Internal Medicine Faculty Publications   Internal Medicine

1-5-2022

Tracking Sentiments toward Fat Acceptance over a Decade on Twitter

Sadie Bograd
Paul Laurence Dunbar High School

Benjamin Chen
Paul Laurence Dunbar High School

Ramakanth Kavuluru
University of Kentucky, ramakanth.kavuluru@uky.edu

Repository Citation

Tracking Sentiments toward Fat Acceptance over a Decade on Twitter

Tracking sentiments toward fat acceptance over a decade on Twitter

Sadie Bograd [ORCID]
Paul Laurence Dunbar High School, Lexington, KY USA

Benjamin Chen
Paul Laurence Dunbar High School, Lexington, KY, USA

Ramakanth Kavuluru
Department of Internal Medicine, University of Kentucky, Lexington, KY, USA

Abstract
The fat acceptance (FA) movement aims to counteract weight stigma and discrimination against individuals who are overweight/obese. We developed a supervised neural network model to classify sentiment toward the FA movement in tweets and identify links between FA sentiment and various Twitter user characteristics. We collected any tweet containing either "fat acceptance" or "#fatacceptance" from 2010–2019 and obtained 48,974 unique tweets. We independently labeled 2000 of them and implemented/trained an Average stochastic gradient descent Weight-Dropped Long Short-Term Memory (AWD-LSTM) neural network that incorporates transfer learning from language modeling to automatically identify each tweet's stance toward the FA movement. Our model achieved nearly 80% average precision and recall in classifying "supporting" and "opposing" tweets. Applying this model to the complete dataset, we observed that the majority of tweets at the beginning of the last decade supported FA, but sentiment trended downward until 2016, when support was at its lowest. Overall, public sentiment is negative across Twitter. Users who tweet more about FA or use FA-related hashtags are more supportive than general users. Our findings reveal both challenges to and strengths of the modern FA movement, with implications for those who wish to reduce societal weight stigma.

Keywords
Fat acceptance, sentiment analysis, Twitter data, social media, natural language processing, machine learning, text mining

---

Sadie Bograd and Benjamin Chen contributed equally to this manuscript.

Corresponding author:
Ramakanth Kavuluru, Division of Biomedical Informatics (Department of Internal Medicine) University of Kentucky, 725 Rose Street, Lexington KY 40536, USA.
Email: ramakanth.kavuluru@uky.edu

Introduction

The fat acceptance (FA) movement challenges the prevailing stigma against individuals with fat bodies by questioning claims about the negative impacts of increased body weight on health [1–7] and by promoting respect for and acceptance of individuals who are overweight or obese, including self-respect and self-acceptance.[1–9] Claims made by various members of the FA community include that body weight is a relatively immutable condition wherein obesity is a form of natural variation; that weight stigma is more harmful than obesity itself; that weight loss efforts are futile, damaging, or an expression of internalized weight stigma; and that people can be healthy at every size (HAES).[1–7,10–13]

The FA movement has existed for decades in various forms, with some historians dating its inception to 1967 protests in New York and related grassroots organizing.[2,4] Over the past five decades, the FA movement has exhibited a number of manifestations and strategies, such as the creation of organizations like the National Association to Aid Fat Americans (now called the National Association to Advance Fat Acceptance) and the Fat Underground,[1,2,4,6,7,14] the distribution of zines,[1,2,4,6,15] advocacy for legislative changes that prohibit weight discrimination,[4] and the emergence of fat studies in the world of academia.[2,4,14,16] The FA movement has been increasingly active in the so-called "Fatosphere," a network of blogs that express pro-FA viewpoints,[1,4,6,7,12,17,18] and in general social media.[2]

FA is considered by many to alleviate the harmful effects of weight stigma, here defined as negative attitudes toward (e.g., stereotypes that people with obesity are lazy) and actions against (e.g., exclusion and size discrimination) people who are overweight/obese.[19] Weight bias and stigma have been documented across genders,[20–23] body weights,[21–23] and ages[22,24] and in numerous institutions and mediums, including among health professionals[25–31] and dietitians.[32] Prior research indicates that such stigma has significant negative ramifications. Apart from the numerous ethical dilemmas it raises, weight stigma and the internalization of weight bias have been found to be harmful to both mental[33–44] and physical health.[31,33,35,36,40,42–46] Such stigma can also cause discrimination against overweight or obese individuals in various areas,[47,48] including in employment,[49–52] and has particularly pernicious effects in the healthcare industry, resulting in poorer standards of care for individuals with fat bodies.[25–28,31,32,36]

Empirical studies and theoretical discourse indicate that participation in the FA movement increases feelings of empowerment, body esteem, and other positive mental health outcomes, especially for overweight or obese women, helping address the deleterious impacts of weight stigma.[5,6,8,11,17,53] Participation in FA communities, especially online, may help individuals "understand, negotiate and, at times, reject" a marginalized identity.[18] Beyond the individual benefits of participation in the FA movement, the spread of FA ideologies may counteract the societal harms of widespread weight stigma, for example, in the healthcare sector. However, the FA movement is not well-studied and has generated significant backlash among the general public.[13,54–56]

Given that one goal of the FA movement is the spread of FA sentiment to a larger segment of the population (i.e., in order to facilitate the inclusion and acceptance of fat bodies in a wide variety of spaces, FA ideology must extend beyond the current members of the FA movement), and that FA has numerous ethical, sociological, and public health implications, it is critical that we have an understanding of how people outside of specific activist user groups feel about the movement. As with any other social movement, understanding what proportion of individuals support or oppose a given movement and how individuals with various stances tend to talk about a movement can lead to rich

insights that are valuable in and of themselves and that can inform FA activists, researchers, medical professionals, and policymakers.

However, despite the potential impact of the movement and the significant volume of media discussing it, there appears to be no widespread analysis of current sentiment on the FA movement. Although some limited research has been conducted on discussions of obesity in social media (SM),[57] there has been no research on societal perceptions of the FA movement specifically, nor any attempt to classify individual SM posts/users as supporting or opposing this movement or to identify common characteristics among pro- and anti-FA posts. This lack of research limits our understanding of how FA has changed over time and what challenges individuals may face in joining and participating in FA; understanding societal perceptions of FA is a crucial first step to encouraging support for FA and reducing weight stigma among the general public.

Our study aimed to remedy this dearth of quantitative research by classifying FA discourse on Twitter, a widely used SM site. We employed an Average stochastic gradient descent Weight-Dropped Long Short-Term Memory (AWD-LSTM) neural network trained with language modeling-based transfer learning methods to efficiently analyze large volumes of Twitter data. By doing this, we were able to develop a broad understanding of the evolution of sentiment on Twitter regarding the FA movement and of common user/post characteristics of pro- and anti-FA discourse.

The use of computational methodologies in general, and sentiment analysis in particular, to analyze social media discourse is not novel. Stance detection, or the automatic classification of a text as supporting/opposing a given subject, is a fairly widespread suite of techniques[58] that has been used to analyze SM users' opinions on everything from e-cigarettes[59] to politics.[60] Stance detection methods include support vector machines, logistic regression, decision trees, and recent advances in deep neural networks, including convolutional neural networks and long short-term memory networks. That said, we emphasize that this study is unique in its focus on the FA movement, an under-studied movement with significant implications for the mental and physical well-being of millions of overweight individuals. Our goal was to analyze the evolution of public FA discourse on SM so that any derived insights can inform the development of social media-based strategies to reduce weight stigma.

Materials and Methods

Our project was completed using Python, along with various APIs and packages. Twitter's premium API service was used to download all tweets from Jan 1, 2010 to December 31 2019 that matched the phrase "fat acceptance" or tag #fatacceptance. We removed those tweets that contained fewer than five words (not counting URLs and hashtags). After removing those, we ended up with 84 192 tweets (48 974 not counting retweets). 3959 of these tweets contained media (images, videos, or GIFs) and 33 240 contained links (including retweets). 36 445 of the unique tweets came from users who had fewer than 10 FA-related tweets, while 12 529 came from users with 10 or more FA-related tweets. This is the final dataset that is explored in the rest of this effort. Before we proceed, we disclose that during the process of obtaining funding that helped execute this project, the institutional IRB deemed that this type of research does not meet the definition of human subjects and thus does not require additional IRB review based on these two criteria: (1) The data is publicly available; and (2) there is no interaction or intervention with subjects. Nevertheless, we only report aggregate metrics in the manuscript without referring to any particular tweets or users.

Our goal was to classify each tweet as supporting the FA movement, opposing it, or holding an unclear/neutral opinion toward it. Our overall objective was to use this classification scheme to study trends across time and characterize messages that belong to the supporting and opposing groups. Manual annotation of nearly 85K tweets is impractical, and hence we set out to build a semi-

supervised machine learned model that was first trained to perform language modeling on unlabeled data, then fine-tuned to classify from annotated data and predict on unseen instances. This way, we were able to scale the annotation through the model.

At this juncture, we needed to build a smaller annotated dataset to train the model. After a manual examination of the dataset, we developed guidelines on what tweet traits constitute (1) supporting, (2) opposing, or (3) neutral tweets, so that we could independently label the tweets with minimal disagreement (Table 1).

Table 1. Labeling guidelines created and followed for the annotation process.

Label	Guidelines for annotating FA-related tweets	Example Tweets
Supporting	Explicitly states support for fat acceptance (or health at every size) movement. OR Implicitly expresses support for movement by: arguing that obesity is not unhealthy, that dieting is unproductive, or that fat shaming is immoral/unhelpful without qualifiers; OR criticizing public figures, media, etc., that promote dieting or perceived fat shaming, or that practice fat acceptance "wrong"; OR advertising lectures, blogs, products, events, etc., about body positivity, intuitive eating, anti-dieting, HAES, and related topics; OR using (numerous) pro-fat acceptance hashtags like #allbodiesaregoodbodies #bodyposi #intuitiveeating #haes	Finding fat acceptance was so integral to rebuilding my mental health and my will to fucking LIVE. I was close to the brink, back then Progress to date - stopping painting now as eyes hurt!! Work in progress - "Unapologetically Yours", [name], watercolour, 770mm x 570mm, 21/06/19 #fatpositive #bodydiversity #fatacceptance #fatandbeautiful #bodypositive #bodypositivity #bodyconfidence @[name] The women who started the nonprofit "The Body Positive". Total erasure of fat acceptance work going back decades
Opposing	Argues that fat/obese people are gluttonous, lazy, or otherwise morally lacking. OR Believes that fat acceptance encourages fatness/discourages people from trying to improve their health. OR Expresses that fat shaming is bad, but that being fat is still unhealthy/fat acceptance is similarly unproductive. OR Shares links to articles, videos, and other media that criticize the movement. OR Makes jokes about fat acceptance being "the only movement without movement," fat acceptance activists dying/becoming ill, etc.	I hate fat acceptance, and I'm fat, that's why I'm trying to lose weight. Keep me shamed so it will help me lose lbs [link] I liked a @YouTube video [link] Why I Hate 'Fat Acceptance' Let me start by saying that both fat Acceptance and fat shaming is bullshit. #fatshaming #fatacceptance
Unclear/neutral	Discusses the perspectives of others on fat acceptance without taking a stance themselves. OR Appears to be attempted objective/genuinely unsure about the movement and expresses desire to discuss or learn more about it. OR Describes an event that is pro-fat acceptance without expressing a view on it, especially for larger news organizations. OR Is only tangentially related to fat acceptance, uses incomprehensible language, or is clearly a bot	"Huge" is advance for fat acceptance in Hollywood [link] Doctors warn that spin the bottle can cause fat acceptance in those under 15 Probably because you don't really see anorexia being promoted by the general public, but more of an unfortunate consequence of the way most models look. You have a fat acceptance movement, but you don't have an anorexia acceptance movement

To increase the diversity of the annotated dataset, we decided to make the dataset from unique tweets by excluding retweets. However, retweets were included in the final analyses, given that retweets quantify the amplification of particular themes. Among unique tweets, we (first and second authors) selected a random set of 100 tweets from the dataset to independently label them into the three categories. Using these two sets of 100 labels, we calculated Cohen's kappa and came to an agreement on the disagreeing labels, using these to refine our labeling guidelines. We repeated this process until we achieved a Cohen's kappa >0.7, which took four rounds of labeling. Finally, we took a random sample of 2000 unique tweets and labeled them individually, reaching a Cohen's kappa of 0.81; we subsequently consolidated the label disagreements via a face-to-face discussion. As per rules of thumb for inter-annotator agreement, kappa of 0.81 indicates "almost perfect" agreement.[61]

We experimented with a set of machine learning algorithms and chose a state-of-the-art AWD-LSTM recurrent neural network-based model, ULMFiT (Universal Language Model Fine-tuning for Text Classification),[62] that relies on language-model-based transfer learning to train classifiers. ULMFiT leverages a language modeling-based objective (based on the AWD-LSTM) to come up with better contextualized representations of words in an input document to aid in downstream classification tasks. This essentially handles ambiguities associated with polysemy and homonymy and imbues more contextualized signals into the prediction task. We compared the ULMFiT model with more traditional linear (logistic regression, support vector machines) and nonlinear (random forests, nearest neighbors) baseline classifiers.

The full annotated dataset was split into 1400 training tweets, 300 validation tweets, and 300 test set tweets. We reserve the test set tweets until the very end to ensure performance is assessed on unseen tweets. Using the fast.ai package provided by the authors of the ULMFiT paper, we trained a language model on all 48 974 unlabeled tweets (not counting retweets) in the dataset. Then, we created a classifier using that language model, trained on the 1400 training tweets,[*] and checked the validation error using the 300 validation tweets. We tweaked the number of epochs, the learning rate, dropout rate, and the weight decay of the language model and classifier to reach the lowest error rate possible on the validation dataset. Then, we took that model and tested it with the 300 test tweets to get the performance metrics (accuracy, precision, recall, and F-score[†]) of the model. We did this experiment 60 different times with different random train/validation/test splits of the dataset to arrive at confidence intervals for the metrics reported. Figure 1 provides a visualization of this process.

The classifier built with the full annotated dataset was then applied to the full dataset of nearly 85k tweets (including retweets). We grouped tweets on a weekly and yearly basis to study proportions of supporting and opposing messages across time over a decade (from 2010 to 2019). We used a Python script to determine the presence of various links, images, and hashtags to examine how the percentage of supporting/opposing tweets changed among tweets with specific media or hashtags. We also calculated the average number of likes and retweets for supporting and opposing tweets. Taking advantage of a Twitter user's account creation date, for each user in our dataset, we computed the ratio of the number of stance-expressing tweets (opposing or supporting) to the count

---

* Our model also appended the tweeter's biography to each of their tweet's contents as the biography seemed to contain complementary signals that helped with the classification process (based on validation experiments).

† Precision is the percentage of tweets that the model predicted as belonging to a class are in fact from that class based on ground truth labels. Recall is the percentage of tweets belonging to a class in the input that were correctly identified as such by the model. F-score is the harmonic mean of the precision and recall. Accuracy is the proportion of correctly classified tweets among all input tweets.
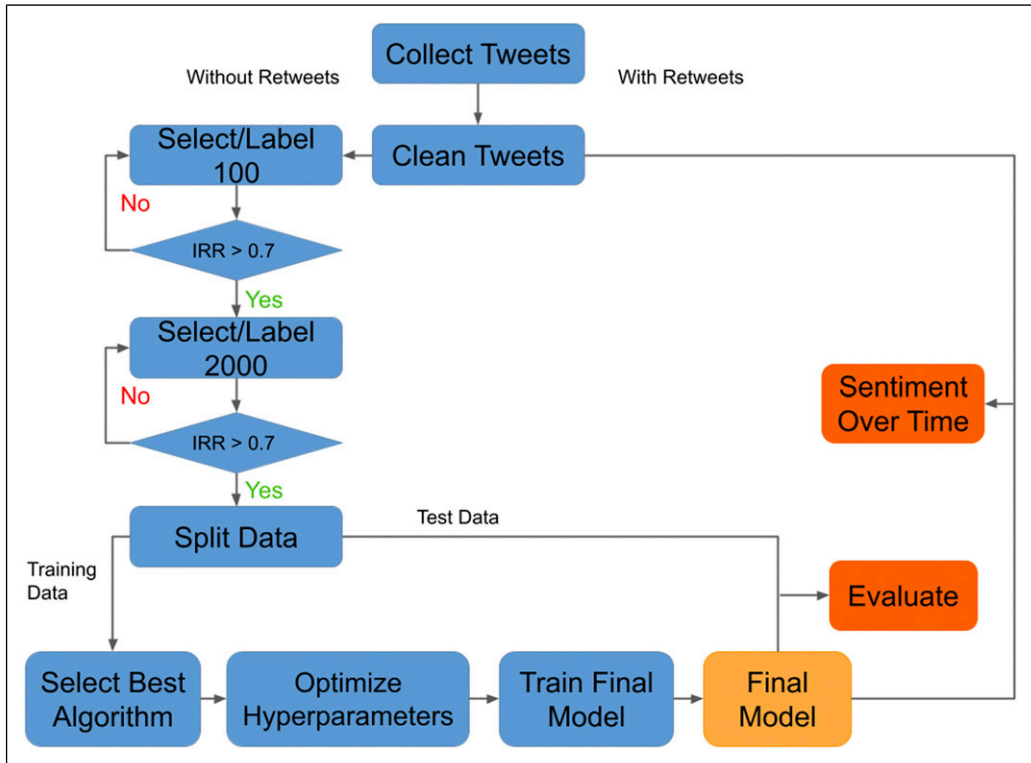
Figure 1. High level flowchart of dataset creation and FA sentiment model evaluation.

of the neutral tweets they authored. Our goal was to assess if users take more opposing/supporting stances as they stay longer on Twitter. To this end, we binned users based on the account age in years (time between account creation and first FA-related tweet) and examined the rate of stance-expressing tweets in each bin. We also grouped users based on the frequency of their FA-related tweeting activity to assess if tweeting more (or less) correlates with a particular stance. We identified the frequency with which certain terms/phrases appeared in specific years; these terms were identified during the manual labeling process as potentially representative of shifts in FA discourse.

Results

Before conducting analyses of the entire database of FA tweets, we tested our model's performance on the 300 test tweets. The ULMFiT model achieved a 95% confidence interval for accuracy of 75.86% ± 0.60%. Other tested algorithms achieved lower accuracies (Table 2), and as such we do not discuss them further. However, accuracy is not very informative when all classes are not equally relevant. As we care about the opposing and supporting classes more than the neutral class, examining class-level metrics is more important. These results are as shown in Table 3 for the ULMFiT model. We compute precision (positive predictive value), recall (sensitivity), and their harmonic mean (popularly called *F*-score) for each class. If we focus exclusively on the opposing and supporting classes, the scores are better: the mean macro average of precision of the two classes is

Table 2. Initial performance metrics for all models with a single run.

Model	Accuracy	F1 (macro)
Stratified guess	0.373	0.282
Decision tree	0.557	0.504
AdaBoost	0.663	0.585
K nearest neighbors	0.717	0.638
Random forest	0.720	0.638
Logistic regression	0.740	0.646
Support vector machine (RBF)	0.757	0.657
Support vector machine (SGD)	0.757	0.681
ULMFiT	0.842	0.782

Table 3. Average performance metrics of the ULMFiT model for the supporting, opposing, and neutral classes based on 60 random dataset splits.

	Precision	Recall	F-score
Supporting	85.64% ± 1.09%	66.58% ± 1.42%	74.68% ± 0.86%
Opposing	71.82% ± 0.73%	93.46% ± 0.80%	81.16% ± 0.49%
Unclear/neutral	77.64% ± 3.32%	33.73% ± 1.82%	46.28% ± 1.86%
Avg. of supporting and opposing	78.73% ± 0.54%	80.02% ± 0.70%	77.92% ± 0.64%

78.73%, and the mean average recall is 80.02%[‡]. This aligns with what we see in the confusion matrix (Figure 2): supporting and opposing tweets were barely ever labeled as neutral, while neutral tweets were commonly mislabeled as supporting or opposing. This explains the performance improvement we see after disregarding the neutral class. However, it is not entirely clear why the classifier struggles with the neutral class.

Having ascertained the accuracy of the models, the ULMFiT model was applied to the broader dataset of 84 192 tweets (including retweets) to determine changes in FA sentiment on Twitter over time and among user groups.

Overall, 60% of tweets were identified as opposing the FA movement (or negative), 33% as supporting the FA movement (or positive), and 7% as unclear/undetermined. On a yearly basis, the overall percentage of tweets identified as negative increased from 2010 (21.3% negative) to 2016 (73.7% negative) (Table 4), then decreased to approximately 64% negative in the following years. Since 2014, more tweets are labeled by the model as negative than as positive. Additionally, the number of FA tweets per year increased over time: 2773 tweets were collected from 2010 as opposed to 22 643 from 2019 (Table 4), and of the 102 weeks in which more than 250 tweets were collected, only one occurred before 2014 and only four before 2015 (Figure 3).

On a week-by-week basis, of the 522 weeks for which tweets were collected, there were 18 weeks in which >500 tweets were collected. Of those 18 weeks, 12 were predominantly negative

---

‡ Averaging of class-specific metrics (precision, recall, F-score) of certain classes of interest is often called macro averaging because it gives equal weight to each constituent class in the average. This is in contrast with micro averaging where errors are grouped in a single confusion matrix across all classes leading to a weighting scheme that aligns with class sizes.
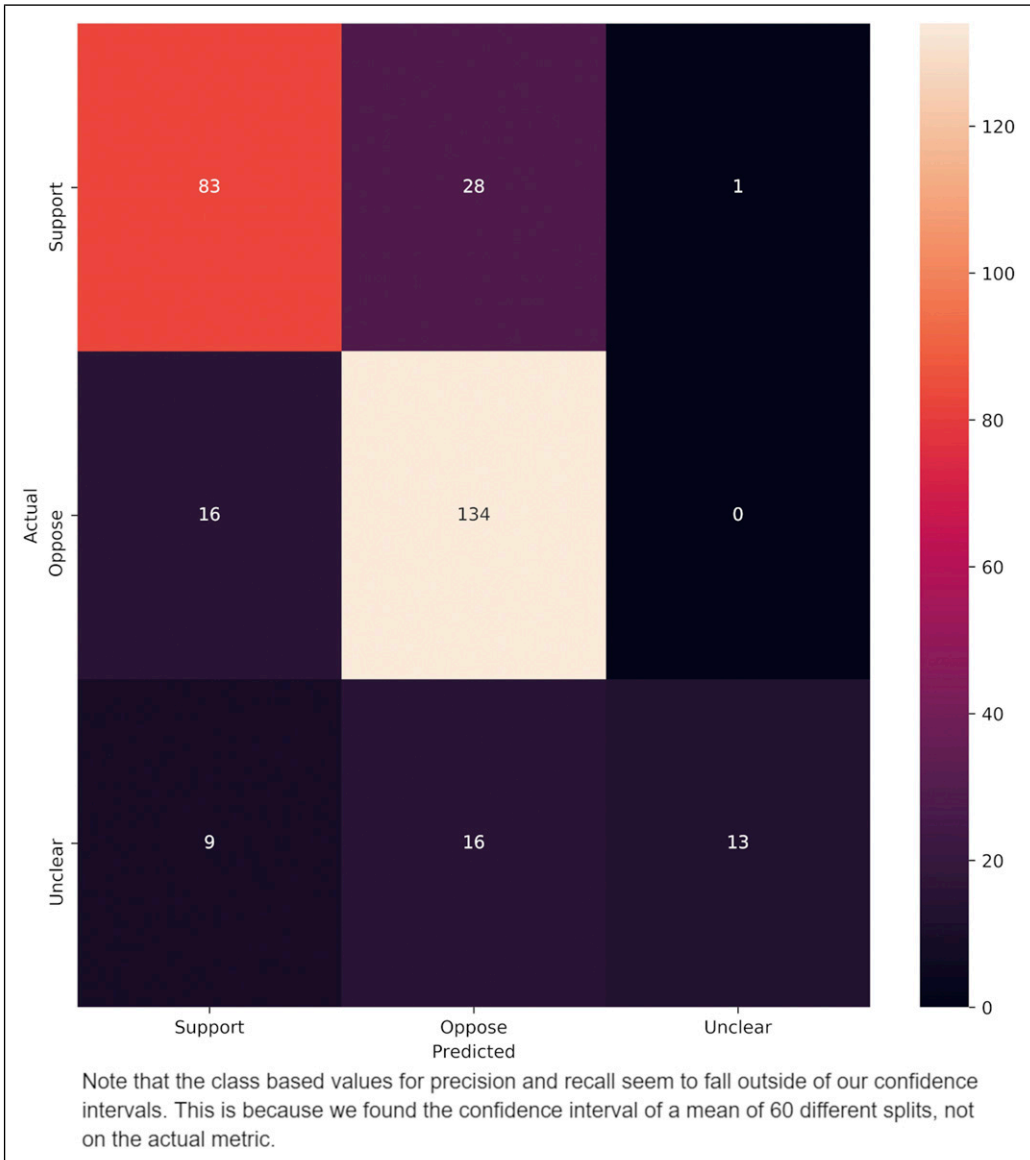
Figure 2. Confusion matrix of errors in 3-way classification of tweets used in our analyses.

(i.e., the majority of the tweets posted that week were labeled as negative by the model), five were predominantly positive, and one was predominantly neutral. Of the 102 weeks in which more than 250 tweets were collected, 82 were majority negative and 18 were majority positive. Additionally, all of the weeks in which there were over 250 tweets and the majority were positive occurred during or after 2017 (Figure 3).

36.3% of non-neutral tweets that included links were labeled by the model as supporting the FA movement, while 63.7% were labeled as opposing. Of the non-neutral tweets that included links,

Table 4. Sentiment distribution of tweets per year.

Year	% supporting	% opposing	% neutral/unclear
2010	36.4 (n = 1008)	21.3 (n = 592)	42.3 (n = 1173)
2011	63.0 (n = 1221)	28.5 (n = 552)	8.6 (n = 166)
2012	58.2 (n = 1529)	34.7 (n = 913)	7.1 (n = 187)
2013	50.4 (n = 1656)	37.6 (n = 1235)	11.9 (n = 392)
2014	43.2 (n = 2311)	51.0 (n = 2729)	5.7 (n = 306)
2015	21.4 (n = 2834)	64.2 (n = 5787)	4.3 (n = 391)
2016	20.1 (n = 1750)	73.7 (n = 6392)	6.1 (n = 526)
2017	31.4 (n = 3927)	62.8 (n = 7852)	5.8 (n = 726)
2018	30.3 (n = 4665)	64.2 (n = 9879)	5.5 (n = 850)
2019	30.6 (n = 6929)	64.5 (n = 14602)	4.9 (n = 1112)
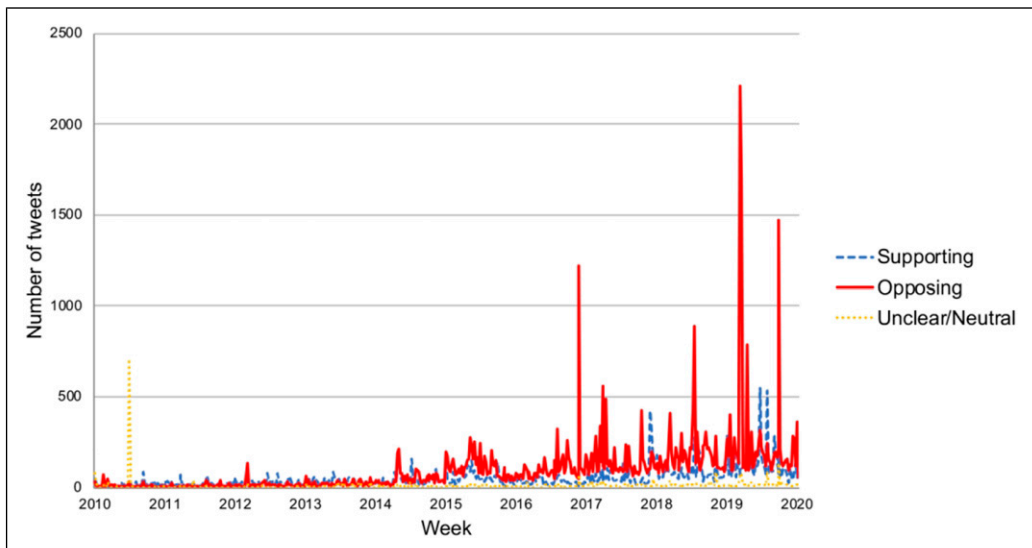
Figure 3. Weekly distribution of FA tweets split across sentiment classes.

Table 5. Sentiment distribution by elements within tweets that indicate a supporting/opposing view.

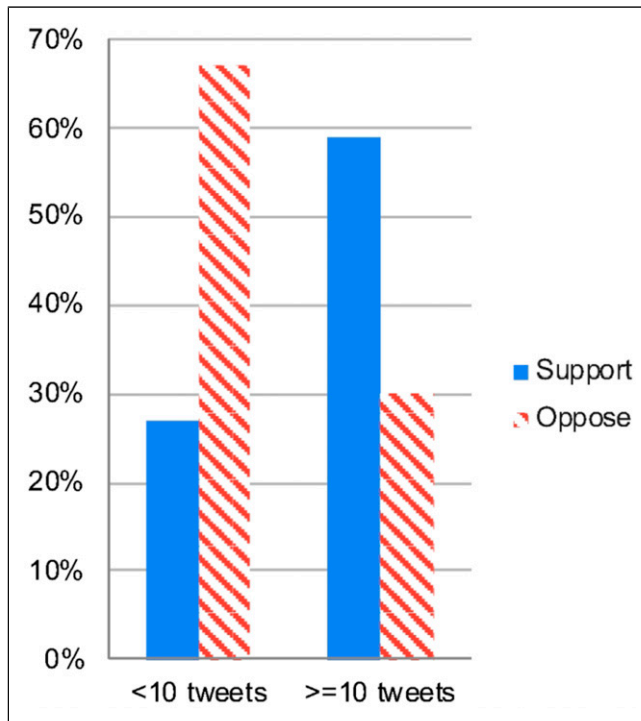| Tweet content elements	% supporting	% opposing
Links (n = 33240)	36.3	63.7
Media (images, videos, GIFs) (n = 3959)	16.4	83.6
YouTube (n = 7006)	14.2	85.8
Tumblr (n = 328)	69.0	31.0
#haes (n = 1868)	66.5	33.5
#fatacceptance (n = 13807)	55.3	44.7
"fat acceptance" (n = 64447)	30.7	69.3

Figure 4. User stances about FA based on tweet activity on the topic.

14.2% of those that linked to YouTube were labeled positive and 69.0% of those that linked to Tumblr were labeled positive. Additionally, 16.4% of non-neutral tweets that included media (images, videos, and GIFs) were labeled as positive. Tweets that included the hashtag #haes (meaning "health at every size") or #fatacceptance were more likely to be labeled positive than those that simply included the term "fat acceptance" (66.5% and 55.3% positive, respectively, vs. 30.7% positive) (Table 5).

Users who tweeted more about FA were more likely to support the movement. 27% of users with fewer than 10 FA tweets supported the movement, while users with 10 or more FA tweets supported FA 59% of the time (Figure 4). Pro-FA tweets also tended to be slightly more popular. On average, supportive tweets had 2.46 likes and 0.65 retweets, while opposing tweets had 2.06 likes and 0.47 retweets. Furthermore, users who had been on Twitter longer had more polarized opinions, insofar as the likelihood of supporting or opposing FA as opposed to feeling neutral generally increased as the age of a user's account increased. For example, the ratio of supportive tweets to neutral/unclear tweets for users whose accounts had existed for <1 year was 1.95:1. For users whose accounts had existed for 10 years, this ratio was 21.6:1. Similarly, the ratio of negative opinions to neutral/unclear opinions went from 5.4:1 for accounts existing for <1 year to 18.2:1 for accounts existing for 10 years (Table 6).

Manual examination of the dataset indicated that the release of the television show Huge in 2010 might have contributed to an increase in neutral tweets during that year, as many news organizations tweeted some variation of "Huge marks advance for fat acceptance in Hollywood." Computational

Table 6. Ratio of supporting/opposing to neutral/unclear tweets by account years on Twitter.

Twitter account age (years)	Supporting/unclear or neutral	Opposing/unclear or neutral	Supporting + opposing/Unclear or neutral
0	1.95	5.44	7.39
1	3.19	6.44	9.63
2	3.27	5.93	9.21
3	5.30	9.78	15.09
4	9.08	17.03	26.11
5	8.55	15.01	23.56
6	9.87	16.40	26.27
7	11.53	19.02	30.55
8	14.98	19.69	34.67
9	12.99	14.25	27.24
10	21.64	18.16	39.81
11	13.15	8.50	21.65
12	40.50	21.50	62.00

analysis of the dataset appeared to confirm this impression. 867 tweets from 2010, or 31.3% of all FA tweets that year, included the word "Huge" (capitalization-sensitive). 380 tweets from 2010, or 13.7% of all FA tweets that year, included the specific phrase "huge marks advance" (not capitalization-sensitive). The latter number likely does not capture all tweets related to the TV show due to variations in formatting; as such, the true number of tweets related to the show is likely between 380 and 867.

We also conducted a straightforward location-based analysis of our dataset. There were 860 unique tweets that provided a location in our full tweet dataset. 256 were foreign, and 604 were from the US. Taking the tweets from the US, then subtracting opposing tweets from supporting tweets gave a net sentiment score for each state. That net sentiment was then compared to adult obesity rates (https://stateofchildhoodobesity.org/adult-obesity/) by fitting a linear regression model. A negative correlation was found ($\beta = -0.312$, $p < 0.05$). Interestingly, this implies that the more a state supports the FA movement, the lower that state's obesity rates. This lends some credibility to the idea that the FA movement may be more effective than fat shaming at preventing obesity. Given that location information is only available for 604 US tweets (1.2% of the full dataset), one could argue that this sample is too sparse to reliably assess correlations of this nature. Unfortunately, this is a limitation of Twitter, as users must opt in to provide their locations, meaning the proportion of tweets with location data attached is always extremely low.

Discussion

The application of our model to the broader database of FA tweets revealed multiple insights. Ultimately, although positive communities exist online, specifically in blogs (which comprise the so-called "fatosphere"), the results of our model indicate that there is no broad acceptance of the fat acceptance movement on Twitter, a space where a multitude of views can be expressed. Nuanced debates about the fat acceptance movement are typically outweighed by more extreme anti-FA discourse, especially in recent years in our dataset; since 2014, the majority of FA tweets every year

were labeled "opposing" by our model. Our findings suggest that various "spikes" in FA discourse on a week-by-week basis could be attributable to specific events; for example, the release of FA-related TV show Huge appears to have caused the increase in neutral tweets in 2010 ($n = 1173$ tweets were neutral, or 42.3% of all FA tweets in 2010), as 380 tweets from 2010 referenced the phrase "Huge marks advance for fat acceptance in Hollywood" and 867 tweets used the word "Huge" (capitalization-sensitive). However, we cannot speculate on the reasons for this general shift in FA discourse without further research. Whatever the cause of these changes in sentiment, the model strongly indicates that the majority of Twitter users discussing the FA movement are opposed to it. Although we are unable to determine whether users tweeting about FA are representative of all Twitter users or the American public, our findings align with current research into the prevalence of weight stigma.

These findings about popular sentiment toward the FA movement have some negative implications for people who would benefit from being part of the FA movement. They indicate that difficulties may exist for overweight or obese individuals who would benefit from locating a supportive pro-FA community. This is true not only because the majority of FA tweets are negative but also because the amount of new anti-FA content is more likely to spike on a week-by-week basis, as suggested by there being significantly more weeks in which there were over 250 or 500 new negative tweets than weeks in which there were over 250 or 500 new positive ones. Such results could suggest that the average Twitter user is less likely to find pro-FA content without intentionally seeking it out. While the FA movement was created to combat weight stigma and bias, and to foster a positive community for people who experience discrimination in their daily lives, discussions of FA on Twitter largely take the form of further anti-fat criticism—there appears to be more content on Twitter about why the FA movement is bad than there is content created by members of the FA movement itself.

That said, the use of hashtags appears to have assisted in the development of pro-FA communities, as tweets referencing #fatacceptance and #haes are mostly positive (55.3% and 66.5% positive, respectively). These percentages may be even higher for less-used or more niche hashtags. Furthermore, users who tweet more about FA are more likely to support the movement. This could suggest that a significant portion of the positive FA discourse on Twitter comes from FA activists and others who are highly involved in the movement, whereas negative FA discourse may come from people who are less invested in the subject. Pro-FA tweets also tend to receive more likes and retweets than anti-FA tweets, highlighting that pro-FA Twitter users often find a receptive audience within the FA movement, even as they are subject to general anti-fat criticism from other SM users. This all indicates the existence of reliable sources of pro-FA content for individuals looking for supportive content and communities so long as they know how to find them, even if the majority of all FA content is negative.

People seeking to find supportive communities and raise support for FA may experience more difficulties on YouTube and fewer on Tumblr, as the majority of tweets linking to YouTube were labeled negative (85.8%), and the majority linking to Tumblr were labeled positive (69.0%). As a general concept, this discrepancy could be indicative of the different user bases of these social media sites. However, further speculation about the cause of this discrepancy is outside the scope of this study and presents a potential site for further research.

This research has some limitations. We constructed a dataset using only two annotators and 2000 tweets. As such, the training dataset may not have been sufficiently large to reflect the diversity of our dataset of 48 974 different tweets (excluding retweets). The average model precision and recall were around 80% and an ideal classifier would have had results in the high nineties. However, on nuanced tasks such as sentiment analysis, this is inherent in the nature of machine learning methods

that are indispensable to scale predictions to thousands of instances.[§] Therefore, while we believe the general trends discovered by the model to be accurate, the exact percentages of tweets in support of and opposition to FA may be subject to some minor variations. In addition, the community of regular Twitter users cannot be taken as a proxy for the general public, and it is unclear in this specific instance how reflective the opinions expressed by FA-related tweets are of those of broader society.

Still, our model offers insight into common perspectives on the FA movement, and may be useful for FA activists and supporters, medical experts, policymakers, and all others with an interest in the FA movement or in combating weight stigma. This study highlights that the FA movement has not yet achieved widespread societal acceptance, and that often-fatphobic anti-FA content is more common than pro-FA content; at the same time, it demonstrates that members of the FA movement have been able to create generally supportive communities through the use of hashtags and FA activist accounts. Both of these findings have implications for future efforts to reduce weight stigma. Independently, too, the development of this model displays the powerful potential of the application of machine learning methods to social science questions.

Contributorship

SB conceived the study, wrote the first draft of the manuscript barring the methods section, and conducted pertinent literature review. BC designed and implemented the classification models needed and wrote the first draft of the methods section. RK guided the study, focusing on methods and evaluation, and edited the manuscript. All three authors contributed to analyses and interpretation of results.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iD

Sadie Bograd ⓘ https://orcid.org/0000-0002-8018-7906

References

1. Afful AA and Ricciardelli R. Shaping the online fat acceptance movement: talking about body image and beauty standards. *J Gend Stud* 2015; 24(4): 453–472.
2. Bolden SE. *Unsettling Boundaries: (Pre-)Digital Fat Activism, Fatphobia, and Enclave Ambivalence*. Syracuse, NY: Syracuse University, 2018.

§ We note that these scores are closer to the inter-annotator agreement of 0.81, which was achieved after multiple small-batch annotations with a clear set of guidelines. This leads us to believe that even if two humans annotated the full dataset, they would not have agreed with each other 100% of the time. Furthermore, given the size of our training dataset (2000 tweets), extremely high scores would often lead to generalizability issues on unseen examples. As such, the model scores reported are reasonable/acceptable.

3. Cain P. *Responding to the Resistance: A Critical Discursive Analysis of Women's Engagement with Health at Every Size and Fat Acceptance Messages*. Murdoch, WA: Murdoch University, 2014.
4. Cooper C. *What's Fat Activism?*, 2008.
5. Donaghue N. and Clemitshaw A. 'I'm totally smart and a feminist…and yet I want to be a waif': Exploring ambivalence towards the thin ideal within the fat acceptance movementand yet I want to be a waif': Exploring ambivalence towards the thin ideal within the fat acceptance movement. *Women's Stud Int Forum* 2012; 35(6): 415–425.
6. Forthun JL. *Weighing in: Coping with Stigmatization through the Fat Acceptance Movement and Weight Loss Surgery.* Linfield College, 2012.
7. Robinson KM. *The Fat Acceptance Movement Contesting Fatness as Illness, 1969-1998*. PhD thesis. Madison, WI: The University of Wisconsin - Madison, 2014.
8. McKinley NM. Resisting body dissatisfaction: fat women who endorse fat acceptance. *Body Image* 2004; 1(2): 213–219.
9. Striley KM and Hutchens S. Liberation from thinness culture: Motivations for joining fat acceptance movements. *Fat Stud* 2020; 9(3): 296–308.
10. Bacon L and Aphramor L. Weight science: evaluating the evidence for a paradigm shift. *Nutr J* 2011; 10: 9.
11. Matacin ML and Simone M. Advocating for fat activism in a therapeutic context. *Women Therapy* 2019; 42(1–2): 200–215.
12. Rabin RC. In the fatosphere, big is in, or at least accepted. The New York Times, January 22,2008.
13. Saguy AC and Riley KW. Weighing both sides: morality, mortality, and framing contests over obesity. *J Health Polit Policy Law* 2005; 30(5): 869–923.
14. Rothblum E. D. Why a journal on fat studies? *Fat Stud* 2012; 1(1): 3–5.
15. Wann M. Fat oppression as a source of inspiration: a case history of the FAT!SO? Book. *Fat Stud* 2017; 6(1): 67–76.
16. Watkins PL, Farrell AE and Doyle Hugmeyer A. Teaching fat studies: from conception to reception. *Fat Stud* 2012; 1(2): 180–194.
17. Dickins M, Thomas SL, King B, et al. The role of the fatosphere in fat adults' responses to obesity stigma. *Qual Health Res* 2011; 21(12): 1679–1691.
18. Smith N, Wickes R and Underwood M. Managing a marginalised identity in pro-anorexia and fat acceptance cybercommunities. *J Sociol* 2013; 51(4): 950–967.
19. Eisenberg D, Noria S, Grover B, et al. ASMBS position statement on weight bias and stigma. *Surg Obes Relat Dis* 2019; 15(6): 814–821.
20. Fikkan JL and Rothblum ED. Is fat a feminist issue? Exploring the gendered nature of weight bias. *Sex Roles* 2012; 66(9): 575–592.
21. Puhl RM, Himmelstein MS and Quinn DM. Internalizing weight stigma: prevalence and sociodemographic considerations in US adults. *Obesity* 2018; 26(1): 167–175.
22. Pearl RL, Himmelstein MS, Puhl RM, et al. Weight bias internalization in a commercial weight management sample: prevalence and correlates. *Obes Sci Pract* 2019; 5(4): 342–353.
23. Hilbert A, Baldofski S, Zenger M, et al. Weight bias internalization scale: psychometric properties and population norms. *PLoS One* 2014; 9(1): e86303.
24. Latner JD and Stunkard AJ. Getting worse: the stigmatization of obese children. *Obes Res* 2003; 11(3): 452–456.
25. Flint S. Obesity stigma: prevalence and impact in healthcare. *Br J Obes* 2015; 1: 14–18.
26. Forhan M and Salas XR. Inequities in healthcare: a review of bias and discrimination in obesity treatment. *Can J Diabetes* 2013; 37(3): 205–209.

27. MacLean L, Edwards N, Garrard M, et al. Obesity, stigma and public health planning. *Health Promot Int* 2009; 24(1): 88–93.

28. Phelan SM, Burgess DJ, Yeazel MW, et al. Impact of weight bias and stigma on quality of care and outcomes for patients with obesity. *Obes Rev* 2015; 16(4): 319–326.

29. Schwartz MB, Chambliss HON, Brownell KD, et al. Weight bias among health professionals specializing in obesity. *Obes Res* 2003; 11(9): 1033–1039.

30. Teachman B and Brownell K. Implicit anti-fat bias among health professionals: is anyone immune? *Int J Obesity* 2001; 25(10): 1525–1531.

31. Tomiyama AJ, Carr D, Granberg EM, et al. How and why weight stigma drives the obesity 'epidemic' and harms health. *BMC Med* 2018; 16(1): 123.

32. Diversi TM, Hughes R and Burke KJ. The prevalence and practice impact of weight bias amongst Australian dietitians. *Obes Sci Pract* 2016; 2(4): 456–465.

33. Durso LE and Latner JD. Understanding self-directed stigma: development of the weight bias internalization scale. *Obesity* 2008; 16(S2): S80–S86.

34. Hatzenbuehler ML, Keyes KM and Hasin DS. Associations between perceived weight discrimination and the prevalence of psychiatric disorders in the general population. *Obesity* 2009; 17(11): 2033–2039.

35. Pearl RL and Puhl RM. Weight bias internalization and health: a systematic review. *Obes Rev* 2018; 19(8): 1141–1163.

36. Puhl RM and Heuer CA. Obesity stigma: important considerations for public health. *Am J Public Health* 2010; 100(6): 1019–1028.

37. Puhl RM, Moss-Racusin CA and Schwartz MB. Internalization of weight bias: implications for binge eating and emotional well-being*. *Obesity* 2007; 15(1): 19–23.

38. Durso LE, Latner JD and Ciao AC. Weight bias internalization in treatment-seeking overweight adults: Psychometric validation and associations with self-esteem, body image, and mood symptoms. *Eat Behaviors* 2016; 21: 104–108.

39. Hain B, Langer L, Hünnemeyer K, et al. Translation and validation of the German version of the weight self-stigma questionnaire (WSSQ). *Obes Surg* 2015; 25(4): 750–753.

40. Hilbert A, Braehler E, Haeuser W, et al. Weight bias internalization, core self-evaluation, and health in overweight and obese persons. *Obesity* 2014; 22(1): 79–85.

41. Lent MR, Napolitano MA, Wood GC, et al. Internalized weight bias in weight-loss surgery patients: psychosocial correlates and weight loss outcomes. *Obes Surg* 2014; 24(12): 2195–2199.

42. Pearl RL and Puhl RM. The distinct effects of internalizing weight bias: an experimental study. *Body Image* 2016; 17: 38–42.

43. Pearl RL, White MA and Grilo CM. Weight bias internalization, depression, and self-reported health among overweight binge eating disorder patients. *Obesity* 2014; 22(5): E142–E148.

44. Latner JD, Durso LE and Mond JM. Health and health-related quality of life among treatment-seeking overweight and obese adults: associations with internalized weight bias. *J Eat Disord* 2013; 1: 3.

45. Pearl RL, Wadden TA, Hopkins CM, et al. Association between weight bias internalization and metabolic syndrome among treatment-seeking individuals with obesity. *Obesity* 2017; 25(2): 317–322.

46. Latner JD, Barile JP, Durso LE, et al. Weight and health-related quality of life: the moderating role of weight discrimination and internalized weight bias. *Eat Behaviors* 2014; 15(4): 586–590.

47. Andreyeva T, Puhl RM and Brownell KD. Changes in perceived weight discrimination among Americans, 1995-1996 through 2004-2006. *Obesity* 2008; 16(5): 1129–1134.

48. Puhl RM, Andreyeva T and Brownell KD. Perceptions of weight discrimination: prevalence and comparison to race and gender discrimination in America. *Int J Obes* 2008; 32(6): 992–1000.

49. Maranto CL and Stenoien AF. Weight discrimination: a multidisciplinary analysis. *Empl Responsib Rights J* 2000; 12(1): 9–24.

50. Paul RJ and Townsend JB. Shape up or ship out? Employment discrimination against the overweight. *Empl ResponsibRights J* 1995; 8(2): 133–145.
51. Roehling MV. Weight-based discrimination in employment: psychological and legal aspects. *Pers Psychol* 1999; 52(4): 969–1016.
52. Roehling MV, Roehling PV and Pichler S. The relationship between body weight and perceived weight-related employment discrimination: The role of sex and race. *J Vocat Behav* 2007; 71(2): 300–318.
53. Gailey JA. Fat shame to fat pride: fat women's sexual and dating experiences. *Fat Stud* 2012; 1(1): 114–127.
54. Glen LR. Big Trouble: Are eating disorders the Lavender Menace of the fat acceptance movement? *Bitch: Feminist Response to Pop Culture*, 2008.
55. Shriver L. Lionel shriver: my brother is eating himself to death. The GuardianSect. Lifestyle, 2009.
56. Worth T. Is the Fat Acceptance Movement Bad for Our Health? CNN, 2010. Available from:, https://www.cnn.com/2010/HEALTH/01/06/fat.acceptance/index.html.
57. Chou WYS, Prestin A and Kunath S. Obesity in social media: a mixed methods analysis. *Transl Behav Med* 2014; 4(3): 314–323.
58. Küçük D. Can F. stance detection: a survey. *ACM Comput Surv* 2020; 53(1). 1–37.
59. Kaunang CPS, Amastini F and Mahendra R (eds). *Analyzing Stance and Topic of E-Cigarette Conversations on Twitter: Case Study in Indonesia*. IEEE 11th Annual Computing and Communication Workshop and Conference. CCWC; , 27-30 Jan; 2021.
60. Lai M, Cignarella AT, Hernández Farías DI, et al. Multilingual stance detection in social media political debates. *Comput Speech Lang* 2020; 63: 101075.
61. Landis JR and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1): 159–174.
62. Howard J and Ruder S (eds) Universal Language Model Fine-tuning for Text Classification. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers)*, Volume 1; 2018.