Characterizing the Visual Social Media Environment of Eating Disorders

Samsara N. Counts, Justine-Louise Manning, Robert Pless
Department of Computer Science
George Washington University

***Abstract*—Eating disorders are often exacerbated by exposure to triggering images on social media. Standard approaches to filtering of social media by detecting hashtags or keywords are difficult to keep accurate because those migrate or change over time. In this work we present proof-of-concept demonstrations to show that Deep Learning classification algorithms are effective at classifying images related to eating disorders. We discuss some of the challenges in this domain and show that careful curation of the training data improves performance substantially.**

I. Introduction

According to the National Association of Anorexia Nervosa and Associated Disorders [8], at least thirty million people in the United States suffer from an eating disorder (ED) [16]. ED are complex medical conditions with the highest mortality rate of any mental disorder [12] that affect people of all genders, races, and walks of life. ED are pervasive and do not discriminate based on race, religion, gender, or socioeconomic status. ED are often a lifelong struggle with approximately two-thirds of patients never achieving a full and sustained remission.

With the expansion of social media on the internet, there has been a rise in online communities centered around different topics and lifestyles across various online platforms, especially those surrounding interests that are perceived to be "alternative" or taboo [17]. In a similar fashion, people with eating disorders have turned online to journal and document their lives [13]. On the extreme side of the spectrum, people with ED have formed communities [2], [18] where they create and share content pursuing the ultimate goal of someone with an eating disorder: achieving their ideal, thin body type. In a way, they are promoting ED; indeed, the terms *proana* (pro-Anorexia) and *promia* (pro-Bulimia) are frequently used as identifiers within these communities. Henceforth, we refer to these communities as *pro-ED*. These websites and blogs often contain "thinspiration:" images of people with body types they aspire to look like [2], [18], [7], [20]. The pictures they post often feature subjects that are significantly underweight to an unhealthy degree, highlighting bodily features posters believe are desirable [20], [9].

ED are fundamentally body image disorders; hence, images are critically important to this community. Therefore, it is critical that clinicians and family members be able to identify websites containing images associated with promotion of ED in order to prevent accidental or intentional exposure to these triggers. This research aims to automatically identify such triggering material, with the ultimate goal of designing software tools to help improve ED patient health outcomes.

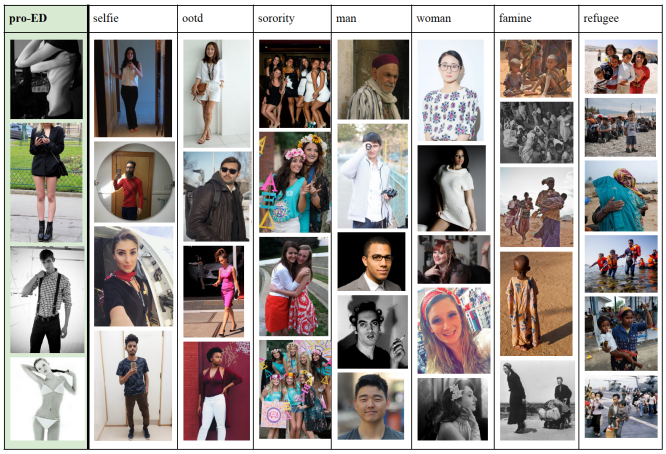

Fig. 1. Eating disorders are often driven by social-media imagery; this paper seeks to build classifiers that detect these images. On the left are examples that were posted in ways promoting EDs, and on the right are 7 sub-categories chosen to have similar demographics and poses, but consist of images not intended to glamorize ED.

The specific contributions of this paper are threefold:

- First, a demonstration that modern CNN-based classification tools can classify images posted with hashtags that relate to eating disorders,
- Second, a characterization of related, but not pro-ED content that is visually similar to pro-ED content and a demonstration that more careful curation of a dataset allows for finer classification, and
- Third, an evaluation of this classifier in the context of characterizing the content of a whole blog, which may be the functional component of a clinician or patient tool to improve an ED patient's chance of recovery.

II. Background

A comprehensive review of the nature of pro-ED social media posts contextualized for a technical audience is given in Pater et al [20] and Wang et al [23].

One important notion in the realm of eating disorder communities online is the distinction between pro-ED images and images of people with ED/associated with an ED hashtag [4]. The difference between them is the *intention* behind posting. For example, a person with an eating disorder may post a

978-1-5386-5541-2/18/$31.00

selfie to their blog where they may have posted about their ED before. However, that selfie is not necessarily pro-ED unless explicitly labeled as such. Therefore, though we include pro-ED and other images in our training set, we do not exclusively detect pro-ED content because, without the full context of an image, we cannot know whether it is pro-ED.

However, by taking a post's context into account, Chancellor et al [4] created a multimodal classifier that detects deviant eating disorder (usually, pro-ED) posts. Text surrounding an image is collected for context, used as a proxy for intention. Hence, their classifier attempts to distinguish between proana/pro-ED images and other images, including those of people with ED. To create a training dataset, they used seed tags and snowball sampling to discover new tags and images. They included images removed from Tumblr for violating community guidelines. Their multimodal deep learning approach combines the text surrounding a particular post with the image(s) associated with that post. For the text-centric part of their classifier, they used word2vec and skip-gram to embed hashtags, then computed each word's nearest neighbors. For the image analysis, they used the deep learning architecture AlexNet pre-trained on ImageNet to learn numerical image features which are used to classify pro-ED content from Tumblr. Rather than make an end-to-end architecture, they chose to learn the last layers of each modality jointly with all subsequent multimodal layers.

Outside of the contributions of Chancellor et al. in [4], past work in detecting ED posts, communities, and users on social media has applied machine learning to numerical and text-based features [23], [7], [5], [11].

He and Luo [11] use decision tree-based algorithms on Tumblr and Twitter data with text features to classify whether a particular post was pro-ED. Wang et al. examine the evolution of pro-ED networks online in depth, exploring community structures and interactions among individuals [23]. They quantify differences between the sample of ED users and users without eating disorders and use those features to build predictive models. Finally, De Choundary used a Support Vector Machine-based model to detect whether a Tumblr post has anorexia-related content and whether a post is pro-ED or pro-recovery [7]. They used four categories of numerical measures as individual features of the prediction model: social, affective, linguistic style, and cognitive processes, each consisting of a custom set of its own sub-features.

However, despite the encouraging results described above, as Chancellor et al. revealed, text-based moderation of pro-ED communities was largely ineffective on Instagram [6]. In an analysis of over 30,000 pro-ED posts and different ED hashtags over time, they demonstrated that lexical variation of hashtags grows as response to platform moderation. Hence, text-based pro-ED content moderation on Instagram is largely ineffective. Not only was there a rise in lexical variation of pro-ED hashtags, the amount of pro-ED posts grows as a response to content moderation, user interactions and participation in these communities increased.

III. Dataset curation

To create a dataset to train a classifier and test its performance, we gather images that are both pro-ED and not pro-ED. For both classes, we gather images from social media platforms and use hashtags included with the images as labels for our data. Hashtags are user-defined and thus provide us with a means of accessing images that are representative of the pro-ED community as defined by its members. The images in our final dataset came from Twitter, Tumblr, Flickr, and Google image searches collected over a period of six weeks. We removed duplicate images, but in all other respects, the dataset was unedited.

proana	bonespo	anorexic
thinspiration	ana	ednos
thinspo	mia	thighgap
bulimic	eatingdisorder	thynspo
bulimia	eatingdisorders	promia
anamia	ed	

Fig. 2. A list of the pro-ED hashtags used to find social media imagery related to the promotion of eating disorders

To gather pro-ED imagery, we used the Twitter and Flickr APIs to find images associated with #proana, a known identifier of a strongly pro-ED community [20], [11], [7], [5]. We used this hashtag as our starting point. Exploring posts with this hashtag led us to a collection of related terms. We found later that our findings were already present in the set of hashtags from [4], shown in Figure 2, and ultimately decided to use those as our pro-ED keywords.

Most images posted online are not pro-ED imagery. To ensure the classifier is learning relevant features and to create interesting test cases it is important to choose related alternative categories. Choosing these alternative categories has a substantial effect on the classifier performance, so we share a collection of approaches that we tried.

Baseline Dataset

We first create a set of non pro-ED imagery by selecting images from 3 additional hashtags, #ootd, #selfie, and #greek. Our ad-hoc analysis suggests that images with these hashtags include similar demographics and poses as the pro-ED imagery, encouraging the classifier to hone in on the visual features relevant to ED rather than the features they have in common and demographics.

Blog Test

For an initial end-to-end evaluation we identified 10 additional Tumblr accounts with a variety of content and image styles on Tumblr. We include five that we judged to have high pro-ED content, 4 blogs without pro-ED content, and 1 fitness inspiration (fitspo) blog. The fitspo community is somewhat pro-ED–posters often post about their ideal body types, show progress pictures, and discuss dieting. The key distinction is that, in fitspo blogs, subjects of posted pictures often appear to be more physically fit and less extremely thin. In addition, posters often focus on healthy eating and post pictures of healthy meals, rather than discussing extreme diets and negativity towards the current self [22]. Hence, the fitspo blog is an interesting test case on which to evaluate the classifier.

		Baseline Dataset	
Label	Sub-category	Training	Validation
Not-pro-ED	selfie	4306	944
	ootd	5906	999
	greek	4381	400
Pro-ED		14779	2397
Total		29372	4740

TABLE I. IMAGE CLASSES AND SUB-CATEGORIES IN OUR BASELINE DATASET, ALONG WITH THE NUMBER OF IMAGES WE USED IN OUR TRAINING AND VALIDATION CLASSES.

Stress Test

A second evaluation serves as a stress test for the classifier. For the test, we gathered all the images posted to twenty websites with content we thought the classifier might fail on because the content came from challenging categories. These included 10 sites from non-pro-ED categories (NPE) such as fashion photography, ballet, yoga and boudoir photography. The proana/pro-ED (PE) set contained 10 sites, and is designed to explore potential failure models including blogs featuring subjects who are genders not well-represented in the training data.

All websites used as image sources were from a few standard [1] blogging/sharing platforms. Tumblr, Twitter, Flickr, Lookbook, Instagram, Wordpress, Pinterest, Blogspot, or, in a few cases, Google Image search results.

Final Dataset

An improved classifier emphasized training data with subcategories of the *not-pro-ED* class that help to define the boundary between, for example, boudoir and pro-ED imagery, based on a large collection of images captured from related topics. These topics were chosen because they are similar in demographics and photographic style to pro-ED content, but come from different categories. Figure III shows the final set of alternative categories we used, and the number of images that were available for training and validation.

Data Pre-processing and Challenges

Social media imagery includes a fair amount of duplication. In online pro-ED communities, this usually consists of reblogs (on sites like Tumblr) or retweets of inspirational photos. Duplication across platforms also occurs when users take images from one social media site and re-post them on another.

To pre-process the data, we eliminated any images with overlaid text that obstructed any central figure(s) of the image. Second, we discarded any images that were not photographs of people. For example, we discarded images containing animated characters or people wearing extreme costumes. Third, we programmatically eliminated any duplicate images with a quick average hashing function [3].

One challenge of this particular domain is the possibility of class overlap between pro-ED and not-pro-ED images. A lot of pro-ED images are not originally created by users; instead, users take images from social media timelines, fashion websites, or other sites and (re-)post the images with pro-ED and/or ED-related tags. Thus it is sometimes extremely difficult and/or impossible to discern where a given image came from and with what intent it has been posted in the past. While we acknowledge that the presence of a hashtag is a weak signal of user intent, for the purpose of our data collection we used the classes defined by the hashtag(s) that come with an image to give us insight its ground-truth label.

TABLE II. TRAINING AND TEST SPLIT PER CATEGORY - MODEL 2

Class	Sub-category	Training	Validation
Not-pro-ED	famine	1332	354
	woman	4057	973
	war	5855	1488
	ootd	6354	1559
	man	6348	1513
	boudoir	3876	960
	sorority	6031	1554
	fashion	5679	1437
	selfie	6827	1749
	refugee	2046	515
Pro-ED		42583	10692
Total		86732	21740

We randomly chose 21740 (20%) of these images to serve as test data and trained the ResNet Deep Learning neural network [10] to classify the remaining training images into the categories of **pro-ED** and **not-pro-ED**.

IV. CLASSIFIER AND TRAINING

Deep Learning classification algorithms are defined by the network architecture, the network initialization, the training data, and the learning parameters.

In this study we use a standard deep learning architecture, known as ResNet-50 [10], implemented on the PyTorch platform [19]. Following standard practice ("All state-of-the-art detection methods use classifier pre-trained on ImageNet" [21]) we initialize the network by pre-training on the millions of images in the ImageNet [15] dataset to train the network to detect generally interesting image features. We then fine-tune the network to optimize performance on our image data and labels. The learning rate was set to .001 and the fine-tuning was continued for 15 epochs.

V. RESULTS

A. Baseline Results

Results on the baseline dataset show promise in identifying pro-ED imagery. Figure 3 shows the confusion matrix between the pro-ED and not-pro-ED categories, showing that images are largely classified correctly.

B. Blog classification

The blog test comprises of data captured from 10 blogs on Tumblr, some of which where pro-ED, some not pro-ED and one blog that was "fitspo" and perhaps in between. Table III shows results of a classifier trained on the baseline dataset in evaluating images from these blogs. These blogs varied in how many images they included, but results were very consistent that the pro-ED blogs had more than 70% of their images

		Predicted		
	Model 2 Validation n=21740	Pro-ED	Not-pro-ED	
Actual	Pro-ED	8545	2147	1132
	Not-pro-ED	2658	8390	11048
		11203	10537	

Fig. 3. Classification results for the categories pro-ED and not-pro-ED on the validation data for the final dataset.

ID	Label	Blog Topic	n	%Pro-ED
1	pro-ED		70	73%
2	pro-ED		6103	88%
3	pro-ED		3316	89%
4	pro-ED		1463	83%
5	fitspo	vegan food, pilates	660	53%
6	not-pro-ED	luxury cars	108	21%
7	not-pro-ED	mathematics jokes	77	4%
8	not-pro-ED	black and white digital illustrations	1463	10%
9	not-pro-ED	actors, drawings, and landscapes	627	9%
10	not-pro-ED	comic book art/graphic design	535	7%

TABLE III. THE TUMBLR BLOG TEST SET INCLUDES IMAGES FROM FOUR PRO-ED, FIVE NOT-PRO-ED BLOGS, AND ONE FITSPO BLOG THAT STRADDLES THESE CATEGORIES.

judged to be pro-ED. The not-pro-ED blogs had less than 21% of their images judged to be pro-ED.

Finally, the fitspo blog had about 50% of the images classified as pro-ED, a number in between the other categories and reflecting the borderline nature of the fitspo hashtag.

C. Stress Test Results

The stress test dataset includes images from a variety of blogs, both pro-ED blogs from unusual demographics, and not-pro-ED blogs including those focused on boudoir photography, fashion models, and ballet.

Figure 4 shows results for a classifier based both on the baseline dataset (Model 1) and the final dataset (Model 2) that explicitly includes examples of boudoir, fashion, and other similar but not-pro-ED categories. The presence of this additional data is apparent in the results. Model 1 was very good at correctly classifying pro-ED imagery; in fact, all 10 pro-ED blogs had at least 72% of their images classified as pro-ED. However, 7 of the 10 related but not-pro-ED categories also had at least 72% of the images classified as pro-ED.

In contrast, Model 2 performs better overall because it was trained on data that included substantially more variation in the not-pro-ED but related classes. With Model 2, only three of the not-pro-ED sites had more than 50% of their images labeled pro-ED. Additionally, only one pro-ED site was mislabelled, despite the dramatic improvement in correctly labelling not-pro-ED sites.

The difference in the two categories is more clear to see if the data is simplified into pro-ED imagery (from all pro-ED sites) and not-pro-ED imagery (from all not-pro-ED sties). In this case, training on the initial dataset gives 89% accuracy when labelling images that are actually pro-ED, but only 37% accuracy at classifying pictures in the not-pro-ED category.

					Model 1		*Model 2*	
ID	Label	Source	NPE	PE	Total	NPE%	PE%	NPE%	PE%
1	NPE	Lookbook - plus-sized white woman	42	22	64	65.6%	34.4%	81.25%	18.75%
2	NPE	Lookbook - thin white woman	21	56	77	27.3%	72.7%	18.18%	81.82%
3	NPE	Instagram - ballerinas	30	138	168	17.9%	82.1%	25%	75%
4	NPE	Lookbook - men of color	74	14	88	84.1%	15.9%	94.32%	5.68%
5	NPE	Google search: "African-American fashion models"	11	171	182	6.0%	94.0%	63.74%	36.26%
6	NPE	Google search: "famine"	13	165	178	7.3%	92.7%	90.45%	9.55%
7	NPE	Lookbook of a white man	247	44	291	84.9%	15.1%	94.16%	5.84%
8	NPE	Vogue runway photoshoot, 2018	10	28	38	26.3%	73.7%	78.95%	21.05%
9	NPE	Boudoir photography	0	74	74	0.0%	100%	37.84%	62.16%
10	NPE	Instagram - yoga poses	40	104	144	27.8%	72.2%	46.25%	43.75%
11	PE	Wordpress - female owner, tips & thinspo	0	16	16	0.0%	100%	12.5%	87.50%
12	PE	Blog of a male anorexic	101	574	675	15.0%	85.0%	26.4%	73.6%
13	PE	Blog - single owner, thinspo	0	12	12	0.0%	100%	0.0%	100.0%
14	PE	Wordpress of a proana teenager in the U.K.	7	70	77	9.1%	90.9%	19.48%	80.52%
15	PE	Overblog - traditional thinspo, anon owner	2	20	22	9.1%	90.9%	22.73%	77.27%
16	PE	Pinterest search: "proana"	6	141	147	4.1%	95.9%	23.81%	76.19%
17	PE	Instagram page - one female owner, thinspo	1	66	67	1.5%	98.5%	7.46%	92.54%
18	PE	Wordpress - thinspo with blonde women	0	10	10	0.0%	100%	20.0%	80.0%
19	PE	Blogspot - ED diary of a college student	0	21	21	0.0%	100%	57.14%	42.86%
20	PE	Instagram page - single owner, thinspo	8	77	85	9.4%	90.6%	17.65%	82.35%
		Total			**2436**				

Fig. 4. Stress Test Results

		Predicted		
	Model 1 Stress test n=2436	Pro-ED	Not-pro-ED	
Actual	Pro-ED	1007	125	1132
	Not-pro-ED	816	488	1304
		1823	615	

		Predicted		
	Model 2 Stress test n=2436	Pro-ED	Not-pro-ED	
Actual	Pro-ED	864	268	1132
	Not-pro-ED	291	1013	1304
		1155	1281	

Fig. 5. Confusion matrix from the challenging stress test data for the Model 1 classifier (left) and the Model 2 classifier (right). The Model 2 classifier is much better at correctly classifying not-pro-ED imagery because it is trained on data from more of the not-pro-ED but similar categories.

Figure 5 shows these summary classification results. The biggest difference between the models can be seen in the classification accuracy of images that were from not-pro-ED sources.

VI. DISCUSSION

In this section we discuss insights from some of our failure modes, and ethical considerations that affect how we report and share this work.

A. Understanding the failures

Our Model 2 is trained on more than 100,000 images from a diverse set of pro-ED sources and a set of not-pro-ED sources chosen because they have visually similar imagery.

Fig. 6. Images representative of incorrectly & correctly classified images from sources 9 and 1, respectively (cropped to emphasize central figure and protect identities of subjects)

Fig. 7. Images representative of incorrectly & correctly classified images from blog 12

While the classifier overall has good performance, there remain a substantial number of pro-ED false positives. One potential explanation for this could be class overlap and noisiness of the dataset: for example, a pro-ED selfie is still a selfie. Another explanation is the possibility that images unrelated to ED get re-purposed by the pro-ED community and posted to pro-ED tags, as detailed below.

We now discuss misclassifications in the stress test that we feel shed light into what the classifier is actually learning.

Blog 1 (Lookbook) and source 9 (boudoir) both feature women with a variety of body types, including a fair amount of plus-sized women. Image subjects in 9 wear lingerie and are showing a lot of skin, whereas in 1 they wear normal attire (see 6). It is interesting that the classifier labeled source 9 as pro-ED, likely due to the amount of skin they are showing. It follows that the classifier may still be confusing the presence of more exposed skin and posing with an image being pro-ED. Interestingly, unlike blog 1, the Lookbook blog 2 was classified as pro-ED, though the subjects in both were showing a comparable amount of shit in pictures. One potential explanation for this is that the classifier is still picking up on relative thinness of body parts and somehow weighing that feature more than the fashion-oriented nature of the images on blog 1.

Source 6, the Google search "famine", includes extremely indistinct figures with only a wrist and ankles visible. The relative proportion of the visible body parts indicate a low BMI, though the figures are well-covered. Perhaps this coverage contributed to its classification as not-pro-ED, similar to the lack of coverage of subjects in source 9 may have influenced its (incorrect) pro-ED classification.

Source 12 is blog of a man with anorexia, a demographic that is underrepresented in our training. Uniquely, his blog featured only pictures of one person (presumably the owner of the blog himself), most in fairly casual settings. For example, there were a fair amount of mirror pictures, silly faces, and selfies. Though the classifier made the correct decision overall, it is interesting to compare the types of images it classified correctly and the types of images it did not. For example, several true positive results from this individual's blog were similar to the photo below of a person's thighs (7):

This type of image is particularly common in the pro-ED world, and there are numerous pictures of women holding the exact same pose [20], a good amount of which are represented in our dataset. This demonstrates that our classifier is picking up on more of the canonical pro-ED images. However, of the 100 false negatives, most fell into the following categories: selfies with only his face in view or odd contortions of the man's body displaying the thinness of his limbs and bones (demonstrated in 7). Those images contrast greatly with the poised, stylish composition of the majority of images in our pro-ED dataset. These results, coupled with the findings in [24], suggest that the ED-related image posts of male sufferers from ED are visually different from those of their female counterparts. This suggests that, when creating a dataset for a similar purpose in the future, one should go to more effort to make sure images from all kinds of ED blogs are well represented.

Instagram profiles 3 and 10 featured images of ballerinas and a yoga practitioner, respectively. As a whole, the majority of image subjects in both accounts were super posed and thin. Hence, their pro-ED classification makes sense. Also there is certainly the possibility of class overlap; as [2] found, it is fairly common for the pro-ED community to re-post photos of thin athletes such as ballerinas as pro-ED images.

Sources 2 and 8 include pictures of professional fashion models: one black, one white. Source 5 features the results of a Google image search for "African-American fashion model." Though the images featuring African-American subjects were correctly classified, compared to the performance of sources 2 and 8, it seems they are more likely to be classified as not-pro-ED (despite overall improvement on misclassifying fashion images) as those of white people. One potential explanation is that maybe the classifier is still conflating the darker pigments of a subject's skin with being less pro-ED. This could also be reflected in source 6, the results of a Google searches for "famine," which is classified correctly as not-pro-ED to a high degree (90.45%).

The one pro-ED blog we misclassified (by a small margin), blog 19, has very few images (12 total) that are a mix of thumbnails, webpage-specific icons, and about half pro-ED images. As is common on some social media sites, the thumbnails seem to be the profile pictures of other users. In

the future, this performance suggests that any sort of software tool that makes use of the classifier to analyze blogs needs to pre-process out irrelevant website-specific icons and content (like other users' profile pictures) to eliminate noise.

B. Ethical Considerations

All data collected for our training and test datasets was public data. By convention, and following the precedent set by [20], there is no consent for use required from users who post data to public websites. However, we have chosen to not show or publish a dataset with images from the social media that we considered. While those images may be public now, they are often of people whose age is difficult to ascertain, and may be images that people would want to forget in the future. Furthermore, the subjects and/or posters of those images did not consent to re-posting of their content for another purpose. Instead, we share the hashtags and the process through which we collected images to support future comparable research. To give a sense of the visual content of the images in the dataset, we include public domain images representative of the originals gathered from the sites Pexels, free-images.com, Wikimedia Commons, and Unsplash.

There are also significant challenges in using social media as a dataset with respect to the generalization of the results across demographic groups. There are inherent biases in who uses social media and which platforms they use [1], as well as what demographics are represented in imagery shared related to eating disorders [24]. Our work to date relates to analyzing the social media pages and content itself. However, if there is a goal to use a classifier for a more diagnostic system, it would be important to train it on data that includes images featuring subjects from a wider variety of nationalities, genders and ages, or there is likely to be wide variation in the system's accuracy.

VII. Conclusion

We have reported on an end-to-end approach that learns to recognize imagery associated with eating disorder-related hashtags. Because images are tightly related to eating disorder communities online, building tools that recognize relevant images may be a more robust approach than recognizing the set of hashtags related to these communities at a given point in time.

Our future work involves using the classifier to support a variety of software tools for people with eating disorders or healthcare providers that treat people with eating disorders. These could take the form of browser plug-ins that could characterize the sites that someone visits for analysis later, or to block sights that might be triggering in order to improve patient health outcomes. The analysis of images could be done within the browser itself [14], or a central server could be set up to provide a real-time evaluation of sites.

References

[1] S. Aaron and A. Monica. Social media use in 2018. *Pew Research Center*, 2018.

[2] D. Borzekowski, S. S, W. JL, and P. R. e-ana and e-mia: A content analysis of pro-eating disorder web sites. *American journal of public health*, 100(8):1526–34, 2010. PMID: 20558807.

[3] J. Buchner. The imagehash python library. https://pypi.org/project/ImageHash/. Accessed: 2018-04-30.

[4] S. Chancellor, Y. Kalantidis, J. A. Pater, M. De Choudhury, and D. A. Shamma. Multimodal classification of moderated online pro-eating disorder content. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3213–3226, New York, NY, USA, 2017. ACM.

[5] S. Chancellor, Z. J. Lin, and M. De Choudhury. "this post will just get taken down": Characterizing removed pro-eating disorder social media content. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 1157–1162, New York, NY, USA, 2016. ACM.

[6] S. Chancellor, J. A. Pater, T. Clear, E. Gilbert, and M. De Choudhury. #thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 1201–1213, New York, NY, USA, 2016. ACM.

[7] M. De Choudhury. Anorexia on tumblr: A characterization study. In *Proceedings of the 5th International Conference on Digital Health 2015*, DH '15, pages 43–50, New York, NY, USA, 2015. ACM.

[8] Eating disorder statistics. http://www.anad.org/education-and-awareness/about-eating-disorders/eating-disorders-statistics/. Accessed: 2018-11-23.

[9] J. Ghaznavi and L. D. Taylor. Bones, body parts, and sex appeal: An analysis of #thinspiration images on popular social media. *Body Image*, 14:54 – 61, 2015.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[11] L. He and J. Luo. what makes a pro eating disorder hashtag: Using hashtags to identify pro eating disorder tumblr posts and twitter users. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 3977–3979, Dec 2016.

[12] A. J, M. AJ, W. J, and N. S. Mortality rates in patients with anorexia nervosa and other eating disorders: A meta-analysis of 36 studies. *Archives of General Psychiatry*, 68(7):724–731, 2011.

[13] A. S. Juarascio, A. Shoaib, and C. A. Timko. Pro-eating disorder communities on social networking sites: A content analysis. *Eating Disorders*, 18(5):393–407, 2010. PMID: 20865593.

[14] A. Karpathy. Convnetjs: Deep learning in your browser (2014). *URL http://cs. stanford. edu/people/karpathy/convnetjs*, 2014.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[16] D. Le Grange, S. A. Swanson, S. J. Crow, and K. R. Merikangas. Eating disorder not otherwise specified presentation in the us population. *International Journal of Eating Disorders*, 45(5):711–718, 2012.

[17] B. Mehra, C. Merkel, and A. P. Bishop. The internet for empowerment of minority and marginalized users. *New Media & Society*, 6(6):781–802, 2004.

[18] A. Oksanen, D. Garcia, and P. Räsänen. Proanorexia communities on social media. *Pediatrics*, 137(1):e20153372, 2016.

[19] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch, 2017.

[20] J. A. Pater, O. L. Haimson, N. Andalibi, and E. D. Mynatt. "hunger hurts but starving works": Characterizing the presentation of eating disorders online. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 1185–1200, New York, NY, USA, 2016. ACM.

[21] J. Redmon and A. Farhadi. Yolo9000: better, faster, stronger. *arXiv preprint*, 2017.

[22] M. Tiggemann, O. Churches, L. Mitchell, and Z. Brown. Tweeting weight loss: A comparison of #thinspiration and #fitspiration communities on twitter. *Body Image*, 25:133 – 138, 2018.

[23] T. Wang, M. Brede, A. Ianni, and E. Mentzakis. Detecting and characterizing eating-disorder communities on social media. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, pages 91–100, New York, NY, USA, 2017. ACM.

[24] T. Wooldridge, C. Mok, and S. Chiu. Content analysis of male participation in pro-eating disorder web sites. *Eating Disorders*, 22(2):97–110, 2014. PMID: 24359281.