Envisioning the Future of Health Informatics and Digital Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI250055

Using C4.5 Algorithm to Gain Insights on Stakeholder Engagement and Use of Artificial Intelligence on Social Media in Dementia Caregiving Disparity Research

Sunmoo YOON[a,1], Frederick SUN[b], Peter BROADWELL[c], Robert CRUPI[b], Melissa PATTERSON[d], Tess D. POTTINGER[a], Milea KIM[b] and Ncole DAVIS[e]
[a] *General Medicine, Columbia University, New York, NY*
[b] *New York Presbyterian Hospital, New York, NY*
[c] *Center for Interdisciplinary Digital Research, Stanford University, Stanford, CA*
[d] *Division of Geriatric Medicine and Aging, Columbia University, New York, NY*
[e] *School of Nursing, Clemson University, Clemson, SC*
ORCID iD:, Sunmoo Yoon https://orcid.org/0000-0001-9413-8980, Peter Broadwell https://orcid.org/0000-0003-4371-9472

Abstract. We applied machine learning techniques to build models that predict perceived risks and benefits of using artificial intelligence (AI) algorithms to recruit African American informal caregivers for clinical trials and general health disparity research via social media platforms. In a U.S. sample of 572 family caregivers of a person with Alzheimer's disease and related dementias (ADRD), our application of the J48 algorithm (C4.5) revealed an interesting trend. African American family members of a person with ADRD were more likely to see the benefits of using AI on social media to ease the burden of recruitment, regardless of age, ethnicity, gender, and level of education. However, white family caregivers, particularly those aged 25-34 with graduate degrees, were more cautious and prone to perceive risks of using AI on social media for recruitment in research. This caution underscores the need for further research and understanding in this area.

Keywords. social media, dementia caregiving, health disparity, AI

1. Introduction

African Americans in the U.S. experience a 64% higher rate of progression to Alzheimer's disease and related dementias (ADRD) compared to non-Hispanic Whites [1]. Care for a person with ADRD is usually provided by informal caregivers such as family or friends. Caregiving is exceptionally demanding compared to other chronic conditions [1]. Understanding the social determinants of health and their impact on caregiver outcomes is essential for addressing health disparities in caregiving. Effective recruitment strategies are crucial for obtaining a representative sample and ensuring rigor in scientific studies, such as those investigating new caregiving intervention methods for underserved groups of dementia family caregivers [2]. While community-based research

---

[1] Corresponding Author, Sunmoo Yoon, General Medicine, Department of Medicine, Columbia University Irving Medical Center, 630W 168 Street, PH105, New York, NY, 10032, USA; E-mail: sy2102@cumc.columbia.edu

strategies are considered the gold standard for recruitment and retention, they often fall short in practice. Engaging potential participants in the community requires research staff with high competency and strong interpersonal skills to build trust within a limited timeframe, and offering compelling incentives is not always feasible. The authors discovered that social media, especially X (formerly Twitter), is attracting growing interest from underserved populations. X's features, such as user-verified checkmarks to distinguish legitimate accounts, show potential in enhancing recruitment, community engagement, and retention efforts [3].

Furthermore, AI automation has the potential to achieve scientific rigor by overcoming the barriers of labor-intensive traditional methods (e.g., time and geography) and providing better representative sampling [2]. However, leveraging such an AI algorithm presents both risks and benefits. Understanding and engaging stakeholders is imperative to identify these risks and benefits effectively. Our study aims to apply the C4.5 algorithm to examine attitudes toward perceived risks and benefits of observing caregiving stakeholders' characteristics and using demographic detection algorithms to assist recruitment for a clinical trial via social media. This is intended to provide insights for developing engagement strategies for ADRD caregiving disparity research in the U.S.

2. Methods

The study was approved by the Institutional Review Board (IRB). We applied a data mining process to a dataset, a survey of 572 American family caregivers of a person with ADRD conducted from April 8th to May 2nd, 2024, to build a model to understand caregiving stakeholder characteristics and the use of AI on social media in ADRD caregiving disparity research. We used the R software, version 4.3.3, and Weka version 3.8.6. Demographic and caregiving background variables were selected by caregiving and behavior experts as relevant to AI acceptance and implementation based on a literature review [4].

The outcome variable was the dichotomized perceived risks and benefits. The respondents were asked to rank the perceived risks and benefits on a five-point Likert scale (strongly agree [greater benefits], moderately agree, neutral, moderately disagree, strongly disagree [greater risks]); in the subsequent summary analyses, these responses were collapsed to "Yes" or "No" to answer the following two scenario questions: Scenario A [The use of AI for clinical trial recruitment on social media]: I think the potential gain (benefit) is more significant than the potential loss (risk) of using AI algorithms when recruiting African American caregivers for a clinical trial over social media. Scenario B [The use of AI for Facebook, X, Reddit, and WhatsApp for data collection and research]: I think the potential gains (benefits) are more significant than the potential loss (risks) of using AI algorithms to recruit African American caregivers using Facebook, Reddit, YouTube, WhatsApp and Twitter for data collection and health disparity research.

We iteratively applied the C4.5 algorithm to build the prediction models for perceived risks and benefits for family caregivers of a person with ADRD to gain insights on stakeholder engagement strategies and the use of AI on social media in dementia caregiving disparity research. For cross-validation (10-fold), the dataset was randomly divided into training and evaluating datasets for model validation before applying the algorithms. We chose the final models based on the model's predictive accuracy, interpretability, and meaningfulness, and the area under the receiver operating characteristic curve (AUC). Lastly, dementia caregiving disparity and informatics experts interpreted the models according to the meaningfulness and applicability of the model for insights into stakeholder engagement strategies. Resources are available on GitHub and OSF.io (https://osf.io/qruf3).

3. Results

A total of 572 family caregivers of individuals with ADRD in the U.S. took part in an online survey. More than half of the respondents were aged 35 or older (51.22%, n=293). The study had a majority of male respondents (73.29%, n=417). The majority of the participants (82.17%, n=470) identified themselves as Black or African American family members of a person with ADRD. One in ten participants identified as Hispanic. Most participants (72.11%, n=411) did not have a graduate degree. Additionally, less than half of the participants were taking care of their parents or grandparents (39.86%, n=228).

Regarding the respondents' views on the use of AI for clinical trial recruitment on social media, 55.4% of those who were 18-24 years old did not agree that the risks were greater than the benefits, followed by 22.3% of those who were 25-34 years old. Missing values were removed from the analysis, with 3 respondents (0.5%) preferring not to report their gender and 4 respondents (0.7%) not reporting their education (Table 1)

Table 1. Statistics for Study Participants (n=572)

		The use of AI for clinical trial recruitment on social media		**The use of AI for Facebook, X, Reddit, and WhatsApp for data collection and research**	
	Total (N=572)	**Agree (N=401)**	**Not (N=171)**	**Agree (N=415)**	**Not (N=157)**
Age		[X^2 p= 0.001]		[X^2 p= 0.001]	
18-24	37 (6.5)	23 (5.7)	14 (8.2)	20 (4.8)	17 (10.8)
25-34	242 (42.3)	142 (35.4)	100 (58.5)	155 (37.3)	87 (55.4)
35+	293 (51.2)	236 (58.9)	57 (33.3)	240 (57.8)	53 (33.8)
Gender		[X^2 p= 0.0852]		[X^2 p= 0.287]	
Male	417 (72.9)	281 (70.1)	136 (79.5)	295 (71.1)	122 (77.7)
Female	152 (26.6)	117 (29.2)	35 (20.5)	118 (28.4)	34 (21.7)
Race		[X^2 p= 0.0218]		[X^2 p= 0.001]	
African American	470 (82.2)	341 (85.0)	129 (75.4)	361 (87.0)	109 (69.4)
White	102 (17.8)	60 (15.0)	42 (24.6)	54 (13.0)	48 (30.6)
Education		[X^2 p= 0.1]		[X^2 p= 0.00229]	
≤High school	19 (3.3)	8 (2.0)	11 (6.4)	7 (1.7)	12 (7.6)
≤Bachelor	390 (68.2)	273 (68.1)	117 (68.4)	298 (71.8)	92 (58.6)
≤Graduate	159 (27.8)	117 (29.2)	42 (24.6)	107 (25.8)	52 (33.1)
Relation		[X^2 p= 0.642]		[X^2 p= 0.649]	
Parents or grandparents	228 (39.9)	163 (40.6)	65 (38.0)	170 (41.0)	58 (36.9)
Others	269 (47.0)	181 (45.1)	88 (51.5)	187 (45.1)	82 (52.2)
Spouse	75 (13.1)	57 (14.2)	18 (10.5)	58 (14.0)	17 (10.8)

In Scenario A [The use of AI for clinical trial recruitment on social media], it was found that family members of individuals with ADRD who were younger than 18 or older than 35 were more likely to see the benefits of using AI algorithms to recruit African American caregivers. However, among Black or African American family caregivers aged 25-34, those with a high school diploma, those who attended college without graduating, or those with a technical degree and not living with the care recipient were less likely to see the benefits and more likely to see the risks of using such algorithms. Similarly, non-Hispanic white or Caucasian family caregivers aged 25-34 were also more likely to see the risks rather than the benefits of using AI for clinical trial recruitment (Figure 1a).

In Scenario B [The use of AI for Facebook, X, Reddit, and WhatsApp for data collection and research], it was found that Black or African American family members of individuals with ADRD were more likely to support the use of AI algorithms on social media for data collection and caregiving disparity research, regardless of age, ethnicity, gender, and education. On the other hand, among white or Caucasian family members of individuals with ADRD aged 25-34 with graduate degrees, there was a higher likelihood of not supporting the use of such algorithms and perceiving them as risky (Figure 1b).

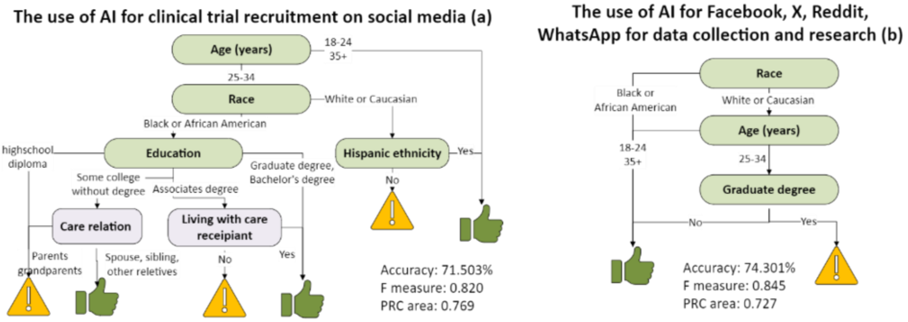

Figure 1. Prediction models for perceived gain or risks of the use of AI on social media for recruitment in a clinical trial (left) and in general (right)

4. Discussion and Conclusion

Our analysis using the J48 algorithm (C4.5) found a pattern of acceptance of the use of AI on social media to assist in recruiting participants for clinical trials and health disparity research. We found that African American family members of individuals with ADRD were more likely to recognize the benefits of using such an algorithm to facilitate recruitment, regardless of their age, ethnicity, gender, level of education or caregiving factors. This discovery highlights the potential of AI to ease the burden of recruitment to caregiving research communities, ultimately reducing health disparities [5]. Traditional recruitment methods using human labor are limited by barriers such as geographic location, time constraints, and biases inherited from characteristics of partner organizations in caregiving disparity research. AI has the potential to overcome these barriers and achieve broader generalizability [5].

Conversely, white family caregivers, particularly those in the 25-34 age group with graduate degrees, were more cautious of using AI systematically to reach potential targets on social media for recruitment. This caution emphasizes the need for in-depth research on motivation and concerns. Creative risk-takers, whose behaviors are regulated by elevated dopamine levels, are often motivated by excitement and passion for innovation [6]. Their confidence in risky situations leads to societal changes. On the other hand, safety-oriented individuals proactively seek safety and anticipate hazards. Because their strict adherence to rules can sometimes impede innovation, future work on deeper understanding of the motivations underlying these attitudes will be crucial [6].

The positive attitude towards AI-enhanced algorithms to ease the burden of recruitment and data collection for health disparity research among African American family members of a person with ADRD indicates that social media platforms like Facebook, X, Reddit, and WhatsApp can serve as valuable tools in engaging these communities [3]. However, we found concerns about using such technology among younger, educated White caregivers. The future work may include resolution efforts to carefully listen to these groups and strategically plan to address the concerns. For the other group, practicing perseverance and being patient will be a virtue that transforms innovation into hope during the transformation period [6].

Additionally, different social media platforms present unique challenges that must be navigated to ensure effective recruitment and data collection for caregiving disparity research [5]. For example, security and privacy settings on platforms like Facebook and WhatsApp differ from the more open nature of You-Tube and X necessitating tailored social-media community engagement approaches. Moreover, while demographic detection algorithms can improve recruitment, there is a risk of reinforcing stereotypes or alienating certain groups. This points to the need for careful implementation of these algorithms and sensitivity to the unique preferences of various demographic groups and caregiving situational context [5]. Our sampling strategy and medium-level predictive accuracy limit the generalizability of our study's findings, and a multisite study is needed in the future. Nevertheless, our goal was transparency over predictive accuracy.

In conclusion, this study highlights the potential of AI to enhance recruitment and data collection strategies in dementia caregiving research, particularly by engaging African American caregivers who recognize its benefits. However, the more cautious stance observed among younger, educated White caregivers’ points to the need for further research to address their concerns and foster trust in AI technologies. By promoting inclusive stakeholder engagement and addressing concerns related to privacy, ethics and transparency researchers can leverage AI to support equitable participation in dementia studies.

5. Acknowledgements

U.S. federal grants R01AG060929 and R01AG060929-04S1 (PI: Yoon)

References

[1] Mehta KM, Yeo GW. Systematic review of dementia prevalence and incidence in United States race/ethnic populations. Alzheimer’s. Dement. 2017;13:72-83.

[2] Broadwell P, Davis N, Yoon S. Using AI to develop a lexicon-based african american tweet detection algorithm to inform culturally sensitive twitter-based social support interventions for african american dementia caregivers. Stud Health Technol Inform. 2022 Jan;289:1-4.

[3] Broadwell P. et al. Analyzing Twitter-Based Social Networks of Support Communities for Hispanic and African American Family Caregivers of Persons with Dementia. Studies in health technology and informatics. 2023 Jun 6;305:155.

[4] Cruz A, Daum C, Comeau A, Salamanca JD, McLennan L, Neubauer N, Liu L. Acceptance, adoption, and usability of information and communication technologies for people living with dementia and their care partners: a systematic review. Disability and Rehabilitation. 2023 May;18(4):443-57

[5] Ali SH, Foreman J, Capasso A, Jones AM, Tozan Y, DiClemente RJ. Social media as a recruitment platform for a nationwide online survey of COVID-19 knowledge, beliefs, and practices in the United States: methodology and feasibility analysis. BMC Med. Res. Methodol. 2020 May;20(116) 1-11

[6] AlNuaimi BK, Singh SK, Ren S, Budhwar P, Vorobyev D. Mastering digital transformation: The nexus between leadership, agility, and digital strategy. J. Bus. Res 2022 Jun;145, 636-48