**Methods to Establish Race or Ethnicity of Twitter Users: A Scoping Review**

Su Golder PhD1, Robin Stevens2, Karen O’Connor MSc3, Richard James4, Graciela Gonzalez-Hernandez PhD3

1Department of Health Sciences, University of York, York, United Kingdom

2Annenberg School for Communication & Journalism, University of Southern California, Los Angeles, CA, USA

3Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

4University of Pennsylvania Biomedical Library, University of Pennsylvania, Philadelphia, PA, USA

Corresponding Author:

Su Golder BSc (hons), MSc, FRSA, PhD

Senior Research Fellow

Department of Health Sciences

University of York

York, YO10 5DD

Tel: 01904 321904

Email: [su.golder@york.ac.uk](mailto:su.golder@york.ac.uk)

Twitter: @SuGolder

Manuscript Count: 4499

TWEET:

**Methods to Establish Race or Ethnicity of Twitter Users: A Scoping Review** @SuGolder @GracielaGon

**Abstract**

Background: A growing amount of health research uses social media data. Those critical of social media research often cite that it may be unrepresentative of the population, but the suitability of social media data in digital epidemiology is more nuanced. Identifying the demographics of social media users can help establish representativeness.

Objectives: We sought to identify the different approaches or combination of approaches to extract race or ethnicity from social media and report on the challenges of using these methods.

Methods: We present a scoping review to identify the methods used to extract the race or ethnicity of Twitter users from Twitter datasets. We searched 17 electronic databases from date of inception to the 15th May 2021 and carried out reference checking and hand searching in order to identify relevant articles. Sifting of each record was undertaken independently by at least two researchers with any disagreement discussed. Studies were required to have extracted race and/or ethnicity of Twitter users with either manual or computational methods or a combination of both.

Results: From 1249 records sifted, we identified 67 that met our inclusion criteria. The majority focus on US based users and English language tweets. A range of data were used including Twitter profile metadata such as names, pictures, information from bios (including self-declarations), or location and/or content of the tweets. A range of methodologies were used including manual inference, linkage to census data, commercial software, language/dialect recognition and machine learning (ML) and/or Natural Language Processing (NLP). Not all studies evaluated their methods. Those that did found accuracy to vary from 45% to 93%with significantly lower accuracy identifying non-white race categories. The inference of race or ethnicity raises important ethical questions which can be exacerbated by the data and methods used. The comparative accuracy of different methods is also largely unknown.

Conclusion: There is no standard accepted approach or current guidelines for extracting or inferring race or ethnicity of Twitter users. Social media researchers must use careful interpretation of race or ethnicity and not over-promise what can be achieved, as even manual screening is a subjective, imperfect method. Future research should establish the accuracy of methods to inform evidence-based best practice guidelines for social media researchers, and be guided by concerns of equity and social justice.

Keywords: Twitter, Social Media, Race, Ethnicity

**INTRODUCTION**

Twitter data are increasingly used as a surveillance and data collection tool in health research. When millions of users post on Twitter, it translates to a vast amount of publicly accessible timely data about a variety of attitudes, behaviors, and preferences in a given population. While this data was not originally intended as a repository of individual information, Twitter data has been retrofitted in infodemiology to investigate population level health trends [1-15]. Researchers often use Twitter data in consort with other sources to test the relationship between online discourse and offline health behavior, public opinions and disease incidence.

The appeal of Twitter data is clear. Twitter is one of the largest public facing social media platform, with an ethnically diverse user base [16, 17] of more than 68 million US Twitter users, with Black users accounting for 26% of that base [18]. This diverse user basegives researchers access to people they may have difficulty reaching using more traditional approaches [19]. However, the promising insights that can be derived from Twitter data are often limited by what is missing, specifically basic socio-demographic information of each Twitter user. Demographic attributes of users is often required within health research for subpopulation analyses and to explore differences and identify inequity. Without evidence of the distal and proximal factors that lead to racial and ethnic health disparities, it is impossible to address and correct these drivers. Insights from social media data can be used to inform service provision, as well as to develop targeted health messaging by understand public perspectives from a diverse populations.

However, in order to use social media and digital health research to address disparities, we need to know not only what is said on Twitter, but also *who* is saying what [20]. While others have discussed extracting or estimating features such as location, age, gender, language, occupation and class, no comprehensive review of the methods used to extract race or ethnicity has been conducted [20]. Extracting race and ethnicity of Twitter users is particularly important to identify trends, experiences and attitudes of racially and ethnically diverse populations [21]. As race is a social construction, not a genetic categorization [22, 23], that practice of defining race and ethnicity in health research has been an ongoing, evolving challenge. Traditional research does have the advantage of knowing the identity of the person in the study and allowing them to systematically identify their racial and ethnic identity. In digital health research, [22, 23] determining a user’s race or ethnicity by extracting data from a user’s Twitter profile, meta data, or tweets, is a process that is inevitably challenging, complex and not without ethical questions.

Furthermore, while Twitter is used for research internationally, an international comparative study of methods to determine race or ethnicity is difficult, practically impossible, given societies use different standardized categories that describe their own populations [24]. A common approach in the US is based off the US Census Bureau practice to allow participants to identify with as many as five to six large racial groupings (Black, White, Asian Pacific Islander, Native, Other), while separately choosing one ethnicity (Hispanic)[25]. However, race and ethnicity variables continue to be misused in study design or when drawing conclusions. For example, race or ethnicity is often incorrectly treated as a predictor of poor health, rather than as a proxy for the impact that being a particular race or ethnicity have on that person’s experience with the health system [26]. Simply put, health disparities are driven by racism, not race [27-29]. While race or ethnicity affiliation are important factors in understanding diverse populations, digital research must tread lightly and thoughtfully in both the collection and assignment of race or ethnicity.

The lack of basic socio-demographic data about Twitter users has led researchers to apply a variety of approaches to better intuit characteristics of the people behind each tweet. The breadth of the landscape of approaches in extracting race or ethnicity is currently unknown. Our overall aim is to summarize and assess the range of computational and manual methods used in research based on Twitter data to determine the race or ethnicity of Twitter users.

**METHODS**

We conducted a comprehensive scoping review of extraction methods and offer recommendations and cautions related to these approaches. [30]We selected Twitter as it is currently the most commonly used social media platform in healthcare research, and it has some unique intrinsic characteristics that drive the methods used for mining it. Thus, we felt that the methods, type of data and the social media platforms used are related in such a way that comparing methods for different social media would add too many variables and would not be truly comparing like with like. A detailed protocol was designed for the methods to be used in our scoping review, but we are unable to register scoping reviews on PROSPERO. We report our methods according to the PRISMA scoping review statement [30].

#### Inclusion criteria

We devised strict inclusion criteria for our review based on the PICOS format (Population, Intervention, Comparators, Outcomes, Study Design). Although this is not a review of effectiveness we felt that the PICOS question breakdown [31] was still the most appropriate breakdown available for our question format [31]. The inclusion criteria were as follows;

**Population- P**: We included only datasets of Twitter users. Studies were eligible for inclusion if they collected information to extract or infer race or ethnicity directly from the users’ tweets, their profile details (such as the users’ photo or avatar, their name, location, and biography (bio), and/or their followers. We excluded studies extracting race or ethnicity from social media platforms other than Twitter, or from unspecified social media platforms, or that used multiple social media platforms that included Twitter but the data relating to Twitter was not presented separately.

**Intervention – I:** Studies were included where the methods to extract or infer race or ethnicity data of Twitter users were stated. Articles that used Machine Learning (ML), Natural Language Processing (NLP), human in the loop or other computationally assisted methods to predict race or ethnicity of users were included, as were manual or non-computational methods, including photo recognition or linking to census data. We excluded studies for which we were unable to determine the methods used or that extracted data solely on other demographic characteristics such as age, gender or geographical location.

**Comparator – C:** The use of a comparison of the methods used was not required.One method could be compared to another method (such as a gold standard) or no comparison could be undertaken.

**Outcome- O:** The extraction or inference of race or ethnicity of Twitter users was the primary or secondary outcome of the paper. . As this was a scoping review in which we aimed to demonstrate the full landscape of the literature no particular measurement of performance of the method used was required in our included studies.

**Study Design - S:** Any type of research study design was considered relevant. Discussion papers, commentaries and letters were excluded.

**Limits:** No date, language or publication type restrictions were applied to the inclusion criteria. However, no potentially relevant articles were identified in any non-English languages and the time period by default was since 2006, the year of the inception of Twitter.

#### Search strategy

A database search strategy was derived from combining three facets; facet one consisted of free-text terms related to Twitter (Twitter OR Tweet\* OR Tweeting OR Retweet\* OR Tweep\*), facet two consisted of terms for race or ethnicity, and facet three consisted of terms for methods of prediction, such as ML, NLP, and artificial intelligence related terms (Supplementary table 1). All ethnology-related subject terms were adapted for different database taxonomies and syntax, with standard methods of prediction subject terms in Medline and other database indexing. The methods of prediction term facet was expanded using a comprehensive list of specific text analysis tools and software names extracted from Hinds and Joinson 2018 [32], which includes a comprehensive list of automated machine-learning processes used in predicting demographic markers in social media. Additional terms were added from a related paper [33].

#### Sources searched

A wide range of bibliographic and grey literature databases were selected for searching covering computer science, health and social sciences. The databases (Table 1) were last searched on the 15th May 2021, with no date or other filters applied.

**Table 1: Databases searched with number of records retrieved**

|  |  |
| --- | --- |
| **Database Name** | **Total number of results** |
| ACL Anthology | Screened first 50 records from 2 searches |
| ACM Digital Library | 150 |
| Cinahl | 200 |
| Conference Proceedings Citation Index – Science (CPCI-S) | 84 |
| Conference Proceedings Citation Index – Social Science (CPCI-SS) | 7 |
| Emerging Sources Citation Index (ESCI) | 41 |
| Google Scholar | (screened first 100 records from 2 searches) |
| IEEE Xplore | 186 |
| Library and Information Science Abstracts | 120 |
| LISTA | 79 |
| OpenGrey | 0 |
| Proquest Dissertations and Theses – UK and Ireland | 195 |
| PsycINFO | 72 |
| PubMed | 84 |
| Science Citation Index (SCI) | 56 |
| Social Science Citation Index (SSCI) | 111 |
| Zetoc | 50 |

Reference checking of all the included studies and any related systematic reviews identified by the searches was conducted. We browsed Journal of Medical Internet Research as this is a key journal in this field as well as hand searched two relevant conferences the International AAAI Conference on Weblogs and Social Media (ICWSM) and ACL proceedings.

Citations were exported to a shared Endnote library and duplicates were removed. The deduplicated records were then imported into Rayyan to facilitate blind independent screening by the authors. Using the inclusion criteria at least two screeners from the research team screened each record independently, with disputes on inclusion discussed and a consensus decision reached.

Only the first 50 records from ACL and the first 100 records from a Google Scholar search were screened during two searches (11/03/2020 and 24/05/2021) as these records are displayed in order of relevance and it was felt that after this number no relevant articles were being identified [12, 21, 34-98].

#### Data Extraction

For each included article we extracted the following data on an excel spreadsheet:

year of publication, study country and language, race or ethnicity categories extracted (such as, for race - Black, White, Asian or for ethnicity - Hispanic, European), and paper type (journal, conference or thesis). We also extracted details on extraction methods (such as classification models or software used), features and predictors used in extraction (tweets, profile, pictures), number of Twitter users, number of Tweets or images used, performance measures to evaluate methods used (validation) and results of any evaluation (such as accuracy). Any performance measure metrics were reported as stated within the included studies themselves. All extracted data was checked by two reviewers.

#### Quality Assessment

There is no formally approved quality assessment tool for this type of study. Since this is a scoping review we did not carry out any formal assessment. However, we did assess any validation carried out and whether the methods are reproducible.

#### Data Analysis

We summarized the stated performance of those papers that included validation. However, we could not compare approaches using the stated performance, as performance measures and validation approaches varied considerably. There is also no recognized gold standard dataset for comparison.

### **RESULTS**

A total of 1735 records were entered into an Endnote library and duplicates were removed leaving 1249 records for sifting (Figure 1). 1080 records were excluded based on title and abstract screening alone. In total 169 references were deemed potentially relevant by either of the independent sifters (RS, GG, RJ, SG and KO). The full-text of these articles were screened independently and 67 studies [12, 21, 34-98] met our inclusion criteria and 102 references were excluded [32, 79, 99-198]. The main reason for exclusion was that although the abstract indicated that demographic data were collected, this did not include race or ethnicity (most commonly, other demographic attributes were collected such as gender, age or location). The other reasons for exclusion were that the researchers included collected demographic data through surveys or questionnaires administered via Twitter (but not from data posted on Twitter) or that the researchers used a social media platform other than Twitter.

**Figure 1: Flow diagram for included studies**

#### Characteristics of the included studies

The majority of the studies stated or implied that they were based solely or predominantly in the US and limited to English language bios or tweets. Six studies were multi-national [40, 43, 58, 68, 85, 88], one was UK based [61] and another based in Qatar [57] and only eight studies (12%) extracted data from tweets in multiple languages [34, 40, 54, 57, 58, 68, 85, 88] (Supplementary table 2).

The most common race examined was White (87%, 58/67), followed by Black/African- American (84%, 56/67), Asian (67%, 45/67), and ethnicity as Hispanic/Latino (64%, 43/67) (Figure 2)

**Figure 2: Percentage of studies with each race or ethnicity**

Some studies treated race as a binary classification such as African-American or not, or African-American or White, while others created a multiclass classifier of three or four classes, or a combination of classes. Six studies went above four classes, however, these often included ethnicity/nationality classifiers as well as race [40, 50, 56, 68, 85, 97].

The data objects from Twitter used to extract the race or ethnicity varied, with the use of profile pictures or the Twitter user’s name being most common. Others also used tweets in the users’ timeline, information from Twitter bios or the Twitter users’ location. Most studies used more than one data object from the Twitter data. In addition the datasets within the studies varied in size between 392 and 168,000,000, with those using manual methods having smaller datasets ranging from just 392 [52] to 4900 [67].

Unfortunately, performance was only measured in some studies (Table 2). Metrics used to report results were particularly varied for those studies using ML or NLP and included: F1 score (which combines precision and recall), accuracy, area under the curve (AUC) or mean average precision. Table 2 lists the methods, features and the reported performance of the top model from each study.

#### Manual Screening

Twelve studies used manual techniques to classify Twitter users into race or ethnicity categories [21, 38, 42, 51-53, 59, 67, 89-92]. These studies generally combined qualitative interpretation of recent tweets, information in user bios making an affirmation of racial or ethnic identity, or photographs/images in the user timeline or profile.

In most cases, tweets were first identified by text matching based on terms of interest to the research topic, for instance, having a baby with a birth defect [52], commenting on a controversial topic [59, 91], or use of potentially gang- or drug-related language [42]. Researchers then identified the tweet authors and in most cases assigned race or ethnicity through hand-coding based on profile and timeline content. Some studies coded primarily by identifying self-identifying statements of race used in a tweet or in users’ bios such as people stating that they are a ‘Black American’ [51, 52, 90, 92] or hashtags [38] (such as #BlackScientist). Others coded exclusively based on the research team’s attribution of racial identity through examination of profile photographs [21, 59] or avatar [89]. Some authors coded primarily with self-declarations, with secondary indicators such as profile pictures, language, usernames or other content [42, 53, 67, 90, 91]. In most cases it appears reasonable to infer that coding was done by the study authors or members of their research teams, with the exception of those using the crowdsourcing marketplace Amazon’s Mechanical Turk (AMT) [21, 92].

Agreement between coders was sometimes measured but the validity and accuracy measurements were not generally included. One study [67], however, documented a 78% reliability for coding race compared to census demographics, with Black and White users being coded accurately 90% of the time and Hispanic or Asian users being accurately coded between 45-60% of the time. The high accuracy for Black users was based on the higher likelihood of Black users to self-identify.

#### Census Driven Prediction

Another approach to predict race or ethnicity is through utilizing demographic information from national census and census-like data and transferring it to the social media cohort. The US based studies largely used census-based race and ethnicity categories: Asian and Pacific Islander, Black/African American, Latino/Hispanic, Native American and White. The one U.K. based study included the categories British and Irish, West European, East European, Greek or Turkish, South East Asian, Other Asian, African & Caribbean, Jewish, Chinese, and other minority [85].

We identified 14 studies [41, 50, 54, 56, 62, 65, 72, 73, 76, 79, 85-87, 97] that utilized Census geographic data, census surname classification or a combination of both. Six studies incorporated census geographic data [41, 54, 65, 76, 85, 86]. For example, Blodgett 2016[41] created a simple probabilistic model to infer user’s ethnicity by matching geotagged tweets to census block information. They averaged the demographic values of all the tweets by the user and assumed this to be a rough proxy of the user’s demographics. Stewart 2014 [76] collected tweets tagged with geolocation (longitude and latitude) information. The zip code of the user was derived from this geolocation information and matched to demographic information found in the ZIP Code Tabulation Area (ZCTA) defined by the census bureau. This information was used to find correlation between ethnicity and African American Vernacular English syntax [76].

Other studies used the census derived name classification system to determine race or ethnicity based on user names. We identified twelve studies that predicted user race or ethnicity using surnames [50, 56, 62, 65, 72, 73, 79, 85-87, 97, 189]. Surnames were used to assign race or ethnicity using either US census based name classification system or less commonly, an author in-house generated classification system. Of these studies, eight relied solely on user last names [50, 56, 62, 65, 72, 73, 86, 87]. Of those that reported validating the system, validation methods of this name-based system alone were not reported, but four studies reported accuracy between 71.80% and 81.25% [65, 72, 73, 85]. Of note, one study reported vastly different accuracy in predicting whiteness vs. blackness (94% predicting white users vs. 33% predicting African-American/Black users)[85]. The remaining studies augmented named based predictions with aggregate demographic data from the American Community Survey or equivalent. For example, statistical and text mining methods were used to extract surnames from Twitter profiles, combining this information with census block information based on geolocated tweets to assess the probability of the user’s race or ethnicity [62]. These studies did not report validation or accuracy.

#### Ad-hoc ML and/or NLP

Twenty four papers [35-37, 39, 40, 48, 49, 63, 66, 68-70, 74, 78, 80-84, 93-96, 98] utilized some ML and/or NLP to automatically classify users by their race or ethnicity. ML and NLP methods were used to process the data made available by the Twitter user, such as profile images, tweets text, and location of residence. These studies almost invariably consisted of larger cohorts with considerable variation in the specific methods employed.

Supervised ML models (where some annotated data is used to ‘train’ the system) were used in 12 of the 24 studies. The models used include support vector machine (SVM) [40, 48, 63], gradient boosted decision tree [69, 70], and regression models [35, 36, 39, 78, 199].

Semi-supervised (where a large set of unannotated data is also used for training the system, in addition to annotated data), or fully unsupervised models using neural networks or regression were used for classification in 10 of the 24 studies [35, 37, 68, 74, 80, 81, 83, 94-96].

Two studies used an ensemble of previously published race or ethnicity classifiers by processing the data through four extant models and using a majority rule approach to classify users based on the output of each classifier [82, 93].

The ML models use features, or data inputs, to predict the desired output. Features derived from textual information in the user’s profile description such as name or location were used by some studies [36, 37, 40, 62, 69, 70, 81, 83, 94, 95]. Other studies included features related to images, including but not exclusively profile images [48, 69, 70, 189] and the facial features in those images [68]. Some studies used linguistic features to classify a user’s race or ethnicity [39, 40, 48, 49, 63, 69, 70, 74, 78, 80, 83, 94-96, 199]. Specific linguistic features used in the models include n-grams [40, 48, 74, 93-96], topic modeling [48, 63, 80], sentiment and emotion [78] and self-reports [69, 70, 83]. Information about a user’s followers, or network of friends, were included as features in some studies under the assumption that members of these networks have similar traits [36, 39, 48, 49, 93].

Labeled datasets are used to train and test supervised and semi-supervised ML models and also validate the output of unsupervised learning methods. Some of the studies used previously created data sets that contained demographic information, such as the MORPH database of images [189], a database of mugshots [40], or manually annotated data from previous studies [81, 83]. Others created ground truth datasets from surveys [199] or by semi-automatic means, such as matching Twitter users to voter registrations [39], using extracted self-identification from user profiles or tweets [69, 70, 83], or used celebrities with known ethnicities[68]. Manual annotation of Twitter users was also used based on profile meta data [36, 37, 48, 78], self-declarations in the timeline [63, 84] or user images [37, 96]. Table 2 summarizes the best performing ML approach, features used and the reported results for each study that utilized automatic classification methods. In the table, classifier is the number of race or ethnicity classification groups, ML model is the top performing algorithm reported and features are the variables used in the predictions.

**Table 2: Top system performance within studies using ML/NLP. Result metrics are reflected here as reported in the original publications**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Study | Classifier | ML Model | Features | Results Reported | | |
|  |  |  |  | Accuracy | F1 Score | AUCb |
| Pennacchiotti 2011[70] | binary | GBDTa | images, text, topics, sentiment | -- | 0.66 | -- |
| Pennacchiotti 2011[69] | binary | GBDT | images, text, topics, sentiment, network | -- | 0.70 | -- |
| Bergsma 2013[40] | binary | SVMa | names; name clusters | 0.85 | -- | -- |
| Ardehaly 2017[37] | binary | DLLPa | text, images | -- | 0.95 (image)  0.92 (text) | -- |
| Volkova 2018[78] | binary | LRa | text, sentiment, emotion | -- | -- | 0.97 |
| Wood-Doughtry 2018[81] | binary | CNNa | name | 0.73 | 0.72 | -- |
| Saravanan, 2017[74] | ternary | CNN | text | NR | NR | NR |
| Ardehaly 2017[35] | ternary | DLLP | text, images | -- | 0.84 (image)  0.83 (text) | -- |
| Gunaranthne 2019[96] | ternary | CNN | text | -- | 0.88 | -- |
| Wood-Doughtry 2018[81] | ternary | CNN | name | 0.62 | 0.43 | -- |
| Culotta 2016[49] | quaternary | REa | network, text | -- | 0.86 | -- |
| Chen 2015[48] | quaternary | SVMa | n-grams, topics, self declarations, image | 0.79 | 0.79 | 0.72 |
| Markson, 2017[63] | quaternary | CNN | synonym expansion, topics | 0.76 | -- | -- |
| Wang 2016[189] | quaternary | CNN | images | 0.84 | -- | -- |
| Xu, 2016[84] | quaternary | SVM | synonym expansion, topics | 0.76 | -- | -- |
| Ardehaly 2015[36] | quaternary | MLRa | census, name, network, tweet language | 0.83 | -- | -- |
| Mohammady 2014[66] | quaternary | LR | census, image tweets | 0.82 | 0.81 | -- |
| Barbera, 2016[39] | quaternary | LR w/ENa | tweets, emojis, network | 0.81 | -- | -- |
| Wood-Doughty 2020[83] | quaternary | CNN | name, profile metadata, text | 0.83 | 0.46 |  |
| Preotiuc\_Pietro 2018[199] | quaternary | LR w/EN | text, topics, sentiment, POS tagging, name, perceived race labels, ensemble |  |  | 0.88 AAc, 0.78 Latino, 0.83 Asian, 0.83 White |
| Muller 2021[93] | quaternary | CNN | Text, accounts followed |  | 0.25 Asian, 0.63 AAc/black, 0.28 Hispanic, 0.90 White |  |
| Bergsma 2013[40] | Multinomial > 4 | SVM | name; name clusters | 0.81 | -- | -- |
| Nguyen 2018[68] | Multinomial > 4 | NNa | images | 0.53 | -- | -- |

a CNN, convolutional neural network; DLLP, deep learning from label proportions; EN, elastic net; GBDT, gradient boosted decision tree; LR, logistic regression; MLR, multinomial logistic regression NN, neural network; RE, regression, SVM, support vector machine

b AUC, area under curve

c AA, African American

The data from Twitter is inherently imbalanced in terms of race or ethnicity groups. In ML, it is important to attempt to mitigate the effects of the imbalance as the models have difficulty learning from few examples and will tend to classify to the majority class and ignore the minority class. Few studies addressed the imbalance directly. Some opted to make the task binary, focusing only on their group of interest versus all others [69, 70, 96], or only the majority classes [40, 78]. Others chose modified performance metrics that account for imbalance when reporting their results [35, 63, 84]. One group, who classified based on images, supplemented their training set from an additional data source for the minority classes [35, 37]. Only two studies experimented with comparator models trained on balanced data sets. In [83], they under sampled the majority class in their training sets, and [199] oversampled the minority classes. In both cases, the overall performance of the models decreased in accuracy from 0.83 to 0.41 (on their best performing unbalanced model) and 0.84 to 0.68 [199], as the performance boost from the models superior performance on the majority class was eradicated.

#### Off-the-shelf Software

Seventeen studies [12, 34, 43-47, 55, 57, 58, 60, 61, 64, 71, 75, 77, 88] used off-the-shelf software packages to derive race or ethnicity. Ten studies [34, 46, 47, 55, 57, 58, 60, 64, 71, 77] used Face++[200] and five studies [12, 43-45, 75] used Demographics Pro[201], and two used Onomap[202] software to determine Ethnicity [61, 88]. Face++ is a validated machine-learning face detection service to analyze features with confidence levels for the inferred race attributes. Specifically, it uses deep-learning to identify whether profile pictures contain a single face and then the race of that face (limited to Asian, Black, and White) and does not infer ethnicity (e.g. Hispanic) [200]. Demographics Pro estimates demographic characteristics based on Twitter behavior/usage using NLP, entity identification, image analyses, and network theory [201]. Onomap is a software tool for the classification of names [202]. Three of the studies that used Face++ used the same baseline dataset [47, 64, 77] and one used a partial subset of this same dataset [71].

Two of the studies that used Face++ [34, 60] did not measure its performance at all. Another study [46] simply stated that Face++ could identify race with 99% confidence or higher for 9% of total users. Two studies [55, 57] used Face++ along with other methods. One of these studies used Face++ in conjunction with Demographics which uses given name or full name from a database that contains US census for demographics. This study simply measured the percentage of twitter users for which they could extract race data (46% college students and 92% of role models) but did not measure the performance of Face++ [55]. The other study [57] built a classifier model on top of using Face++ and recorded an accuracy of 83.8% when compared to users who stated their own nationality.

Four studies [47, 64, 71, 77] (with the same dataset in full or in part) used the average confidence level reported by Face++ for race which was 85.97 ± 0.024%, 85.99 ± 0.03%, 86.12 ± 0.032% respectively with a confidence interval of 95%. When one of these studies [47] carried out their own accuracy assessment they found an accuracy score of 79% for race when compared to 100 manually annotated pictures. Huang 2020 [58] also carried out their own accuracy assessment and found that Face++ achieved an averaged accuracy scores of 88.4% for race when compared to 250 manually annotated pictures.

Five studies [12, 43-45, 75] used Demographics Pro and whilst they reported on Demographics Pro success in general they did not directly report any metrics of its success. The two studies using Onomap provided no validation of the software [61, 88].

In light of our results we have compiled our recommendations for best practice which are summarized in figure 3 and is examined further in our discussion.

**Figure 3: Summary of our best practice recommendations**

**DISCUSSION**

#### Principal Findings

As there are no guidelines currently published or even best practice guidance, it is no surprise that we found researchers use a variety of methods for estimating race or ethnicity of Twitter users. We identified four categorizations for the methods used; manual screening, census-based prediction, ad-hoc ML and/or NLP and off-the-shelf software. These methods all exhibit particular strengths as well as inherent biases and limitations.

Comparing the validity of methods for the purpose of deriving race or ethnicity is difficult as classification models differ not only in approach, but in the definition of the classification of race or ethnicity itself [111, 203, 204]. There was also a distinct lack of evaluation or validation of methods used. Those that did measure performance of the methods used found accuracy to vary from 45% to 93%with significantly lower accuracy identifying non-white race categories.

This review shielded little light on the performance of commercial software. Previous empirical comparisons of facial recognition APIs have found that Face++ achieves 93.0% accuracy [205] and works comparatively better for lighter male skins [206]. The included studies in our review suggest lower accuracy. Whilst data on accuracy was not forthcoming in any of the included studies using Demographics Pro [201]. Even when performance was assessed, the methodology used may be biased if there are issues with the ‘gold standard’ used to train the model.

In addition to the four over-arching methods used, the studies also varied in terms of the features used to determine or define race or ethnicity. Further the reliability of the features used to determine or define race or ethnicity for this purpose is questionable. Specifically, the use of twitter users’ profile pictures, names, and locations, the use of unvalidated linguistic features attributed to racial groups (such as, slang words, African-American Vernacular English, Spanglish, or Multicultural London English) and the use of training data that are prone to perpetuate biases (e.g. police booking photos or mug shots) were all of particular concern. Issues related to methods used

Approaches that include, or solely rely on profile pictures to determine race or ethnicity can introduce bias. First, not all users have a photograph as their profile picture, nor is it easy to determine that a picture used is that of the user. A study of the feasibility of utilizing Face++, found only 30.8% of Twitter users had a detectable single face in their profile. A manual review of the automatically detected faces determined 80% could potentially be of the user (i.e. not a celebrity) [207]. Human annotation may introduce additional bias, studies have found systematic biases in the classification of people into race or ethnic groups based on photographs [208, 209]. Furthermore, it is known that humans have an inclination to perceive one’s own race more readily than others [210, 211]. Thus, race or ethnicity among the annotation team has an impact on the accuracy of their race or ethnicity labels, potentially skewing the sample labels towards the race or ethnicity of the annotators [212, 213]. Given ML and NLP methods are trained on these datasets, the human biases transfer to automated methods, leading to poorly-supervised ML and training which has been shown to result in discrimination by the algorithm [214-216]. These concerns did not appear to be interrogated by the study designers. Without exception, they present categorization of persons into race or ethnicity, assuming that a subjective reading of facial features or idiomatic speech is a gold standard both for coding of race or ethnicity and for training and evaluation of automated methods.

Other methods, such as using geography or names as indicators of race, could also be unreliable. One could argue that the demographic profile for a geographic region is a better representation of the race or ethnicity in the demographic environment, rather than individual’s race or ethnicity. Problems of using postcodes or location to decipher individual social determinants are well documented [217]. The use of census data from too large an area may skew results. Amongst the studies reviewed, some use census block data which is granular, while others extrapolate from larger areas, such as city or county level data. For example, Saravanan 2017 [74] inferred the demographics of users in a city as a certain ethnic group based on that city having a large population of that group, however, no fine grained analysis was done either for the city chosen or for geolocation of the Twitter user. Thus, the validity of their assumption that a user in LA County is of Mexican descent [74] is therefore, questionable at best. As these data was then used to create a ‘race or ethnicity’ dictionary of terms used by that group to train their model, the questionable assumption further taints downstream applications and results. Models also do not consider the differences between the demographics of twitter users versus the general demographics of the population.

Additionally, Census demographic data that uses names is also questionable, due to name taking in marriage and indiscernible names.

The practice of using a Twitter user’s self-reported race or ethnicity would provide a label with high confidence, but restrict the amount of usable data and introduce a margin of error depending on the method used to extract such self-reports. For example, in a sample of 14 million users less than 0.1% of users matched precise regular expression created to detect self-reported race or ethnic identity [127]. Another study used mentions of keywords related to race or ethnicity in a user’s bio, however, limited validation was conducted to ensure that mention was actually related to the user’s race or ethnicity [69, 70]. This lack of information gathered from profile information leads to a sampling bias in the training of the models [152].

Some models trained on manually annotated data did not have high inter annotator agreement, for example Chen 2015 [48] crowdsourced annotation agreement measured at 0.45. This can be interpreted as weak agreement with the percent of reliable data being 15-35% [218]. Training a model on such weakly labeled data produces uncertain results.

It is not possible to assume the accuracy of black box proprietary tools and algorithms. The only race or ethnicity measure that seems empirically reliable is self-report, but this has considerable limitations: so, the faulty methods continue to underpin digital health research, and researchers are likely to become increasingly dependent on them. The ‘gold standard’ data required to know the demographic characteristics of the twitter user is difficult to ascertain.

Methods that we highlight as best practice include asking Twitter users directly. This can be achieved, for example, by asking respondents of a traditional survey for both their demographic data and their Twitter handle so that the data can be linked [199]. This was undertaken in the NatCen Social Research British Social Attitudes Survey 2015 – and this has the added benefit of allowing the study of the accuracy of further methods for deriving demographic data [20]. Contacting Twitter users may also provide a gold standard but be impractical given the current terms of use of Twitter that might consider such contact a form of spamming.

[74, 205, 206, 217]. A limitation of extracting race or ethnicity from social media is the necessity to oversimplify the complexity of racial identity. Categories were often limited to Black, White, Hispanic, and Asian. Note that ‘Hispanic’ is considered ‘ethnicity’ by the US Census, but most studies in ML used it as a ‘race’ category, more so than Asian (due to low numbers in this category). Multiple more racial identities exist, particularly from an international perspective, and this overlooks multiracial or primary and secondary identities. In addition, inferred identities may differ from self-identities, raising further issues.

Given the sensitive nature of the data, it is important as a ‘best practice’ for the results of studies that derive race or ethnicity from Twitter data to be reproducible for validation and future use. Reproducibility of most of the studies in this review would be difficult or impossible, as only five studies linked to available code or data [40, 49, 81, 83, 107]. Furthermore, there was limited information on the coding of training data. None of the studies detailed their annotation schema nor made available annotation guidelines. Detailed guidelines as a ‘best practice’ may allow for recreating or extending datasets in situations where the original data may not be shared or where there is data loss over time. This is particularly true of data collected from Twitter, where the terms of use require that shared data sets consist of only tweet IDs, not tweet texts, and that best efforts to delete IDs from the data set if the original tweet is removed or made private by the user be in place. Additional restrictions are placed on special use cases for sensitive information, prohibiting the storage of such sensitive information if detected or inferred about the user. Twitter explicitly states that information on racial or ethnic origin cannot be “derived or inferred” for an individual Twitter user, and allows academic research studies to use only aggregate level data for analysis [219]. This policy it may be argued is more likely to be targeted at commercial activities.

#### Strengths and Limitations

We did not limit our databases searches and other methods by study design, yet we were unable to identify any previous reviews on the subject. To our knowledge, therefore, this is the first review of the methods used to extract race or ethnicity from social media. We identified studies from a range of disciplines and sources and were able to categorize and summarize the methods used. We were unable, however, to obtain information on the methodologies used by private sector companies that have created software for this purpose. Marketing and targeted advertising is common on social media, and is likely to use race as part of their algorithms to derive the target users.

We did not limit our included papers to those for which the extraction of race or ethnicity was the primary focus. Whilst this can be conceived as a strength it also meant that the reporting of the methods used was often poor. The accurate recreation of the data lost was hampered by not knowing how decisions were made in the original studies, including what demographic definitions of race or ethnicity were used or how accuracy was determined. This limited our assessment of the included studies. Few studies validated the methods or conducted an error analysis to assess how often race is misapplied and those that did rarely used the most appropriate gold standard. This made it difficult to directly compare the results of the different approaches.

#### Future directions

Researchers in future studies should interrogate their methodological approaches to estimating race or ethnicity, offering careful interpretations that acknowledge the significant limits of these approaches and their impact on the interpretation of results. This may include reporting results as a range that communicates the inherent uncertainty of the classification model. Social media data may be best used in combination with other information. In addition, we must always be mindful that race is a proxy measure for the much larger impact of being a particular race or ethnicity in a society. As a result, the variability associated with race and ethnicity might reveal more about the effects of racism and social stratification than individual user attributes. To conduct this work ethically and rigorously, we recommend several practices that can help reduce bias and increase reproducibly.

We recommend acknowledgment of the bias of the researchers that can influence the conceptualization of implementation of the study. Incorporating this reflexivity, as common in qualitative research, allows for the identification of potential blind spots that weaken the research. One way to address homogenous research teams is through the inclusion of experts in race or ethnicity or in those communities being examined. These biases can also be reduced through the inclusion of members of study populations in the research process as experts and advisors[220]. Though big data from social media can be collected without ever connecting with the people who contributed the data, it does not remove the ethical need for researchers to include representative perspectives in research processes. Examples of patient engaged research and patient-centered outcomes research, community-based participatory research and citizen science (public participation in scientific research) within the health and social sciences amply demonstrate the instrumental value and ethical obligation of intentional efforts to involve non-scientist partners in co-creation of research [220] The quality of data science can be improved by seriously heeding the imperative, “Nothing about us without us” [220]. Documenting and establishing the diversity competence attributes of the research team should become a standard. Emphasizing the importance of diverse teams within the research process will contribute to social and racial justice in other ways than just the improvement and reliability of research.

In terms of the retrieved data, the most reliable (though imperfect) method of ascertaining race is when users self-identify their racial affiliation. Further research on overcoming the availability limitations this introduces to sample sizes may be warranted. Indeed a hybrid model with automation methods and manual extraction may be preferred. For example, automation methods could be better developed to identify potential self-declarations in a user profile or timeline which could then be manually interpreted.

Finally, we call for greater reporting of validation by our colleagues. Without error analysis, the computational techniques won’t tell you when bias is present. We also need further research to establish whether any bias is systematic or random, that is, do inaccuracies favor one direction or another.

**CONCLUSIONS**

We identified major concerns that impact the reliability of the methods and bias the results. There are also ethical concerns throughout the process, particularly with inference of race or ethnicity as opposed to extraction of self-identity. However, the potential usefulness of using social media research requires thoughtful consideration of the best ways to estimate demographic characteristics like race and ethnicity [111]. This is particularly important given the increased access to Twitter data [203, 204].

We therefore propose several approaches to improve the extraction of race or ethnicity from social media including representative research teams, and a mixture of manual and computational methods as well as future research on methods to reduce bias.

### Competing Interests

The authors declare that there are no competing interests

### Author Contribution

SG, RS, KO, RJ and GG contributed equally to this article. RS and GG proposed the topic and main idea. SG and RJ were responsible for the literature search. SG, RS, KO, RJ and GG were all responsible for the study selection and the data extraction. SG wrote the initial draft. SG, RS, KO, RJ and GG commented on and revised the paper. SG made the final version. All authors contributed to the final draft of the paper.

### Funding Statement

This work was supported by National Institutes of Health (NIH) National Library of Medicine under grant number NIH NLM 1R01 (PI Gaciela Gonzalez-Hernandez, with co-applicants Karen O’Connor and Su Golder) and National Institutes of Health (NIH) National Institute of Drug Abuse R21 DA049572-02 to Robin Stevens. NIH National Library of Medicine funded this research but were not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

### Data Availability

The included studies are available online and the extracted data is contained in supplementary table 2. A preprint of this article is also available; Golder S, Stevens R, O'Connor K, James R, and Gonzalez-Hernandez G. 2021. Who Is Tweeting? A Scoping Review of Methods to Establish Race and Ethnicity from Twitter Datasets. SocArXiv. February 14. doi:10.31235/osf.io/wru5q.

## References

1. Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. Br J Clin Pharmacol. 2015;80(4):878-88. PMID: 26271492. doi: 10.1111/bcp.12746.

2. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform. 2015 Apr;54:202-12. PMID: 25720841. doi: 10.1016/j.jbi.2015.02.004.

3. Bhattacharya M, Snyder S, Malin M, Truffa MM, Marinic S, Engelmann R, et al. Using Social Media Data in Routine Pharmacovigilance: A Pilot Study to Identify Safety Signals and Patient Perspectives. Pharmaceutical Medicine. 2017 2017/06/01;31(3):167-74. doi: 10.1007/s40290-017-0186-6.

4. Convertino I, Ferraro S, Blandizzi C, Tuccori M. The usefulness of listening social media for pharmacovigilance purposes: a systematic review. Expert Opin Drug Saf. 2018 Nov;17(11):1081-93. PMID: 30285501. doi: 10.1080/14740338.2018.1531847.

5. Golder S, Smith K, O'Connor K, Gross R, Hennessy S, Gonzalez-Hernandez G. A Comparative View of Reported Adverse Effects of Statins in Social Media, Regulatory Data, Drug Information Databases and Systematic Reviews. Drug Saf. 2021 Feb;44(2):167-79. PMID: 33001380. doi: 10.1007/s40264-020-00998-1.

6. Bychkov D, Young S. Social media as a tool to monitor adherence to HIV antiretroviral therapy. J Clin Transl Res. 2018;3(Suppl 3):407-10. PMID: 30873489.

7. Kalf RR, Makady A, Ten Ham RM, Meijboom K, Goettsch WG. Use of Social Media in the Assessment of Relative Effectiveness: Explorative Review With Examples From Oncology. JMIR Cancer. 2018 Jun 8;4(1):e11. PMID: 29884607. doi: 10.2196/cancer.7952.

8. Golder S, O'Connor K, Hennessy S, Gross R, Gonzalez-Hernandez G. Assessment of Beliefs and Attitudes About Statins Posted on Twitter: A Qualitative Study. JAMA Netw Open. 2020;3(6):e208953-e. PMID: 32584408. doi: 10.1001/jamanetworkopen.2020.8953.

9. Golder S, Bach M, O'Connor K, Gross R, Hennessy S, Gonzalez Hernandez G. Public Perspectives on Anti-Diabetic Drugs: Exploratory Analysis of Twitter Posts. JMIR Diabetes. 2021 Jan 26;6(1):e24681. PMID: 33496671. doi: 10.2196/24681.

10. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Monitoring Online Discussions About Suicide Among Twitter Users With Schizophrenia: Exploratory Study. JMIR Ment Health. 2018 Dec 13;5(4):e11483. PMID: 30545811. doi: 10.2196/11483.

11. Howie L, Hirsch B, Locklear T, Abernethy AP. Assessing the value of patient-generated data to comparative effectiveness research. Health Aff (Millwood). 2014 Jul;33(7):1220-8. PMID: 25006149. doi: 10.1377/hlthaff.2014.0225.

12. Cavazos-Rehg PA, Krauss MJ, Costello SJ, Kaiser N, Cahn ES, Fitzsimmons-Craft EE, et al. "I just want to be skinny.": A content analysis of tweets expressing eating disorder symptoms. PLoS One. 2019;14(1):e0207506. PMID: 30650072. doi: 10.1371/journal.pone.0207506.

13. Ahmed W, Bath PA, Sbaffi L, Demartini G. Novel insights into views towards H1N1 during the 2009 Pandemic: a thematic analysis of Twitter data. Health Info Libr J. 2019 Mar;36(1):60-72. PMID: 30663232. doi: 10.1111/hir.12247.

14. Cook N, Mullins A, Gautam R, Medi S, Prince C, Tyagi N, et al. Evaluating Patient Experiences in Dry Eye Disease Through Social Media Listening Research. Ophthalmol Ther. 2019 Sep;8(3):407-20. PMID: 31161531. doi: 10.1007/s40123-019-0188-4.

15. Roccetti M, Salomoni P, Prandi C, Marfia G, Mirri S. On the interpretation of the effects of the Infliximab treatment on Crohn’s disease patients from Facebook posts: a human vs. machine comparison. Network Modeling Analysis in Health Informatics and Bioinformatics. 2017 2017/06/26;6(1):11. doi: 10.1007/s13721-017-0152-y.

16. Madden M LA, Cortesi S, Gasser U, Duggan M, Smith A, Beaton M. Teens, Social Media, and Privacy. Pew Internet & American Life Project 2013; Available from: <http://www.pewinternet.org/2013/05/21/teens-social-media-and-privacy/>.

17. Chou W-yS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. Journal of medical Internet research. 2009;11(4):e48-e. PMID: 19945947. doi: 10.2196/jmir.1249.

18. Pew Research Center. Social Media Use in 2018. 218; Available from: <https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/>.

19. Bowleg L, Teti M, Malebranche DJ, Tschann JM. "It's an Uphill Battle Everyday": Intersectionality, Low-Income Black Heterosexual Men, and Implications for HIV Prevention Research and Interventions. Psychol Men Masc. 2013 Jan 1;14(1):25-34. PMID: 23482810. doi: 10.1037/a0028392.

20. Sloan L. Social Science 'Lite'? Deriving demographic proxies from Twitter. In: Sloan L Q-HA, editor. The SAGE Handbook of Social Media Research Methods2017.

21. McCormick TH, Lee H, Cesare N, Shojaie A, Spiro ES. Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing. Sociological Methods & Research. 2017;46(3):390-421. PMID: 29033471. doi: 10.1177/0049124115605339.

22. Smedley A, Smedley BD. Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race. Am Psychol. 2005 Jan;60(1):16-26. PMID: 15641918. doi: 10.1037/0003-066x.60.1.16.

23. Yudell M, Roberts D, DeSalle R, Tishkoff S. NIH must confront the use of race in science. Science. 2020 Sep 11;369(6509):1313-4. PMID: 32913094. doi: 10.1126/science.abd4842.

24. Davenport L. The Fluidity of Racial Classifications. Annual Review of Political Science. 2020 2020/05/11;23(1):221-40. doi: 10.1146/annurev-polisci-060418-042801.

25. U.S. Census Bureau. U.S. Census Bureau. QuickFacts. United States. 2021; Available from: <https://www.census.gov/quickfacts/fact/note/US/RHI625219>.

26. Zuberi T. Thicker than blood : how racial statistics lie. Minneapolis: University of Minnesota Press; 2001.

27. RR H, J K. Examining racism in health services research: A disciplinary self-critique. Health Serv Res. 2020 Oct;55 Suppl 2(Suppl 2):777-80. PMID: 32976632. doi: 10.1111/1475-6773.13558.

28. Jenkins W, Schoenbach V, Rowley D, Ford C. Overcoming the Impact of Racism on the Health of Communities: What We Have Learned and What We Have Not. In: Ford C, Griffith D, Bruce M, Gilbert K, editors. Racism: Science & Tools for the Public Health Professional2019.

29. Jones CP. Toward the Science and Practice of Anti-Racism: Launching a National Campaign Against Racism. Ethn Dis. 2018;28(Suppl 1):231-4. PMID: 30116091. doi: 10.18865/ed.28.S1.231.

30. Tricco A, Lillie, E, Zarin, W, O'Brien, KK, Colquhoun, H, Levac, D, Moher, D, Peters, MD, Horsley, T, Weeks, L, Hempel, S et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467-73. PMID: 30178033. doi: 10.7326/M18-0850.

31. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021): Cochrane; 2021. Available from: [www.training.cochrane.org/handbook](file:///\\hscifs6\usershomedir$\spg3\Documents\2020\Race%20on%20Twitter\www.training.cochrane.org\handbook).

32. Hinds J, Joinson AN. What demographic attributes do our digital footprints reveal? A systematic review. PLoS One. 2018 2018;13(11):e0207112. PMID: 30485305.

33. Abubakar U, Bashir, S. A., Abdullahi, M. B., Adebayo, O. S. Comparative Study of Various Machine Learning Algorithms for Tweet Classification. Journal on Computer Science. 2019;6(4):12-24. doi: 10.26634/jcom.6.4.15722.

34. An J, Weber I. # greysanatomy vs.# yankees: Demographics and Hashtag Use on Twitter. Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016). 2016.

35. Ardehaly E, Culotta A. Lightly Supervised Machine Learning for Classifying Online Social Data (Dissertation). Ann Arbor: Illinois Institute of Technology. 2017 2017(10269108):159. PMID: 1916585479.

36. Ardehaly EM, Culotta A. Inferring latent attributes of Twitter users with label regularization. Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Denver, Colorado: Association for Computational Linguistics. 2015:185–95.

37. Ardehaly EM, Culotta A. Co-training for demographic classification using deep learning from label proportions. IEEE International Conference on Data Mining Workshops, ICDMW. 2017. doi: <https://doi.org/10.1109/ICDMW.2017.144>.

38. Auguste D, Polman JL, Miller SL. A Data Science Approach to STEM (Science, Technology, Engineering and Math) Identity Research for African American Communities. Ann Arbor: University of Colorado at Boulder. 2019 2019(13860170):148. PMID: 2231590809.

39. Barbera P. Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data. Work Paper NYU <http://pablobarberacom/static/less-is-morepdf>. 2017.

40. Bergsma S, Dredze M, Van Durme B, Wilson T, Yarowsky D. Broadly Improving User Classification via Communication-Based Name and Location Clustering on Twitter. HLT-NAACL Atlanta, Georgia, pp 1010–1019. 2013.

41. Blodgett SL, Wei J, O'Connor B. Twitter Universal Dependency Parsing for African-American and Mainstream American English. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), July 2018 Melbourne, Australia: Association for Computational Linguistics. 2018:1415-25.

42. Borradaile G, Burkhardt B, LeClerc A. Whose tweets are surveilled for the police: an audit of a social-media monitoring tool via log files. FAT\* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, January 2020 Barcelona, Spain: Association for Computing Machinery. 2020 2020:570–80.

43. Cavazos-Rehg P, Krauss M, Grucza R, Bierut L. Characterizing the followers and tweets of a marijuana-focused Twitter handle. Journal of Medical Internet Research. 2014 Jun-27;16(6):e157. PMID: 24974893.

44. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. Journal of Adolescent Health. 2015 Feb;56(2):139-45. PMID: 25620299.

45. Cavazos-Rehg PA, Zewdie K, Krauss MJ, Sowles SJ. "No High Like a Brownie High": A Content Analysis of Edible Marijuana Tweets. American Journal of Health Promotion. 2017 May;32(4):880-6. PMID: 29214836.

46. Cesare N, Spiro ES, Stovel K. United We Tweet?: A Quantitative Analysis of Racial Differences in Twitter Use. Ann Arbor: Thesis (PhD)University of Washington. 2017 2017(10686866):181. PMID: 2015186423.

47. Chakraborty A, Messiaso J, Benevenutoo F, Ghosh S, Ganguly N, Gummadi KP. Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations. 11th AAAI International Conference on Web and Social Media (ICWSM 2017). 2017.

48. Chen X, Wang Y, Agichtein E, Wang F. A Comparative Study of Demographic Attribute Inference in Twitter. ICWSM Proceedings of the Ninth International AAAI Conference on Web and Social Media Palo Alto, CA, pp 590–593. 2015.

49. Culotta A, Ravi NK, Cutler J. Predicting Twitter User Demographics using Distant Supervision from Website Traffic Data. Journal of Artificial Intelligence Research. 2016 2016;55:389-408.

50. De Choudhury M. Tie Formation on Twitter: Homophily and Structure of Egocentric Networks. IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, Boston, MA, 2011, pp 465-470, doi: 101109/PASSAT/SocialCom2011177. 2011 9-11-Oct.:465-70.

51. Firmansyah FM, Jones JJ. Did the Black Panther Movie Make Blacks Blacker? Examining Black Racial Identity on Twitter Before and After the Black Panther Movie Release. Lecture Notes in Computer Science. 2019;11864:66-78.

52. Golder S, Chiuve S, Weissenbacher D, Klein A, O'Connor K, Bland M, et al. Pharmacoepidemiologic Evaluation of Birth Defects from Health-Related Postings in Social Media During Pregnancy. Drug Saf. 2018 Mar;42(3):389-400. PMID: 30284214.

53. González YM, Cutter SL. Leveraging Geotagged Social Media to Monitor Spatial Behavior during Population Movements Triggered by Hurricanes (Doctoral dissertation). Ann Arbor: University of South Carolina. 2019 2019(13881663):1-180. PMID: 2312896199.

54. Haffner M. A spatial analysis of non-English Twitter activity in Houston, TX. Transactions in Gis. 2018 Aug;22(4):913-29.

55. He L, Murphy L, Luo JB, Berendt B, Bringmann B, Fromont E, et al. Using Social Media to Promote STEM Education: Matching College Students with Role Models. Machine Learning and Knowledge Discovery in Databases Berlin: Springer-Verlag Berlin. 2016;9853:79-95.

56. Hswen Y, Hawkins JB, Sewalk K, Tuli G, Williams DR, Viswanath K, et al. Racial and Ethnic Disparities in Patient Experiences in the United States: 4-Year Content Analysis of Twitter. Journal of Medical Internet Research. 2020;22(8):e17048. PMID: 32821062. doi: 10.2196/17048.

57. Huang W, Weber I, Vieweg S. Inferring nationalities of Twitter users and studying inter-national linking. HT '14: Proceedings of the 25th ACM conference on hypertext and social media, September 2014. 2014:237–42

58. Huang X, Xing L, Dernoncourt F, Paul MJ. Multilingual Twitter Corpus and Baselines for Evaluating Demographic Bias in Hate Speech Recognition. Proceedings of the 12th Language Resources and Evaluation Conference Marseille, France: European Language Resources Association. 2020:1440–8.

59. Karlsen AS, Scott KD, Dueñas G, Gelbukh A, Rodriguez-Diaz CA, Mancera S. Making sense of Starbucks' anti-bias training and the arrests of two African American men : A thematic analysis of Whites' Facebook and Twitter comments Automatic Detection of Regional Words for Pan-Hispanic Spanish on Twitter. Discourse, Context & Media. 2019 2019;32:404-16. doi: 10.1016/j.dcm.2019.100332.

60. Kteily NS, Rocklage MD, McClanahan K, Ho AK. Political ideology shapes the amplification of the accomplishments of disadvantaged vs. advantaged group members. Proceedings of the National Academy of Sciences of the United States of America. 2019 Jan-29;116(5):1559-68. PMID: 30642960.

61. Longley PA, Adnan M. Geo-temporal Twitter demographics. International Journal of Geographical Information Science. 2016 2016;30:369-89. doi: 10.1080/13658816.2015.1089441.

62. Luo F, Cao G, Mulligan K, Li X. Explore spatiotemporal and demographic characteristics of human mobility via Twitter : A case study of Chicago. Applied Geography. 2016 2016;70:11-25. doi: 10.1016/j.apgeog.2016.03.001.

63. Markson CR. Detecting user demographics in twitter to inform health trends in social media (Dissertation). New Jersey Institute of Technology. 2017.

64. Messias J, Vikatos P, Benevenuto F. White, man, and highly followed: gender and race inequalities in Twitter. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'17), August 2017 Leipzig, Germany: Association for Computing Machinery 2017 2017:266–74.

65. Mislove A, Lehmann S, Ahn Y-Y, J-P.J. O, Rosenquist N. Understanding the Demographics of Twitter Users. Fifth International AAAI Conference on Weblogs and Social Media. 2011.

66. Mohammady E, A C. Using county demographics to infer attributes of Twitter users. ACL Joint Workshop on Social Dynamics and Personal Attributes in Social Media. 2014.

67. Murthy D, Gross A, Pensavalle A. Urban Social Media Demographics : An Exploration of Twitter Use in Major American Cities. Journal of Computer Mediated Communication. 2016 2016;21:33-49. PMID: RN374903031.

68. Nguyen V, Tran M, Luo J. Are French Really That Different? Recognizing Europeans from Faces Using Data-Driven Learning. 24th International Conference on Pattern Recognition (ICPR), Beijing. 2018 20-24-Aug.:2729-34.

69. Pennacchiotti M, Popescu A-M. Democrats, republicans and starbucks afficionados: user classification in twitter. KDD '11: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, August 2011 San Diego, California, USA: Association for Computing Machinery. 2011 2011:430–8.

70. Pennacchiotti M, Popescu A-M. A Machine Learning Approach to Twitter User Classification. Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21. 2011:281–8.

71. Reis JCS, Kwak H, An J, Messias J, Benevenuto F. Demographics of News Sharing in the U.S. Twittersphere. HT '17: Proceedings of the 28th ACM Conference on Hypertext and Social Media, July 2017, Prague, Czech Republic: Association for Computing Machinery. 2017 2017:195–204.

72. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. A Study of the Demographics of Web-Based Health-Related Social Media Users. Journal of Medical Internet Research. 2015 Aug-6;17(8):e194. PMID: 26250986.

73. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. Demographic-Based Content Analysis of Web-Based Health-Related Social Media. Journal of Medical Internet Research. 2016 Jun-13;18(6):e148. PMID: 27296242.

74. Saravanan M. Determining Ethnicity of Immigrants using Twitter Data. MISNC '17: Proceedings of the 4th Multidisciplinary International Social Networks Conference, July 2017 Bangkok, Thailand: Association for Computing Machinery. 2017 2017:Article 7.

75. Sowles SJ, Krauss MJ, Connolly S, Cavazos-Rehg PA. A Content Analysis of Vaping Advertisements on Twitter, November 2014. Preventing Chronic Disease. 2016 Sep-29;13:E139. PMID: 27685432.

76. Stewart I. Now We Stronger than Ever: African-American English Syntax in Twitter. Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics Gothenburg, Sweden: Association for Computational Linguistics. 2014:31-7.

77. Vikatos P, Messias J, Manoel M, Benevenuto F. Linguistic Diversities of Demographic Groups in Twitter. HT '17: Proceedings of the 28th ACM Conference on Hypertext and Social Media, July 2017, Prague, Czech Republic: Association for Computing Machinery. 2017 2017:275–84.

78. Volkova S, Backrach Y. Inferring Perceived Demographics from User Emotional Tone and User-Environment Emotional Contrast. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), August 2018 Berlin, Germany: Association for Computational Linguistics. 2018:1567–78.

79. Wang W, Chi G. Who Are You? Estimating Demographics of Twitter Users. In PAA 2017 Annual Meeting PAA, 2017. 2017.

80. Wang Y, Li Y, Luo J. Deciphering the 2016 U.S. Presidential Campaign in the Twitter Sphere: A Comparison of the Trumpists and Clintonists. 10th International AAAI Conference on Web and Social Media. 2016.

81. Wood-Doughty Z, Andrews N, Marvin R, Dredze M. Predicting Twitter User Demographics from Names Alone. Proceedings of the Second Workshop on Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media, pages 105–111 New Orleans, Louisiana, June 6, 2018 Association for Computational Linguistics. 2018.

82. Wood-Doughty Z, Smith M, Broniatowski DA, Dredze M. How Does Twitter User Behavior Vary Across Demographic Groups? Proceedings of the Second Workshop on Natural Language Processing and Computational Social Science, pages 83–89, Vancouver, Canada, August 3, 2017 Association for Computational Linguistics. 2017.

83. Wood-Doughty Z, Xu P, Liu X, Dredze M, editors. Using Noisy Self-Reports to Predict Twitter User Demographics. Proceedings of the Ninth International Workshop on Natural Language Processing for Social Media; 2021; Association for Computational Linguistics; in.

84. Xu S, Markson C, Costello KL, Xing CY, Demissie K, Llanos AA. Leveraging Social Media to Promote Public Health Knowledge: Example of Cancer Awareness via Twitter. JMIR Public Health Surveillance. 2016 Jan-Jun;2(1):e17. PMID: 27227152.

85. Ye J, Han S, Hu Y, Coskun B, Liu M, Qin H, et al. Nationality Classification Using Name Embeddings. CIKM '17: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, November 2017 Singapore, Singapore: Association for Computing Machinery. 2017 2017:1897–906.

86. Yin J, Chi G, Hook JV. Evaluating the Representativeness in the Geographic Distribution of Twitter User Population. GIR'18: Proceedings of the 12th Workshop on Geographic Information Retrieval, November 2018 Article No: 6 WA, USA: Association for Computing Machinery. 2018 2018:1-2.

87. Rivas R. Automated Analysis of User-Generated Content on the Web [Ph.D.]. Ann Arbor: University of California, Riverside. 2021.

88. Adnan M, Longley PA, Khan SM. Social dynamics of Twitter usage in London, Paris, and New York City. First Monday. 2014 10.5210/fm.v19i5.4820;19(5).

89. Coleman L. "We're a Part of This City Too": An Examination of the Politics of Representation of D.C. Native via #DCNativesDay. Social Media + Society. 2021;7. doi: 10.1177/2056305120984446.

90. Saha K, Yousuf, A., Hickman, L., editor. A Social Media Study on Demographic Diferences in Perceived Job Satisfaction. Proc ACM Hum-Comput Interact; 2021: ACM.

91. Hong T, Wu, J., Wijaya, D., Xuan, Z., Fetterman, J. JUUL the heartbreaker: Twitter analysis of cardiovascular health perceptions of vaping. Tobacco Induced Diseases. 2021;19. doi: 10.18332/tid/130961.

92. Jiang J, Vosoughi, S., editor. Not judging a user by their cover: Understanding harm in multi-modal processing within social media research. 2nd International Workshop on Fairness, Accountability, Transparency and Ethics in Multimedia; 2020: ACM; 2020.

93. Mueller A, Wood-Doughty Z, Amir S, Dredze M, Nobles AL, editors. Demographic Representation and Collective Storytelling in the Me Too Twitter Hashtag Activism Movement. Proceedings of arXiv ACM, New York, NY, USA, 27 pages 2020.

94. Aguirre C, Harrigian K, Dredze M, editors. Gender and Racial Fairness in Depression Research using Social Media. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume; 2021: Association for Computational Linguistics.

95. Aguirre C, Dredze M, editors. Qualitative Analysis of Depression Models by Demographics. Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access; 2021: Association for Computational Linguistics.

96. Gunarathne P, Rui H, Seidmann A, editors. Racial Discrimination in Social Media Customer Service: Evidence from a Popular Microblogging Platform. HICSS; 2019.

97. Ye J, Skiena S. The Secret Lives of Names? Name Embeddings from Social Media. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery &amp; Data Mining; Anchorage, AK, USA: Association for Computing Machinery; 2019. p. 3000–8.

98. Ardehaly EM, Culotta A, Raghavan V, Aluru S, Karypis G, Miele L, et al. Mining the Demographics of Political Sentiment from Twitter Using Learning from Label Proportions. IEEE International Conference on Data Mining Workshops, ICDMW 2017. 2017:733-8. doi: 10.1109/ICDMW.2017.144.

99. An J, Ciampaglia GL, Grinberg N, Joseph K, Mantzarlis A, Maus G, et al. Reports of the Workshops Held at the 2017 International AAAI Conference on Web and Social Media. AI Magazine. 2017 Winter;38(4):93-8. PMID: 1987348159. doi: 10.1609/aimag.v38i4.2772.

100. Anindya IC. Understanding and Mitigating Privacy Risks Raised by Record Linkage [Ph.D.]. . Ann Arbor: The University of Texas at Dallas; 2020.

101. Bardier C. Detecting Electronic Cigarette User Disparity Behaviors: An Infovelliance Study on Twitter. . Ann Arbor, University of California, San Diego: 49.2020.

102. Basterra L, Worthington T, Rogol J, Brown D. Socio-Temporal Trends in Urban Cultural Subpopulations through Social Media. New York: IEEE; 2017 2017. 341-6 p. ISBN: 978-1-5386-1848-6.

103. Beretta V, Maccagnola D, Cribbin T, Messina E. An Interactive Method for Inferring Demographic Attributes in Twitter. HT '15: Proceedings of the 26th ACM Conference on Hypertext & Social Media. 2015 2015:113–22.

104. Bergsma S, Van Durme B. Using conceptual class attributes to characterize social media users. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2013:710-20.

105. Bi B, Shokouhi M, Kosinski M, Graepel T. Inferring the demographics of search users: social data meets search queries. International World Wide Web Conference

Committee (IW3C2). 2013 2013:131–40. doi: 10.1145/2488388.2488401.

106. Blevins T, Kwiatkowski R, Macbeth J, McKeown K, Patton D, Rambow O. Automatically processing tweets from gang-involved youth: towards detecting loss and aggression. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016:2196-206.

107. Blodgett SL, Wei J, O’Connor B. A dataset and classifier for recognizing social media english. Proceedings of the 3rd Workshop on Noisy User-generated Text, September 2017 Copenhagen, Denmark: Association for Computational Linguistics. 2017:56-61.

108. Bokanyi E, Kondor D, Dobos L, Sebok T, Steger J, Csabai I, et al. Race, religion and the city: twitter word frequency patterns reveal dominant demographic dimensions in the United States. Palgrave Communications. 2016 2016;2:9.

109. Brinkman N, Jacobi L. Racial Identities on Social Media: Projecting Racial Identities on Facebook, Instagram, and Twitter. Ann Arbor: Minnesota State University, Mankato; 2018.

110. Burnap P, Colombo G, Amery R, Hodorog A, Scourfield J. Multi-class machine classification of suicide-related communication on Twitter. Online Soc Netw Media. 2017 Aug;2:32-44. PMID: 29278258.

111. Cesare N, Grant C, Nguyen Q, Lee H, Nsoesie EO. How well can machine learning predict demographics of social media users? arXiv preprint arXiv:170201807. 2017.

112. Chan MS, Winneg K, Hawkins L, Farhadloo M, Jamieson KHA, D. Legacy and social media respectively influence risk perceptions and protective behaviors during emerging health threats: A multi-wave analysis of communications on Zika virus cases. Social Science & Medicine. 2018;212:50-9. doi: 10.1016/j.socscimed.2018.07.007.

113. Chenworth M, Perrone J, Love JS, Greller HA, Sarker A, Chai PR. Buprenorphine Initiation in the Emergency Department: a Thematic Content Analysis of a #firesidetox Tweetchat. J Med Toxicol. 2020 Jan-2. PMID: 31898154.

114. Cheong M, Lee V. Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base. SWSM '09: Proceedings of the 2nd ACM workshop on Social web search and mining. 2009 2009:1–8. doi: 10.1145/1651437.1651439.

115. Chi G, Giles L, Kifer D, Van Hook J, Yin J. Predicting Twitter User Demographics as a First Step in Big Data for Population Research: Developing Unsupervised, Scalable Methods Using Real-Time, Large-Scale Twitter Data. 2017 International Population Conference. 2017.

116. Claude F, Konow R, Ladra S. Fast compressed-based strategies for author profiling of social media texts. CERI '16: Proceedings of the 4th Spanish Conference on Information Retrieval. 2016 2016:Article 14. doi: 10.1145/2934732.2934744.

117. Compton R, Lee C, Lu TC, De Silva L, Macy M, Glass K, et al. Detecting future social unrest in unprocessed Twitter data "Emerging Phenomena and Big Data". New York: IEEE; 2013. 56-60 p. ISBN: 978-1-4673-6213-9; 978-1-4673-6214-6.

118. Cui Y, He Q. Augmenting Household Travel Survey and Travel Behavior Analysis using Large-Scale Social Media Data and Smartphone GPS Data. Ann Arbor: State University of New York at Buffalo; 2019.

119. Dai HY, Hao JQ. Mining social media data for opinion polarities about electronic cigarettes. Tobacco Control. 2017 Mar;26(2):6. doi: 10.1136/tobaccocontrol-2015-052818.

120. Daughton AR, Paul MJ. Identifying Protective Health Behaviors on Twitter: Observational Study of Travel Advisories and Zika Virus. Journal of Medical Internet Research. 2019 May;21(5):16. PMID: 31094347. doi: 10.2196/13090.

121. DeJohn AD, Schulz EE, Pearson AL, Lachmar EM, Wittenborn AK. Identifying and Understanding Communities Using Twitter to Connect About Depression: Cross-Sectional Study. JMIR Mental Health. 2018 Nov;5(4):10. PMID: 30401662. doi: 10.2196/mental.9533.

122. Diaz F, Gamon M, Hofman JM, Kiciman E, Rothschild D. Online and Social Media Data As an Imperfect Continuous Panel Survey. PLoS One. 2016 2016;11(1):e0145406. PMID: 26730933.

123. Ding Q, Cui Q. Using Social Media to Evaluate Public Acceptance of Infrastructure Projects. Ann Arbor: University of Maryland, College Park; 2018.

124. Eisenstein J. Phonological factors in social media writing. Proceedings of the Workshop on Language Analysis in Social Media. 2013:11-9.

125. Farrington S. A Case Study on Black Twitter’s Reactions to the Framing of Blacks in Dove’s 2017 Facebook Advertisement. Ann Arbor, University of South Florida: 59.2020.

126. Filho RM, Almeida JM, Pappa GL. Twitter Population Sample Bias and its impact on predictive outcomes: a case study on elections. 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2015 2015:1254–61. doi: 10.1145/2808797.2809328.

127. Filippova K. User demographics and language in an implicit social network. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012 2012:1478–88.

128. Frederick E, Lim CH. A world of one-way and two-way streets: Exploring the nuances of fan-athlete interaction on Twitter. Ann Arbor: Indiana University; 2012.

129. Georgiou T, Abbadi AE, Yan X. Privacy Cyborg: Towards Protecting the Privacy of Social Media Users. IEEEXplore. 2017 19-22-April:1395-6.

130. Ghazouani D, Lancieri L, Ounelli H, Jebari C. Assessing socioeconomic status of Twitter users: A survey. Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019). 2019:388-98.

131. Gibbons J, Malouf R, Spitzberg B, Martinez L, Appleyard B, Thompson C, et al. Twitter-based measures of neighborhood sentiment as predictors of residential population health. PLoS One. 2019 2019;14(7):e0219550. PMID: 31295294.

132. Gilchrist-Herring NS. An Analysis of Attitudes Towards Transgender Individuals Utilizing Social Media Usage, Ethnicity, Gender, Age Range, and Level of Education. . Ann Arbor, Capella University: 116.2020.

133. Gilchrist-Herring NS. An Analysis of Attitudes Towards Transgender Individuals Utilizing Social Media Usage, Ethnicity, Gender, Age Range, and Level of Education [Ph.D.]. Ann Arbor: Capella University; 2020.

134. Giorgi S, Yaden DB, Eichstaedt JC, Ashford RD, Buffone AEK, Schwartz HA, et al. Cultural Differences in Tweeting about Drinking Across the US. Int J Environ Res Public Health. 2020 Feb-11;17(4). PMID: 32053866.

135. Gong W, Lim EP. Profiling Social Media Users with Selective Self-Disclosure Behavior. Ann Arbor: Singapore Management University (Singapore); 2016.

136. Gong Z, Song D. Towards Secure and Privacy-Preserving Online Social Networking Services. Ann Arbor: University of California, Berkeley; 2015.

137. Gundecha P, Ranganath S, Feng Z, Liu H. A tool for collecting provenance data in social media. KDD '13: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. 2013 2013:1462–5. doi: 10.1145/2487575.2487713.

138. Guo GM, Zhu FD, Chen EH, Liu Q, Wu L, Guan C. From Footprint to Evidence: An Exploratory Study of Mining Social Data for Credit Scoring. ACM Transactions on the Web. 2016 Dec;10(4):38. doi: 10.1145/2996465.

139. Gupta H, Lam T, Pettigrew S, Tait RJ. The association between exposure to social media alcohol marketing and youth alcohol use behaviors in India and Australia. BMC Public Health. 2018 Jun-13;18(1):726. PMID: 29895264.

140. Haffner M. A place-based analysis of #BlackLivesMatter and counter-protest content on Twitter. GeoJournal. 2019 Oct;84(5):1257-80. doi: 10.1007/s10708-018-9919-7.

141. Ikeda K, Hattori G, Ono C, Asoh H, Higashino T. Twitter user profiling based on text and community mining for market analysis. Knowledge-Based Systems. 2013 Oct;51:35-47. doi: 10.1016/j.knosys.2013.06.020.

142. Irel, ME, Chen Q, Schwartz HA, Ungar LH, Albarracin D. Action Tweets Linked to Reduced County-Level HIV Prevalence in the United States: Online Messages and Structural Determinants. AIDS Behav. 2016 Jun;20(6):1256-64. PMID: 26650382.

143. Jha D, Singh R. SMARTS: the social media-based addiction recovery and intervention targeting server. Bioinformatics. 2020 Oct-24. PMID: 31647520.

144. Jimenez S, Dueñas G, Gelbukh A, Rodriguez-Diaz CA, Mancera S. Automatic detection of regional words for pan-hispanic spanish on twitter. Ibero-American Conference on Artificial Intelligence. 2018:404-16.

145. Jones T. Toward a Description of African American Vernacular English Dialect Regions Using ``Black Twitter''. American speech. 2015 2015;90:403-40. PMID: RN604432960.

146. Jørgensen A, Hovy D, Søgaard A. Learning a POS tagger for AAVE-like language. Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016:1115-20.

147. Kang Y, Zeng X, Zhang Z, Wang Y, Fei T. Who are happier? Spatio-temporal Analysis of Worldwide Human Emotion Based on Geo-Crowdsourcing Faces. 2018 22-23-March:1-8.

148. Kent JD, Capello HT. Spatial patterns and demographic indicators of effective social media content during the Horsethief Canyon fire of 2012. Cartography and Geographic Information Science. 2013 2013;40(2):78-89. doi: 10.1080/15230406.2013.776727.

149. Khan S, Zheleva E. Debiasing 2016 Twitter Election Analysis via Multi-Level Regression and Poststratification (MRP). Ann Arbor: University of Illinois at Chicago; 2019.

150. Kostakos P, Pandya A, Kyriakouli O, Oussalah M, editors. Inferring Demographic Data of Marginalized Users in Twitter with Computer Vision APIs. European Intelligence and Security Informatics Conference (EISIC); 2018 24-25 Oct. 2018.

151. Kotze E, Senekal B. Employing sentiment analysis for gauging perceptions of minorities in multicultural societies: An analysis of Twitter feeds on the Afrikaner community of Orania in South Africa. The Journal for Transdisciplinary Research in Southern Africa. 2018 Nov;14(1):11. doi: 10.4102/td.v14i1.564.

152. Kumar D, Ukkusuri SV. Enhancing demographic coverage of hurricane evacuation behavior modeling using social media. Journal of Computational Science. 2020;45:101184. doi: 10.1016/j.jocs.2020.101184.

153. Lachlan KA, Spence PR, Lin X. Expressions of risk awareness and concern through Twitter: On the utility of using the medium as an indication of audience needs. Computers in Human Behavior. 2014;35:554-9. doi: 10.1016/j.chb.2014.02.029.

154. Lama Y, Chen T, Dredze M, Jamison A, Quinn SC, Broniatowski DA. Discordance Between Human Papillomavirus Twitter Images and Disparities in Human Papillomavirus Risk and Disease in the United States: Mixed-Methods Analysis. J Med Internet Res. 2018 Sep-14;20(9):e10244. PMID: 30217792.

155. Lee LA, Jimenez-Munoz G. Virtual homespace:(Re)constructing the body and identity through social media. Ann Arbor: State University of New York at Binghamton; 2016.

156. Lee-Won RJ, White TN, Potocki B. The Black catalyst to tweet : the role of discrimination experience, group identification, and racial agency in Black Americans' instrumental use of Twitter. Information, communication & society. 2018 2018;21:1097-115. doi: 10.1080/1369118X.2017.1301516.

157. Li J, Ritter A, Hovy E. Weakly supervised user profile extraction from twitter. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014:165-74.

158. Lienemann BA, Unger JB, Cruz TB, Chu KH. Methods for Coding Tobacco-Related Twitter Data: A Systematic Review. Journal of Medical Internet Research. 2017 Mar;19(3):18. PMID: 28363883. doi: 10.2196/jmir.7022.

159. Lin Y-R. Assessing Sentiment Segregation in Urban Communities. SocialCom '14: Proceedings of the 2014 International Conference on Social Computing. 2014 2014:1–8. doi: 10.1145/2639968.2640066.

160. Long T, Tew M. As Seen On Twitter: African-American Rhetorical Traditions Gone Viral. Ann Arbor: Eastern Michigan University; 2012.

161. Luong TBT. Human activity recognition: A data-driven approach. Dissertation Abstracts International Section A: Humanities and Social Sciences. 2016 2016;76(8):No Pagination Specified. PMID: 2016-17334-281.

162. Lwowski B, Rios A. The risk of racial bias while tracking influenza-related content on social media using machine learning. J Am Med Inform Assoc. 2021 Mar 18;28(4):839-49. PMID: 33484133. doi: 10.1093/jamia/ocaa326.

163. Magdy A, Ghanem TM, Musleh M, Mokbel MF, editors. Understanding language diversity in local twitter communities. Proceedings of the 27th ACM Conference on Hypertext and Social Media; 2016.

164. Maheshwari T, Reganti AN, Chakraborty T, Das A. Socio-Ethnic Ingredients of Social Network Communities. CSCW '17 Companion: Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 2017:235–8. doi: 10.1145/3022198.3026322.

165. Meng HW, Kath S, Li DP, Nguyen QC. National substance use patterns on Twitter. Plos One. 2017 Nov;12(11):15. PMID: 29107961. doi: 10.1371/journal.pone.0187691.

166. Montasser O KD. Predicting Demographics of High-Resolution Geographies with Geotagged Tweets. . In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence AAAI Press, Vancouver, Canada, pp 1460–1466. 2017.

167. Nguyen Msph T, Adams N, Huang D, Glymour M, Allen A, Nguyen Q. State-level racial attitudes and adverse birth outcomes: applying natural language processing to Twitter data to quantify state context for pregnant women. JMIR Public Health and Surveillance. 2020;6:e17103. PMID: 32298232. doi: 10.2196/17103.

168. Mulders D, Bodt Cd, Bjell J, Pentl AS, Verleysen M, Montjoye Yd. Improving individual predictions using social networks assortativity. 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM). 2017 28-30-June:1-8. doi: 0.1109/WSOM.2017.8020023.

169. Nelson JK, Quinn S, Swedberg B, Chu WH, MacEachren AM. Geovisual Analytics Approach to Exploring Public Political Discourse on Twitter. Isprs International Journal of Geo-Information. 2015 Mar;4(1):337-66. doi: 10.3390/ijgi4010337.

170. Nguyen QC, Kath S, Meng HW, Li D, Smith KR, VanDerslice JA, et al. Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. Appl Geogr. 2016 Aug;73:77-88. PMID: 28533568.

171. Nguyen QC, Li D, Meng HW, Kath S, Nsoesie E, Li F, et al. Building a National Neighborhood Dataset From Geotagged Twitter Data for Indicators of Happiness, Diet, and Physical Activity. JMIR Public Health Surveill. 2016 Oct-17;2(2):e158. PMID: 27751984.

172. Novak AN, Johnson KC, Pontes M. LatinoTwitter: Discourses of Latino civic engagement in social media. First Monday. 2016;21(8).

173. Odlum M, Cho H, Broadwell P, Davis N, Patrao M, Schauer D, et al. Application of Topic Modeling to Tweets as the Foundation for Health Disparity Research for COVID-19. Studies in Health Technology and Informatics. 2020;Volume 272: The Importance of Health Informatics in Public Health during a Pandemic:24-7. PMID: 32604591. doi: 10.3233/SHTI200484.

174. Oktay H, Firat A, Ertem Z. Demographic breakdown of Twitter users: An analysis based on names. Academy of Science and Engineering (ASE), Big Data, Socialcom, Cybersecurity Conference. 2014.

175. Orsolini L, Papanti GD, Francesconi G, Schifano F. Mind navigators of chemicals' experimenters? A web-based description of e-psychonauts. Cyberpsychol Behav Soc Netw. 2015 May;18(5):296-300. PMID: 25965863.

176. Pick J, Sarkar A, Rosales J. Social Media Use in American Counties: Geography and Determinants. Isprs International Journal of Geo-Information. 2019 Sep;8(9):25. doi: 10.3390/ijgi8090424.

177. Polimis K, Crowder K, Lee H. Developing Computational Approaches to Investigate Health Inequalities. Ann Arbor: University of Washington; 2017.

178. Priante A, Hiemstra D, Saeed A, van den Broek T, Ehrenhard M, Need A. #WhoAmI in 160 Characters? Classifying Social Identities Based on Twitter Profile Descriptions. Proceedings of the First Workshop on NLP and Computational Social Science; Austin, Texas: Association for Computational Linguistics; 2016. p. 55-65.

179. Riederer CJ, Zimmeck S, Phanord C, Chaintreau A, Bellovin SM. “I don’t have a photograph, but you can have my footprints.”: Revealing the Demographics of Location Data. OSN '15: Proceedings of the 2015 ACM on Conference on Online Social Networks. 2015 2015:185–95. doi: 10.1145/2817946.2817968.

180. Roberts MJ, Perera M, Lawrentschuk N, Romanic D, Papa N, Bolton D. Globalization of continuing professional development by journal clubs via microblogging: a systematic review. J Med Internet Res. 2015 Apr-23;17(4):e103. PMID: 25908092.

181. Roy S, Ghosh P. A Comparative Study on Distancing, Mask and Vaccine Adoption Rates from Global Twitter Trends. Healthcare (Basel). 2021 Apr 21;9(5). PMID: 33919097. doi: 10.3390/healthcare9050488.

182. Rummo PE, Cassidy O, Wells I, Coffino JA, Bragg MA. Examining the Relationship between Youth-Targeted Food Marketing Expenditures and the Demographics of Social Media Followers. Int J Environ Res Public Health. 2020 Mar-3;17(5). PMID: 32138342.

183. Runge K. “Social” Science, Spider Goats and American Science Audiences: Investigating the Effects of Interpersonal Networks on Perceptions of Emerging Technologies [Ph.D.]. Ann Arbor: The University of Wisconsin - Madison; 2017.

184. Sijtsma B, Qvarfordt P, Chen F, editors. Tweetviz: Visualizing tweets for business intelligence. Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval; 2016.

185. Singh M, Singh A, Bansal D, Sofat S. An Analytical Model for Identifying Suspected Users on Twitter. Cybernetics and Systems. 2019:383-404. doi: 10.1080/01969722.2019.1588968.

186. Tomeny TS, Vargo CJ, El-Toukhy S. Geographic and demographic correlates of autism-related anti-vaccine beliefs on Twitter, 2009-15. Social Science & Medicine. 2017 Oct;191:168-75. PMID: 28926775. doi: 10.1016/j.socscimed.2017.08.041.

187. Tulloch JS. An appraisal of health datasets to enhance the surveillance of Lyme disease in the United Kingdom. Ann Arbor: The University of Liverpool (United Kingdom); 2019.

188. Vydiswaran VGV, Romero DM, Zhao XY, Yu DH, Gomez I, Lu JX, et al. Uncovering the relationship between food-related discussion on Twitter and neighborhood characteristics. Journal of the American Medical Informatics Association. 2020 Feb;27(2):254-64. PMID: 31633756. doi: 10.1093/jamia/ocz181.

189. Wang Y, Feng Y, Luo J, Zhang X. Voting with Feet: Who are Leaving Hillary Clinton and Donald Trump. IEEE International Symposium on Multimedia (ISM), San Jose, CA. 2016 11-13-Dec.:71-6. doi: 10.1109/ISM.2016.0022.

190. Weeg C, Schwartz HA, Hill S, Merchant RM, Arango C, Ungar L. Using Twitter to Measure Public Discussion of Diseases: A Case Study. JMIR Public Health Surveill. 2015 Jun-26;1(1):e6. PMID: 26925459.

191. Wright MCN, Adams T. #KnowBetterDoBetter: An Examination of Twitter Impact on Disaster Literacy. Ann Arbor: Howard University; 2019.

192. Yazdavar AH, Mahdavinejad MS, Bajaj G, Romine W, Sheth A, Monadjemi AH, et al. Multimodal mental health analysis in social media. PLoS One. 2020 2020;15(4):e0226248. PMID: 32275658.

193. Ying QF, Chiu DM, Venkatramanan S, Zhang X. Profiling OSN Users Based on Temporal Posting Patterns. WWW '18: Proceedings of The Web Conference 2018. 2018 2018:1451–6. doi: 10.1145/3184558.3191592.

194. Yuan F, Li M, Zhai W, Qi B, Liu R. Social Media Based Demographics Analysis for Understanding Disaster Response Disparity. Construction Research Congress 2020: Computer Applications2020. p. 1020-8.

195. Zhang Z, Bors G. “Less is more”. Online Information Review. 2019 2019--;43(1):213-37. PMID: 2337998364.

196. Zhao P, Jia J, An Y, Liang J, Xie L, Luo J. Analyzing and Predicting Emoji Usages in Social Media. Www '18. 2018:327–34. doi: 10.1145/3184558.3186344.

197. Zhong Y, Yuan NJ, Zhong W, Zhang F, Xie X. You Are Where You Go: Inferring Demographic Attributes from Location Check-ins. WSDM '15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. 2015 2015:295–304. doi: 0.1145/2684822.2685287.

198. Jiang YQ, Li ZL, Ye XY. Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level. Cartography and Geographic Information Science. May;46(3):228-42. doi: 10.1080/15230406.2018.1434834.

199. Preotiuc-Pietro D, Ungar L. User-Level Race and Ethnicity Predictors from Twitter Text. In Proceedings of the 27th International Conference on Computational Linguistics, Aug, 2018, Santa Fe, New Mexico, USA, Association for Computational Linguistics, 1534-1545, <https://wwwaclweborg/anthology/C18-1130>. 2018.

200. Face++. Face++. 2022 [cited 2022 24th Feb]; Available from: <https://www.faceplusplus.com/>.

201. DemographicsPro. Powerful Audience Demographics. 2022 [cited 2022 24th Feb]; Available from: <https://www.demographicspro.com/>.

202. Onomap. Onomap is Changing. 2022 [cited 2022 24th Feb]; Available from: <https://www.onomap.org/>.

203. Twitter Inc. Academic research: Preparing for the Academic Research application: Learn everything there is to know about applying for the Academic Research product track. <https://developer.twitter.com/en/solutions/academic-research/application-info>. 2021.

204. Reuters WSC. Twitter Grants Academics Full Access to Public Data, but Not for Suspended Accounts. Jan. 26, 2021. <https://www.usnews.com/news/technology/articles/2021-01-26/twitter-grants-academics-full-access-to-public-data-but-not-for-suspended-accounts>

2021.

205. Jung S, An J, Kwak H, Salminen J, Jansen B. Assessing the Accuracy of Four Popular Face Recognition Tools for Inferring Gender, Age, and Race. ICWSM. 2018.

206. Buolamwini J, Gebru T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. 2018;81:77--91.

207. Jung SG, An, J., Kwak, H., Salminen, J. & Jansen, B.J. Inferring Social Media Users’ Demographics from Profile Pictures: A Face++ Analysis on Twitter Users. Proceedings of The 17th International Conference on Electronic Business (pp 140-145) ICEB, Dubai, UAE, December 4-8. 2017.

208. Gong S., Liu X., A.K. J. Jointly De-Biasing Face Recognition and Demographic Attribute Estimation. In: Vedaldi A., Bischof H., Brox T., Frahm JM. (eds) Computer Vision – ECCV 2020. . Lecture Notes in Computer Science, vol 12374 Springer, Cham. 2020:330-47. doi: 10.1007/978-3-030-58526-6\_20.

209. Fu S, He H, Hou Z. Learning Race from Face: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014;36(12):2483-509. doi: 10.1109/TPAMI.2014.2321570.

210. Goldinger SD, He Y, Papesh MH. Deficits in cross-race face learning: insights from eye movements and pupillometry. J Exp Psychol Learn Mem Cogn. 2009;35(5):1105-22. PMID: 19686008. doi: 10.1037/a0016548.

211. Meissner C, Brigham J. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. Psychology, Public Policy and Law. 2001;7:3-35.

212. Jofre A, Berardi V, Brennan KPJ, Cornejo A, Bennett C, Harlan J. Crowdsourcing Image Extraction and Annotation: Software Development and Case Study. Digital Humanities Quarterly. 2020;14.

213. King RD, Johnson BD. A Punishing Look: Skin Tone and Afrocentric Features in the Halls of Justice. American Journal of Sociology. 2016 2016/07/01;122(1):90-124. PMID: 29873458. doi: 10.1086/686941.

214. Cavazos J, Philips PJ, Castillo C, O'Toole AJ. Accuracy comparison across face recognition algorithms: Where are we on measuring race bias? IEEE Journals & Magazine IEEE Transactions in Biometrics, Behavior and Identity Science 2020 doi: 101109/TBIOM20203027269. 2020.

215. Buolamwini J, Gebru T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: Sorelle AF, Christo W, eds. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. Proceedings of Machine Learning Research: PMLR, 2018:77--91. 2018.

216. Torralba A, Efros AA. "Unbiased look at dataset bias," CVPR 2011, Providence, RI, 2011, pp. 1521-1528, doi: 10.1109/CVPR.2011.5995347. 2011.

217. Moscrop A, Ziebland S, Bloch G, Iraola JR. If social determinants of health are so important, shouldn’t we ask patients about them? BMJ. 2020;371:m4150. PMID: 33234506. doi: 10.1136/bmj.m4150.

218. McHugh M. Interrater reliability: the kappa statistic. Biochemia medica. 2012;22. PMID: 23092060.

219. Inc T. Developer terms: More about restricted uses of the Twitter APIs. 2022 [cited 2022 4th March]; Available from: <https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases>.

220. Alsaied T, Allen KY, Anderson JB, Anixt JS, Brown DW, Cetta F, et al. The Fontan outcomes network: first steps towards building a lifespan registry for individuals with Fontan circulation in the United States. Cardiol Young. 2020 Aug;30(8):1070-5. PMID: 32635947. doi: 10.1017/s1047951120001869.