

What Twitter Profile and Posted Images Reveal About Depression and Anxiety

Sharath Chandra Guntuku¹, Daniel Preotiuc-Pietro², Johannes C. Eichstaedt¹, Lyle H. Ungar¹

¹University of Pennsylvania, ²Bloomberg LP
 {sharathg@sas, jeich@sas, ungar@cis}.upenn.edu, dpreotiucpie@bloomberg.net

Abstract

Previous work has found strong links between the choice of social media images and users' emotions, demographics and personality traits. In this study, we examine which attributes of profile and posted images are associated with depression and anxiety of Twitter users. We used a sample of 28,749 Facebook users to build a language prediction model of survey-reported depression and anxiety, and validated it on Twitter on a sample of 887 users who had taken anxiety and depression surveys. We then applied it to a different set of 4,132 Twitter users to impute language-based depression and anxiety labels, and extracted interpretable features of posted and profile pictures to uncover the associations with users' depression and anxiety, controlling for demographics. For depression, we find that profile pictures suppress positive emotions rather than display more negative emotions, likely because of social media self-presentation biases. They also tend to show the single face of the user (rather than show her in groups of friends), marking increased focus on the self, emblematic for depression. Posted images are dominated by grayscale and low aesthetic cohesion across a variety of image features. Profile images of anxious users are similarly marked by grayscale and low aesthetic cohesion, but less so than those of depressed users. Finally, we show that image features can be used to predict depression and anxiety, and that multitask learning that includes a joint modeling of demographics improves prediction performance. Overall, we find that the image attributes that mark depression and anxiety offer a rich lens into these conditions largely congruent with the psychological literature, and that images on Twitter allow inferences about the mental health status of users.

Introduction

Depression continues to be under-diagnosed with only 13-49% receiving minimally adequate treatment (Wang et al. 2005). This is caused by a complex set of factors such as imperfect screening methods, social stigma associated with the diagnosis, and lack of cost-effective resources for and access to diagnosis and treatment. Efforts to detect depression predominantly rely on online and phone surveys. However, these surveys are resource intensive, both in terms of cost and time, and insufficiently reach all at risk (Suchman 1962). The single time point of assessment also means there are

Copyright © 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

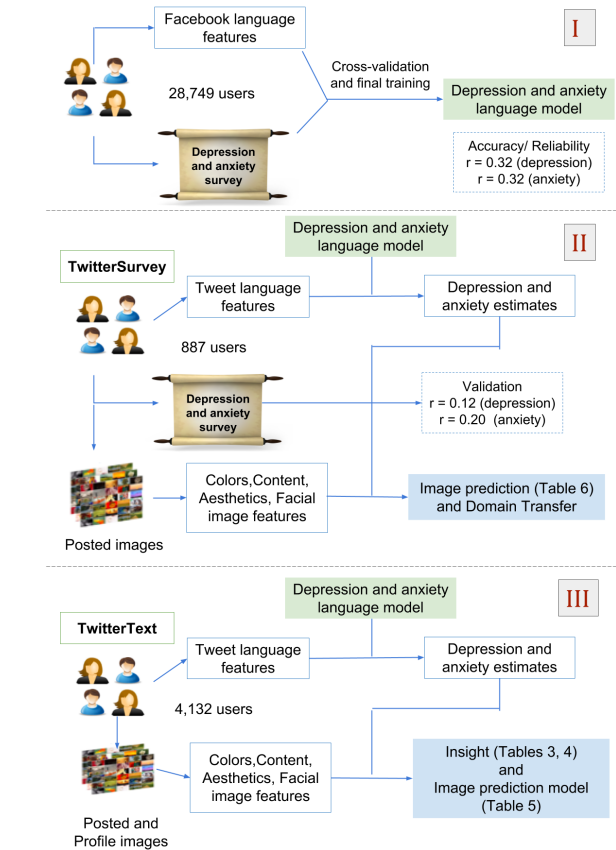


Figure 1: Overview of data flow, models and analysis in the paper

often gaps between data collection and depression onset. Automated analysis of user generated content can potentially provide methods for early detection of depression (Guntuku et al. 2017b). If an automated process could detect elevated depression or anxiety levels in a person, that individual could be targeted for a more thorough assessment (and provided with digitized forms of support and treatment), alleviating many of the constraints associated with traditional assessment methods.

The last few years have seen an increased interest in studying mental health through social media. Recent studies have investigated the association of language and social media

Study	# Users	Traits	Image Type	Image Features
Our Work (Reece and Danforth 2017) (Andalibi, Ozturk, and Forte 2015)	887 + 4,132 166 –	Depression & Anxiety Depression Depression	Twitter Posted & Profile Images Instagram Photos 500 ‘#depression’ Instagram Photos	Color, Facial, Aesthetics, Content, VGG-Net Colors Manual Annotation
(Ferwerda and Tkalčić 2018) (Nie et al. 2018) (Samani et al. 2018) (Farnadi et al. 2018) (Guntuku et al. 2017a) (Segalin et al. 2017) (Liu et al. 2016) (Ferwerda, Schedl, and Tkalčić 2016) (Skowron et al. 2016) (Guntuku et al. 2016) (Guntuku et al. 2015) (Al Moubayed et al. 2014) (Celli, Bruni, and Lepri 2014)	193 2238 300 5670 4132 + 161 11,736 66,502 113 62 300 123 829 112	Personality (continuous) Perceived Personality (continuous) Personality (continuous) Personality (binary) Personality (continuous) Personality (continuous & binary) Personality (continuous) Personality (binary) Personality (binary) Personality (continuous) Personality (continuous) Personality (binary) Personality (binary)	Instagram Photos Web Portrait Images Twitter and Flickr Posts, Likes, & Profiles Facebook Profile Images Posted, liked images and text on Twitter Facebook Profile Images Twitter Profile Images Instagram Photos Instagram Photos Liked (‘Fave’) images on Flickr Selfies on Weibo Face Images Facebook Profile Images	Content Facial, Social information Colors, Content, VGG-Net Facial, Text, Likes Color, Bag of Imagg tags, VGG-Net Aesthetics, BOVW, VGG-Net, IATO Color, Facial Colors, #Faces, Filters Colors Colors, semantic features, aesthetics Color, Aesthetics, BOVW, Emotions Eigenfaces Bag-of-Visual-Words (BOVW)

Table 1: Summary of data and methods used in previous work analyzing images of individual users.

use patterns with several mental illnesses, including stress, depression and suicidality. Such studies can be grouped into two categories: a) studies aimed at detecting mental illnesses and their symptoms from social media (the focus of this paper); and b) studies that examine how the use of social media contributes to or alleviates distress in users. Most prior works explored the use of linguistic attributes to predict mental health conditions with reasonable success. (Guntuku et al. 2017b) review the use of social media text-based features in detecting mental health conditions.

With the ubiquity of camera-enabled mobile devices and growth of image-based social media platforms (such as Twitter, Instagram or Snapchat) social media content is increasingly in image form (Burdick et al. 2012), and preliminary work suggests that users are increasingly communicating their emotions using images (Reece and Danforth 2017). Current computer vision algorithms allow us to automatically analyze large scale data sets and extract shallow color features as well as semantic content such as objects or scenes. Previous work has studied the variation in colors and use of filters of posted images in depressed individuals and has not studied the diverse content associated with posted images or the characteristics of profile images for users with depression and anxiety (Reece and Danforth 2017).

Moreover, prior works typically study mental health conditions in isolation, not exploiting the underlying implicit correlations in conditions and demographic factors. Tasks with underlying commonalities have been shown to benefit from multi-task learning, e.g. action recognition (Wang et al. 2016), chronic disease prediction (Nie et al. 2015). Particularly user demographics (age and gender) and the co-morbidity between mental health conditions have been used to diagnose patients by clinicians (Bijl, Ravelli, and Van Zessen 1998). We use multi-task learning to take these strong associations into account.

In summary, we examine the following questions:

- What meaningful, relevant and interpretable patterns do images that users post and choose as profile pictures reveal about users’ depression and anxiety?
- What effect does joint learning of demographics along with co-morbid mental health conditions have on prediction performance?
- Do text-predicted labels help when predicting survey-based ground-truth labels from images?

Related Work

Researchers have used images to study personality, measured using the Big Five model (McCrae and John 1992), based on profile pictures (with facial features). Others have also used posted images. One of the earliest works predicted self-assessed personalities of 100 users using their Facebook profile images (Celli, Bruni, and Lepri 2014) with ~65% accuracy using bag-of-visual-words (BoVW) features. Random portraits from the web (Nie et al. 2014) and existing face recognition data sets (Al Moubayed et al. 2014) were also used to model users’ personality. Recently, aesthetic features (Datta et al. 2006; Machajdik and Hanbury 2010) (apart from facial features) were also used to predict personality on a ~66,000 user data set (Liu et al. 2016) from Twitter. Further, (Segalin et al. 2017) used multiple sets of features extracted from Facebook images of 11,736 users and built computational models which have better predictive power than human raters in predicting personality traits, specifically Extraversion and Neuroticism.

However, in the health domain manifestation of mental health conditions in individual users based on social media images is under explored, despite recent work being done on studying public health of communities (Manikonda and De Choudhury 2017; Chancellor et al. 2017; Garimella, Alfayad, and Weber 2016a). Table 1 presents a summary of the relevant works: number of users, traits, image types studied and features used.

(Andalibi, Ozturk, and Forte 2015) examined the distribution of themes in 500 images with the hash-tag #depression uncovering themes posted by depressed users. However, limited to a small set of images, further work is required to study the generalisability of the findings. (Reece and Danforth 2017) looked for the markers of depression in Instagram posts. However, the study aggregates images by days of posting instead of participants even though depression was measured a single time as a trait using a questionnaire. It also looks at a limited set of image features on posted images.

The aim of our work is to use social media images to study how depression and anxiety are related to the content of images that people post or choose as profile pictures on social media. We explore if incorporating author attributes such as age and gender in the joint modeling of depression and anxiety can improve prediction performance. Further, we examine if weak text-predicted labels can improve performance

in predicting reliable survey based outcomes.

Data

We use a Facebook dataset (where we have language data but not image data from users) and two Twitter datasets (where we have both language and image data from users) in this study (overview in Figure 1). This study received approval from the University of Pennsylvania Institutional Review Board (IRB).

The Facebook data is from a previous study (Schwartz et al. 2014), consisting of 28,749 users who had taken the IPIP NEO-PI-R survey (Costa and McCrae 2008) that contains the Depression and Anxiety Facets of the Neuroticism Factor. These users also consented to share access to their status updates which was used to build a text-regression model. The text model was trained using 1-,2-, and 3-grams used by at least 5% of users (resulting in 10,450 ngrams), 2000 LDA derived topics, and 64 LIWC features extracted from status updates and a linear regression with L2 ridge penalization on 10% principal component features. In the original validation, the model achieved a Pearson correlation of $r = .32$ predictive performance, which is considered a high correlation in psychology, especially when measuring internal states (Meyer et al. 2001). We did not have access to image data for the Facebook cohort.

TwitterSurvey was collected by deploying a survey on Qualtrics¹ (a platform similar to Amazon Mechanical Turk), comprising several demographic questions (age, gender, race, education, and income) and the Beck’s Depression Inventory (Beck, Steer, and Brown 1996) to measure depression and Anxiety facet from the International Personality Item Pool proxy for the NEO Personality Inventory Revised (NEO-PI-R) (Costa and McCrae 2008). Users received an incentive for their participation, and we obtained their informed consent to access their Twitter posts. All users were based in the US. Out of 887 users who took the survey, 766 users posted 241,683 images. We excluded users who posted fewer than 20 images, resulting in 560 users with 145,436 posted images. We used continuous depression and anxiety scores (descriptive statistics in Table 2) and, for regression, we standardized the scores by converting them to have mean 0 and standard deviation 1. As the distribution of all psychological traits is approximately normal with basically no exceptions, we preferred not to threshold continuous distributions, to have the largest amount of useful variance available when considering covariances and correlations. In Figure 2, we visualize random images from users with top and bottom 25% percentile of both depression and anxiety scores. The distribution of scores are shown in Table 2.

TwitterText is a domain related dataset which has self-reported age and gender for all users. We used it to be able to control for and use demographics for multi-task modeling of mental health conditions. This dataset was used in previous works by (Preoțiuc-Pietro et al. 2017; Guntuku et al. 2017a). It is an order of magnitude larger data set consisting of 4132 Twitter users. Since we do not have depression and anxiety computed via surveys for this data

TwitterSurvey			TwitterText		
	Dep Score	Anx Score		Dep Score	Anx Score
Min	-1.323	-2.390	Min	-2.644	-3.052
25%	-0.758	-0.693	25%	-0.719	-0.729
50%	-0.194	0.155	50%	-0.084	-0.077
75%	0.629	0.791	75%	0.702	0.662
Max	3.851	1.852	Max	3.492	3.967

Table 2: Descriptive statistics of outcomes in both data sets. Scores are z-normalized

set, we use the language-prediction model from Facebook to impute depression and anxiety scores. We downloaded the 3200 most recent user tweets for each user, leading to a data set of 5,547,510 tweets, out of which 700,630 posts contained images and 1 profile image each across 3498 users. We exclude the texts associated with the tweets which contain images when predicting depression and anxiety to limit any potential confound. Then, for our analysis, we excluded users who posted less than 20 photos.

Feature Extraction

Since we are dealing with both posted images and profile pictures, we extract different sets of features to capture the representations associated with both. From posted images, we extract colors, aesthetics and image content related features, and from profile images, in addition to the three sets, we also extract facial-related features considering the literature linking facial expressions to emotions and mental states (Gur et al. 1992).

Colors

The colors of an image represent the most notable features to a human. Research has shown that colors can invoke emotions (Wexner 1954), psychological traits (Huang, Wang, and Wu 2006) or, on social media, reflect the personality of the person posting the image (Skowron et al. 2016) and even mental health states (Reece and Danforth 2017).

Colors can be expressed in various color spaces. We use the HSV (Hue–Saturation–Value), which provides a more intuitive representation of colors for humans (Bigun 2006). A pixel in the HSV space is characterized by three numbers: (1) *Hue*: the color type ranging between 0 and 360 degrees e.g., 0 is red, 60 is yellow; (2) *Saturation*: the intensity of the color ranging from 0 to 1 e.g., 0 represents no color and is a shade of gray; (3) *Value*: the brightness of the color ranging from 0 to 1 e.g., 0 represents black.

Using the images converted to the HSV space, we first determine whether an image is *grayscale*. A picture is considered grayscale if there is no pixel for which hue can be calculated with accuracy i.e., $V \in [0.15, 0.95]$, $S > 0.2$ (Ke, Tang, and Jing 2006). We ignore grayscale images (216 profile images, 4.99% of posted images) from subsequent color analysis only, as their presence may bias the results. We compute *saturation* and *brightness* as the average saturation and value respectively of the pixels in the image. An experimental study of colors established the relationship between saturation and brightness and the dimensional model of affect containing three factors: $Pleasure = .69 \cdot V + .22 \cdot S$, $Arousal = -.31 \cdot V + .60 \cdot S$, $Dominance = -.76 \cdot V + .32 \cdot S$ (Mehrabian and Russell 1974;

¹www.qualtrics.com/Survey-Software

Valdez and Mehrabian 1994). We compute the standard deviation of the HSV values, which we will use only for prediction.

Using the hues of an image, we compute the *hue count* of a photo as a measure of its simplicity (Ke, Tang, and Jing 2006). Professional photos usually have a lower number of unique hues, although each color may be rich in tones, leading to a simpler composition. To compute the hue count defined in (Ke, Tang, and Jing 2006), we obtain the 20-bin hue histogram from each pixel for which we can accurately compute its hue (see above) and compute the number of bins containing at least 5% of the pixels of the maximum bin size. As the hue count distribution is skewed towards a low number of hues, we log-scale this value for our analysis.

We also compute a 6-bin histogram which splits pixels into the primary and secondary colors and a 12-bin histogram to count the proportion of pixels from the primary, secondary and tertiary colors. Finally, we compute percentage of warm and cold color pixels using the hue values: *Warm*: $H \in [285, 75]$, *Cold*: $H \in [105, 255]$. A combination of contrast and hue count was used as a proxy for sharpness. After extracting features from individual posted images, we aggregate them to the users using mean pooling.

Aesthetics

Using pair-wise ranking of image pairs as well as the image attribute and content information, (Kong et al. 2016) propose to learn aesthetics. Two base networks of the Siamese architecture for each of the two images (in the pair) adopt the AlexNet configurations with the final fully connected layer removed. The base network is fine-tuned using aesthetics data with an Euclidean Loss regression layer followed by the Siamese network ranking the loss for every sampled image pairs. The fine-tuned network is used as a preliminary feature extractor. These features are then used for an attribute prediction task which is trained in a multi-task manner by combining the rating regression Euclidean loss, attribute classification loss and ranking loss. Finally, a content classification softmax layer is added to predict a predefined set of category labels. The categories are defined as: ‘balancing element’ - whether the image contains balanced elements; ‘content’ - whether the image has good/interesting content; ‘color harmony’ - whether the overall color of the image is harmonious; ‘object emphasis’ - whether the image emphasizes foreground objects; ‘rule of thirds’ - whether the photography follows rule of thirds; ‘vivid color’ - whether the photo has vivid color, not necessarily harmonious color; ‘repetition’ - whether the image has repetitive patterns; ‘symmetry’ - whether the photo has symmetric patterns; ‘depth of field’ - whether the image has shallow depth of field; ‘lighting’ - whether the image has good/interesting lighting; ‘motion blur’ - whether the image has motion blur. The various categories labeled are photographic attributes and image content information which help regularize the photo aesthetic score, which is modeled as a complex non-linear combination of each of the categories. The network also outputs an aesthetic score. After extracting features from individual posted images, we aggregate them to the users using mean pooling.

Image Content

Images posted in tweets can have diverse content beyond faces. Therefore, apart from color features, as used in prior works (Reece and Danforth 2017) and (Ferwerda, Schedl, and Tkalčič 2016)), we use automatic content analysis techniques to generate tags for these images. We labeled all images with the Imagga Tagging API² and generated for each image a *bag-of-tags* out of the top-10 predicted tags, following the developers’ recommendations. We removed all tags that occurred less than 200 times in our data set, leaving us with 1,299 distinct tags. Imagga was successfully used in previous research (Garimella, Alfayad, and Weber 2016b).

We found multiple tags which were very similar and usually co-occurred together (e.g., glass, drink, beverage; waves, shoreline, seascape, seaside). We therefore reduce the feature space by learning tag clusters that contain frequently co-occurring tags. This decreases sparsity and increases interpretability. We use a clustering procedure that was originally applied to words in tweets, which produced very accurate prediction results (Lampos et al. 2014). First, we compute the Normalised Pointwise Mutual Information (NPMI) between all pairs of tags (Bouma 2009). NPMI measures the degree to which two tags are likely co-occur in the same context (image) and takes a maximum value of 1 if two tags always co-occur and a value of 0 if they occur as according to chance. We use NPMI as a similarity measure; negative values are replaced with 0 as in (Lampos et al. 2014) and compute a tag \times tag similarity matrix. We feed this matrix to the spectral clustering algorithm, a hard-clustering method appropriate for generating non-convex clusters (Ng, Jordan, and Weiss 2002; Shi and Malik 2000; von Luxburg 2007), which performs a graph partitioning on the Laplacian of the similarity matrix. The number of clusters needs to be specified in advance. We use 400 clusters throughout this study, based on preliminary experiments.

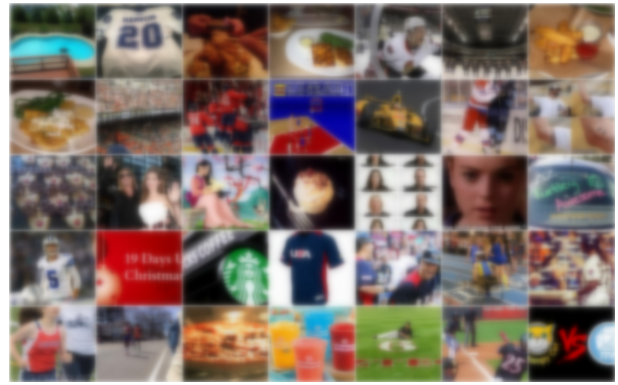
Once these clusters of semantically similar tags are created, we represent each image as a vector containing the normalized number of times each tag cluster is detected. For each user, we derive a feature vector of *image content topics* as the normalized number of times each topic cluster is present in the tweet-embedded images. We also calculated the percentage of image posts and percentage of posts with people (based on Imagga tags).

VGG-Net We use a pre-trained version of the 19-layer version of the VGG-Net image classifier based on convolutional neural networks (Simonyan and Zisserman). This classifier achieved the best results in the ImageNet Large Scale Visual Recognition Challenge 2014 in the object classification and localization challenge. It predicts a class probability for the 1,000 objects in the ImageNet tagset. We also extract the features from the last fully connected layer ($fc7$) for prediction. After extracting features from individual posted images, we aggregate them to the users using mean pooling.

²<http://docs.imagga.com/#auto-tagging>



(a) High level of depression and anxiety



(b) Low level of depression and anxiety

Figure 2: A random sampling of images posted by users with top and bottom 25% percentile of both depression and anxiety scores, blurred for privacy reasons.

Face-related

For profile photo analysis, we use the Face++³ and EmoVu⁴ APIs for facial feature extraction, as profile photos on Twitter usually contain one or more faces. Previous research has shown facial features can predict personality of users (Liu et al. 2016) and has used Face++ as a proxy for extracting user demographics from profile images (An and Weber 2016; Zagheni et al. 2014).

Facial Presentation & Expression This category contains facial features that try to capture the self-presentation characteristics of the user. Features include the face ratio (the size of the face divided by the size of the image), whether the face features any type of glasses (reading or sunglasses), the closeness of the subject’s face from the acquisition sensor provided by EmoVU’s attention measurement, the eye openness and the 3D face posture, which includes the pitch, roll and yaw angle of the face.

For facial emotion detection, we use Ekman’s model of six discrete basic emotions: anger, disgust, fear, joy, sadness and surprise, which were originally identified based on facial expressions (Ekman and Friesen 1971). We use the EmoVU API to automatically extract these emotions, as well as neutral expression (Batty and Taylor 2003), from the largest detected face in each profile image. The six basic emotions can be categorized as either positive (joy and surprise) or negative (anger, disgust, fear, sadness). Along with the basic emotional expressions, EmoVU also gives composite features: Expressiveness is the highest value of the six basic emotions; Negative and positive mood are calculated as the maximum value of the positive and negative emotions respectively; valence is the average of the negative mood and positive mood. Also, we add the smiling degree provided by Face++.

Analysis

We perform univariate correlation tests between each feature and mental health condition to uncover the associated image

³<http://faceplusplus.com/>

⁴<http://emovu.com/>

features. We control for age and gender using partial correlation so that our analysis is not skewed by any potential demographic bias in the data. Additionally, in content analysis, as depression and anxiety are highly inter-correlated ($r=.67$), we control for the other trait in order to isolate the unique attributes of each dimension. We adjust for multiple comparisons using Benjamini-Hochberg multi-test correction.

Results for our analysis on posted images using image colors and aesthetic features on TwitterText data set are presented in Tables 3a, using 50 Imagga topics in Table 4, and analysis on profile images are presented in 3b. The same set of experiments on mental health outcomes was conducted on the TwitterSurvey data set, but were no longer significant when controlling for multiple comparisons. This shows the need for these behaviors to be studied using larger samples. However, we validated the language prediction model to predict scores on TwitterSurvey, which yielded a correlation of .12 and .20 with depression and anxiety respectively. Compared to other psychological outcomes such as personality, these correlations are consistent e.g. (Segalin et al. 2017; Reece and Danforth 2017; Jaidka et al. 2018).

Posted Images

In Figure 2, we visualize random images from users with top and bottom 25% percentile of both depression and anxiety scores. Users scoring high in both depression and anxiety have a preference to post images which are grayscale, less intense in colors (low saturation) and in the diversity of colors (low hue count), low in arousal and low on all attributes of aesthetics. Further, users high in depression post images which emphasize foreground objects and are low in brightness. Content analysis shows that users high in both traits are posting images containing text (*3d, alphabet..*), images of animals (*animals, cat..*), while users low in depression and anxiety post images of sports (*action, audience..*), nature (*artistic, astrology*), every day things (*apartment, bed..*), meals (*cuisine, delicious..*), motor vehicles (*automobile, car..*), outdoor activities (*bicycle, bike..*). We note that the associations are stronger for most topics for depression, indicating this outcome is more strongly associated with image

Feature	Demographics		MentalHealth	
	Gender	Age	Dep	Anx
Colors				
Grayscale	.157	-.237	.301	.258
Brightness	-.113	-.099	-.109	
Saturation	-.167	.317	-.178	-.112
Pleasure	.216	-.103	.093	.060
Arousal	-.170	.169	-.155	-.097
Dominance	-.154	.052	.052	.011
Hue Count	-.151	.057	-.229	-.127
Warm Colors	.156	-.091		.083
Aesthetics	Gender	Age	Dep	Anx
Color Harmony		.201	-.151	-.135
Motion Blur	-.153	.260	-.090	-.142
Lighting	-.062	.236	-.212	-.151
Content		.176	-.109	-.068
Repetition	-.121	.157	-.096	-.087
Depth of Field	-.067	.154	-.102	-.061
Vivid Color	-.162	.283	-.275	-.189
Symmetry	.120	.282	-.274	-.189
Object Emphasis	.136		.093	
Balancing Element	.039	.089		
Overall Aesth. Score	-.067	.239	-.155	-.124
Meta	Gender	Age	Dep	Anx
% Image posts	.049	-.136	.128	.276
% Posts with People		-.195		.066

(a) Posted Images

Feature	Demographics		MentalHealth	
	Gender	Age	Dep	Anx
Colors				
Brightness			.043	.045
Contrast	.033	-.08		-.011
Hue				.044
Image Composition	Gender	Age	Dep	Anx
Hue Count				.045
Visual Weight			.043	.045
Image Type	Gender	Age	Dep	Anx
One Face	-.532	-.300	.142	.077
Num. of Faces	.165	-.242	-.121	
Facial Presentation	Gender	Age	Dep	Anx
Reading Glasses	.091	-.581	.064	.064
Sunglasses		-.758	.068	.071
Pitch Angle	.274	-.518		
Yaw Angle				.052
Face Ratio			.136	.083
Facial Expressions	Gender	Age	Dep	Anx
Smiling			-.185	.074
Anger	-.258	.349	.082	
Disgust	-.431	-.137		
Fear	-.281	.148		
Joy	.135	-.066	-.149	
Neutral		.237	.149	.064
Expressiveness		-.082	-.143	
Negative Mood	-.166	-.078		
Positive Mood	.126		-.136	

(b) Profile Images

Table 3: Pearson correlations between color and aesthetic features extracted from posted and profile images and mental health conditions, and with age and gender (coded as 1 for female, 0 for male) separately and used age as a continuous variable. Correlations for mental health conditions are controlled for age, gender and other mental health condition. Positive correlations are highlighted with green ($p < .01$, two-tailed t-test) and negative correlations with red ($p < .01$, two-tailed t-test). Correlations which are not significant are not presented.

posting preferences.

Apart from these themes, anxious users specifically are seen to post portraits and selfies (tagged as *adult*, *attractive*..), corporate environments (*businessman*, *corporate*..), and depressed users are seen to post more animal related images (*ape*, *aquatic_mammal*..). Whereas people low in depression score post vacation related images (*beach*, *coast*..).

Profile Images

Choice of profile pictures uncovers more insight into the behavior of users with mental health conditions; while depressed users preferred images which are not sharp and which do not show face, anxious users usually chose sharper images with multiple people in them.

Prediction

Finally, we investigate the accuracy of visual features in predicting the mental health conditions. We use linear regression with ElasticNet regularization (Zou and Hastie 2005) as our prediction algorithm and report results on 10 fold cross-validation (sampled such that users in the training set are not present in the test set). Performance is measured using Pearson correlation (and Mean Squared Error in brackets) across the 10 folds.

We also explore whether incorporating author attributes such as age and gender in the joint modeling of depression and anxiety can improve prediction performance. We achieve this by introducing a different set of regularization constraints in the ElasticNet optimization function, a method known as $l2/1$ norm regularization (Liu, Ji, and Ye 2009) (denoted as MT in the results). In single-task learning, we used image features as predictors and each mental health variable as outcome in a separate regression with ElasticNet regularization. For multi-task learning, we used both demographics (age and gender) and mental health outcomes (depression or anxiety) in the same regression using L1/L2 mixed-norm as regularizer to exploit the underlying implicit correlations in mental health and demographic factors. The performance on the TwitterText data set is shown in Table 5 and on the TwitterSurvey data set is shown in Table 6. For single-task (ST), we tried support vector regression and L1 and L2 regularization for linear regression and found no significant difference in results. To be consistent in comparing the same methods in ST and MT, we used ElasticNet regularization and linear regression for both. Combination of several feature sets is also tested and termed as Combination in the Tables.

We find that multi-task learning offers significant gains over single task learning for posted images (see Table 5a). While VGG penultimate features do not perform as well as others at predicting depression, multi-task learning boosts the performance by almost 80%. MT shows an advantage for both depression ($r = .619$ for MT vs $.554$ for ST) and anxiety ($r = .580$ vs $.532$) in Table 5a. While colors and aesthetics do not see a big jump from single-task to multi-task learning, deep learning based methods see a drastic improvement. However, this is not the case with profile images; multi-task learning shows very similar performance as single-task learning.

On profile images, image-based demographic predictions

r_{dep}	r_{anx}	Image Tag Clusters	r_{dep}	r_{anx}	Image Tag Clusters
(More depressed and anxious)			(Less depressed and anxious)		
.196	.163	3d, alphabet, book, capital, document, font, pen, text, typescript	.243	.151	action, active, aerobics, athlete, balance_beam, ball, ballplayer, barbell, baseball
.120	.124	affenpinscher, american_staffordshire_terrier, and, ap-penzeller, arctic_fox, australian_terrier	.233	.129	audience, championship, cheering, competition, crowd, event, flag, lights, match
—	.122	adult, attractive, beard, brunette, caucasian, face, fashion, glamour, gorgeous	.057	.114	artistic, astrology, astronomy, atmosphere, azure, buildings, business_district
.147	.109	animals, cat, domestic_cat, egyptian_cat, eye, feline, fur	.134	.100	apartment, bed, bedclothes, bedroom, bedroom_furniture, blanket
.141	.096	angora, badger, bunny, cavy, easter, footed_ferret, fox_squirrel, guinea_pig	.096	.083	asphalt, avenue, broom, cleaning_implement, express-way, highway
—	.062	businessman, businesspeople, confident, corporate, executive, handsome, manager	—	.076	button, buttons, circle, design, glossy, graphic, icon, icons, internet
—	.059	baby, boy, boyfriend, brother, buddy, child, childhood, children, couple	.053	.073	alp, alpine, alps, autumn, canyon, cascade, cold, creek, crystal
.110	.057	african_chameleon, african_crocodile, agama, agamid, alligator, american_alligator	.079	.064	aroma, caffeine, cappuccino, china, coffee, coffee_mug, crockery, cup
.063	—	bank, banking, bill, book_jacket, cash, comic_book, currency, finance, financial	.057	.056	cuisine, delicious, dinner, dish, gourmet, lunch, meal, meat, plate, restaurant, tasty, vegetable
.052	—	aquatic_mammal, dugong, eared_seal, electric_ray, great_white_shark, hammerhead, ray	.064	.052	automobile, car, convertible, drive, motor, motor_vehicle, speed, sports_car
.050	—	ape, capuchin, chimpanzee, gorilla, macaque, marmoset, monkey, orangutan, primate	.054	.048	bicycle, bike, built, carriage, cart, horse_cart, mountain_bike, minibike, moped
			.052	.051	bagel, baked, baked_goods, bakery, bread, breakfast, bun, burger, cheeseburger, cracker
			.061	—	bay, beach, coast, coastline, fence, idyllic, island, obstruction, ocean, palm, paradise, relaxation, resort

Table 4: Pearson correlations between Image tag clusters extracted from posted pictures and mental health conditions. All correlations are significant at $p < .05$, two-tailed t-test, Benjamini-Hochberg corrected. Results for depression and anxiety are controlled for age and gender. Tags are sorted by occurrence in our data set within a cluster. ‘—’ indicates that the tag cluster is not significantly associated with Depression (dep) and/or Anxiety (anx).

outperform other features, due to the dependence of mental health conditions on age, gender and race, followed by facial expressions and the aesthetic attributed of the profile image. Predicting anxiety using profile images is more accurate than predicting depression, suggesting that a single profile image might be able to provide more information about anxiety. However, a set of posted images are required to predict depression with a reasonable accuracy.

We then examine if using text-predicted outcomes as proxies for survey labels can improve performance in predicting more reliable survey based outcomes (not shown in any Table). We evaluated the performance of models trained on TwitterText when tested on TwitterSurvey, and found that the performance ($r = .164$ using aesthetics) is similar to the best performing model trained and tested on TwitterSurvey ($r = .167$) for depression, but however outperforms the corresponding model for anxiety ($r = .223$ when transfer learning from TwitterText to TwitterSurvey vs. $r = .175$ when trained and tested on TwitterSurvey). This shows that text-predicted labels can be used as proxies in analyzing image-posting behavior when studying mental health.

Discussion

In this paper, we explored how depression and anxiety traits can be automatically inferred by just looking at images that users post and set as profile pictures. We compared five different visual feature sets (extracted from posted images and profile pictures) and the findings about image features associated with mental illness in large part confirm previous findings about the manifestations about depression and anxiety.

Feature set	# Feat	Depression		Anxiety	
		ST	MT	ST	MT
Colors	44	.446 (.811)	.449 (.802)	.441 (.815)	.446 (.803)
Aesthetics	10	.434 (.818)	.434 (.810)	.380 (.866)	.377 (.857)
Imagga	500	.443 (.836)	.509 (.742)	.426 (.837)	.483 (.766)
VGG Penultimate (fc7)	4096	.343 (1.022)	.603 (.644)	.351 (.890)	.555 (.693)
VGG Classes	1000	.438 (.821)	.520 (.731)	.442 (.811)	.505 (.747)
Combination	5	.554 (.689)	.619 (.613)	.532 (.715)	.580 (.661)

(a) Posted Images

Feature set	# Feat	Depression		Anxiety	
		ST	MT	ST	MT
Colors	44	.084 (.997)	.101 (.989)	.133 (.986)	.146 (.978)
Image Composition	10	.038 (1.006)	.046 (.997)	.056 (1.001)	.059 (.996)
Image Type	5	.069 (1.001)	.070 (.995)	.046 (1.002)	.049 (.997)
Image Demographics	5	.254 (.939)	.255 (.922)	.403 (.839)	.403 (.846)
Facial Presentation	7	.056 (1.001)	.056 (.984)	.059 (1.000)	.061 (1.007)
Facial Expressions	14	.170 (.976)	.174 (.938)	.133 (.989)	.137 (.979)
Aesthetics	12	.111 (.993)	.114 (.986)	.100 (.994)	.108 (.988)
Combination	7	.305 (.880)	.311 (.873)	.427 (.815)	.429 (.813)

(b) Profile Images

Table 5: **TwitterText**: Prediction results for mental health conditions with all features. Performance is measured using Pearson correlation (and MSE in parenthesis) in 10-fold cross-validation. **ST** represents Single-Task learning and **MT** represents Multi-Task learning

ety.

Regarding depression, perhaps with the highest face validity, in profile images of depressed users the facial expressions show fewer signs of positive mood (less joy and smiling), and are appear more neutral and less expressive. Previous findings suggest that social desirability and self-presentation biases generally discourage the sharing of negative information about the self (such as negative emotion in a society that values positivity), instead, negative emotion is often mani-

Feature set	# Feat	Depression		Anxiety	
		ST	MT	ST	MT
Colors	44	.070 (1.06)	.113 (.996)	.120 (1.03)	.086 (1.02)
Aesthetics	10	.065 (1.07)	.068 (1.01)	.038 (1.00)	.130 (.998)
Imagga	500	.157 (1.06)	.160 (1.00)	.104 (1.23)	.128 (1.01)
VGG Penultimate (fc7)	4096	.062 (1.06)	.143 (.999)	.118 (1.16)	.126 (1.02)
VGG Classes	1000	.036 (1.11)	.045 (.999)	.140 (1.01)	.148 (1.00)
Combination	5	.151 (1.01)	.167 (.975)	.167 (1.01)	.175 (1.00)

Table 6: **TwitterSurvey**: Prediction results for mental health conditions with all features on posted images. Performance is measured using Pearson correlation (and MSE in parenthesis) in 10-fold cross-validation. **ST** represents Single-Task learning and **MT** represents Multi-Task learning

festated as a lack of expression of positive emotion (Dibble and Levine 2013), as evident in this study. Additionally, depressed individuals profile pictures are marked by the fact that they are more likely to contain a single face (that of the user), rather than show the user surrounded by friends. In fact, focus on the self is one of the most consistent markers of negative emotionality and depression in particular (Tackman et al. 2018); in language studies of depression, the use of first personal singular pronouns has emerged as one of the most dependable markers of depression (Edwards and Nicholas S 2017) the single user profile picture may very well be its analogue in the world of image posts.

In posted images, depressed individuals post pictures that are generally less aesthetically pleasing, extending across features such as color harmony, lighting, content, repetition, depth of field, vividness of color and symmetry in the pictures. These findings suggest perceptual symptoms of depression, leading to picture choices that lack coherence and traditional markers of beauty – likely because such lack of order is experienced to be more representative of the psychological reality of depression, which is often without orientation, order and appreciation of beauty. Finally, the nominally highest correlation among all image features associated with depression is observed in the increased use of grayscale images, which may again be the visual expression of the undifferentiated, monotone emotional experience that marks depression. The automatically extracted image features that track the emotional circumplex (pleasure and arousal) fail to fully capture the depressed experience, as depressed posted images show less arousal (very much as expected in depression), but nominally more pleasure.

In terms of Imagga tag clusters of posted images, depressed users choose images with more text content and animals, neither of which is an obvious marker of depression. Again depression is more strongly manifested in what posted image content does *not* cover, which includes a variety of markers of the psychologically well-adjusted life: sports (specifically team sports), recreational activities that include outdoor activities, driving, bicycles and beaches and vacation, in addition few references to food and beverages.

Regarding anxiety, the facial expression in profile picture do not show fewer positive emotions, in fact, nominally more smiling is observed. Frequently, anxiety includes elements of social anxiety, which may interact with the social desirability biases inherent in social media sharing to generate a facade of well-being anxious people are frequently referred to as “the worried well;” their profile images show a much

less clearer set of associations that could distinguish them from the norm (all correlations $< .1$). Similar to depressed users, anxious users images are more grayscale and lack aesthetic cohesion across a variety of image features here too suggesting a lack of responsiveness to traditional markers of beauty and emotional balance. Again, the emotional circumplex image features (pleasure and arousal) fail to fully capture the anxious experience, as anxious individuals are generally understood to experience more (not less) arousal, and less (not more) pleasure.

In terms of Imagga tag clusters of posted images, the profile in large part overlaps with that of depression with generally lower effect sizes compared to normative users, suggesting that anxiety is not as observable in the image choices as depression. Anxious users additionally seem to post marginally more content related to family and work; aspects of social anxiety may here manifest as the wish to signal fulfillment of typical role obligations.

We observe that multi-task learning improves the performance of mental health assessment confirming the need to include the personal characteristics of users while building automatic models (Degens et al. 2017). Even though profile images do not offer high performance in terms of predicting mental health conditions, they offer insights as described above. Combined with the findings from posted images, such cues could be used in improving the automatic assessment of mental health conditions.

A recent meta-analysis of predicting Big-Five personality from digital traces (Settanni, Azucar, and Marengo 2018) shows that the Big-Five personality dimensions can be predicted from social media traces at accuracies ranging from $r = .29$ to $.40$ (generally in cross-validation frameworks on the same data), which roughly match the highest correlations that are generally found between psychological traits and observable behaviors in the large psychological literature ($r \approx .3$). Depression is more nuanced and a less broad characteristic than any of the Big-Five personality traits, therefore, observing a cross-validated prediction performance of $r = .32$ suggests very decent model performance. When the same Facebook model is applied to the TwitterSurvey dataset, the observed correlations with the TwitterSurvey survey estimates are lower ($r = .12$ and $.20$ for depression and anxiety, respectively). This is likely due to the fact that Twitter and Facebook have differences in language, both in terms of vocabulary (e.g. emoticons) and subject matter use (Jaidka et al. 2018; Zhong et al. 2017; Guntuku et al. 2019). Importantly, these highly significant correlations ($p < .01$) nevertheless demonstrate that the Facebook prediction models encode significant mental health information that can be used to estimate the mental health status of Twitter users.

Limitations and Future Work

Our work has some limitations: first, though we show that the weak text-labeled mental health scores on TwitterText are reliable, by testing a model trained on TwitterText on TwitterSurvey, where ground truth was collected using surveys, further work on creating large-scale datasets is required to uncover other dimensions of depressed and anxious behavior using stronger measures. Further, we use only one

profile picture per user in our analysis. It would be a promising research direction to conduct a study with experience sampling where images across time are collected along with users' response to questionnaires to study the temporal orientation of mental health conditions and how their social media imagery changes as a consequence, potentially using smartphone based data collection (Singh and Agarwal 2016). Further, we consider only the image-based features due to the initial premise, which we've seen to be true in our results, that visual imagery can uncover several strong signals for mental health conditions. Users' social network features, coupled with image posting behavior and also text, could further enhance the understanding about the manifestation of depression and anxiety (Hong et al. 2017).

Another direction of work could assess if predictive/automatic assessments of mental health could outperform third-party human assessments, as has been seen in other works (Segalin et al. 2017). Also, we did not collect the date of first onset of depression from the users; this data could be used to test if social-media based machine assessments could be used as an early screening tools to help clinicians in identifying at patients at risk of mental health disorders, which usually manifest early on platforms like social media (De Choudhury, Counts, and Horvitz 2013). Finally, we also do not consider predicting depression and anxiety as a classification task as it is not easy to binarise mental health conditions using median or quartile-split (Beck, Steer, and Brown 1996) and a continuous score can often be helpful in personalizing the diagnosis and treatment for patients, and also serve as a risk assessment tool for providers.

As a benchmark, the diagnosis and differentiation of mental health conditions is difficult and costly. The gold standard of clinical psychological diagnosis used in the therapy effectiveness literature are Structured Clinical Interviews for the Diagnostic and Statistical Manual of Mental Disorders (SCIDs). They involve a trained clinicians interviewing the patient in person for at least about an hour, using a guide that serves as a decision tree. Passive approaches that analyze digital traces left online are at this point substantially less powerful in their ability to detect the presence of mental health conditions, in part because they are unobtrusive (unlike validated clinical surveys like the PHQ-9, which have 80%+ sensitivity and 90%+ specificity compared to the SCID baselines (Gilbody et al. 2007)). In terms of improving the unobtrusive detection of mental health conditions, the next step for the field appears to be multi-modal approaches that combine the relative power of different feature spaces. Specifically, text-based analyses of social media ought be combined with those of text messages and other instant messaging communication, and with analyses of image content being shared, as in this study. Importantly, the more domains of communication are covered outside of social media, the more such data feeds can cover lack of information being shared on social media, as may occur during sensor depression. In addition, phone or wearable data can give estimates as to the users activity levels and sleeping hours (which are closely related to diagnosis criteria for depression). Further, the results in this study do not imply causality, but are correlational.

Our paper studies the significance of using visual traces on

social media to glean insight and develop predictive models to automatically assess users' mental health conditions. While there are several flip sides of such technologies when used for incorrect motives such as pre-screening by insurance or employers to discriminate against specific individuals, this research is very useful to develop techniques of providing summarized feedback both to social media users and their clinicians about social media activity to help them quantify the extent to which their mental health condition has shown its footprint in the digital world; we hypothesize that such a consent-based sharing can help in more productive therapy sessions, apart from the opportunity that the user has to self-regulate. Thus, data collection, processing and dissemination of resulting models has to keep the privacy of users as top priority and should be for discovering new insights into the manifestation of mental health conditions and to better assist clinicians.

Conclusion

We analyzed image posting and profile picture preferences using interpretable Imagga tag clusters, colors, aesthetic and facial features with the aim of identifying the way and extent to which they reveal users' mental health conditions based on the images they post and select as profile pictures on Twitter. For example, images posted by depressed and anxious users tend to be dominated by grayscale, low arousal images lacking in aesthetic sense, while anxious users are characterized by posting more images compared to regular users and users with depression. The choice of profile pictures uncovers that depressed users prefer images which are not sharp and which do not contain a face, anxious users usually chose sharper images with multiple faces in them. Results indicate that multi-task learning gives significant improvements in performance for modelling mental health conditions jointly with demographics (here age and gender), factors which clinicians usually consider while diagnosing patients. Further, models trained on larger data sets using text-predicted outcomes show reliable performance when predicting more reliable survey based mental health outcomes. Our results offer new insights and a promising avenue for future mental health research of individuals.

References

- Al Moubayed, N.; Vazquez-Alvarez, Y.; McKay, A.; and Vinciarelli, A. 2014. Face-Based Automatic Personality Perception. MM.
- An, J., and Weber, I. 2016. # greysanatomy vs.# yankees: Demographics and Hashtag Use on Twitter. ICWSM.
- Andalibi, N.; Ozturk, P.; and Forte, A. 2015. Depression-related imagery on instagram. In *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative work & social computing*, 231–234. ACM.
- Batty, M., and Taylor, M. J. 2003. Early Processing of the Six Basic Facial Emotional Expressions. *Cognitive Brain Research* 17(3).
- Beck, A. T.; Steer, R. A.; and Brown, G. K. 1996. Beck depression inventory-ii. *San Antonio* 78(2):490–8.

- Bigun, J. 2006. *Vision with Direction: A Systematic Introduction to Image Processing and Vision*.
- Bijl, R. V.; Ravelli, A.; and Van Zessen, G. 1998. Prevalence of psychiatric disorder in the general population: results of the netherlands mental health survey and incidence study (nemesis). *Social psychiatry and psychiatric epidemiology* 33(12):587–595.
- Bouma, G. 2009. Normalized (Pointwise) Mutual Information in Collocation Extraction. GSCL.
- Burdick, A.; Drucker, J.; Lunenfeld, P.; Presner, T.; and Schnapp, J. 2012. *Digital Humanities*.
- Celli, F.; Bruni, E.; and Lepri, B. 2014. Automatic Personality and Interaction Style Recognition from Facebook Profile Pictures. MM.
- Chancellor, S.; Kalantidis, Y.; Pater, J. A.; De Choudhury, M.; and Shamma, D. A. 2017. Multimodal classification of moderated online pro-eating disorder content. In *CHI*.
- Costa, P., and McCrae, R. 2008. *The Revised NEO Personality Inventory (NEO-PI-R)*. SAGE Publications Ltd.
- Datta, R.; Joshi, D.; Li, J.; and Wang, J. Z. 2006. Studying Aesthetics in Photographic Images Using a Computational Approach. ECCV.
- De Choudhury, M.; Counts, S.; and Horvitz, E. 2013. Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference, WebScience*, 47–56.
- Degens, N.; Endrass, B.; Hofstede, G. J.; Beulens, A.; and André, E. 2017. what i see is not what you get: why culture-specific behaviours for virtual characters should be user-tested across cultures. *AI & society* 32(1):37–49.
- Dibble, J. L., and Levine, T. R. 2013. Sharing good and bad news with friends and strangers: Reasons for and communication behaviors associated with the mum effect. *Communication Studies* 64(4):431–452.
- Edwards, T., and Nicholas S, H. 2017. A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality* 68:63–68.
- Ekman, P., and Friesen, W. V. 1971. Constants across Cultures in the Face and Emotion. *Journal of Personality and Social Psychology* 17(2).
- Farnadi, G.; Tang, J.; De Cock, M.; and Moens, M.-F. 2018. User profiling through deep multimodal fusion. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*. ACM.
- Ferwerda, B., and Tkalcic, M. 2018. You are what you post: What the content of instagram pictures tells about users personality. In *The 23rd International on Intelligent User Interfaces*.
- Ferwerda, B.; Schedl, M.; and Tkalčič, M. 2016. Using Instagram Picture Features to Predict Users' Personality. MMM.
- Garimella, V. R. K.; Alfayad, A.; and Weber, I. 2016a. Social Media Image Analysis for Public Health. CHI.
- Garimella, V. R. K.; Alfayad, A.; and Weber, I. 2016b. Social Media Image Analysis for Public Health. CHI.
- Gilbody, S.; Richards, D.; Brealey, S.; and Hewitt, C. 2007. Screening for depression in medical settings with the patient health questionnaire (phq): a diagnostic meta-analysis. *Journal of general internal medicine* 22(11):1596–1602.
- Guntuku, S. C.; Qiu, L.; Roy, S.; Lin, W.; and Jakhetiya, V. 2015. Do others perceive you as you want them to?: Modeling personality based on selfies. In *Proceedings of the 1st International Workshop on Affect & Sentiment in Multimedia*, 21–26. ACM.
- Guntuku, S. C.; Zhou, J. T.; Roy, S.; Weisi, L.; and Tsang, I. W. 2016. Who likes what, and why? insights into personality modeling based on imagelikes'. *IEEE Transactions on Affective Computing*.
- Guntuku, S. C.; Lin, W.; Carpenter, J.; Ng, W. K.; Ungar, L. H.; and Preoțiu-Pietro, D. 2017a. Studying personality through the content of posted and liked images on twitter. In *Proceedings of the 2017 ACM on web science conference*, 223–227. ACM.
- Guntuku, S. C.; Yaden, D. B.; Kern, M. L.; Ungar, L. H.; and Eichstaedt, J. C. 2017b. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences* 18:43–49.
- Guntuku, S. C.; Buffone, A.; Jaidka, K.; Eichstaedt, J.; and Ungar, L. 2019. Understanding and measuring psychological stress using social media.
- Gur, R. C.; Erwin, R. J.; Gur, R. E.; Zwi, A. S.; Heimberg, C.; and Kraemer, H. C. 1992. Facial emotion discrimination: Ii. behavioral findings in depression. *Psychiatry research* 42(3):241–251.
- Hong, R.; He, C.; Ge, Y.; Wang, M.; and Wu, X. 2017. User vitality ranking and prediction in social networking services: A dynamic network perspective. *IEEE Transactions on Knowledge and Data Engineering* 29(6):1343–1356.
- Huang, K.-Q.; Wang, Q.; and Wu, Z.-Y. 2006. Natural Color Image Enhancement and Evaluation Algorithm Based on Human Visual System. *Computer Vision and Image Understanding* 103(1).
- Jaidka, K.; Guntuku, S. C.; Buffone, A.; Schwartz, H. A.; and Ungar, L. 2018. Facebook vs. twitter: Differences in self-disclosure and trait prediction. In *Proceedings of the International AAAI Conference on Web and Social Media*.
- Ke, Y.; Tang, X.; and Jing, F. 2006. The Design of High-level Features for Photo Quality Assessment. CVPR.
- Kong, S.; Shen, X.; Lin, Z.; Mech, R.; and Fowlkes, C. 2016. Photo aesthetics ranking network with attributes and content adaptation. In *ECCV*.
- Lamos, V.; Aletras, N.; Preoțiu-Pietro, D.; and Cohn, T. 2014. Predicting and Characterising User Impact on Twitter. EACL.
- Liu, L.; Preoțiu-Pietro, D.; Riahi, Z.; Moghaddam, M. E.; and Ungar, L. 2016. Analyzing Personality through Social Media Profile Picture Choice. ICWSM.
- Liu, J.; Ji, S.; and Ye, J. 2009. Multi-task feature learning via efficient l_{2,1}-norm minimization. UAI, 339–348.
- Machajdik, J., and Hanbury, A. 2010. Affective Image

- Classification Using Features Inspired by Psychology and Art Theory. *MM*.
- Manikonda, L., and De Choudhury, M. 2017. Modeling and understanding visual attributes of mental health disclosures in social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI, 170–181.
- McCrae, R. R., and John, O. P. 1992. An Introduction to the Five-factor Model and its Applications. *Journal of Personality* 60.
- Mehrabian, A., and Russell, J. A. 1974. *An Approach to Environmental Psychology*. MIT Press.
- Meyer, G. J.; Finn, S. E.; Eyde, L. D.; Kay, G. G.; Moreland, K. L.; Dies, R. R.; Eisman, E. J.; Kubiszyn, T. W.; and Reed, G. M. 2001. Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist* 56(2):128.
- Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2002. On Spectral Clustering: Analysis and an Algorithm. *NIPS*.
- Nie, J.; Cui, P.; Yan, Y.; Huang, L.; Li, Z.; and Wei, Z. 2014. How your Portrait Impresses People?: Inferring Personality Impressions from Portrait Contents. In *Proceedings of the 22nd ACM International Conference on Multimedia*, MM, 905–908.
- Nie, L.; Zhang, L.; Yang, Y.; Wang, M.; Hong, R.; and Chua, T.-S. 2015. Beyond doctors: Future health prediction from multimedia and multimodal observations. In *Proceedings of the 23rd ACM international conference on Multimedia*, 591–600. ACM.
- Nie, J.; Wei, Z.; Li, Z.; Yan, Y.; and Huang, L. 2018. Understanding personality of portrait by social embedding visual features. *Multimedia Tools and Applications* 1–20.
- Preoțiu-Pietro, D.; Liu, Y.; Hopkins, D.; and Ungar, L. 2017. Beyond binary labels: political ideology prediction of twitter users. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, 729–740.
- Reece, A. G., and Danforth, C. M. 2017. Instagram photos reveal predictive markers of depression. *EPJ Data Science*.
- Samani, Z. R.; Guntuku, S. C.; Moghaddam, M. E.; Preoțiu-Pietro, D.; and Ungar, L. H. 2018. Cross-platform and cross-interaction study of user personality based on images on twitter and flickr. *PLoS one* 13(7):e0198660.
- Schwartz, H. A.; Eichstaedt, J.; Kern, M. L.; Park, G.; Sap, M.; Stillwell, D.; Kosinski, M.; and Ungar, L. 2014. Towards assessing changes in degree of depression through facebook. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 118–125.
- Segalin, C.; Celli, F.; Polonio, L.; Kosinski, M.; Stillwell, D.; Sebe, N.; Cristani, M.; and Lepri, B. 2017. What your facebook profile picture reveals about your personality. In *Proceedings of the 2017 ACM on Multimedia Conference*, 460–468. ACM.
- Settanni, M.; Azucar, D.; and Marengo, D. 2018. Predicting individual characteristics from digital traces on social media: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking* 21(4):217–228.
- Shi, J., and Malik, J. 2000. Normalized Cuts and Image Segmentation. *Transactions on Pattern Analysis and Machine Intelligence* 22(8).
- Simonyan, K., and Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*.
- Singh, V. K., and Agarwal, R. R. 2016. Cooperative phoneotypes: exploring phone-based behavioral markers of cooperation. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 646–657. ACM.
- Skowron, M.; Tkalčič, M.; Ferwerda, B.; and Schedl, M. 2016. Fusing Social Media Cues: Personality Prediction from Twitter and Instagram. *WWW Companion*.
- Suchman, E. A. 1962. An analysis of ‘bias’ in survey research. *Public Opinion Quarterly* 102–111.
- Tackman, A. M.; Sbarra, D. A.; Carey, A. L.; Donnellan, M. B.; Horn, A. B.; Holtzman, N. S.; Edwards, T. S.; Pennebaker, J. W.; and Mehl, M. R. 2018. Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of personality and social psychology*.
- Valdez, P., and Mehrabian, A. 1994. Effects of Color on Emotions. *Journal of Experimental Psychology* 123(4).
- von Luxburg, U. 2007. A Tutorial on Spectral Clustering. *Statistics and Computing* 17(4).
- Wang, P. S.; Lane, M.; Olfson, M.; Pincus, H. A.; Wells, K. B.; and Kessler, R. C. 2005. Twelve-month use of mental health services in the united states: results from the national comorbidity survey replication. *Archives of general psychiatry* 62(6):629–640.
- Wang, K.; Zhai, S.; Cheng, H.; Liang, X.; and Lin, L. 2016. Human pose estimation from depth images via inference embedded multi-task learning. In *Proceedings of the 2016 ACM on Multimedia Conference*, 1227–1236. ACM.
- Wexner, L. B. 1954. The Degree to which Colors (hues) are Associated with Mood-tones. *Journal of Applied Psychology* 38(6).
- Zagheni, E.; Garimella, V. R. K.; Weber, I.; and State, B. 2014. Inferring International and Internal Migration Patterns from Twitter Data. *WWW Companion*.
- Zhong, C.; Chan, H.-w.; Karamshuk, D.; Lee, D.; and Sasstry, N. 2017. Wearing many (social) hats: How different are your different social network personae? *arXiv preprint arXiv:1703.04791*.
- Zou, H., and Hastie, T. 2005. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B*.