Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 6292–6299
Marseille, 11–16 May 2020

Inferring Social Media Users' Mental Health Status from Multimodal Information

Zhentao Xu, Verónica Pérez-Rosas, Rada Mihalcea
University of Michigan
Ann Arbor MI, USA
{frankxu,vrncapr,mihalcea}@umich.edu

Abstract

Worldwide, an increasing number of people are suffering from mental health disorders such as depression and anxiety. In the United States alone, one in every four adults suffers from a mental health condition, which makes mental health a pressing concern. In this paper, we explore the use of multimodal cues present in social media posts to predict users' mental health status. Specifically, we focus on identifying social media activity that either indicates a mental health condition or its onset. We collect posts from Flickr and apply a multimodal approach that consists of jointly analyzing language, visual, and metadata cues and their relation to mental health. We conduct several classification experiments aiming to discriminate between (1) healthy users and users affected by a mental health illness; and (2) healthy users and users prone to mental illness. Our experimental results indicate that using multiple modalities can improve the performance of this classification task as compared to the use of one modality at a time, and can provide important cues into a user's mental status.

Keywords: mental health, social media, multimodal analysis, machine learning

1. Introduction

Mental health is a pressing health concern worldwide. According to WHO Composite International Diagnostic Interview (WMH-CIDI), in the United States alone, around one in every four people has experienced a mental illness at some point in their lives (Demyttenaere et al., 2004).

Studies addressing mental health in the clinical domain have largely relied on patient's self-reported experiences, behaviors reported by relatives or friends, and a mental status examination via psychology testing. Recently, several studies have highlighted the potential to learn about people's health and well-being through linguistic and behavioral analyses of their social media activity (Burke et al., 2010; Moorhead et al., 2013). The use of social media as a data source provides a less intrusive and more scalable approach to observe and analyze a wide range of self-reported behaviors, thus allowing to capture an individual's thinking, personalities, and more importantly, mental health status. In addition, with the availability of mobile devices and increasing internet speed, more people are willing to share their mood and daily activities on social networks, hence making it possible to obtain high-quality self-reported data around mental health concerns, such as depression, self-harm activity, or eating disorders.

In this work, we analyze users' behavior on the Flickr platform, a popular online community where users share personal photographs. Our choice of using Flickr as a data source is motivated by its high-levels of user activity and the multimodal nature of their posts.[1] Flickr posts consist of an image and its metadata along with an associated caption and tags, which makes them a rich source of textual, visual, and other meta information. We seek to explore the multimodal cues present in Flickr posts to infer users' mental health status. We consider mental health as a broad concept as frequently several mental illnesses occur simultaneously. With this broad categorization, we intend to capture general insights into behavioral changes experienced by Flickr users that are related to the onset or the presence of mental health distress regardless of the specific disorder being experienced.

More specifically, we focus our efforts on two classification tasks. First, we aim to distinguish between healthy users and users affected by a mental health illness. Second, we attempt to discriminate between healthy users and mental illness prone users by identifying their mental illness onset, defined as the time when they start using posts tags related to mental illnesses. We explore a large set of multimodal features that attempt to capture linguistic, visual, and behavioral aspects of users' posting activity. We analyze feature significance using an effect size analysis to identify which cues from which modality are indicative of mental health status. We also conduct several learning experiments where we explore the predictive power of the different visual, language, and posting activity features for the two classification tasks.

2. Related Work

Several computational approaches have been proposed to study mental health via the analysis of users' behavior in social media, which provides a naturalistic way to observe and analyze individuals' daily activities and thoughts. Mental health has been studied in different social media outlets, such as Twitter, Instagram, Flickr, and Facebook. A large fraction of these studies has been focused on analyzing patterns in users' language and social activity. Among them, authors in (Park et al., 2012) explored language differences in the way depression is talked about in Twitter and found that individuals with major depressive symptoms show a significant increase in the number of anger and negative words used in their tweets as compared to the gen-

[1] Flick currently has over 90 million monthly users https://expandedramblings.com/index.php/flickr-stats/

eral Twitter population. Authors in (De Choudhury et al., 2013) crowd-sourced depression screening questionnaires and users' Twitter feeds and analyze their language, expressed emotions, and network activity to identify depression markers that predict the onset of depressive behaviors. Similarly, in (Moreno et al., 2011) authors showed that status updates on Facebook can reveal symptoms of major depressive episodes in college students. Authors in (De Choudhury et al., 2016) explored the linguistic structure, linguistic style matching and interaction behaviors of Reddit users to identify markers of shifts to suicidal ideation in subreddits discussing mental health and suicide support. Lexical, behavioral and topical changes on eating-disorder posts were studied on Instagram user activity (Reece and Danforth, 2017). Another study analyzed patterns of internet usage such as the average packets per flow, chat octets, and Internet packets and found that they are correlated to depression symptoms (Kotikalapudi et al., 2012).

More recently, studies have also incorporated other modalities that can help assess mental health status, such as visual content. Authors in (Reece and Danforth, 2017) applied color analysis and face detection in user-generated images to derive visual features that were jointly used with language and user activity features to identify predictive markers of depression in Instagram posts.

Also related to this line of work is research conducted on the analysis of user dimensions, e.g., sentiment, emotion, personality, and demographics from language and visual data. Authors in (Wang and Li, 2015) addressed the identification of sentiment in social media images using both, visual information and contextual network information, including friends' comments and users' descriptions. The work presented in (Wendlandt et al., 2017) explored the relation between user-generated images and their captions to predict personality and gender. Their paper presented correlational methods to explore the predictive power of several visual and textual features in the prediction of user attributes and showed that methods that rely on both, image and textual information outperform models build using only one modality at the time.

Our work follows a similar approach to analyze a large set of visual, language, and post metafeatures derived from Flickr posts to infer users' health status.

3. Data Collection

In order to collect Flickr posts related to mental health, we started by automatically identifying images tagged with mental health topics. First, we search for posts containing the #mentalillness tag using the Flickr's official API.[2] We examine post activity associated with this tag between 01/02/2001 and 03/01/2018 and identify 12,056 unique image tags. From the set we remove tags used by less than 20 users and tags containing non-ASCII characters, thus remaining with 82 tags that we further filter based on their relatedness to the #mentalillness tag (#MI). We use Pointwise Mutual Information (PMI) to rank the relatedness of each tag pair as shown in Equation 1, where #MI stands for the #mentalillness tag.

$$PMI(t, \#MI) = \log_2 \frac{|\{posts\_w/\_t\_and\_\#MI\}|}{|\{posts\_w/\_t\}| \cdot |\{posts\_w/\_\#MI\}|} \quad (1)$$

We then select the 30 top-ranked tags and manually verify the resulting set to remove tags deemed as too general, such as #illness, #mental, and #sad as well as unrelated tags such as #portrait, and #art. After this filtering step, we obtain a final set consisting of 15 tags, including #mentalillness, #bipolardisorder, #schizophrenia, #self-harm, #depression, #bipolar, #psychosis, #insanity, #anxiety, #depressed, #insane, #disorder, #suicide, #despair. Next, we use these tags to collect user's posts as described below.

Mental Illness Posts. We collect posts from users who have used these tags for at least ten different days between 01/01/2001 and 03/01/2018. For each user, we randomly select 20 posts generated during this time interval, thus remaining with 14,781 total posts from 770 users. In the remainder of the paper, we denote this set as mental illness set MI.

Pre-mental Illness Posts. To analyze users' behavior before they use tags related to mental illness, we collect samples of their previous posting activity. We first identify the date in which they started using mental illness tags and consider it as the potential onset of mental illness. Next, we collect 20 random posts posted during the year before this date. The resulting set, which consists of 11,828 posts from 658 users is then used as the pre-mental illness set, $PREMI$.

Healthy Users Posts. Additionally, we collect a set of posts from healthy users using a subset of the YFCC100M dataset by (Thomee et al., 2016). YFCC100M is a public corpus containing Flickr posts from 581,099 users. As a preprocessing step, we verify the entire posting activity of every user in the corpus and exclude users who use any of our mental-illness tags and also users with less than 10 days posting activity. Given a large number of available posts from each user, we opted to randomly select one post from each of them to generate a set of posts from healthy users, HE. The final set consists of 15,000 posts. Figure 1 shows an overview of the data collection process. Sample images posted by each group, along with their associated captions and tags, are shown in Figure 2.

Privacy Considerations. To address ethical and privacy concerns during the data collection, we use only public data available in Flickr and remove any identifiable personal information before conducting further analyses. The sample visual content shown in this paper was originally posted under the creative commons license.

4. Multimodal Features

To enable our experiments, we derive several sets of features from three different modalities present in Flickr posts. More specifically, we extract visual features from the image, linguistic features from the post caption, and post metafeatures derived from the image metadata and post views statistics.

[2] https://www.flickr.com/services/api/ accessed via the Flickrapi library.

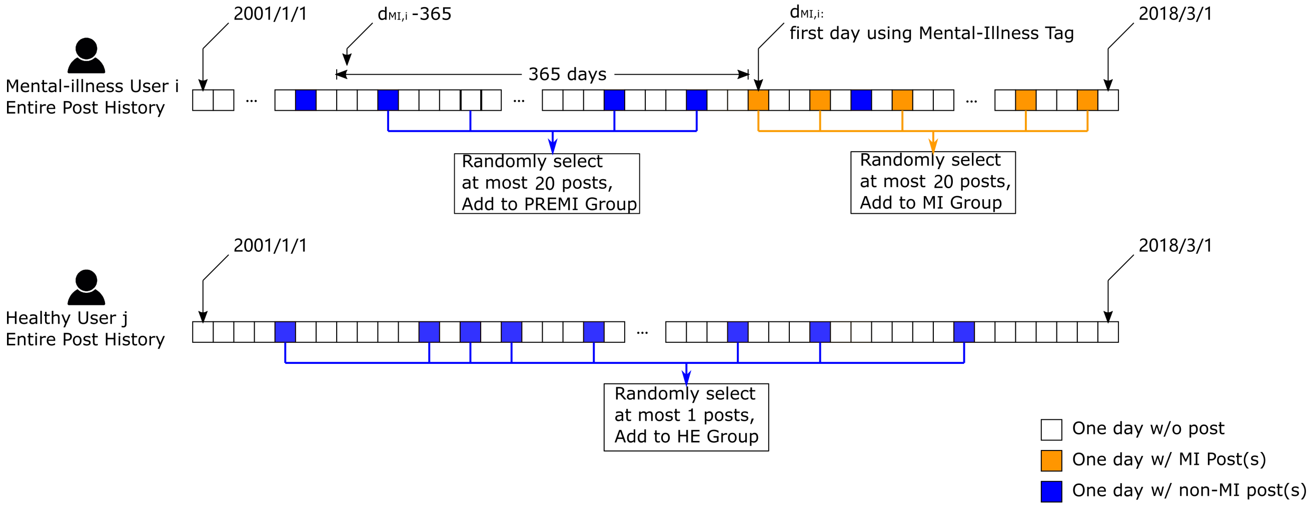
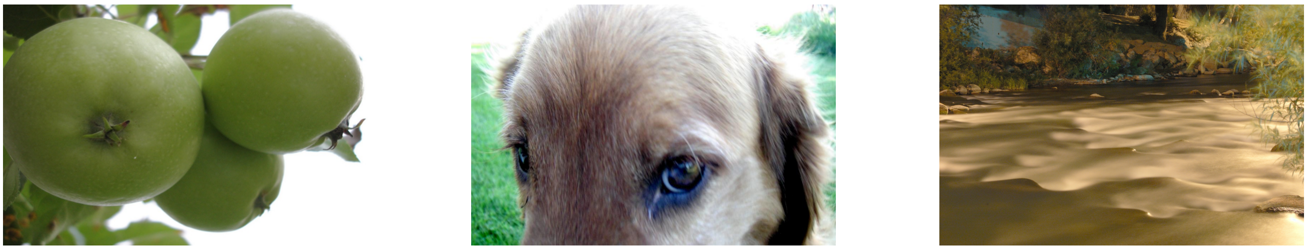


Figure 1: An overview of the timeline used for mental health data collection. Healthy posts (HE), mental illness posts (MI), pre-mental illness (PREMI) posts.



(a) How do you like them? #apples, #green, #growing

(b) Forlorn. #dog, #emotion, #sadness, #eye, #tears

(c) 45 seconds in the life of the Truckee River. #depression, #bipolar

Figure 2: Sample Flickr posts from (a) healthy; (b) pre-mental illness; and (c) mental illness users.

4.1. Visual Features

We explore several visual features to identify visual aspects associated with the mental health status of Flickr users. We extract low-level features such as color distribution, brightness and emotion matrices, texture, and static and dynamic lines. In addition, we extract high-level visual features that provide overall cues of the objects and scenes present in the images. These features have been previously found correlated with psychological traits and with user's sentiment (De Choudhury et al., 2016; Wendlandt et al., 2017).

Color Distribution. Previous studies have shown that color choices are related to emotion (Naz and Epps, 2004; Redi et al., 2015). To characterize the color distribution in the images, we use the Probabilistic Latent Semantic Analysis model by (Van De Weijer et al., 2009). This method assigns a color to each pixel in the image from a set of eleven colors namely, black, blue, brown, grey, green, orange, pink, purple, red, white, and yellow. We derive visual features using the percentage of each color across the Flickr image.

Brightness and Emotion. Previous studies have shown that color properties such as brightness and saturation are related to depressive states (Reece and Danforth, 2017). We extract the average and variance values of the of HSV color model, i.e., hue (H), saturation (S) and value (V, brightness), from each image. We also calculate the average intensity using $I = 0.3r + 0.59g + 0.11b$ where r, g, b are the average values of the red, green, and blue channels separately. In addition, we use the brightness and saturation values to derive estimates of affect dimensions such as pleasure, arousal, and dominance (Valdez and Mehrabian, 1994). These features are calculated as follows: $Pleasure = 0.69v + 0.22s$, $Arousal = -0.31v + 0.60s$, and $Dominance = -0.76v + 0.32s$.

Texture. Image texture provides information about the spatial arrangement of color and intensity in an image and has been found related to emotional affect (Stockman and Shapiro, 2001; Toet et al., 2011). We incorporate several metrics of image texture, including contrast, correlation, energy, and homogeneity. These features are obtained from the image gray-level co-occurrence matrix.

Static and Dynamic Lines. The ratio between numbers of static lines and slant lines, which can reflect the stability and dynamism of an image, has also been found to relate to affective responses (Machajdik and Hanbury, 2010). We explore adding image dynamics information into our analysis by detecting significant line patterns in images. To identify lines, we first apply the Hough transform and classify a line as static if they are in the following ranges: $-15° < \theta < 15°$ or $75° < \theta < 105°$; and as slant otherwise. We then derive three features, indicating the proportion of static and dynamic lines in the image.

Scenes. The environment or scene where a photo is taken can provide clues into people's emotional and mental status. We use the Places365-CNN models (Zhou et al., 2017) to detect scenes in the images. Place365-CNN is an image scene detector that can estimate the presence of 365 common scenes, along with their attributes. To represent scene information, we obtain a vector that includes the probability distribution of the 365 scenes and their attributes as provided by the classifier.

Faces. The number of faces in an image has been found correlated to social behavior (Wendlandt et al., 2017). To incorporate this visual cue, we use the CascadeObjectDetector function in the MATLAB's vision toolbox to detect the number of faces present in each image. The raw faces count is used directly as a feature.

Objects. We use AlexNet (**?**) to perform object detection in the images. AlexNet is a deep neural network for object detection trained on the ImageNet dataset. For any image, it can predict the object among 1000 classes with relatively high accuracy. We use the pre-trained AlexNet module in MATLAB for this task. To extract these features, we reshape all images to a standard size of 227x227x3 and then feed them into the deep neural net. The obtained features are then represented as a 1000-dimension one-hot vector, where each element represents the corresponding object.

Visual Features and Users' Health Status. To explore potential relationships between the visual features described above and Flickr users' health status, we measure the relationship between each type of feature and the healthy (HE), mental illness (MI) and pre-mental illness (PREMI) groups using Cohen's d effects size. Cohen's d measures how many standard deviations two groups differ by, and is calculated by dividing the mean difference by the pooled standard deviation. The results are shown in Table 1. In this table, positive effect sizes indicates that healthy users prefer the feature, while negative effect size indicate that the corresponding mental illness group prefer the feature.

Our analysis reveals interesting differences between the three groups. Particularly, we observe important effect differences in the color distribution of images in mental illness posts as users show a preference for the black color, while images in healthy posts contain more blue and green. This observation is in line with previous research findings of dark colors being often regarded as depressing (Carruthers et al., 2010) and green colors being evocative of positive emotions such as relaxation and comfort (Naz and Epps, 2004). Moreover, users in the mental illness group show a preference for less saturated images (higher dominance and arousal), high contrast textures, and static lines. In addition, they are likely to post photos taken indoors and with fewer faces on them. In contrast, healthy users prefer to post photos with higher pleasure scores, and their images include more lines, show more dynamic patterns, and include more faces on them. Most likely because healthy users are more active and social than users suffering from mental distress. Finally, we observe high overlap between visual features preferred by the mental illness and the pre-mental illness groups. Interestingly, we notice that features belonging to the texture, static lines, and faces feature sets have higher effect sizes for images posted by the pre-mental illness users than the ones posted by users presumably suffering from a mental illness.

Table 1: Visual features with significant difference ($p < 0.05$) between healthy and mental illness groups (left column); healthy and pre-mental illness groups (right column).

Mental Illness		Pre-mental Illness	
Image Feature	Effect Size	Image Feature	Effect Size
Color Distribution			
Black	-0.262	Black	-0.141
White	-0.185	Purple	-0.070
Brown	0.131	Pink	-0.061
Blue	0.224	Blue	0.101
Green	0.242	Green	0.142
Brightness and Emotion Matrices			
Dominance	-0.181	Dominance	-0.119
Arousal	-0.179	Arousal	-0.118
Mean (s)	0.168	Std (h)	0.076
Pleasure	0.183	Pleasure	0.119
Mean (h)	0.190	Mean (h)	0.124
Texture			
Contrast	-0.430	Contrast	-0.471
Energy	-0.188	Energy	-0.050
Homogeneity	0.309	Homogeneity	0.410
Correlation	0.452	Correlation	0.490
Static and Dynamic Lines			
%Statistic Lines	-0.435	%Statistic lines	-0.457
#Static Lines	-0.257	#Static lines	-0.310
Total Lines	-0.248	Total Lines	-0.306
Scenes			
Indoor	-0.452	Indoor	-0.219
Jail cell	-0.339	Jail cell	-0.178
Burial chamber	-0.220	Dressing room	-0.147
Dressing room	-0.214	Beauty salon	-0.146
Lagoon	0.162	Shower	-0.138
Valley	0.174	Basement	-0.118
Scene Attributes			
Enclosed area	-0.470	Enclosed area	-0.241
No horizon	-0.352	No horizon	-0.206
Man-made	-0.316	Man-made	-0.182
Leaves	0.266	Aged	-0.173
Foliage	0.296	Socializing	0.149
Vegetation	0.299	Congregating	0.171
Faces			
Number of faces	0.230	Number of faces	0.317
Objects			
Prison	-0.168	Sportscar	-0.093
Website	-0.159	Bookjacket	-0.079
Bookjacket	-0.125	Lipstick	-0.068
Lakeside	0.115	Sandbar	0.073
Valley	0.123	Valley	0.081

4.2. Language Features

Flickr users usually provide captions along with the photos they share. Captions can provide important clues into people's thoughts and moods. To explore whether they can provide insights into users' mental status, we derive several language features, as described below.

Stylistic Features. We extract surface level stylistic features from the image caption, including the number of tokens and the number of tokens of length 3 and 6. We also extract several readability metrics, including the Automated

Table 2: Language features with significant difference (p¡0.05) between healthy and mental illness groups (left column); healthy and pre-mental illness groups (right column).

Mental Illness		Pre-mental Illness	
Language Feature	Effect Size	Language Feature	Effect Size
Stylistic Features			
Total Tokens	-0.653	Tokens len ≥ 3	-0.542
Tokens len ≥ 3	-0.650	Total Tokens	-0.525
Tokens len ≥ 2	-0.649	Tokens len ≥ 4	-0.525
Readibility (CLI)	-0.643	Readibility (CLI)	-0.515
Readibility (LWF)	-0.571	Readibility (LWF)	-0.462
Readibility (DW)	-0.428	Readibility (DW)	-0.303
Readibility (ARI)	-0.422	Readibility (FKG)	-0.232
Ngrams			
The	-0.257	The	-0.222
To	-0.226	Dsc	0.165
Depression	-0.223	To	-0.151
How+to	-0.121	How+to	-0.078
State+hospital	-0.111	In+the	-0.076
Suicide+squad	-0.100	Dsc+jpg	0.067
Part-of-speech ngrams			
Preposition	-0.319	Adjective	-0.239
Determiner	-0.300	Determiner	-0.234
Adjective	-0.294	Preposition	-0.222
Determiner+Noun	-0.217	Adjective+Noun	-0.197
Adjective+Noun	-0.209	Determiner+Noun	-0.185
LIWC Features			
Function	-0.579	Function	-0.408
NegativeEmotion	-0.428	BioProc	-0.245
BioProc	-0.384	Drives	-0.240
Article	-0.384	Time	-0.235
Drives	-0.335	You	-0.232

Readability Index (ARI), Simple Measure of Gobbledygook (SMOG), Coleman-Liau Index(CLI), Linsear Write Formula (LWF), Difficult words (DW) and Flesch-Kincaid Grade (FKG).

Ngrams. We also extract unique words (ngrams) and unique word pairs (bigrams) from captions. After removing ngrams that occurred less than 20 times we obtain 955 unique unigrams and 173 unique bigrams. Similarly, we extracted Part-Of-Speech (POS) tags ngrams. All features are normalized by their word counts.

LIWC Features. We also obtain a set of features from the Linguistic Inquiry and Word Count (LIWC 2011) lexicon (Pennebaker and King, 1999). LIWC focuses on broad categories such as language composition, as well as emotional, cognitive, and social processes.The features consists of the percentage of each of the LIWC word classes present in the caption's text.

Language Features and Users' Health Status. We conduct an effect size analysis in textual features derived from the image's captions. Results are shown in Table 2. Results show that the language used by users might be suffering from mental illnesses seem to be more complex as they use longer words and a larger number of words in their captions. Moreover, this group uses more negative emotion words and questions in their captions.

Table 3: Post features with significant difference ($p < 0.05$) between healthy and mental illness groups (left column); healthy and pre-mental illness groups (right column).

Mental Illness		Pre-mental Illness	
Meta Feature	Effect Size	Meta Feature	Effect Size
Post views	-0.791	Post views	-0.557
Exif	-0.596	Exif	-0.360

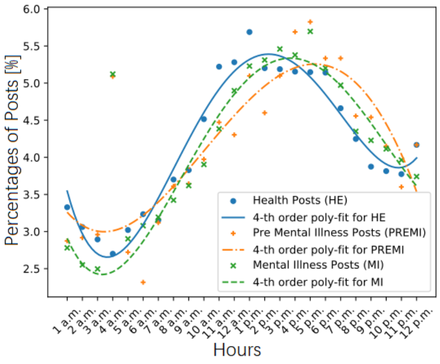


Figure 3: Temporal Distribution Comparison among Groups.

4.3. Metadata Features

Temporal Features. Previous research showed that people under mental distress have different temporal patterns from healthy users because they might suffer from insomnia (Wang et al., 2017). To analyze how this could be also reflecting in Flickr users activity, we extract the local hour in which the image was posted using the upload time. The feature is represented with a one-hot encoding vector.

Post Views. The number of post views can reflect users' social relation and activity status as generally posts of a person who is active on social media are more likely to be viewed by others. Thus, we also use the number of post views as a feature.

Image File (exif) Features. Photos taken by digital cameras usually include metadata in an exif (exchangeable image file format) file. However when an image is externally modified the exif data is usually lost. In this case, we hypothesize that the user has spent more effort in preparing the image before posting it. We represent this information with a binary feature that indicates whether the image has exif information.

Post Features and Users' Health Status. Figure 3 shows a different temporal pattern in postings among groups. We observe that users suffering from mental illness have more posting activity during the afternoon and evening as compared to healthy users. Table 3 shows the effect size analysis on the post features. We observe that users suffering from a mental illness are more likely to modify their images before posting them (35.99% vs. 11.6% effect size in the exif feature). This suggests that they prefer darker or monochromatic images, which requires post-processing by software and thus results in exif data loss. We observe that posts from users suffering from mental illness have significantly higher views than healthy people. This can be potentially due to repeated views from highly connected

Table 4: Classification F-scores for healthy users (HE) and users suffering from a mental illness (MI) using four algorithms and visual, language and post metafeatures (meta).

Modalities	DT		SVMLinear		AB		NeuralNet	
	F-HE	F-MI	F-HE	F-MI	F-HE	F-MI	F-HE	F-MI
Visual	0.683	0.664	0.660	0.671	0.720	0.715	0.722	0.720
Language	0.660	0.670	0.734	0.662	0.746	0.680	0.746	0.674
Meta	0.770	0.774	0.760	0.701	0.788	0.779	0.791	0.768
Visual +Language	0.686	0.701	0.731	0.713	0.781	0.764	0.779	0.758
Visual + Meta	0.804	0.786	0.737	0.734	0.827	0.821	0.807	0.767
Language + Meta	0.799	**0.800**	**0.784**	0.741	0.836	0.828	0.823	0.795
All Features	**0.810**	0.793	0.771	**0.754**	**0.852**	**0.847**	**0.857**	**0.850**

Table 5: Classification F-scores for posts from healthy users (HE) and posts from users prone to mental illness (PREMI) using four algorithms and visual, language and post metafeatures (meta).

Modalities	DT		SVMLinear		AB		NeuralNet	
	F-HE	F-PREMI	F-HE	F-PREMI	F-HE	F-PREMI	F-HE	F-PREMI
Visual	0.638	0.676	0.614	0.619	0.672	0.672	0.673	0.676
Language	0.634	0.634	0.670	0.561	0.669	0.622	0.681	0.627
Metadata	0.697	0698	0.677	0.622	0.719	0.704	0.708	0.650
Visual +Language	0.680	0.665	0.647	0.620	0.715	0.706	0.711	0.701
Visual + Metadata	0.716	0.731	0.644	0.637	0.753	0.743	0.741	0.732
Text + Metadata	0.723	0.725	**0.688**	0.623	0.745	0.732	0.726	0.688
All Features	**0.733**	**0.739**	0.672	**0.646**	**0.767**	**0.755**	**0.756**	**0.749**

close-knit networks (De Choudhury et al., 2013). However, given the limited availability of data in our study, we are not able to test this assumption.

5. Predicting User's Mental Health Status from Multimodal Posts

During our experiments, we seek to incorporate multimodal insights into users' mental health from visual, language and posts features derived from Flickr posting activity.

5.1. Preprocessing and Experimental Setup

To enable our experiments, we first obtain a balanced number of posts from each group of users (11,828 posts) by randomly removing the excess posts. We normalize all features to zero mean and unit variance. During the experiments, the evaluations are conducted at the user level using 10-fold cross-validation. Thus, posts from the same user appear in either training or test sets but not both. We use the following classifiers: Decision Tree (DT, max depth = 4); Adaptive Boost (AB); Support Vector Machine with a linear kernel (SVMLinear).[3] In addition, we also experiment with a two-layer neural network classifier implemented using the TensorFlow library (Abadi et al., 2015). More specifically, we used a rectified linear unit (ReLU) as the activation function for the hidden layers, and a sigmoid function to constrain the output probability to be between zero and one. A threshold of 0.5 is then applied to conduct the binary classification. In order to experiment with different feature combinations, we added a channel valve after the first hidden layer to regulate the use of each feature set. Thus, its output is equal to the input if the valve is turned on and zero otherwise.

5.2. Predicting Mental Health Status using Multimodal Observations

To evaluate the predictive power of the visual, language, and metadata features, both individually and jointly, we conduct several experiments using the feature sets described in Section 4. The features are derived from either images (visual), captions (language), or post metafeatures (meta), and the combinations between them by simply concatenating the different feature vectors. Classification results to discriminate between posts generated by healthy users and users suffering from a mental illness are shown in Table 4. Similarly, classification results for healthy users and users prone to mental illness are shown in Table 5.

Overall, we found that the majority of the classifiers performed the best when using all modalities jointly as compared to using single modalities. This suggests that combining visual, language, and post information can provide improved performance on the prediction of a user's mental health status.

In terms of task performance, the multimodal approach was more effective when distinguishing between posts from the mental illness and healthy groups. A potential explanation for this is that the behaviors from users in the pre-mental illness group are more similar to the healthy group thus making it difficult to establish clearer differences between the two groups. Nonetheless, we observe improved performances when jointly using the different feature sets while distinguishing between pre-mental illness posts versus healthy posts.

[3] We use the default parameters in scikit-learn for the Adaptive Boost and Support Vector Machine classifiers.

Regarding the classification method, the best F-HE and F-MI scores are archived by the neural classifiers that make use of all features. Interestingly, the worst F-HE and F-MI scores in neural classifiers are obtained when using language features. We hypothesize that this was caused by the sparseness of the language feature vectors. To address this issue, our future work includes using vector representations instead of hand-engineered features.

6. Conclusion

In this paper, we explored the use of multimodal cues present in social media posts to predict a user's mental health status. Our paper makes several contributions. First, we collected a new dataset containing social media posts from Flickr discussing mental health before and after the onset of mental-illness. Second, using this dataset we conducted an extensive analysis of visual, language, and post features and identified behavioral differences in posts generated by healthy users, users suffering from a mental-illness, and users prone to mental-illnesses.

Our findings suggest behavioral differences between the two groups as compared to the healthy group. Specifically, we found that individuals suffering from mental illnesses and individuals prone to mental illness prefer, at different extent, posting darker images with high contrasts showing indoors scenes and fewer faces as compared to healthy users. Finally, through several experiments, we showed that the derived features are useful for the prediction of users' mental health status. Moreover, we showed that the combination of visual, language, and posts information can improve the performance of this task as compared to the use of cues from a single modality at a time.

7. Bibliographical References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Burke, M., Marlow, C., and Lento, T. (2010). Social network activity and social well-being. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1909–1912, New York, NY, USA. ACM.

Carruthers, H. R., Morris, J., Tarrier, N., and Whorwell, P. J. (2010). The manchester color wheel: development of a novel way of identifying color choice and its validation in healthy, anxious and depressed individuals. *BMC medical research methodology*, 10(1):12.

De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. (2013). Predicting depression via social media. *ICWSM*, 13:1–10.

De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., and Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 2098–2110. ACM.

Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., Gasquet, I., Kovess, V., Lepine, J., Angermeyer, M. C., Bernert, S., Morosini, P., Polidori, G., et al. (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the world health organization world mental health surveys. *Jama*, 291(21):2581–2590.

Kotikalapudi, R., Chellappan, S., Montgomery, F., Wunsch, D., and Lutzen, K. (2012). Associating depressive symptoms in college students with internet usage using real internet data. *IEEE Technology and Society Magazine*, 31(4):73–80.

Machajdik, J. and Hanbury, A. (2010). Affective image classification using features inspired by psychology and art theory. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 83–92. ACM.

Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., and Hoving, C. (2013). A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *J Med Internet Res*, 15(4):e85, Apr.

Moreno, M. A., Jelenchick, L. A., Egan, K. G., Cox, E., Young, H., Gannon, K. E., and Becker, T. (2011). Feeling bad on facebook: Depression disclosures by college students on a social networking site. *Depression and anxiety*, 28(6):447–455.

Naz, K. and Epps, H. (2004). Relationship between color and emotion: A study of college students. *College Student J*, 38(3):396.

Park, M., Cha, C., and Cha, M. (2012). Depressive moods of users portrayed in twitter. In *Proceedings of the ACM SIGKDD Workshop on healthcare informatics (HI-KDD)*, volume 2012, pages 1–8. ACM New York, NY.

Pennebaker, J. W. and King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296.

Redi, M., Quercia, D., Graham, L. T., and Gosling, S. D. (2015). Like partying? your face says it all. predicting the ambiance of places with profile pictures. *arXiv preprint arXiv:1505.07522*.

Reece, A. G. and Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1):15.

Stockman, G. and Shapiro, L. G. (2001). *Computer Vision*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition.

Thomee, B., Shamma, D. A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., and Li, L.-J. (2016). Yfcc100m: The new data in multimedia research. *Communications of the ACM*, 59(2):64–73.

Toet, A., Henselmans, M., Lucassen, M. P., and Gevers, T. (2011). Emotional effects of dynamic textures. *i-Perception*, 2(9):969–991.

Valdez, P. and Mehrabian, A. (1994). Effects of color on emotions. *Journal of experimental psychology: General*, 123(4):394.

Van De Weijer, J., Schmid, C., Verbeek, J., and Larlus, D. (2009). Learning color names for real-world applications. *IEEE Transactions on Image Processing*, 18(7):1512–1523.

Wang, Y. and Li, B. (2015). Sentiment analysis for social media images. In *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*, pages 1584–1591. IEEE.

Wang, Y., Tang, J., Li, J., Li, B., Wan, Y., Mellina, C., O'Hare, N., and Chang, Y. (2017). Understanding and discovering deliberate self-harm content in social media. In *Proceedings of the 26th International Conference on World Wide Web*, pages 93–102. International World Wide Web Conferences Steering Committee.

Wendlandt, L., Mihalcea, R., Boyd, R. L., and Pennebaker, J. W. (2017). Multimodal analysis and prediction of latent user dimensions. In *International Conference on Social Informatics*, pages 323–340. Springer.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.