

Short Paper

Automatically Identifying Twitter Users for Interventions to Support Dementia Family Caregivers: Annotated Data Set and Benchmark Classification Models

Ari Z Klein¹, PhD; Arjun Magge¹, PhD; Karen O'Connor¹, MS; Graciela Gonzalez-Hernandez², PhD

¹Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

²Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA, United States

Corresponding Author:

Ari Z Klein, PhD

Department of Biostatistics, Epidemiology, and Informatics

Perelman School of Medicine

University of Pennsylvania

Blockley Hall, 4th Fl.

423 Guardian Dr.

Philadelphia, PA, 19104

United States

Phone: 1 310 423 3521

Email: ariklein@pennmedicine.upenn.edu

Abstract

Background: More than 6 million people in the United States have Alzheimer disease and related dementias, receiving help from more than 11 million family or other informal caregivers. A range of traditional interventions has been developed to support family caregivers; however, most of them have not been implemented in practice and remain largely inaccessible. While recent studies have shown that family caregivers of people with dementia use Twitter to discuss their experiences, methods have not been developed to enable the use of Twitter for interventions.

Objective: The objective of this study is to develop an annotated data set and benchmark classification models for automatically identifying a cohort of Twitter users who have a family member with dementia.

Methods: Between May 4 and May 20, 2021, we collected 10,733 tweets, posted by 8846 users, that mention a dementia-related keyword, a linguistic marker that potentially indicates a diagnosis, and a select familial relationship. Three annotators annotated 1 random tweet per user to distinguish those that indicate having a family member with dementia from those that do not. Interannotator agreement was 0.82 (Fleiss kappa). We used the annotated tweets to train and evaluate support vector machine and deep neural network classifiers. To assess the scalability of our approach, we then deployed automatic classification on unlabeled tweets that were continuously collected between May 4, 2021, and March 9, 2022.

Results: A deep neural network classifier based on a BERT (bidirectional encoder representations from transformers) model pretrained on tweets achieved the highest F_1 -score of 0.962 (precision=0.946 and recall=0.979) for the class of tweets indicating that the user has a family member with dementia. The classifier detected 128,838 tweets that indicate having a family member with dementia, posted by 74,290 users between May 4, 2021, and March 9, 2022—that is, approximately 7500 users per month.

Conclusions: Our annotated data set can be used to automatically identify Twitter users who have a family member with dementia, enabling the use of Twitter on a large scale to not only explore family caregivers' experiences but also directly target interventions at these users.

(*JMIR Aging* 2022;5(3):e39547) doi: [10.2196/39547](https://doi.org/10.2196/39547)

KEYWORDS

natural language processing; social media; data mining; dementia; Alzheimer disease; caregivers

Introduction

More than 6 million people in the United States have Alzheimer disease and related dementias, and the burden is projected to double by 2060 [1]. Alzheimer disease is the sixth leading cause of death in the United States [2], and only 8% of people with dementia do not receive help from family members or other informal care providers [3], amounting to more than 11 million family or other unpaid caregivers in 2020 [4]. Caregivers of people with dementia are impacted physically, cognitively, socially, mentally, and financially. For instance, compared with noncaregivers, they are more vulnerable to disease due to chronic stress [5] and have lower durations and quality of sleep [6]. Compared with non-dementia caregivers, they are more likely to experience a decline in cognition [7] and social network size [8]. They are also more likely to experience depression compared with noncaregivers [9] and non-dementia caregivers [10], and depressive symptoms in dementia caregivers are associated with increased health care use and costs [11]. In addition to the increased costs of their personal health care, family caregivers of people with dementia pay for much of the recipient's total care costs, with the costs being significantly higher for people with dementia than without dementia [12].

A range of traditional interventions has been developed to support family caregivers of people with dementia [13]; however, most of them have not been implemented in practice and remain largely inaccessible [14]. Recent systematic reviews have concluded that internet-based interventions are valued by family caregivers of people with dementia for their easy access [15] and can have beneficial effects on caregivers' health [16]. While recent studies [17-23] have shown that family caregivers of people with dementia use Twitter to discuss their experiences, to the best of our knowledge, methods have not been developed to enable the use of Twitter as a platform for internet-based interventions. Given that nearly 1 of every 4 adults in the United States uses Twitter [24], Twitter may present a novel opportunity to reach family caregivers on a large scale, such as through user-targeted advertisements providing information about dementia, caregiving, resources, or services. The objective of this study was to develop an annotated data set and benchmark classification models for automatically identifying a cohort of Twitter users who have a family member with dementia.

Methods

Ethical Considerations

The data used in this study were collected in accordance with the Twitter Terms of Service. The Institutional Review Board of the University of Pennsylvania reviewed this study (protocol number: 828972) and deemed it exempt human subjects research under 45 CFR §46.101(b)(4) for publicly available data sources.

Data Collection and Annotation

Between May 4 and May 20, 2021, we collected 67,060 publicly available tweets from the Twitter streaming application programming interface (API) that are in English, are not retweets, and include both a dementia-related keyword (eg, *dementia*, *youngdementia*, *#yod*, *#ftd*, *alzheimer's*, *alz*,

alzheimersdisease, *mild cognitive impairment*) and a linguistic marker that potentially indicates a diagnosis (eg, *diagnosed*, *diagnosis*, *has*, *got*, *developed*, *with*, *from*). The full list of API search terms is available in [Multimedia Appendix 1](#). We then searched these tweets for references to select familial relationships ([Multimedia Appendix 2](#)), identifying 10,733 (16%) of the 67,060 tweets. We randomly sampled 1 tweet per user—8846 (82%) of the 10,733 tweets—and developed annotation guidelines ([Multimedia Appendix 3](#)) to help 3 annotators distinguish tweets that indicate having a family member with dementia from those that do not. Among the 8846 annotated tweets, 8346 (94%) were dual annotated, and 500 (6%) were annotated by all 3 annotators. Interannotator agreement, based on the 500 tweets annotated by all 3 annotators, was 0.82 (Fleiss kappa). Upon resolving the disagreements, it was determined that 5946 (67%) of the tweets indicate that the user has a family member with dementia, and 2900 (33%) of the tweets do not.

Automatic Classification

We performed benchmark supervised machine learning experiments to assess the utility of the annotated data set for automatically identifying Twitter users who have a family member with dementia. For the classifiers, we used the LibSVM [25] implementation of support vector machine (SVM) in Weka and SVM and 6 deep neural network classifiers based on BERT (bidirectional encoder representations from transformers): the BERT-Base-Uncased [26], DistilBERT-Base-Uncased [27], RoBERTa-Large [28], BioBERT-Large-Cased [29], Bio+ClinicalBERT [30], and BERTweet-Large [31] pretrained models in the *Flair* Python library. We split the 8846 tweets into 80% (7077 tweets) and 20% (1769 tweets) random sets as training data ([Multimedia Appendix 4](#)) and held-out test data, respectively, stratified based on the distribution of the binary annotated classes. For the SVM classifier, we preprocessed the tweets by normalizing URLs, usernames, digits, and keywords related to dementia ([Multimedia Appendix 1](#)) and familial relationships ([Multimedia Appendix 2](#)), removing nonalphanumeric characters and extra spaces, and lowercasing and stemming [32] the text. We used the Weka NGram Tokenizer to extract n-grams (n=1-3) as features in a bag-of-words representation. We used the radial basis function kernel and set the *cost* at $c=32$. For the BERT-based classifiers, we preprocessed the tweets by normalizing URLs and usernames and lowercasing the text. For training, we used stochastic gradient descent optimization, a batch size of 8, 15 epochs, and a learning rate of 0.001. During training, we fine-tuned all layers of the transformer model with our annotated tweets. To optimize performance, the model was evaluated after each epoch on a 5% split of the training set. To assess the scalability of our approach, we then deployed automatic classification on 198,674 unlabeled tweets, posted by 119,640 users, that were continuously collected from the Twitter streaming API ([Multimedia Appendix 1](#)) between May 4, 2021, and March 9, 2022, and mentioned a select familial relationship ([Multimedia Appendix 2](#)).

Results

Table 1 presents the precision, recall, and F_1 -scores of SVM and 6 deep neural network classifiers for the class of tweets indicating that the user has a family member with dementia, evaluated on a held-out test set of 1769 (20%) of the 8846 manually annotated tweets. The classifier based on a model pretrained on tweets (BERTweet-Large) achieved the highest F_1 -score: 0.962 (precision=0.946 and recall=0.979). When deployed on 198,674 unlabeled tweets, posted by 119,640 users, between May 4, 2021, and March 9, 2022, the BERTweet classifier detected 128,838 tweets indicating that the user has a family member with dementia, posted by 74,290 users—that is, approximately 7500 users per month.

Table 2 presents examples of false positives and false negatives of the BERTweet classifier in the test set. Among the 68 false positives, 36 (47%) refer to people with dementia who are not or may not be select family members (Tweet 1), 8 (12%) report that a family member has a condition other than dementia (Tweet 2), and 5 (7%) merely speculate that a family member has dementia (Tweet 3). Another 8 (12%) of the 68 false positives were a result of manual annotation errors. Among the 25 false negatives, 14 (56%) use deixis or anaphora, requiring additional context in the tweet to understand that a non-first person determiner (eg, “their” in Tweet 4) actually refers to the user, or that a personal pronoun (eg, “she” in Tweet 5) refers to a select family member with dementia. Furthermore, 12 (86%) of these 14 tweets also include references to people who are not family members or do not have dementia. Another 4 (16%) of the 25 false negatives were a result of manual annotation errors.

Table 1. Precision, recall, and F_1 -scores of classifiers for detecting tweets indicating that the user has a family member with dementia.

Classifier	Precision	Recall	F_1 -score
SVM ^a	0.884	0.939	0.910
BERT ^b -Base-Uncased	0.924	0.954	0.938
DistilBERT-Base-Uncased	0.930	0.942	0.936
RoBERTa-Large	0.918	0.982	0.949
BioBERT-Large-Cased	0.907	0.978	0.941
Bio+ClinicalBERT	0.903	0.958	0.930
BERTweet-Large	0.946	0.979	0.962

^aSVM: support vector machine.

^bBERT: bidirectional encoder representations from transformers.

Table 2. Sample false positives and false negatives of a BERTweet classifier for detecting tweets indicating that the user has a select family member with dementia.

Tweet number	Tweet	Actual	Predicted
1	Evelyn has dementia, I know. But when she asked me today how my dad was doing... it still hurt.	–	+
2	We really don't have a clue about what causes Alzheimer's. We don't have a clue about Parkinson's, which is what got my dad, either.	–	+
3	I just listened to the Everywhere at The End of Time, by The Caretaker, and thought about my grandmother. The songs are about dementia, something my grandma wasn't clearly diagnosed with, but it hit hard.	–	+
4	If someone tells u their parent has Alzheimer's please don't say your grandparent or great aunt did too. I appreciate that u can relate to the experience but it is so different. Tell me a different time.	+	–
5	I have a family member who is vulnerable and two children in their late 20s. I didn't want to risk passing virus to her or from her to my family member. My sister made a bubble with her and her carers. She has dementia so she probably hasn't missed me!	+	–

Discussion

Principal Findings

The benchmark performance of automatic classification demonstrates that our annotated data set has utility for accurately identifying Twitter users who have a family member with

dementia, and deploying automatic classification on unlabeled tweets demonstrates that a large cohort of users can be identified. Therefore, our annotated data set enables the use of Twitter to scale up accessible, internet-based interventions directly targeted at family caregivers of people with dementia. Because our approach involves identifying tweets that mention a familial

relationship, it would also enable interventions to be tailored to the care recipient.

Limitations

Our approach to identifying family caregivers assumes that having “close” relatives with dementia would likely imply the users’ involvement in caregiving; however, the users identified in this study may not necessarily be caregivers or may have been caregivers but are no longer. We took this approach because we believe that limiting our identification of caregivers to users who explicitly state that they are providing ongoing

care would underutilize the potential of Twitter for reaching caregivers on a large scale.

Conclusions

This paper presented an annotated data set and benchmark classification models for automatically identifying Twitter users who have a family member with dementia, enabling the use of Twitter on a large scale to not only explore family caregivers’ experiences among their tweets but also directly target interventions at these users.

Acknowledgments

This work was supported by the National Library of Medicine (R01LM011176). The authors thank Ivan Flores for contributing to software applications, and Alexis Upshur and Aiden McRobbie-Johnson for contributing to annotating the Twitter data.

Authors' Contributions

AZK designed the data collection, edited the annotation guidelines, performed the support vector machine classification experiments, conducted the error analysis, and wrote the manuscript. AM performed the deep learning classification experiments, deployed the BERTweet classifier, and edited the manuscript. KO developed the annotation guidelines, annotated the Twitter data, and edited the manuscript. GGH conceptualized and guided the study and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Twitter streaming application programming interface search terms.

[\[TXT File , 3 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Family member keywords.

[\[TXT File , 0 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Annotation guidelines.

[\[PDF File \(Adobe PDF File\), 119 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Training data.

[\[TXT File , 159 KB-Multimedia Appendix 4\]](#)

References

1. Matthews KA, Xu W, Gaglioti AH, Holt JB, Croft JB, Mack D, et al. Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015-2060) in adults aged ≥65 years. *Alzheimers Dement* 2019 Jan 19;15(1):17-24 [[FREE Full text](#)] [doi: [10.1016/j.jalz.2018.06.3063](https://doi.org/10.1016/j.jalz.2018.06.3063)] [Medline: [30243772](https://pubmed.ncbi.nlm.nih.gov/30243772/)]
2. Kochanek K, Xu J, Arias E. Mortality in the United States, 2019. *NCHS Data Brief* 2020 Dec(395):1-8. [Medline: [33395387](https://pubmed.ncbi.nlm.nih.gov/33395387/)]
3. Kasper JD, Freedman VA, Spillman BC, Wolff JL. The disproportionate impact of dementia on family and unpaid caregiving to older adults. *Health Aff (Millwood)* 2015 Oct;34(10):1642-1649 [[FREE Full text](#)] [doi: [10.1377/hlthaff.2015.0536](https://doi.org/10.1377/hlthaff.2015.0536)] [Medline: [26438739](https://pubmed.ncbi.nlm.nih.gov/26438739/)]
4. Alzheimer's Association. 2021 Alzheimer's disease facts and figures. *Alzheimers Dement* 2021 Mar;17(3):327-406. [doi: [10.1002/alz.12328](https://doi.org/10.1002/alz.12328)] [Medline: [33756057](https://pubmed.ncbi.nlm.nih.gov/33756057/)]
5. Fonareva I, Oken BS. Physiological and functional consequences of caregiving for relatives with dementia. *Int Psychogeriatr* 2014 May;26(5):725-747 [[FREE Full text](#)] [doi: [10.1017/S1041610214000039](https://doi.org/10.1017/S1041610214000039)] [Medline: [24507463](https://pubmed.ncbi.nlm.nih.gov/24507463/)]

6. Gao C, Chapagain NY, Scullin MK. Sleep duration and sleep quality in caregivers of patients with dementia: a systematic review and meta-analysis. *JAMA Netw Open* 2019 Aug 02;2(8):e199891 [FREE Full text] [doi: [10.1001/jamanetworkopen.2019.9891](https://doi.org/10.1001/jamanetworkopen.2019.9891)] [Medline: [31441938](https://pubmed.ncbi.nlm.nih.gov/31441938/)]
7. Dassel K, Carr D, Vitaliano P. Does caring for a spouse with dementia accelerate cognitive decline? Findings from the health and retirement study. *Gerontologist* 2017 Apr 01;57(2):319-328. [doi: [10.1093/geront/gnv148](https://doi.org/10.1093/geront/gnv148)] [Medline: [26582383](https://pubmed.ncbi.nlm.nih.gov/26582383/)]
8. Liu C, Fabius CD, Howard VJ, Haley WE, Roth DL. Change in social engagement among incident caregivers and controls: findings from the caregiving transitions study. *J Aging Health* 2021 Jan 23;33(1-2):114-124. [doi: [10.1177/0898264320961946](https://doi.org/10.1177/0898264320961946)] [Medline: [32962491](https://pubmed.ncbi.nlm.nih.gov/32962491/)]
9. Ma M, Dorstyn D, Ward L, Prentice S. Alzheimers' disease and caregiving: a meta-analytic review comparing the mental health of primary carers to controls. *Aging Ment Health* 2018 Nov 05;22(11):1395-1405. [doi: [10.1080/13607863.2017.1370689](https://doi.org/10.1080/13607863.2017.1370689)] [Medline: [28871796](https://pubmed.ncbi.nlm.nih.gov/28871796/)]
10. Sheehan O, Haley W, Howard V, Huang J, Rhodes J, Roth D. Stress, burden, and well-being in dementia and nondementia caregivers: insights from the caregiving transitions study. *Gerontologist* 2021 Jul 13;61(5):670-679 [FREE Full text] [doi: [10.1093/geront/gnaa108](https://doi.org/10.1093/geront/gnaa108)] [Medline: [32816014](https://pubmed.ncbi.nlm.nih.gov/32816014/)]
11. Zhu CW, Scarmeas N, Ornstein K, Albert M, Brandt J, Blacker D, et al. Health-care use and cost in dementia caregivers: Longitudinal results from the Predictors Caregiver Study. *Alzheimers Dement* 2015 Apr 17;11(4):444-454 [FREE Full text] [doi: [10.1016/j.jalz.2013.12.018](https://doi.org/10.1016/j.jalz.2013.12.018)] [Medline: [24637299](https://pubmed.ncbi.nlm.nih.gov/24637299/)]
12. Kelley AS, McGarry K, Bollens-Lund E, Rahman O, Husain M, Ferreira KB, et al. Residential setting and the cumulative financial burden of dementia in the 7 years before death. *J Am Geriatr Soc* 2020 Jun 18;68(6):1319-1324 [FREE Full text] [doi: [10.1111/jgs.16414](https://doi.org/10.1111/jgs.16414)] [Medline: [32187655](https://pubmed.ncbi.nlm.nih.gov/32187655/)]
13. Gaugler JE, Potter T, Prunelli L. Partnering with caregivers. *Clin Geriatr Med* 2014 Aug;30(3):493-515. [doi: [10.1016/j.cger.2014.04.003](https://doi.org/10.1016/j.cger.2014.04.003)] [Medline: [25037292](https://pubmed.ncbi.nlm.nih.gov/25037292/)]
14. Gitlin L, Marx K, Stanley I, Hodgson N. Translating evidence based dementia caregiving interventions into practice: state-of-the-science and next steps. *Gerontologist* 2015 Apr;55(2):210-226 [FREE Full text] [doi: [10.1093/geront/gnu123](https://doi.org/10.1093/geront/gnu123)] [Medline: [26035597](https://pubmed.ncbi.nlm.nih.gov/26035597/)]
15. Hopwood J, Walker N, McDonagh L, Rait G, Walters K, Iliffe S, et al. Internet-based interventions aimed at supporting family caregivers of people with dementia: systematic review. *J Med Internet Res* 2018 Jun 12;20(6):e216 [FREE Full text] [doi: [10.2196/jmir.9548](https://doi.org/10.2196/jmir.9548)] [Medline: [29895512](https://pubmed.ncbi.nlm.nih.gov/29895512/)]
16. Leng M, Zhao Y, Xiao H, Li C, Wang Z. Internet-based supportive interventions for family caregivers of people with dementia: systematic review and meta-analysis. *J Med Internet Res* 2020 Sep 09;22(9):e19468 [FREE Full text] [doi: [10.2196/19468](https://doi.org/10.2196/19468)] [Medline: [32902388](https://pubmed.ncbi.nlm.nih.gov/32902388/)]
17. Yoon S. What can we learn about mental health needs from tweets mentioning dementia on World Alzheimer's Day? *J Am Psychiatr Nurses Assoc* 2016 Nov 01;22(6):498-503 [FREE Full text] [doi: [10.1177/1078390316663690](https://doi.org/10.1177/1078390316663690)] [Medline: [27803262](https://pubmed.ncbi.nlm.nih.gov/27803262/)]
18. Danilovich M, Tsay J, Al-Bahrani R, Choudhary A, Agrawal A. #Alzheimer's and dementia: expressions of memory loss on Twitter. *Topics in Geriatric Rehabilitation* 2018;34(1):48-53. [doi: [10.1097/TGR.0000000000000173](https://doi.org/10.1097/TGR.0000000000000173)]
19. Cheng TY, Liu L, Woo BK. Analyzing Twitter as a platform for Alzheimer-related dementia awareness: thematic analyses of Tweets. *JMIR Aging* 2018 Dec 10;1(2):e11542 [FREE Full text] [doi: [10.2196/11542](https://doi.org/10.2196/11542)] [Medline: [31518232](https://pubmed.ncbi.nlm.nih.gov/31518232/)]
20. Yoon S, Lucero R, Mittelman MS, Luchsinger JA, Bakken S. Mining Twitter to inform the design of online interventions for Hispanic Alzheimer's disease and related dementias caregivers. *Hisp Health Care Int* 2020 Sep 24;18(3):138-143. [doi: [10.1177/1540415319882777](https://doi.org/10.1177/1540415319882777)] [Medline: [31646904](https://pubmed.ncbi.nlm.nih.gov/31646904/)]
21. Mehta N, Zhu L, Lam K, Stall NM, Savage R, Read SH, et al. Health forums and Twitter for dementia research: opportunities and considerations. *J Am Geriatr Soc* 2020 Dec 07;68(12):2881-2889. [doi: [10.1111/jgs.16790](https://doi.org/10.1111/jgs.16790)] [Medline: [32894780](https://pubmed.ncbi.nlm.nih.gov/32894780/)]
22. Bacsu J, O'Connell ME, Cammer A, Azizi M, Grewal K, Poole L, et al. Using Twitter to understand the COVID-19 experiences of people with dementia: infodemiology study. *J Med Internet Res* 2021 Feb 03;23(2):e26254 [FREE Full text] [doi: [10.2196/26254](https://doi.org/10.2196/26254)] [Medline: [33468449](https://pubmed.ncbi.nlm.nih.gov/33468449/)]
23. Yoon S, Broadwell P, Alcantara C, Davis N, Lee H, Bristol A, et al. Analyzing topics and sentiments from Twitter to gain insights to refine interventions for family caregivers of persons with Alzheimer's disease and related dementias (ADRD) during COVID-19 pandemic. *Stud Health Technol Inform* 2022 Jan 14;289:170-173 [FREE Full text] [doi: [10.3233/SHTI210886](https://doi.org/10.3233/SHTI210886)] [Medline: [35062119](https://pubmed.ncbi.nlm.nih.gov/35062119/)]
24. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021 Apr 07. URL: <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/> [accessed 2022-02-25]
25. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2011 Apr;2(3):1-27. [doi: [10.1145/1961189.1961199](https://doi.org/10.1145/1961189.1961199)]
26. Devlin J, Cheng M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. 2019 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); June 2-7, 2019; Minneapolis, Minnesota, US p. 4171-4186. [doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)]

27. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERTmaller, faster, cheaper and lighter. 2019 Presented at: Fifth Workshop on Energy Efficient Machine Learning and Cognitive Computing; December 13, 2019; Vancouver, Canada.
28. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv Preprint posted online July 26, 2019. [[FREE Full text](#)]
29. Lee J, Yoon W, Kim S, Kim D, Kim S, So C, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020 Feb 15;36(4):1234-1240 [[FREE Full text](#)] [doi: [10.1093/bioinformatics/btz682](https://doi.org/10.1093/bioinformatics/btz682)] [Medline: [31501885](https://pubmed.ncbi.nlm.nih.gov/31501885/)]
30. Alsentzer E, Murphy J, Boag W, Weng W, Jindi D, Naumann T, et al. Publicly available clinical BERT embeddings. 2019 Presented at: Proceedings of the 2nd Clinical Natural Language Processing Workshop; June 7, 2019; Minneapolis, Minnesota, USA p. 72-78. [doi: [10.18653/v1/w19-1909](https://doi.org/10.18653/v1/w19-1909)]
31. Nguyen D, Vu T, Nguyen A. BERTweet: a pre-trained language model for English tweets. 2020 Presented at: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; November 16-20, 2020; Online p. 9-14. [doi: [10.18653/v1/2020.emnlp-demos.2](https://doi.org/10.18653/v1/2020.emnlp-demos.2)]
32. Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems* 1980;14(3):130-137. [doi: [10.1108/eb046814](https://doi.org/10.1108/eb046814)]

Abbreviations

- API:** application programming interface
BERT: bidirectional encoder representations from transformers
SVM: support vector machine
-

Edited by J Wang, T Leung; submitted 16.05.22; peer-reviewed by K Verspoor, JW Kwon; comments to author 27.06.22; revised version received 08.07.22; accepted 08.07.22; published 16.09.22

Please cite as:

*Klein AZ, Magge A, O'Connor K, Gonzalez-Hernandez G
Automatically Identifying Twitter Users for Interventions to Support Dementia Family Caregivers: Annotated Data Set and Benchmark Classification Models
JMIR Aging 2022;5(3):e39547
URL: <https://aging.jmir.org/2022/3/e39547>
doi: [10.2196/39547](https://doi.org/10.2196/39547)
PMID:*

©Ari Z Klein, Arjun Magge, Karen O'Connor, Graciela Gonzalez-Hernandez. Originally published in JMIR Aging (<https://aging.jmir.org>), 16.09.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Aging, is properly cited. The complete bibliographic information, a link to the original publication on <https://aging.jmir.org>, as well as this copyright and license information must be included.