



Uncovering the relationships between military community health and affects expressed in social media

Svitlana Volkova*, Lauren E Charles, Josh Harrison and Courtney D Corley

*Correspondence:
svitlana.volkova@pnnl.gov
Data Sciences and Analytics,
National Security Directorate, Pacific
Northwest National Laboratory, 902
Battelle Blvd, Richland, WA 99354,
USA

Abstract

Military populations present a small, unique community whose mental and physical health impacts the security of the nation. Recent literature has explored social media's ability to enhance disease surveillance and characterize distinct communities with encouraging results. We present a novel analysis of the relationships between influenza-like illnesses (ILI) clinical data and affects (i.e., emotions and sentiments) extracted from social media around military facilities. Our analyses examine (1) differences in affects expressed by military and control populations, (2) affect changes over time by users, (3) differences in affects expressed during high and low ILI seasons, and (4) correlations and cross-correlations between ILI clinical visits and affects from an unprecedented scale - 171M geo-tagged tweets across 31 global geolocations. Key findings include: Military and control populations differ in the way they express affects in social media over space and time. Control populations express more *positive* and less *negative* sentiments and less *sadness*, *fear*, *disgust*, and *anger* emotions than military. However, affects expressed in social media by both populations within the same area correlate similarly with ILI visits to military health facilities. We have identified potential responsible cofactors leading to location variability, e.g., region or state locale, military service type and/or the ratio of military to civilian populations. For most locations, ILI proportions positively correlate with *sadness* and *neutral* sentiment, which are the affects most often expressed during high ILI season. The ILI proportions negatively correlate with *fear*, *disgust*, *surprise*, and *positive* sentiment. These results are similar to the low ILI season where *anger*, *surprise*, and *positive* sentiment are highest. Finally, cross-correlation analysis shows that most affects lead ILI clinical visits, i.e. are predictive of ILI data, with affect-ILI leading intervals dependent on geolocation and affect type. Overall, information gained in this study exemplifies a usage of social media data to understand the correlation between psychological behavior and health in the military population and the potential for use of social media affects for prediction of ILI cases.

Keywords: social media analytics; machine learning; natural language processing; emotion detection; sentiment analysis; biosurveillance; influenza

1 Introduction

Social media is an open source, real-time outlet for sharing with others thoughts, actions or feelings. Researchers have taken advantage of this data source as a complement to tradi-

tional disease surveillance in effort to increase the speed of detecting disease outbreaks or other health issues within populations [1]. In addition to using social media for actionable surveillance, the way the information is expressed (i.e., through specific language characteristics) can provide insight into how illness affects individuals emotionally and how they are coping with disease.

Influenza-like illness (ILI) is one of the most widely studied syndromes using social media, most likely because of its seasonal yearly occurrence in a high proportion of populations with easily identifiable symptoms. In addition, ILI is highly contagious (airborne) and can present in a range of severity based on the causal agent, e.g., non-debilitating common cold to a deadly influenza strain. The potential implications of this disease on human health and health care systems are drastic. Normal seasonal influenza outbreaks alone cause up to 500,000 deaths annually worldwide and 50,000 deaths a year in the U.S. [2]. The Centers for Disease Control and Prevention (CDC) has five different types of influenza surveillance, i.e., outpatient illness, virologic, mortality, hospitalization, and geographic spread surveillance systems, and provide official summary reports 1 to 2 weeks after patients are seen. This lag in notification is due to the required chains of data collection, verification, and reporting when dealing with human health data. The time lag is overcome with the use of open source social media data where people ‘report’ their own health conditions yet the validity and credibility of the data is minimized. Regardless, many studies have shown an increase in time to ILI outbreak detection using available online data signals, such as search query logs (e.g., Google Flu Trends [3–5] and Yahoo [6]), health related web page views (e.g., Wikipedia [7, 8]), self reported illness (e.g., crowd-sourced reports (Flu Near You) [9, 10] and social media (Twitter) [11–16]), and combined data sources [17].

To date, studies using social media for ILI surveillance, including those mentioned above, have relied on health-related text from the user, e.g., self-reporting of symptoms, syndrome or actual illness, in their analyses. However, social media analytics can provide insight into the user’s affects (defined as emotions and sentiments) based on the tweet itself. Many studies in socio- and psycholinguistics show that a person’s affects influence their health status, namely *positive* and *negative* affects correlating with good and bad health, respectively [18–21]. Therefore, there is the potential that affects expressed in social media may be indirectly related to the health status of the user. Limited work has been done in this area, namely in effort to predict emotions and mental illness for civilian populations [22, 23] and correlating CDC-estimated vaccination rates and vaccination sentiments identified through social media [24].

In this paper, we focus on affect signals originating from Twitter social media surrounding military bases and compare them to the number of ILI visits recorded at military health facilities in the same area. Military communities present a unique case study because they are semi-closed populations of people who share common location, responsibilities, and way of life. They actively use social media to stay connected with their unit at home or with friends and family when deployed [25]. In addition, ILI has the ability to quickly spread throughout the military population resulting in disruptions in military operations and concern for national security [26].

To the best of our knowledge, this is the first work that studies the relationships between affects expressed via social media and military ILI data. This study is also the largest published study of the effect of influenza on military communities. Past studies targeting

military population affects are limited to small-scale, designed experiments (e.g., investigating emotions in military veterans via questionnaires [27], measuring emotions in military speech [28], studying the relationship between occupational stress and mental illness in the military [29], and evaluating the health-related quality of life of United States (U.S.) military personnel [30, 31]). The largest study on mental-health risk among the U.S. military, done jointly by the U.S. Army and the U.S. National Institute of Mental Health, reports the rate of major depression is five times higher among military as civilians, intermittent explosive disorder is six times higher, and Post-Traumatic Stress Disorder (PTSD) is nearly 15 times higher [32]. By tracking emotion and sentiment changes over time across locations expressed by military personnel and their families via social media may provide a deeper insight and faster response time to changes in mental health for these communities.

Unlike the existing approaches on predicting the dynamics of influenza outbreaks from social media, we focus on studying spatiotemporal variations in emotions and sentiments expressed by military populations and correlate these affect signals with clinical diagnosis data collected from 31 global military areas. Our main contributions include (1) identifying how emotions and sentiments expressed on Twitter differ between military and civilian populations over time and space, (2) identifying differences in affects expressed during high and low ILI seasons for military and control populations, (3) qualitatively assessing and quantitatively estimating correlations between ILI clinical visit data and affects expressed by military populations in social media over time, and (4) investigating whether affects identified from user tweets lead or lag ILI clinical visit data, and thus, can be used to build ILI predictive models.

2 Data

In this section, we present ILI-related clinical visit data for military personnel and their families across 31 geolocations. We describe data collection, sampling, and annotation procedures of social media from both military and civilian populations.

2.1 ILI-related clinical visit data

The ILI clinical data consists of the number of visits to a Defense Medical Information System (DMIS) Identifier (ID) location for symptoms identified as influenza-like illness (ILI) based on the International Statistical Classification of Disease and Health Related Problems (ICD) codes documented in the electronic patient record (Table 1). The DMIS ID facility types identified as reporting ILI symptoms in patients are hospitals, clinics, administration, and dental offices. The patients who visit these facilities include active duty, reserve, and retired members along with their dependents and cadets, recruits and applicants for active duty from the army, navy, marine corps, coast guard, air force, National Oceanic and Atmospheric Administration, and other public health service. This military health data was collected from 31 specific locations (25 U.S. and 6 international). Each location comprised all DMSID IDs within a 25-mile radius around military bases (mean 7 IDs, range 2-19 IDs). The total number of health-related visits to these facilities between 2011 and 2014 was summarized (Table 2) and the percent of ILI visits to total visits was used in the subsequent analyses (mean 3.6%, range 1.1-9.5%).

Table 1 The ICD-9 codes used to describe ILI symptoms and their descriptions

ICD9	Code Description for ILI Symptoms
79.99	Viral infection, NOS
382.9	Unspecified otitis media
460	Acute nasopharyngitis (common cold)
461.9	Acute sinusitis unspecified
465.8	Acute upper respiratory infections of other multiple sites
465.9	Acute upper respiratory infections of unspecified sites
466	Acute bronchitis
486	Pneumonia organism unspecified
487	Influenza with pneumonia
487.1	Influenza with other respiratory manifestations
487.8	Influenza with other manifestations
488	Influenza due to identified agent
490	Bronchitis, NOS
780.6	Fever
786.2	Cough

NOS = Not otherwise specified.

Table 2 The geolocation's facility types, mean ILI-visits, and mean all-visits for 2011-2014

ID	Facility Types ^a				2011-2014 Mean ILI-visits			2011-2014 Mean All-visits		
	Total	H	C	O	Yearly	Weekly	SD	Yearly	Weekly	SD
<i>l</i> ₀	11	3	8	1	76K	1,446	251	2.1M	39,641	4,017
<i>l</i> ₃	13	2	11	0	119K	2,289	610	3.6M	68,985	6,833
<i>l</i> ₄	7	1	6	2	75K	1,455	450	2.5M	48,047	4,539
<i>l</i> ₁₀	1	1	0	0	34K	657	210	1.4M	26,246	2,321
<i>l</i> ₁₁	13	2	11	1	78K	1,503	443	2.7M	51,401	5,246
<i>l</i> ₁₂	7	2	5	0	61K	1,169	432	1.7M	33,198	3,995
<i>l</i> ₁₃	13	3	10	0	90K	1,724	513	3.0M	57,772	6,004
<i>l</i> ₁₄	8	4	4	2	138K	2,644	809	4.6M	88,628	8,623
<i>l</i> ₁₅	10	1	9	0	94K	1,807	641	3.0M	58,327	6,142
<i>l</i> ₁₆	5	2	3	0	34K	643	186	926K	17,726	2,117
<i>l</i> ₁₉	3	1	2	1	33K	640	205	862K	16,494	2,294
<i>l</i> ₂₀	4	1	3	1	55K	1,050	159	1.2M	22,574	2,924
<i>l</i> ₂₁	6	2	4	0	67K	1,284	401	1.9M	36,280	4,190
<i>l</i> ₂₂	9	1	8	1	85K	1,622	505	2.4M	45,055	5,224
<i>l</i> ₂₃	10	1	9	1	107K	2,049	690	3.4M	64,988	6,801
<i>l</i> ₂₅	4	2	2	0	95K	1,824	435	2.9M	55,825	4,916
<i>l</i> ₂₇	12	2	10	1	214K	4,104	1,323	6.5M	124,905	11,165
<i>l</i> ₂₈	2	1	1	0	48K	925	268	1.4M	27,401	2,857
<i>l</i> ₂₉	5	1	4	1	70K	1,331	381	1.8M	35,230	3,533
<i>l</i> ₃₀	5	3	2	0	37K	709	217	1.3M	25,472	3,216
<i>l</i> ₃₁	7	1	6	1	62K	1,185	304	1.7M	32,833	3,828
<i>l</i> ₃₂	4	1	3	1	54K	1,031	294	1.6M	30,666	3,364
<i>l</i> ₃₃	5	1	4	0	45K	866	247	1.5M	27,824	2,881
<i>l</i> ₃₄	5	2	3	0	59K	1,123	310	1.7M	33,391	3,368
<i>l</i> ₃₇	2	0	2	1	27K	525	128	649K	12,429	1,433
<i>i</i> ₂	5	-	-	-	19K	356	63	196K	3,754	817
<i>i</i> ₃	11	-	-	-	16K	302	56	532K	10,185	1,207
<i>i</i> ₁₇	11	-	-	-	7K	127	52	572K	10,967	1,525
<i>i</i> ₂₀	19	-	-	-	17K	329	117	259K	4,956	750
<i>i</i> ₂₅	7	-	-	-	9K	170	54	182K	3,478	390
<i>i</i> ₂₇	4	-	-	-	10K	195	45	383K	7,327	1,063

H - Hospital, C - Clinic, and O - Other (i.e., Administrative and Dental facilities). The average number of total counts per year and the weekly mean with standard deviation (SD) are given for each location. *i* are locations within the U.S. and *l* denotes international locations.

^aFacility types are only available for U.S. locations.

2.2 Social media data

We used social media data to study military populations across different geolocations in the U.S. and internationally. The Twitter data, acquired from a social media vendor and through the public Application Program Interface, was anonymized for usernames, user IDs, and tweet IDs based on a rigorous procedure, i.e., state-of-art encryption algorithm. To ensure privacy of all users in our sampled datasets, our analysis was based only on completely anonymized data and our findings are reported on an aggregate rather than individual level. This study was approved by our institutional review board.

2.2.1 Global Military Twitter Dataset (aka Global Military Dataset)

Geo-tagged social media data has been previously used to identify specific populations and demographic groups in social networks, e.g., geographical communities and urban areas [33], users with specific demographics [34] and personality [35], and new mothers [36].

To study communications generated by military populations in social media, we collected geo-tagged tweets within a 25-mile radius of 25 military locations in the U.S. (*l*) and six international locations (*i*) following standard practices on extracting geolocation coordinates from user meta-data [23]. The tweets were collected from over 167 weeks between November 2011 through December 2014. Our large Twitter sample includes 171 million tweets produced within a 25-mile radius across 31 military locations. We report tweet distribution for each military location, i.e. point of interest, in Figure 1.

2.2.2 U.S. military and non-military Twitter timeline dataset (aka Comparison Dataset)

To identify if differences exist between affects expressed in tweets from military and non-military (i.e., control) populations, we identified Twitter users from each population. Two military and one control user sample datasets were identified from one state in the West Coast (p_3, p_{10}), South East (p_{22}, p_{23}), and South Central U.S. (p_{12}, p_{15}). The most recent methods in the literature that identify specific users in social media rely on searching

Figure 1 Twitter dataset statistics across 31 military locations. The number of tweets collected within a 25-mile radius of military installations for 31 geolocations.

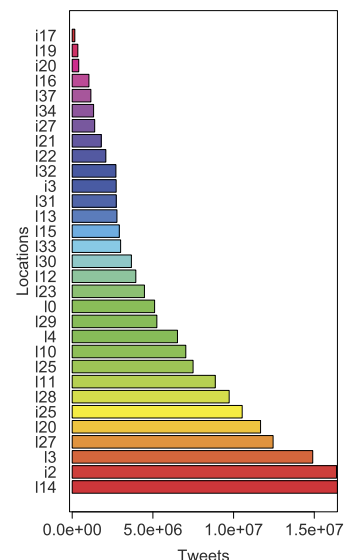


Table 3 The distribution of military T_M , control T_C , and total T tweets per Twitter point of interest

Point	T_M	T_C	T
p_{12}	381,178	335,578	716,756
p_{15}	467,509	424,701	892,210
p_{23}	607,461	605,201	1,212,662
p_{22}	578,843	550,761	1,129,604
p_{10}	378,837	402,298	781,135
p_3	1,615,887	1,739,865	3,355,752
Total	4,029,715	4,058,404	8,088,119

We reuse Twitter point indices from military location indices used in a large military Twitter dataset (at least a portion of tweets extracted from military user timelines was sampled around known military locations e.g., l_{12}).

certain keywords in user profile metadata [23, 37–39]. Our method for sampling sub-populations of users combined geolocation with the more common keywords extraction from user biography fields. The military users were identified from tweets within a 25-mile radius of military facilities with high military to civilian population ratios and from military-specific keywords, e.g., *military*, *corporal*, *army brat*, etc., present in their user profile data. Conversely, the control population users were sampled from a geolocation at least 50 miles away from any military facility in the same state and did not contain any military-specific keywords from our lexicon in their profile. Note: there is a possibility that our control sample includes military users if they never stated their membership in their profile descriptions and resided more than 50 miles away from any military facility at the point of user classification. Once the users were identified as control or military, user timelines containing the most recent 3,200 tweets were collected, regardless of the user's location. The resulting dataset of military and control timelines spans over 313 weeks between Jan 2009 and Dec 2014 and contains eight million tweets from the six initial user identification sites (Table 3). A more detailed explanation on sampling and annotation processes of military vs. control populations is discussed by [40].

3 Methods

3.1 Sentiment and emotion classification models

To predict sentiments and emotions arising from tweets, we used machine learning and natural language processing techniques to build supervised classification models [34] extending recently developed approaches for affect prediction in social media [41–45]. We relied on all tweets produced by military and control populations in specific geolocations to go beyond influenza-related keywords and tweets previously used to predict ILI [15], and capture other linguistic predictors e.g., discourse about the weather, personal well-being, travel, indoor and outdoor activities etc. We trained sentiment models on tweets annotated with three opinion classes downloaded from seven publicly available sentiment datasets [46]. The training data for sentiment classification includes $T_S = 19,555$ tweets in total (35% *positive*, 30% *negative*, and 35% *neutral*) from multiple domains e.g., health, debates, politics etc. To evaluate the predictive power of our sentiment model on tweets from the general domain, we tested our model on the official SemEval-2013 test set [47] of 3,223 tweets and report the $F1 = 0.66$ for 3-way classification. We found that our sentiment model is comparable with the state-of-the-art systems for sentiment classification on Twitter [41, 42].

We trained our emotion models on tweets annotated with basic emotion hashtags that correspond to six Ekman's emotion classes - *joy*, *sadness*, *disgust*, *surprise*, *anger*, and *fear* [34, 43]. Despite the limitation that existing approaches do not disambiguate sarcastic hashtags e.g., *It's Monday #joy* vs. *It's Friday #joy*, they still demonstrate that a hashtag is a reasonable representation of real feelings [48]. Similar to Gonzalez-Ibanez (2011), we collected tweets with the hashtags at the end of the tweet, excluded non-English tweets and tweets with less than three tokens [48]. Moreover, we extended our emotion-hashtag dataset with emotion synonyms collected from WordNet-Affect [49] and Roget's Thesaurus [50] to lower the rate of false positives that might be present in the data due to the sarcasm factor. Overall, we collected $T_E = 52,925$ tweets annotated with *anger* (9.4%), *joy* (29.3%), *fear* (17.1%), *sadness* (7.9%), *disgust* (24.5%), and *surprise* (15.6%).

For emotion and sentiment classification, we trained tweet-level log-linear models with L2 regularization (aka logistic regression) Φ_s (as defined in Eq. (1)) and Φ_e (as defined in Eq. (2)) using scikit-learn toolkit [51] that rely on lexical ngram features extracted from tweets $t_k \in T$ annotated with three opinion $s_i \in s$ and six emotion $e_j \in e$ categories. In addition to lexical features, we extracted a set of syntactic and stylistic markers from tweets including emoticons, elongated words, capitalization, repeated punctuation, and number of hashtags and took into account clause-level negation [34].

$$\Phi_s = \operatorname{argmax}_{s_i} \Pr(S = s_i | t), \quad (1)$$

$$\Phi_e = \operatorname{argmax}_{e_j} \Pr(E = e_j | t). \quad (2)$$

We evaluated our emotion model prediction quality using 10-fold cross validation on our emotion dataset of $T_E = 52,925$ tweets and report weighted F1 = 0.78 for 6-way classification. We found that our emotion model significantly outperforms the existing approaches for emotion classification [43, 52–54].

3.2 Evaluation metrics

3.2.1 ILI and affect proportions

To study the relationships between clinical ILI visit data and affects derived from tweets, we define ILI proportions per location $l \in L$ with each location aggregated over multiple DMISIDs and per week $w \in W$.

Weekly location-specific ILI visit proportions:

$$I_{w,l} = \frac{\# \text{ of weekly ILI visits per location } l}{\text{weekly total visits per location } l}. \quad (3)$$

After applying our emotion and sentiment classification models, every tweet in our dataset was annotated with its predicted affects. We aggregated affect annotations over time and geolocations to obtain weekly location-specific sentiment and emotion proportions as shown below.

Weekly location-specific sentiment proportions over three sentiment classes $\{positive, negative, neutral\} \in s$, $\sum_i s_i = 1$:

$$S_{w,l}^{s_i} = \frac{\# \text{ of weekly tweets labeled with sentiment } s_i \text{ per location } l}{\# \text{ of weekly tweets per location } l}. \quad (4)$$

Weekly location-specific emotion proportions over six Ekman's emotion classes $\{joy, sadness, fear, disgust, surprise, anger\} \in e$, $\sum_j e_j = 1$:

$$E_{w,l}^{e_j} = \frac{\text{\# of weekly tweets labeled with emotion } e_j \text{ per location } l}{\text{\# of weekly tweets per location } l}. \quad (5)$$

We aggregated weekly clinical visit data $I_{w,l}$, sentiment $S_{w,l}^{s_i}$ and emotion $E_{w,l}^{e_j}$ proportions by geolocation to construct the ILI and affect proportion time series. We then applied correlation [55] and cross-correlation analysis [56] over ILI and affect location-specific time-series to study (1) what emotions and sentiments have positive vs. negative correlations with clinical ILI visit data, and (2) whether affects are predictive of clinical ILI visit data, respectively.

To study the relationships between ILI and affects expressed in social media, we used Pearson correlation to measure of the degree of linear dependence between two variables e.g., location-specific sentiment proportions $S_{w,l}^{s_i}$ (simplified as S below) and ILI clinical visit proportions $I_{w,l}$ (simplified as I below):

$$\rho_{S,I} = \frac{\text{cov}(I, S)}{\sigma_S \sigma_I} = \frac{\mathbf{E}[(S - \mu_S)(I - \mu_I)]}{\sigma_S \sigma_I}, \quad (6)$$

where μ_I and σ_I are the mean and standard deviation for I ; and similarly for S ; \mathbf{E} indicates the expected value.

To study the predictive power of emotions and sentiments originating from tweets produced by military populations in social media, we used cross-correlation analysis to measure similarity between two series e.g., location-specific emotion proportions $E_{w,l}^{e_j}$ (shown as E) and clinical ILI visit data $I_{w,l}$ (shown as I) as a function of the lag of one relative to the other:

$$\rho_{I,E}(w) = \frac{\mathbf{E}[(I_w - \mu_I)(E_{w+1} - \mu_E)]}{\sigma_I \sigma_E}, \quad (7)$$

where μ_I and σ_I are the mean and standard deviation of the process (I_w); and similarly for (E_w); \mathbf{E} indicates the expected value.

3.2.2 U.S. military and non-military affect time-series

Similar to Eq. (4) and (5), we constructed military-specific vs. non-military-specific ILI and affect time-series from the 8 million tweets annotated in the Comparison dataset. We then performed a Mann-Whitney U test [57] to investigate whether emotions and sentiments expressed by military vs. control populations during the same time frame are statistically different.

Moreover, to further study the affect differences between military and control populations, we estimated emotion and sentiment military-to-control ratios for every sentiment s_i and emotion e_j class over time across six points of interests. The military-to-control emotion ratio and sentiment ratio are defined similarly. As an example, we present the emotion ratio:

$$R(w, p, e_j) = \frac{\text{\# of weekly military tweets labeled with emotion } e_j \text{ per point } p}{\text{\# of weekly control tweets labeled with emotion } e_j \text{ per point } p}. \quad (8)$$

3.2.3 Emotion and sentiment analysis in low vs. high ILI season

To further evaluate the emotions and sentiments expressed in social media during high and low ILI periods, we identified the top 25 weeks for location-specific highest and lowest ILI visit proportion numbers $I_{w,l}$. Then, we extracted the corresponding emotion and sentiment proportions over the same period $E_{w,l}$ and $S_{w,l}$. We observe that over a period between 2011 and 2014 the highest ILI proportions were reported during winter months, as expected for temperate regions [58].

To estimate whether affects expressed in social media are statistically significantly different between high and low ILI periods, we apply a Mann-Whitney U test to (a) 6 military vs. control affect distributions and (b) 31 location-specific affect distributions.

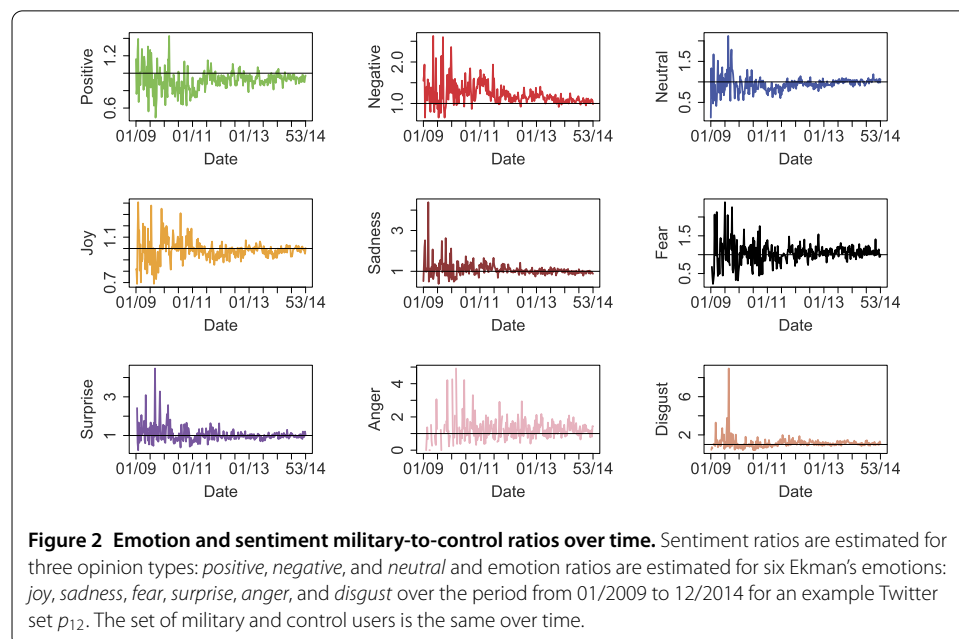
4 Results

We begin this section with our novel findings on sentiment and emotion differences in social media data between military and non-military populations (i.e., the Comparison Dataset). First, we present the general differences identified in sentiments and emotions of the two populations. Second, we discuss the differences between the populations during weeks with high and low ILI visits for a subset of 25 weeks. Next, we present a detailed analysis of the Global Military Dataset. We first focus on the relationships between Twitter point-specific affect and ILI clinical visit time-series evaluated using correlation [55] and cross-correlation analysis [56]. Then, we discuss the affect differences within the Global Military Dataset between 25 weeks with high and 25 weeks with low ILI visits. Finally, we evaluate the predictive power of affects for nowcasting ILI dynamics using the state-of-the-art machine learning models for 31 locations in the Global Military Dataset.

4.1 The variations in affects between military and non-military populations

4.1.1 Differences in emotions and sentiments expressed in the comparison dataset

We observed that emotions and sentiments identified in tweets vary significantly over time for military and control populations (i.e., Comparison Dataset). Figure 2 presents



military-to-control sentiment and emotion ratios (defined in Eq. (8)) over time for an example Twitter set of interest p_{12} . Ratios $R(w, p_{12}, e_j) > 1$, show when military population expresses more of a certain affect vs. $R(w, p_{12}, e_j) \leq 1$, which show when control population expresses the same or more of a specific emotion or sentiment. Overall, the estimated military-to-control ratios over time from the six Twitter sets containing military and control tweet annotations (as shown in Table 3) can be summarized as follows. On average, over 313 weeks (from 11/2009 to 12/2014) across six Twitter points, the military population expresses:

- More *negative* (76% of all weeks military expressed more negative opinions compared to the control) and less *positive* (30% of all weeks military expressed more positive opinions compared to the control) sentiments compared to the control population.
- More *anger* (62%) and *disgust* (60%) than the control population, but less *joy* (36%).

High variance (more spikes) in military-to-control ratios was observed between 2009 and 2011 across all affects in Figure 2. This time period represents the lowest number of tweets produced from military and control populations as shown in Figure 3. Here, a small change in affects extrapolated from tweets can have a large affect on the ratio calculated. Thus, our emotion and sentiment proportion estimates are less confident during these years compared to between 2011 and 2014. Another reason for the high variance seen is that the timeline of tweets collected from a user most likely contain tweets originating from various locations. This highlights the role of location as an important cofactor in affects expressed by military and control populations.

Nevertheless, our findings on military personnel expressing more *negative* emotions and sentiments compared to control population across several regions concurs with the recent report on well-being and mental-health risk among the U.S. military. The large-scale study found that the risk of depression, intermittent explosive disorder, and PTSD is much higher in military than civilian populations in the U.S. [32].

In Figure 2 we demonstrated that military and control populations express different sentiments and emotions in their communications in six separate datasets over time. In order to find whether these differences are statistically significant, we performed a Mann-Whitney U test (Table 4). We found that for the majority of affect types mean differences are statistically significant except for:

- *fear* ($p_{12}, p_{23}, p_{15}, p_{22}, p_{10}$) and *sadness* ($p_3, p_{23}, p_{22}, p_{10}$),
- *surprise* (p_{12}, p_3), *anger* (p_{12}, p_{22}), and *disgust* (p_{12}, p_{15}),
- *joy* and *neutral* sentiment (p_{23}).

Overall, we observed that military populations express significantly different emotions and opinions in social media compared to civilian populations at the same time except for two emotions - *sadness* and *fear*. According to the Mann Whitney U Test, differences in *sadness* and *fear* over time across the majority of areas on interest for military and control populations are not statistically significant.

To provide more insight into our findings on tweet-based aggregate analysis of affects over time, we perform user-based analysis that ignores the temporal component. For that, we aggregate affects expressed by each user between 2011-2014 and contrast the means for each effect type across military and control users. Our results in Table 5 demonstrate that similar to aggregated tweet-based analysis, military and control users significantly differ in the way they express emotions and sentiments in their tweets (except for sadness

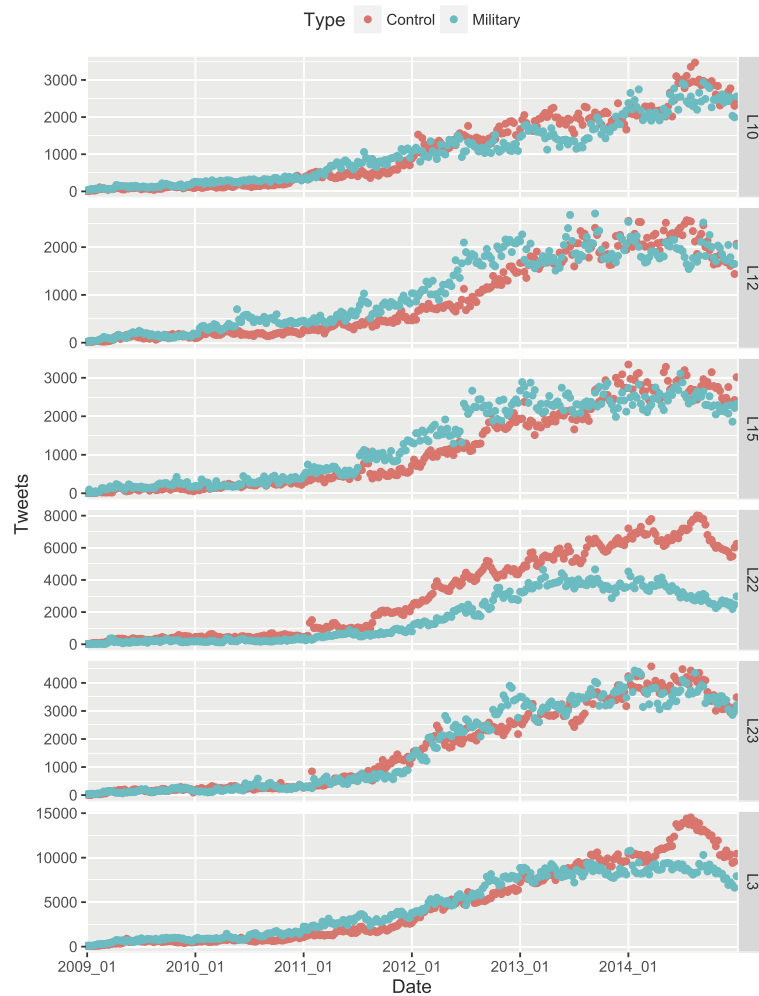


Figure 3 Military and control tweets per from 01/2009 to 12/2014 across six geolocations. Weekly number of tweets vary across locations and over time e.g., between 2009-2011 the number of weekly tweets is significantly lower than between 2011-2014.

emotion and positive sentiment). For the majority of this study, we focus on the population tweet-level analysis of affects over time to understand how they relate to ILI incidence.

4.1.2 Affect differences in high and low ILI seasons between military and non-military populations

Figure 4 presents mean differences for emotions and opinions extracted from military and control populations during high vs. low ILI seasons. We only report statistically significant differences. Moreover, since we perform multiple statistical tests, some will have p -values ≤ 0.05 purely by chance, even if all null hypotheses are really true. Thus, to control the false discovery rate we use Benjamini-Hochberg procedure [59]. Our key significant observations are outlined below.

- Military population, regardless if ILI was high or low, expressed more *anger* and *disgust* emotions, and *negative* sentiment than control population.
- Control population expresses more *joy* emotion and *positive* sentiment compared to military population.

Table 4 Differences in emotion and sentiment proportions between military (m-) and control (c-) populations

Affect	p_{12}	p_{15}	p_{23}	p_{22}	p_3	p_{10}
<i>c-positive</i> ↑	0.399	0.392	0.404	0.406	0.397	0.410
<i>m-positive</i>	0.384	0.385	0.393	0.397	0.389	0.375
<i>c-negative</i>	0.306	0.321	0.323	0.319	0.315	0.298
<i>m-negative</i> ↑	0.334	0.345	0.334	0.334	0.330	0.352
<i>c-neutral</i> ↑	0.296	0.287		0.275	0.289	0.292
<i>m-neutral</i>	0.282	0.270		0.269	0.281	0.273
<i>c-anger</i>		0.015	0.017		0.015	0.014
<i>m-anger*</i>		0.017	0.019		0.018	0.017
<i>c-fear</i>					0.090	
<i>m-fear*</i>					0.096	
<i>c-disgust</i>			0.091	0.091	0.092	0.092
<i>m-disgust</i> ↑			0.092	0.093	0.098	0.098
<i>c-sadness</i>	0.118	0.121				
<i>m-sadness</i> ↑	0.128	0.128				
<i>c-surprise</i>		0.088	0.092	0.091		0.088
<i>m-surprise*</i>		0.092	0.088	0.094		0.086
<i>c-joy</i> ↑	0.588	0.592		0.588	0.591	0.594
<i>m-joy</i>	0.577	0.578		0.583	0.577	0.582

We only report affect proportion means that are statistically significantly different evaluated using Mann-Whitney U Test (p -value ≤ 0.05) (non-significant differences are blank). We mark affects expressed more ↑ by military or control populations across six datasets; affects marked with * vary across sets. We control false positive (FP) discovery rate by applying Benjamini-Hochberg correction with critical value for a FP discovery rate of 0.1.

Table 5 Differences in emotion and sentiment proportions expressed by military and control users between 2011-2014 aggregated over 6 points of interest

Affect	Control Users		Military Users	
	Mean	St. Dev.	Mean	St. Dev.
anger	0.014	0.006	0.016↑	0.007
disgust	0.092	0.029	0.096↑	0.028
fear	0.090	0.024	0.093↑	0.024
joy	0.594↑	0.053	0.587	0.050
surprise	0.096↑	0.023	0.094	0.026
negative	0.313	0.060	0.328↑	0.062
neutral	0.411↑	0.068	0.394	0.067

We only report affect proportion means averaged over military and control users that are statistically significantly different (non-significant differences for sadness emotion and positive sentiment are not reported). We mark affects expressed more ↑ by military or control users. We control false positive (FP) discovery rate by applying Benjamini-Hochberg correction with critical value for a FP discovery rate of 0.1.

- We have not observed statistically significant differences between military and control populations for *fear* and *surprise* emotions in low ILI periods, and *neutral* sentiment during high ILI periods.
- Both military and control populations express more *anger* (p_{10} , p_3) and *negative* (p_3) sentiment during low ILI season; more *joy* (p_3) and positive (p_{10}) sentiment during high ILI period; military population expresses more *disgust* during low ILI period (p_3).

We present a detailed analysis of inter-affect correlations for six points of interests on the West Coast (p_3 , p_{10}), South East (p_{22} , p_{23}), or South Central U.S. (p_{12} , p_{15}) in Figures 9-14 in Additional file 1.

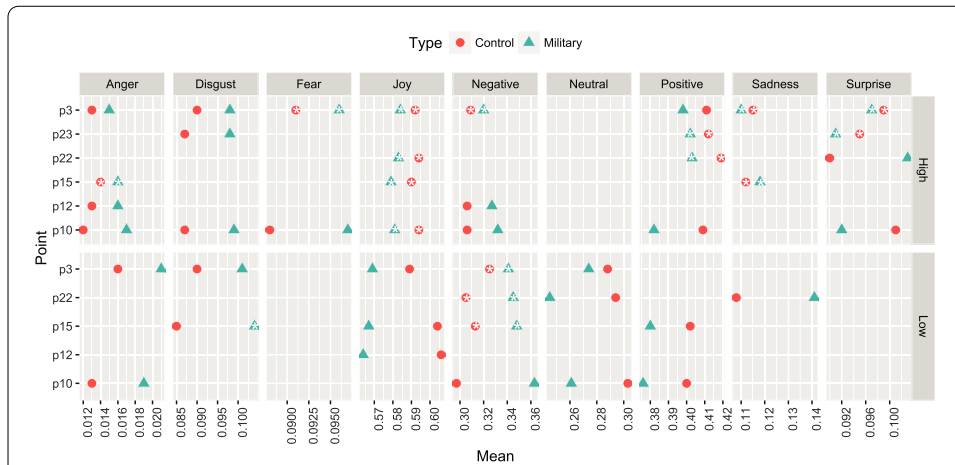


Figure 4 Mean affects during high vs. low ILI seasons across six Twitter sets. We only report differences that are statistically significant from the Mann-Whitney U test (p -value ≤ 0.05); To control false positive discovery rate we apply Benjamini-Hochberg correction: * represent Benjamini-Hochberg critical value with a false discovery rate of 0.2; and the rest of points - critical value with a false discovery rate of 0.1.

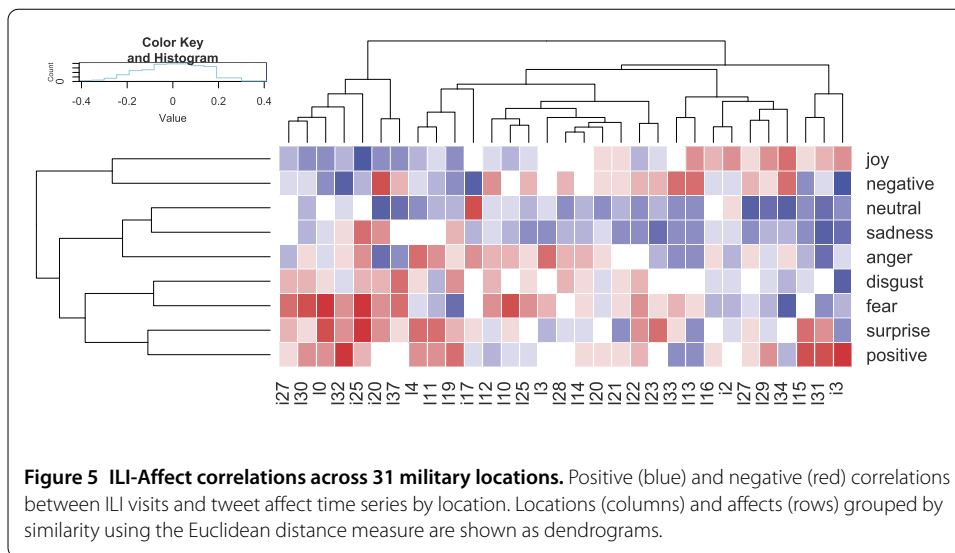


Figure 5 ILI-Affect correlations across 31 military locations. Positive (blue) and negative (red) correlations between ILI visits and tweet affect time series by location. Locations (columns) and affects (rows) grouped by similarity using the Euclidean distance measure are shown as dendrograms.

4.2 The relationships between ILI and affects across global military populations

4.2.1 Correlations between ILI and affect time-series across military locations

In the previous section, we showed that ILI-Affect correlations are different with some general trends observed. To investigate these differences, we present ILI-Affect correlations across 31 military geolocations from our Global Military Dataset (Figure 5). We performed correlation and cross-correlation analysis by subsampling ILI and affect data during influenza seasons (September through May) for several years between November 2011 and December 2014.

Figure 5 demonstrates that across the majority of geolocations ILI positively correlates with *sadness* (except i_{20} , i_{25} , l_{19} and l_{32}) and *neutral* sentiment (except i_3), and negatively correlates with *disgust* (except i_3), *fear* (except l_{34} , i_3 , l_{31} , l_{19}), and *surprise* (except i_3 , l_{13} , l_{31}) emotions, and *positive* sentiment (except l_{34} , l_{13} , l_{33}). We found that the direction of ILI-

Affect correlations vary by locations for *joy* (9 positive and 15 negative), *anger* (11 positive and 17 negative), *negative* (13 positive and 14 negative).

Moreover, the dendrogram that groups military locations (columns) and affects (rows) by similarity using the Euclidean distance measure shows that:

- *Sadness* and *neutral* sentiment positively correlate with ILI visits.
- *Disgust*, *fear*, *surprise* emotions and *positive* sentiment negatively correlate with ILI visits.

The strongest identified affect time series correlations to ILI visits are:

- Positive: *joy* (i_{25}), *negative* sentiment (l_{32} , i_{17} , i_3), *neutral* sentiment (l_{27} , l_{29} , l_{34} , i_{20}), *sadness* (i_3 , l_{31}), *anger* (l_{31}), *disgust* (i_3), *fear* (l_{34} , l_{19}), *surprise* (l_{21} , l_{13}), and *positive* sentiment (l_{33} , l_{13}).
- Negative: *joy* (l_{34}), *negative* sentiment (i_{20}), *neutral* (i_{17}), *sadness* (i_{25} , i_{20}), *anger* (l_3 , l_4), *disgust* (l_{37} , l_{19}), *fear* (i_{25} , l_0 , l_{10}), *surprise* (i_{25}), and *positive* sentiment (i_3 , l_{31} , l_{15} , l_{32}).

4.2.2 Cross-correlations between ILI and affect time-series across military locations

Previous research has shown that the discourse in social media is predictive of influenza outbreaks [15, 17]. In this section, we evaluate whether sentiments and emotions extracted from Twitter communication has the potential to predict influenza dynamics across military geolocations. We apply cross-correlation analysis to assess the predictive power of affects on ILI proportions in our Global Military Dataset. This analysis enable us to identify which affects are seen before ILI (i.e., lead) and those that appear after ILI visits (i.e., lag). We report lead and lag intervals in weeks for every affects across 31 military geolocations in Figure 6.

We observed specific emotions and sentiments reliably lead ILI data and, therefore, are candidates for predicting ILI proportions (shown as blue and green in Figure 6 across geolocations) in lagged regression models for ILI disease forecasting. On the other hand, the affects that lag ILI data may be useful for nowcasting disease. Our key novel findings are the following:

- *Disgust* emotion and all types of sentiment lead ILI proportions for most geolocations between 1 and 4 weeks.

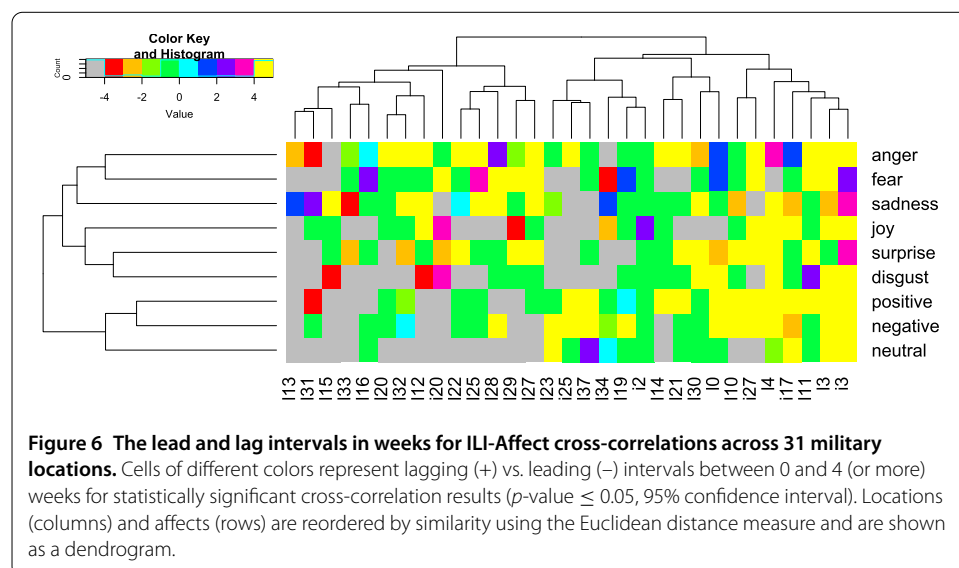


Table 6 Percent of locations that contain the affect as a potential predictive factor of ILI

Affect	West Coast	South East	South Central	International
Anger	20%	53%	25%	33%
Disgust	20%	92%	100%	67%
Fear	40%	69%	75%	50%
Joy	60%	100%	75%	17%
Sadness	80%	38%	50%	83%
Surprise	40%	62%	100%	67%
Negative	20%	69%	100%	50%
Neutral	80%	77%	100%	67%
Positive	0%	85%	75%	33%

These numbers were aggregated from cross-correlation analyses results visualized in Figure 6. Percentages in bold indicate those affects where equal to or greater than 50% of the locations in that region have the affect leading the ILI visit counts.

- *Anger, surprise, fear, sadness, and joy* emotions lead or lag ILI proportions, depending on geolocations, between 1 and 4 weeks.

By combining the 31 locations by U.S. region and separating them from the international locations, we are able to clearly visualize the resulting similarities in leading affects to ILI visit cross-correlation based on geolocation (Table 6). The U.S. is split into three regions, the south east ($n = 13$), south central ($n = 4$) and west coast ($n = 5$), and international ($n = 6$) as one location. Note: two single U.S. locations were unable to be grouped into a region based on geolocation. The differences in affects across locations are exemplified here as many regions have a high percent (80-100) of predictive status for specific affects that other locations have a low percent (0-40). The affects with consistent high percentages are *neutral* in all locations and *joy* in the U.S. and consistent low percentages across all locations is *anger*.

4.2.3 Measuring affect differences in high vs. low ILI seasons across 31 locations

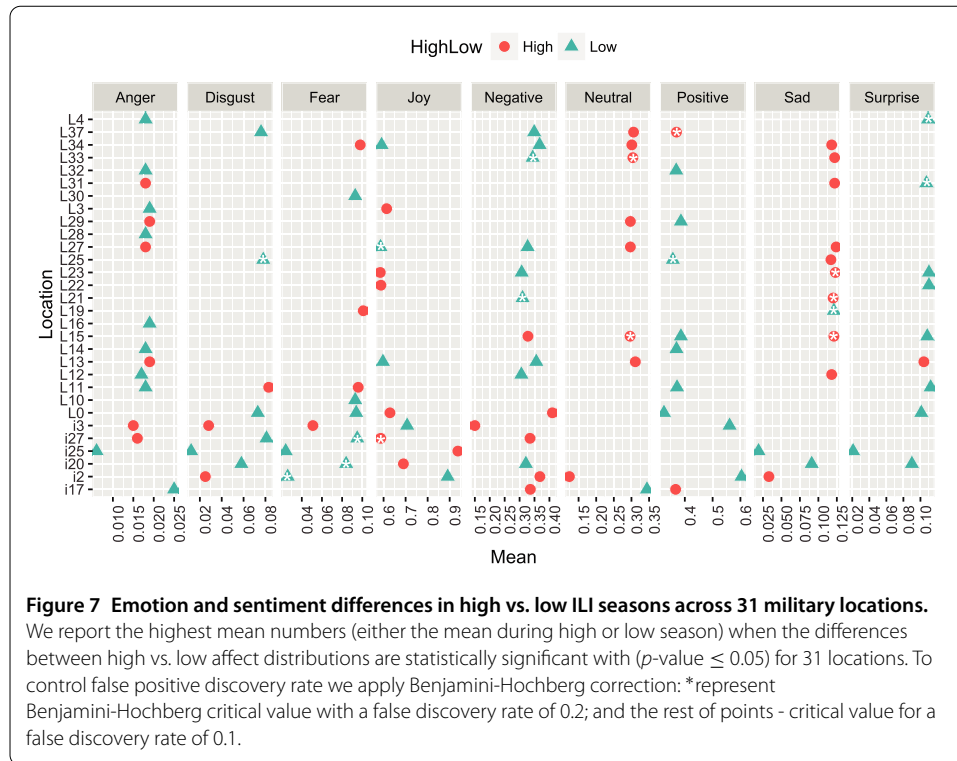
Figure 7 reports the highest mean numbers during high and low seasons for every emotion and sentiment type across 31 locations (i.e., the Global Military Dataset). Our key significant findings in terms of sentiment and emotion means are listed below.

- *Neutral* sentiment and *sadness* emotion are higher during high ILI periods.
- *Positive* sentiment, *anger* and *surprise* are higher during low ILI periods.
- *Negative* sentiment and *disgust, joy and fear* emotion means vary across locations during high and low ILI periods.

4.2.4 Predicting ILI dynamics from affects

We ran preliminary experiments on using affects (emotion and sentiment proportions) to predict location-specific ILI dynamics with machine learning models. We relied on regression models previously used with social media features for ILI dynamics prediction - AdaBoost (DecisionTree regressor), RandomForest, and Linear regressors [17] implemented in scikit-learn [51]. We report results for nowcasting - predicting current week % ILI in Table 7. We evaluate the predictive power of affects using several metrics (i.e., Pearson correlation (CORR), Root Mean Squared Error (RMSE), and Maximum Absolute Percent Error (MAPE)) between predicted and true ILI values as defined in [17]. We train our models on 2012-2013 years and predict for year 2014.

We found that Random Forest and Linear regressors yield higher performance (higher Pearson correlation and lower MAPE and RMSE) compared to AdaBoost models. We show how performance varies across locations. For 39% of locations Linear regressor



demonstrates the highest performance, for 39% of locations Random Forest is the best, and the AdaBoost is the best for 23% of locations.

Our preliminary experiments demonstrate that affects are predictive (depending on location) but not sufficient to accurately predict ILI dynamics (as shown in Figure 8), especially if one wants to make forecasts several weeks in advance. However, if there is no historical ILI data available, affects could serve as predictors of ILI dynamics (Pearson correlation for locations i25, L4, L33, L37, L14, and L3 is between 0.5-0.58). We think that affects combined with other features extracted from tweets can boost model performance. Moreover, recently emerged deep neural networks for sequence prediction, e.g., Long Short-Term Memory models, can potentially boost predictions when learned jointly on text and affect signals. Future work may also include experimenting models that rely on affect and language features for forecasting ILI dynamics several weeks in advance.

5 Discussion

5.1 Variations in affects between military and non-military populations

Military populations are often thought of as small, closed communities that face different combinations of life stressors and challenges than the general public. In support of this concept, we show that emotions and sentiments expressed by military populations in their day-to-day social media discourse differ from surrounding non-military populations in the U.S. In general, military life consists of intense training and subsequent engagement in national security tasks, which put the lives of the military personnel or those that they work with in dangerous and/or chaotic situations. These trauma filled experiences can affect their families directly or indirectly. Twitter is a type of social media where communication is limited in length yet spread to a wide audience rapidly. With only 140 characters allowed, a tweet contains the core essence of what the tweeter is trying to get

Table 7 ILI prediction results (current week estimates for %ILI) for 2014 for 31 locations from affect features using Linear, AdaBoost and Random Forest regressors

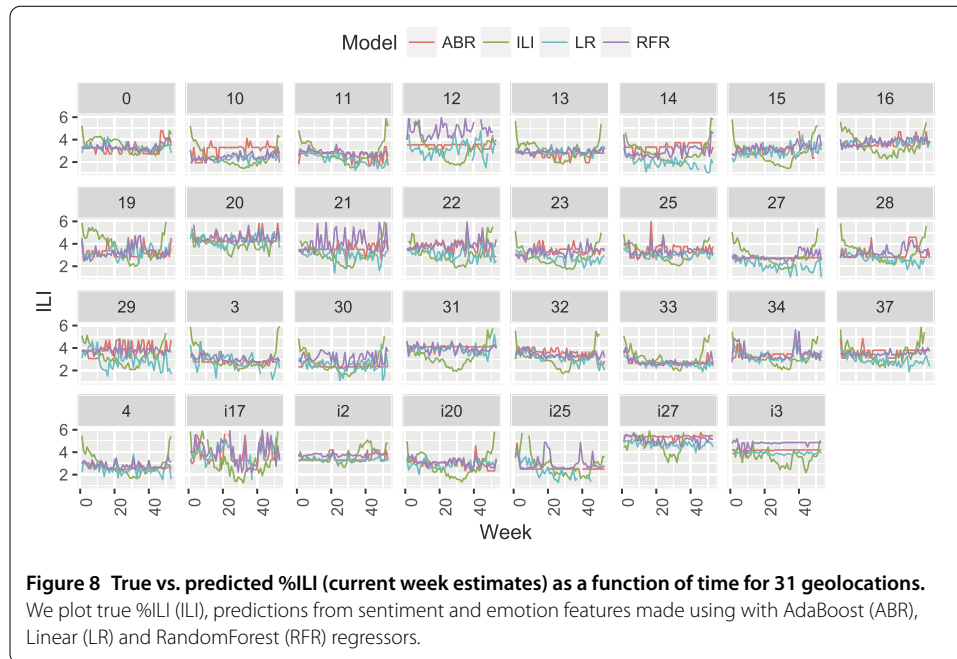
	Linear			AdaBoost			RandomForest		
	CORR	MAPE	RMSE	CORR	MAPE	RMSE	CORR	MAPE	RMSE
L0	0.10	14.5	0.69	0.15	18.2	0.80	0.24	15.40	0.65
L3	0.21	32.0	1.32	0.25	18.4	0.96	0.39	17.90	0.87
L4	0.07	23.4	1.04	0.56	18.9	0.82	0.28	23.60	0.87
L10	0.19	26.9	0.84	-0.60	60.6	1.44	0.09	30.40	0.87
L11	0.22	24.6	1.08	0.13	25.8	1.02	0.18	27.70	0.92
L12	0.10	43.6	1.70	0.08	34.8	1.48	0.12	67.10	2.09
L13	0.28	23.0	1.11	0.28	22.7	1.20	0.21	22.70	1.13
L14	0.19	34.1	1.49	-0.15	37.8	1.20	0.41	28.90	0.92
L15	-0.01	37.7	1.26	-0.01	50.6	1.63	0.26	40.10	1.17
L16	-0.07	26.6	1.30	0.06	24.8	1.29	-0.19	30.50	1.37
L19	0.09	32.1	1.61	0.15	28.1	1.61	-0.28	35.00	1.78
L20	0.16	16.9	1.23	0.28	18.2	1.18	0.21	22.90	1.27
L21	0.06	29.1	1.55	0.25	32.7	1.41	0.03	50.30	1.80
L22	-0.16	29.8	1.62	-0.44	41.1	1.68	0.08	42.70	1.56
L23	0.06	28.3	1.14	-0.12	41.8	1.27	0.06	37.80	1.18
L25	0.24	18.9	0.88	-0.06	28.9	1.11	0.03	24.30	0.98
L27	0.23	32.2	1.57	0.20	22.7	1.24	0.21	23.20	1.20
L28	0.10	22.0	1.39	-0.01	24.0	1.31	-0.21	27.00	1.29
L29	-0.11	29.5	1.67	-0.07	37.0	1.48	-0.16	31.80	1.35
L30	0.16	28.7	1.37	0.11	23.0	1.33	0.23	38.00	1.28
L31	0.12	35.4	1.22	0.13	39.9	1.30	0.11	35.40	1.22
L32	-0.29	25.0	1.45	-0.48	33.7	1.51	0.13	26.00	1.22
L33	0.14	19.6	1.16	0.37	17.3	1.04	0.50	18.00	1.00
L34	0.18	19.8	0.99	-0.10	29.4	1.32	0.16	23.90	1.11
L37	0.09	22.7	1.67	0.29	20.2	1.34	0.49	18.70	1.34
i2	0.12	12.1	0.73	0.28	11.7	0.66	0.13	13.70	0.62
i3	0.33	19.2	0.78	-0.01	25.3	0.95	-0.08	40.20	1.42
i17	0.17	71.0	2.42	-0.09	57.8	1.84	0.15	52.70	1.74
i20	0.09	34.6	1.09	-0.14	38.4	1.21	-0.12	39.30	1.17
i25	0.58	35.1	1.54	-0.03	32.9	1.65	-0.08	33.10	1.50
i27	0.03	20.8	1.67	0.29	23.2	1.44	0.17	21.20	1.49

Models are trained exclusively on emotion and sentiment proportions extracted from tweets in 2012-2013. Models are evaluated to predict %ILI in 2014. The highest CORR and the lowest MAPE and RMSE are highlighted in bold.

across, whether emotional or factual. Along these lines, tweets can provide insight into a person's emotional and sentimental state. This study shows that military populations tend to contain more *negative* and less *positive* sentiments than control populations, in addition to increased emotions of *sadness*, *fear*, *disgust* and *anger*. These findings of more *negative* sentiments, including *disgust* and *anger*, by military and *positive* sentiments by the control population even hold true when specifically looking at high and low times of ILI. These results may be attributed to the hard facts of military life in general and the day-to-day challenges that they face in order to ensure protection of the civilian populations.

5.2 Correlations between ILI and affect time-series across global military locations

Studies in the literature have investigated the relationship between affects and health, many of which show that *positive* affects correlate with good health [18, 19]. In this paper, we examined the correlation between various affects identifiable in tweets and the health status of the subpopulation, containing those tweeters, based on the number of ILI visits to a health facility. We found that people within a 25-mile radius around military bases express sentiment and emotion in tweets that correlate with ILI visits to medical



facilities in the same location. These correlations are mostly location and affect type dependent, which may suggest that location-specific demographics and characteristics may be influencing these differences. Different location specific aspects were investigated including the percent of military population to civilians, the volume of tweets collected, the ratio of ILI visits to tweet volume, and the service type of the bases within each location boundaries. The best generalizations were identified when the U.S. locations were split into southeastern, south central, and west coast. By doing this regionalization, correlations between affects and ILI visits became more apparent.

Although *neutral* sentiment was positively correlated with ILI visits throughout the U.S. (which means that populations are more neutral, in other words, express less opinionated tweets), *positive* and *negative* sentiments correlated differently to ILI visits regionally. In the Southeast, 77% of the military locations' ILI visits were negatively correlated to *negative* sentiment. However, those locations that were positively correlated (23% of SE military) were all identified to be located with in a single state and military service type. The latter positive correlation between ILI and *negative* sentiment was also observed in the South Central and West Coast states with a high proportion of military to civilian populations. For *positive* sentiment, there were varying degrees of negative correlation to ILI visits, e.g., 54% of locations in the southeast and 75% of locations in the south central U.S. For the west coast, those locations with a higher percent of military to civilians showed a similar negative correlation between ILI and *positive* sentiment. In contrast, a single state with multiple locations and service types showed a positive correlation between *positive* sentiment and ILI visits. This degree of variability in correlations based on location stresses the importance of understanding which characteristics may play a role in the sub-populations reaction to health situations. As discussed by Gallo and Matthews (2003), the differences in location's affects may be attributed to the socioeconomic status of the areas and the amount of stress that plays a role in the tweeter's life [21].

The six emotions identified in tweets showed varying degrees of correlation with ILI visits. The most straight forward was a positive correlation with *sadness* found in all regions but with varying degrees of concurrence throughout each area, i.e. 60% of west coast location groups, 85% of the southeast, and 100% of the south central region. There was a negative correlation to *anger* sentiment in 100% of west coast locations and 75% in south central U.S. that contained the areas with the lowest percentage of military to civilian population. Other emotions showed demographic-specific results. *Disgust* correlated positively with military in the air force and negatively with those in the navy, specifically in the Southeast. *Fear* and *disgust* were highest when ILI visits were highest in areas dense with military personnel yet low when ILI was highest in areas with low density of military personnel. *Joy* and *surprise* were very dependent on location, percent of military population, and service type. For example, in the west coast locations, *surprise* was negatively correlated with ILI in 80% of the locations and *joy* was positively correlated in 80% of the locations. In the south central region, military percent played the biggest role with negative correlations between ILI visits and both *surprise* and *joy* in areas with high military to civilian populations along with positive correlation to *surprise* in areas with less military than civilians. Lastly, in the southeastern U.S., service type seems to play the most important role with 62% of the locations negatively correlated with *surprise* and all army service areas positively correlated. In the same respect, all negatively correlated *joy* emotions to ILI visits were observed in the army service in the same area.

In general for the locations studied here, there is a correlation between sentiments and emotions identified in tweets and ILI visits to health facilities within the same areas. The strongest correlations across all U.S. locations are a negative correlation to *anger* and *surprise*, positive correlation to *sadness*, and a mixed bag for *fear*, *disgust*, and *joy*. For sentiments, there was an overwhelming positive correlation to *neutral* statements, negative correlation to *positive* sentiments, and a mixed bag for *negative* sentiments correlating to ILI visits based on which region was analyzed. Very similar results were identified when we looked at affects most present during high vs. low ILI clinical visits across datasets, i.e., *neutral* sentiment and *sadness* were expressed most during high ILI times and *positive* sentiment, *anger*, and *surprise* are expressed during low periods. Interestingly, our findings support the psychology research that identifies *positive* affects correlating to good health. However, the converse isn't naturally shown in our data. Instead, *neutral* sentiments, i.e., a decreased amount of strongly opinionated tweets, characterize the period of ill health, namely ILI, for the population. This lack of *positive* affects when ILI visits are high can be explained by Fredrickson's paper, which emphasizes that positive emotions are not just the absence of *negative* emotions [20].

5.3 Cross-correlations between ILI and affect time-series across 31 military locations

By regressing affects 1 to 5 weeks before the ILI visit counts, we were able to identify which affects in tweets could be used to predict the number of ILI cases. The results were overall variable but when broken down regionally, we were able to identify specific attributes that may be responsible for these differences, e.g. location (i.e., country, region and/or state level), military service type, and percent of military per location. *Neutral* sentiment and *joy* emotion were the only affects that were consistent across all U.S. locations as strong candidates for predictive models of ILI visits whereas *anger* was constantly unreliable.

In the southeast, there were strong trends by state, whereas in the west coast and south central areas, military service and the percent of the population being military played a role in predictive affects. In general, the south eastern and south central regions affects lead ILI visits more often than other studied locations and, therefore, may be a good starting point for predictive investigation of ILI visits.

6 Limitations

We note that our analysis and methodology have several limitations.

First, our social media data collection approach for identifying military versus control populations does not guarantee that the control population is not military. Also, the method used does not allow for any assumptions on the location of the Twitter user since the time series of tweets were collected based on user and not location. Second, our social media data collection approach does not allow us to claim that all tweets originating within a 25-mile radius of military locations have been produced by military personnel or their family members. Third, our affect classification models are capable of predicting affects in tweets with only a certain level of accuracy, which would bring some noise, e.g., mislabeled annotations, to our analysis. Fourth, using ILI location-specific weekly proportions as our gold standard for influenza dynamics in military populations may not necessarily be ideal. Finally, we only study correlations between affects and ILI clinical diagnosis data. As such, we can not make any causal inference regarding these parameters.

Despite these limitations, due to the size of analyzed dataset, we believe our conclusions regarding emotion and sentiment differences between military and control populations, ILI-Affect correlations, and cross-correlations are accurate.

7 Conclusion

We performed a novel large-scale study of automatically inferred emotions and sentiments emanating from 171 million tweets produced by military and non-military associated people across 25 locations in the U.S. and 6 international locations.

Through studying military and non-military associated Twitter communications, we found significant differences in their expressions of sentiments and emotions in social media throughout time. In general in control population communications, *positive* affects are expressed more and *negative* affects less than in military population communications during a year. This underlying theme may be a result of the different lifestyle and responsibilities endured between the military and civilian populations.

In addition, our analysis identified emotions and sentiments in tweets for a given location that correlate with the amount of military ILI visits to medical facilities in the same area regardless of whether or not the tweets are from military or non-military individuals. In general, *positive* affects in tweets correlate with less medical facility visits for ILI symptoms and *neutral* affects are used during times of increased ILI visits.

We then showed that by combing locations into regions, trends in the direction of correlation and the predictive usage of affects, become more apparent and easily identified. In specific instances, we have identified additional characteristics within the regions that help explain trends seen. These include similarities in emotions and sentiments expressed by all locations in a specific state, military service type, and/or locations within a region that have high or low ratios of military to civilian populations.

Our preliminary regression experiments on predicting ILI dynamics showed that affects are predictive but not sufficient to accurately generate current week ILI estimates.

We found that predictive power of sentiments and emotions vary significantly across locations. Nevertheless, we think that affects combined with other features extracted from tweets, location-specific feature selection (e.g., our findings from cross-correlation analysis), as well as deep neural network models for sequence prediction can boost prediction accuracy, and even forecast ILI dynamics several weeks in advance.

Overall, the information gained in this study exemplifies a usage of social media data to understand the correlation between psychological behavior and health in the military population and the potential for use of social media affects for prediction of ILI cases.

Additional material

Additional file 1: Figures 9-14 demonstrate the inter-affect correlations presented in Figure 4 for six Twitter points of interest with military and control split on the West Coast, South East and South Central U.S. (pdf)
Additional file 2: Availability of data and materials. (zip)

Acknowledgements

The authors thank Commander Jean-Paul Chretien, Aaron Kite-Powell and Vivek Khatri, who were at the Armed Forces Health Surveillance Branch, Defense Health Agency, for helpful discussions on the research and experimental design of this study.

Funding

This work was funded by a contract with the Defense Threat Reduction Agency, Joint Science and Technology Office for Chemical and Biological Defense under project number CB10082. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle under Contract DE-AC05-76RL01830.

Abbreviations

CDC, Centers for Disease Control and Prevention; DMIS, Defense Medical Information System; ICD, International Statistical Classification of Disease and Health Related Problems; ID, Identifier; ILI, influenza-like illness; NOS, Not otherwise specified; PTSD, Post-traumatic stress disorder; U.S., United States.

Availability of data and materials

The datasets supporting the conclusions of this article are included in Additional file 2.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CDC SV LCS conceived and designed the experiments. SV performed the experiments. JH collected and pre-processed the data. SV built the models and analysis tools. SV LCS analyzed the data. SV LCS CDC wrote the paper. All authors have read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 15 June 2016 Accepted: 22 May 2017 Published online: 08 June 2017

References

1. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EH, Olsen JM et al (2015) Using social media for actionable disease surveillance and outbreak management: a systematic literature review. *PLoS ONE* 10(10):e0139701
2. WHO (2015) Influenza (Seasonal), Fact Sheet Number 211. <http://www.who.int/mediacentre/factsheets/fs211/en/index.html>
3. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2009) Detecting influenza epidemics using search engine query data. *Nature* 457(7232):1012-1014
4. Cook S, Conrad C, Fowlkes AL, Mohebbi MH (2011) Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS ONE* 6(8):e23610
5. Lazer D, Kennedy R, King G, Vespignani A (2014) The parable of Google Flu: traps in big data analysis. *Science* 343(14):1203-1205
6. Polgreen PM, Chen Y, Pennock DM, Nelson FD, Weinstein RA (2008) Using Internet searches for influenza surveillance. *Clin Infect Dis* 47(11):1443-1448
7. McIver DJ, Brownstein JS (2014) Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. *PLoS Comput Biol* 10(4):e1003581

8. Generous N, Fairchild G, Deshpande A, Del Valle SY, Priedhorsky R (2014) Global disease monitoring and forecasting with Wikipedia. *PLoS Comput Biol* 10(11):e1003892
9. Chunara R, Aman S, Smolinski M, Brownstein JS (2013) Flu near you: an online self-reported influenza surveillance system in the USA. *Online J Public Health Inform* 5(1):e133
10. Crawley AW (2012) Flu near you: comparing crowdsourced reports of influenza-like illness to the CDC outpatient influenza-like illness surveillance network, October 2012 to March 2014. In: CSTE annual conference, pp 1-7
11. Culotta A (2010) Towards detecting influenza epidemics by analyzing Twitter messages. In: Proceedings of the first workshop on social media analytics. ACM, New York, pp 115-122
12. Corley CD, Cook DJ, Mikler AR, Singh KP (2010) Text and structural data mining of influenza mentions in web and social media. *Int J Environ Res Public Health* 7(2):596-615
13. Signorini A, Segre AM, Polgreen PM (2011) The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PLoS ONE* 6(5):e19467
14. Achrekar H, Gandhe A, Lazarus R, Yu SH, Liu B (2012) Twitter improves seasonal influenza prediction. In: HEALTHINF, pp 61-70
15. Paul MJ, Dredze M, Broniatowski D (2014) Twitter improves influenza forecasting. *PLoS currents*, 6
16. Chen L, Tozammel Hossain K, Butler P, Ramakrishnan N, Prakash BA (2014) Flu gone viral: syndromic surveillance of flu on Twitter using temporal topic models. In: IEEE International Conference on Data Mining (ICDM). IEEE, New York, pp 755-760
17. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS (2015) Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS Comput Biol* 11(10):e100451
18. Pressman SD, Cohen S (2005) Does positive affect influence health?. *Psychol Bull* 131(6):925-971
19. Tugade MM, Fredrickson BL, Feldman Barrett L (2004) Psychological resilience and positive emotional granularity: examining the benefits of positive emotions on coping and health. *J Pers* 72(6):1161-1190
20. Fredrickson BL (2000) Cultivating positive emotions to optimize health and well-being. *Prev Treat* 3(1):1a
21. Gallo LC, Matthews KA (2003) Understanding the association between socioeconomic status and physical health: do negative emotions play a role?. *Psychol Bull* 129(1):10
22. De Choudhury M, Counts S, Horvitz E (2013) Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 3267-3276
23. Coppersmith G, Harman C, Dredze M (2014) Measuring post traumatic stress disorder in Twitter. In: Proceedings of ICWSM
24. Salathé M, Khandelwal S (2011) Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol* 7(10):e1002199
25. Bittner L (2014) How is social media used by military families to communicate during deployment?
26. Hodge J, Shanks D (2011) The ability of seasonal and pandemic influenza to disrupt military operations. *J Mil Veterans' Health* 19(4):13-18
27. Reber CAS, Boden MT, Mitragotri N, Alvarez J, Gross JJ, Bonn-Miller MO (2012) A prospective investigation of mindfulness skills and changes in emotion regulation among military veterans in posttraumatic stress disorder treatment. *Mindfulness* 4(4):311-317
28. Tokuno S, Tsumatori G, Shono S, Takei E, Suzuki G, Yamamoto T et al (2011) Usage of emotion recognition in military health care. In: Defense science research conference and Expo, vol DSR, pp 1-5
29. Pflanz S (2001) Occupational stress and psychiatric illness in the military: investigation of the relationship between occupational stress and mental illness among military mental health patients. *Mil Med* 166(6):457-462
30. Boehmer TK, Boothe VL, Flanders WD, Barrett DH (2003) Health-related quality of life of U.S. military personnel: a population-based study. *Mil Med* 168(11):941-947
31. Day AL, Livingstone HA (2001) Chronic and acute stressors among military personnel: do coping styles buffer their negative impact on health? *J Occup Health Psychol* 6(4):348-360
32. Study WV (2014) Rates of many mental disorders much higher in soldiers than in civilians. <http://www.cnn.com/2014/03/03/health/jama-military-mental-health/>
33. Delgado Valdes JM, Eisenstein J, De Choudhury M (2015) Psychological effects of urban crime gleaned from social media. In: Proceedings of ICWSM
34. Volkova S, Bachrach Y (2015) On predicting sociodemographic traits and emotions from communications in social networks and their implications to online self-disclosure. *Cyberpsychol Behav Soc Netw* 18(12):726-736
35. Bachrach Y, Kosinski M, Graepel T, Kohli P, Stillwell D (2012) Personality and patterns of Facebook usage. In: Proceedings of ACM WebSci, pp 24-32
36. Schoenebeck SY (2013) The secret life of online moms: anonymity and disinhibition on YouTubeMom.com. In: Proceedings of ICWSM, pp 555-562
37. Beller C, Knowles R, Harman C, Bergsma S, Mitchell M, Van Durme B (2014) I'm a believer: social roles via self-identification and conceptual attributes. In: Proceedings of ACL, pp 181-186
38. Cui A, Zhang M, Liu Y, Ma S, Zhang K (2012) Discover breaking events with popular hashtags in Twitter. In: Proceedings of CIKM. ACM, New York, pp 1794-1798
39. Starbird K, Maddock J, Orand M, Achterman P, Mason RM (2014) Rumors, false flags, and digital vigilantes: misinformation on Twitter after the 2013 Boston marathon bombing. In: Proceedings of iConference, pp 1-5
40. Pavalanathan U, Datla V, Volkova S, Charles-Smith L, Pirrung M, Harrison J et al (2016) Discourse, health and well-being of military populations through the social media lens. In: Proceedings of the 3rd international workshop on the World Wide Web and population health intelligence (W3PHI), pp 796-803
41. Mohammad SM, Kiritchenko S, Zhu X (2013) NRC-Canada: building the state-of-the-art in sentiment analysis of tweets. In: Proceedings of SemEval, pp 321-327
42. Zhu X, Kiritchenko S, Mohammad SM (2014) NRC-Canada-2014: recent improvements in the sentiment analysis of tweets. In: SemEval, pp 443-447
43. Mohammad SM, Kiritchenko S (2014) Using hashtags to capture fine emotion categories from tweets. *Comput Intell* 31(2):301-326

44. Volkova S, Bachrach Y, Armstrong M, Sharma V (2015) Inferring latent user properties from texts published in social media (demo). In: Proceedings of AAAI, pp 4296–4297
45. Preoiuc-Pietro D, Volkova S, Lampos V, Bachrach Y, Aletas N (2015) Studying user income through language, behaviour and affect in social media. *PLoS ONE* 10(9):e0138717
46. Saif H, Fernandez M, He Y, Alani H Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold. First ESSEM workshop. 2013;
47. Nakov P, Rosenthal S, Kozareva Z, Stoyanov V, Ritter A, Wilson T (2013) SemEval-2013 task 2: sentiment analysis in Twitter. In: Proceedings of SemEval, pp 312–320
48. González-Ibáñez R, Muresan S, Wacholder N (2011) Identifying sarcasm in Twitter: a closer look. In: Proceedings of ACL, pp 581–586
49. Valitutti R (2004) WordNet-affect: an affective extension of WordNet. In: Proceedings of LREC, pp 1083–1086
50. Roget's Thesaurus. <http://www.thesaurus.com/Roget-Alpha-Index.html>
51. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
52. Roberts K, Roach MA, Johnson J, Guthrie J, Harabagiu SM (2012) EmpaTweet: Annotating and detecting emotions on Twitter. In: Proceedings of LREC, pp 3801–3813
53. Wang W, Chen L, Thirunarayan K, Sheth AP (2012) Harnessing Twitter 'big data' for automatic emotion identification. In: Proceedings of SocialCom, pp 587–592
54. Qadir A, Riloff E (2013) Bootstrapped learning of emotion hashtags #hashtags4you. In: WASSA 2013, pp 1–11
55. Pearson K (1895) Note on regression and inheritance in the case of two parents. *Proc R Soc Lond* 58:240–242
56. Cross Correlation Functions and Lagged Regressions. <https://onlinecourses.science.psu.edu/stat510/node/74>
57. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 18(1):50–60
58. Wenger JB, Naumova EN (2010) Seasonal synchronization of influenza in the United States older adult population. *PLoS ONE* 5(4):e10187
59. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc, Ser B, Methodol* 57(1):289–300

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com