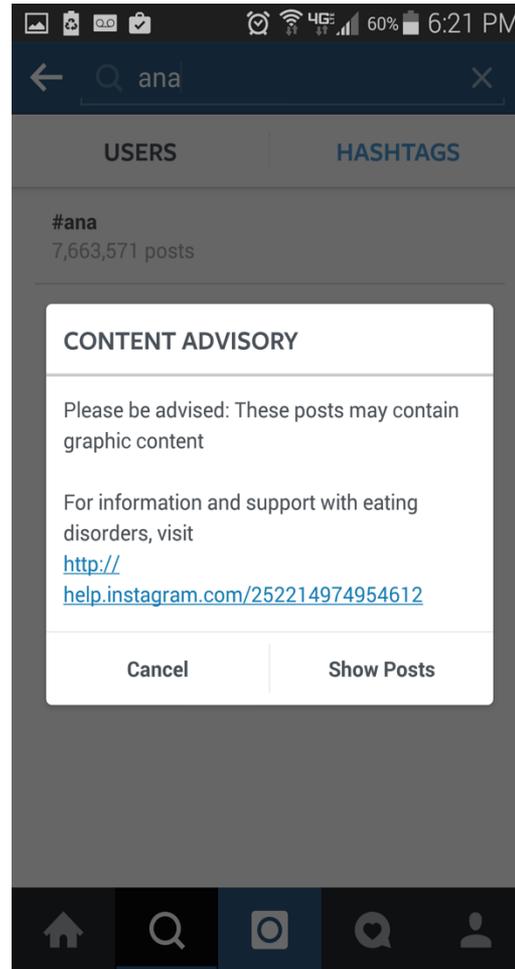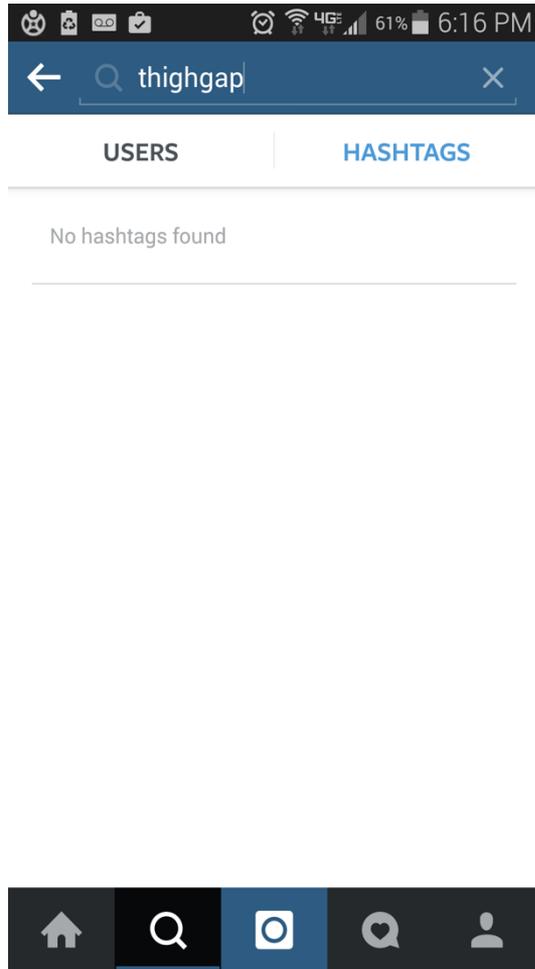
facebook

?

Hi Gerald, a friend thinks you might be going through something difficult and asked us to look at your recent post.

🔒

Only you can see this. Anything you do there will be kept private.

| See Post | Continue |

Google

suicide

| Web | News | Images | Videos | Books | More ▾ |

About 214,000,000 results (0.44 seconds)

Need help? United States:

1 (800) 273-8255

National Suicide Prevention Lifeline

Hours: 24 hours, 7 days a week
Languages: English, Spanish
Website: www.suicidepreventionlifeline.org

A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media

Stevie Chancellor
Georgia Tech
Atlanta, GA, US
schancellor3@gatech.edu

Michael L Birnbaum
Northwell Health
Glen Oaks, NY, US
mbirnbaum@northwell.edu

Eric D. Caine
University of Rochester
Rochester, NY, US
Eric_Caine@urmc.rochester.edu

Vincent M. B. Silenzio
University of Rochester
Rochester, NY, US
vincent.silenzio@rochester.edu

Munmun De Choudhury
Georgia Tech
Atlanta, GA, US
munmund@gatech.edu

ABSTRACT

Powered by machine learning techniques, social media provides an unobtrusive lens into individual behaviors, emotions, and psychological states. Recent research has successfully employed social media data to predict mental health states of individuals, ranging from the presence and severity of mental disorders like depression to the risk of suicide. These algorithmic inferences hold great potential in supporting early detection and treatment of mental disorders and in the design of interventions. At the same time, the outcomes of this research can pose great risks to individuals, such as issues of incorrect, opaque algorithmic predictions, involvement of bad or unaccountable actors, and potential biases from intentional or inadvertent misuse of insights. Amplifying these tensions, there are also divergent and sometimes inconsistent methodological gaps and under-explored ethics and privacy dimensions. This paper presents a taxonomy of these concerns and ethical challenges, drawing from existing literature, and poses questions to be resolved as this research gains traction. We identify three areas of tension: ethics committees and the gap of social media research; questions of validity, data, and machine learning; and implications of this research for key stakeholders. We conclude with calls to action to begin resolving these interdisciplinary dilemmas.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing**; **Social media**; • **Applied computing** → *Psychology*;

1 INTRODUCTION

Last year, Facebook unveiled automated tools to identify individuals contemplating suicide or self-injury [75, 62]. The company claims that they "use pattern recognition technology to help identify posts and live streams as likely to be expressing thoughts of suicide," which then can deploy resources to assist the person in crisis [75]. Reactions to Facebook's suicide prevention artificial intelligence (AI) are mixed, with some concerned about the use of AI to detect suicidal ideation as well as potential privacy violations [86]. Other suicide prevention AIs, however, have been met with stronger public backlash. Samaritan's Radar, an app that scanned a person's friends for concerning Twitter posts, was pulled from production, citing concerns for data collection without user permission [54], as well as enabling harassers to intervene when someone was vulnerable [4].

Since 2013, a new area of research has incorporated techniques from machine learning, natural language processing, and clinical psychology to categorize individuals' moods and expressed well-being from social media data. These algorithms are powerful enough to infer with high accuracy whether an individual might be suffering from disorders such as major depression [28, 19, 84, 73, 78], postpartum depression [26, 27], post-traumatic stress [21], schizophrenia [60, 6], and suicidality [15, 22]. These algorithms can also reveal symptomatology linked to psychiatric challenges,