<u>Viewpoint</u>

A Perspective on Crowdsourcing and Human-in-the-Loop Workflows in Precision Health

Peter Washington, PhD

Information and Computer Sciences, University of Hawaii at Manoa, Honolulu, HI, United States

Corresponding Author:
Peter Washington, PhD
Information and Computer Sciences
University of Hawaii at Manoa
1680 East-West Road
Honolulu, HI, 96822
United States
Email: pyw@hawaii.edu

Abstract

Modern machine learning approaches have led to performant diagnostic models for a variety of health conditions. Several machine learning approaches, such as decision trees and deep neural networks, can, in principle, approximate any function. However, this power can be considered to be both a gift and a curse, as the propensity toward overfitting is magnified when the input data are heterogeneous and high dimensional and the output class is highly nonlinear. This issue can especially plague diagnostic systems that predict behavioral and psychiatric conditions that are diagnosed with subjective criteria. An emerging solution to this issue is crowdsourcing, where crowd workers are paid to annotate complex behavioral features in return for monetary compensation or a gamified experience. These labels can then be used to derive a diagnosis, either directly or by using the labels as inputs to a diagnostic machine learning model. This viewpoint describes existing work in this emerging field and discusses ongoing challenges and opportunities with crowd-powered diagnostic systems, a nascent field of study. With the correct considerations, the addition of crowdsourcing to human-in-the-loop machine learning workflows for the prediction of complex and nuanced health conditions can accelerate screening, diagnostics, and ultimately access to care.

KEYWORDS

Introduction

Crowdsourcing, a term first coined in 2006 [1], is the use of distributed human workers to accomplish a central task. Crowdsourcing exploits the "power of the crowd" to achieve goals that are only feasible with a distributed group of humans collaborating, either explicitly or implicitly, toward a common goal. Crowdsourcing has often been applied to public health surveillance [2], such as for tracking epidemics [3,4], quantifying tobacco use [5], monitoring water quality [6], tracking misinformation [7], and understanding the black-market price of prescription opioids [8]. In the context of health care, crowdsourcing is most often used for public health, a domain that can clearly benefit from scalable and distributed assessments of health status. Although sampling bias can be an issue in epidemiological uses of crowdsourcing [9], approaches that account for these issues have performed quite robustly.

A smaller but potentially transformative effort to apply crowdsourcing to precision health rather than population health has recently emerged. In precision health contexts, the goal is to provide a diagnosis using information labeled by crowd workers. There are several variations to this basic setup. Crowdsourcing workflows for diagnostics can diverge with respect to the underlying task, worker motivation strategies, worker training, worker filtering, and privacy requirements.

Here, I describe the existing research in the relatively small and early but growing field of crowdsourcing for precision health. I then discuss ongoing challenges and corresponding opportunities that must be addressed as this field matures.

Existing Examples of Crowdsourcing in and Adjacent to Health Care

There are relatively few examples of crowdsourcing in precision health. The vast successes of machine learning for health [10-15] and the human labor costs required for crowdsourcing make purely automated approaches more appealing when they are possible and feasible. However, the crowdsourcing approaches that have been tested tend to perform well for prediction tasks that are beyond the scope of current automated approaches, especially in psychiatry and the behavioral sciences.

I want to begin by highlighting successes in science, as they can often be applied to health and have started to lead to improvements in diagnostics. Framing crowdsourcing tasks as "citizen science" opportunities can be an effective incentive mechanism [16]. Oftentimes, these projects are "gamified." Gamification refers to the incorporation of engaging elements into traditionally burdensome workflows, and in particular game-like affordances, to foster increased participation. A combination of large crowd sizes, worker training procedures, and easy identification tasks have led to previous success in the existing gamified citizen science experiments applied to precision health. For example, in a study involving nearly 100,000 crowd workers who scored images on a citizen science platform, cancer was correctly identified with an area under the receive operating characteristic of around 95% [17]. In the BioGames app, users who performed with greater than 99% accuracy in a training tutorial were invited to diagnose malaria [18,19]. It was discovered that with a large crowd size, the aggregated diagnostic accuracy of nonexpert crowd workers approached that of experts [20]. Another citizen science malaria diagnosis application, MalariaSpot, resulted in 99% accuracy in the diagnosis of malaria from blood films [21]. If the annotation task is relatively simple and nonexperts can be trained with minimal onboarding efforts, then citizen science can be an effective and affordable approach.

"Gamified" crowdsourcing for citizen science has also been successful without explicitly requiring workers to undergo a formal training process. Foldit [22-25] and EteRNA [26-31] are 2 games where players with no biology or chemistry background can explore the design space of protein and RNA folding, respectively. These are both NP-hard (ie, computationally complex) problems, and human players in aggregate have designed solutions that outcompete state-of-the-art computational approaches. These solutions have been used to solve health challenges, such as finding a gene signature for active tuberculosis, which can potentially be used in tuberculosis diagnostics [32]. Other gamified experiences have been used to build training libraries for complex classification tasks in precision psychiatry. Notably, GuessWhat is a mobile charades game played between children with autism and their parents [33,34]. While the game provides therapeutic benefits to the child with autism [35], the game simultaneously curates automatic labels of behaviors related to autism through the metadata associated with gameplay [36,37]. These automatically annotated video data have been used to develop state-of-the-art computer vision models for behaviors related

to the diagnosis of autism, such as facial expression evocation [38-41], eye gaze [42], atypical prosody [43], and atypical body movements [44,45].

An alternative incentive mechanism is paid crowdsourcing. The most popular paid crowdsourcing platform, by far, is Amazon Mechanical Turk (MTurk) [46]. While paid crowdsourcing specifically for precision health is a relatively nascent field, the general study of paid crowdsourcing (particularly on MTurk) is quite mature. Studies have explored worker quality management [47], understanding crowd worker demographics [48], the generation of annotations for use in the training of machine learning models [49-53], the rights of crowd workers [54-56], and understanding crowd worker communities and economics [57-59]. Preliminary studies of paid crowdsourcing have yielded mixed success. Around 81% of images were correctly classified on MTurk in a study involving the grading of diabetic retinopathy from images, with workers failing to correctly indicate the level of severity [60]. In a separate binary labeling task for glaucomatous optic neuropathy, workers achieved sensitivity in the 80s but reached a specificity below 50% [61].

In a broader classification task of various medical conditions, workers consistently labeled the "easy" cases while struggling to correctly label and even refusing to label more complicated and nuanced tasks [62]. Clearly, there is a need for extensive innovations to the traditional paid crowdsourcing workflow to translate this methodology to precision health.

I have extensively investigated the utility of paid crowdsourcing for the diagnosis of autism from unstructured home videos, achieving relatively high diagnostic performance [63-66]. In these experiments, untrained annotators watched short videos depicting children with and without autism and answered questions about the behaviors depicted within the videos. These annotations were provided as input into previously developed machine learning models, achieving binary test performance in the 90s across performance metrics due to the reduction of the complex feature space (unstructured videos) into a low-dimensional representation (vectors of a few categorical ordinal values). This pipeline combining crowdsourcing and machine learning can possibly be extended to other diagnostic domains in psychiatry where the input feature space is complex, heterogeneous, and subjective.

Ongoing Challenges and Corresponding Opportunities

Since crowdsourcing for precision health care is an emerging field of study, numerous challenges must be solved for clinical translation to develop. In the proceeding sections, I highlight several areas that are pressing for the field and for which preliminary work has been published.

Worker Identification and Training

While traditional crowdsourcing can work with minimal to no worker training, complex annotation tasks require the identification of qualified workers. I have found that worker identification can occur through the quantification of their

performance on test tasks [67,68] and training promising workers [66]. Such crowd filtration paradigms will require domain-specific procedures. There is ample room to develop new crowdsourcing systems that inherently support natural worker identification and training procedures for crowdsourcing workflows that require well-designed training processes.

Worker Retention

Once proficient workers are identified, continually engaging and retaining these workers is critical. I have found that workers who are repeatedly encouraged by a human (or human-like chatbot) and treated as members of a broader research team tend to enjoy paid work and even ask for more tasks after the completion of the study [69]. Thus, it is possible that the guarantee of job security can lead to long-term worker retention. However, worker retention in unpaid settings that rely on intrinsic motivation will require additional innovations. For example, there exists an opportunity to explore the creation of crowd worker communities to provide a means of intrinsic motivation leading to worker retention.

Task Assignment

Certain workers perform exceptionally well on a subset of tasks while underperforming on other assigned tasks [70,71]. There is an opportunity to develop algorithmic innovations involving the effective and optimal assignment of workers to subtasks in a dynamic manner. Reinforcement learning could be a promising approach but has yet to be explored in such scenarios.

Privacy of Human Participants

Data in psychiatry and behavioral sciences are particularly sensitive. Ensuring that sensitive health information is handled appropriately and that workers' privacy is maintained is essential from an ethical perspective. There are 2 general families of approaches to achieving privacy in crowd-powered health care. First, the data can be modified to obscure sensitive information without removing information required for a diagnosis. I have explored privacy-preserving alterations to video data that obfuscate the identity of participants while maximizing the capacity for workers to annotate behaviors of interest [70,71]. For example, in the case of video analytics on bodily movements, the face can be tracked and blurred, or the body can be converted to a stick figure using a pose-based computer vision library. Sometimes, however, it is impossible to modify the data without severely degrading the diagnostic performance. Therefore, the second family of approaches involves carefully vetting crowd workers, training them, and onboarding them into a secure computing environment. In my previous experiences with this process [40], I discovered that crowd workers were enthusiastic about the prospect of the "job security" that is implied from the thorough vetting procedure and were, therefore, willing to complete extra privacy and security training (in our case, Research, Ethics, Compliance, and Safety training). There is ample room to expand upon these methods and to develop new paradigms and systems for crowdsourcing involving identifiable and protected health information.

Ensuring Reliability and Reproducibility

An intrinsic challenge when incorporating human workers into precision health workflows is the variability in human responses, both within workers and between workers. I have found that while most crowd workers are inconsistent in their annotation patterns, there are workers who provide consistently sensitive and specific annotations across a wide spectrum of data points [67]. It is therefore critical to measure both internal consistency and consistency against a gold standard when recruiting crowd workers for precision health care workflows.

Handling Financial Constraints

The crowdsourcing method with the lowest setup barriers is paid crowdsourcing. In such scenarios, financial constraints can limit the scalability of crowdsourcing workflows. One approach is to migrate from a paid system to a gamified system or another means of providing intrinsic motivation to crowd workers. However, achieving critical mass for large-scale pipelines is likely unattainable for such unpaid solutions. Paid crowd workers who consistently perform well could be recruited as full-time or long-term part-time employees for companies and organizations providing crowd-powered services. Integrating such workflows into a Food and Drug Administration (FDA)–approved process can be challenging, but it is worth exploring if it turns out that crowd-powered solutions for digital psychiatry continue to remain superior to pure-artificial intelligence (AI) approaches in the coming years.

Translation Outside of Research Contexts

While pure machine learning approaches for precision health are beginning to translate to clinical settings through formal FDA approval procedures, the prospect of translating human-in-the-loop methods that integrate crowd workers rather than expert clinicians is daunting, especially in light of the challenges mentioned above. However, if such approaches lead to clinical-grade performance for certain conditions that are challenging to diagnose using machine learning alone, then the extra implementation and regulatory effort required to migrate these methods into production-level workflows are likely to be warranted.

Conclusion

While machine learning for health has enabled and will continue to enable more efficient, precise, and scalable diagnostics for a variety of conditions, such models are unlikely to generalize to more difficult scenarios such as psychiatry and the behavioral sciences, which require the ability to identify complex and nuanced social human behavior. Crowd-powered human-in-the-loop workflows have the potential to mitigate some of these current limitations while still offering a high degree of automation. I invite researchers in the fields of digital phenotyping [72-76], mobile sensing [77-83], affective computing [84-90], and related subjects to consider integrating crowdsourcing and human-in-the-loop approaches into their methods when pure-AI leads to suboptimal performance.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Howe J. The rise of crowdsourcing. Wired. Jun 01, 2006. URL: https://sistemas-humano-computacionais.wdfiles.com/local--files/capitulo%3Aredes-sociais/Howe_The_Rise_of_Crowdsourcing.pdf [accessed 2024-03-29]

2. Brabham DC, Ribisl KM, Kirchner TR, Bernhardt JM. Crowdsourcing applications for public health. Am J Prev Med. 2014;46(2):179-187. [doi: 10.1016/j.amepre.2013.10.016] [Medline: 24439353]

3. Leung GM, Leung K. Crowdsourcing data to mitigate epidemics. Lancet Digit Health. 2020;2(4):e156-e157. [FREE Full text] [doi: 10.1016/S2589-7500(20)30055-8] [Medline: 32296776]

4. Stockham N, Washington P, Chrisman B, Paskov K, Jung JY, Wall DP. Causal modeling to mitigate selection bias and unmeasured confounding in internet-based epidemiology of COVID-19: model development and validation. JMIR Public Health Surveill. 2022;8(7):e31306. [FREE Full text] [doi: 10.2196/31306] [Medline: 35605128]

5. Kraemer JD, Strasser AA, Lindblom EN, Niaura RS, Mays D. Crowdsourced data collection for public health: a comparison with nationally representative, population tobacco use data. Prev Med. 2017;102:93-99. [FREE Full text] [doi: 10.1016/j.ypmed.2017.07.006] [Medline: 28694063]

6. Jakositz S, Pillsbury L, Greenwood S, Fahnestock M, McGreavy B, Bryce J, et al. Protection through participation: crowdsourced tap water quality monitoring for enhanced public health. Water Res. 2020;169:115209. [doi: 10.1016/j.watres.2019.115209] [Medline: 31669904]

7. Ghenai A, Mejova Y. Catching Zika fever: application of crowdsourcing and machine learning for tracking health misinformation on Twitter. arXiv. Preprint posted online Jul 12, 2017. [doi: 10.48550/arXiv.1707.03778]

8. Dasgupta N, Freifeld C, Brownstein JS, Menone CM, Surratt HL, Poppish L, et al. Crowdsourcing black market prices for prescription opioids. J Med Internet Res. 2013;15(8):e178. [FREE Full text] [doi: 10.2196/jmir.2810] [Medline: 23956042]

9. Wazny K. "Crowdsourcing" ten years in: a review. J Glob Health. 2017;7(2):020602. [FREE Full text] [doi: 10.7189/jogh.07.020602] [Medline: 29302322]

10. Chen PHC, Liu Y, Peng L. How to develop machine learning models for healthcare. Nat Mater. 2019;18(5):410-414. [doi: 10.1038/s41563-019-0345-0] [Medline: 31000806]

11. Dua S, Acharya UR, Dua P, editors. Machine Learning in Healthcare Informatics, Volume 56. Berlin, Heidelberg. Springer; 2014.

12. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]

13. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA Jt Summits Transl Sci Proc. 2020;2020:191-200. [FREE Full text] [Medline: 32477638]

14. Shailaja K, Seetharamulu B, Jabbar MA. Machine learning in healthcare: a review. 2018. Presented at: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA); March 29-31, 2018;910-914; Coimbatore, India. [doi: 10.1109/iceca.2018.8474918]

15. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]

16. Das R, Keep B, Washington P, Riedel-Kruse IH. Scientific discovery games for biomedical research. Annu Rev Biomed Data Sci. 2019;2(1):253-279. [FREE Full text] [doi: 10.1146/annurev-biodatasci-072018-021139] [Medline: 34308269]

17. Dos Reis FJC, Lynn S, Ali HR, Eccles D, Hanby A, Provenzano E, et al. Crowdsourcing the general public for large scale molecular pathology studies in cancer. EBioMedicine. 2015;2(7):681-689. [FREE Full text] [doi: 10.1016/j.ebiom.2015.05.009] [Medline: 26288840]

18. Mavandadi S, Dimitrov S, Feng S, Yu F, Sikora U, Yaglidere O, et al. Distributed medical image analysis and diagnosis through crowd-sourced games: a malaria case study. PLoS One. 2012;7(5):e37245. [FREE Full text] [doi: 10.1371/journal.pone.0037245] [Medline: 22606353]

19. Ozcan A. Educational games for malaria diagnosis. Sci Transl Med. 2014;6(233):233ed9. [FREE Full text] [doi: 10.1126/scitranslmed.3009172] [Medline: 24760185]

20. Feng S, Woo M, Chandramouli K, Ozcan A. A game-based platform for crowd-sourcing biomedical image diagnosis and standardized remote training and education of diagnosticians. In: Optics and Biophotonics in Low-Resource Settings, Volume 9314. 2015. Presented at: SPIE BIOS; February 7-12, 2015; San Francisco, CA. [doi: 10.1117/12.2077884]

21. Luengo-Oroz MA, Arranz A, Frean J. Crowdsourcing malaria parasite quantification: an online game for analyzing images of infected thick blood smears. J Med Internet Res. 2012;14(6):e167. [FREE Full text] [doi: 10.2196/jmir.2338] [Medline: 23196001]

XSL•FO
RenderX

22. Cooper S, Khatib F, Makedon I, Lu H, Barbero J, Baker D, et al. Analysis of social gameplay macros in the Foldit cookbook. 2011. Presented at: FDG '11: Proceedings of the 6th International Conference on Foundations of Digital Games; June 29, 2011 - July 1, 2011;9-14; Bordeaux, France. [doi: 10.1145/2159365.2159367]

23. Curtis V. Motivation to participate in an online citizen science game: a study of Foldit. Sci Commun. 2015;37(6):723-746. [doi: 10.1177/1075547015609322]

24. Eiben CB, Siegel JB, Bale JB, Cooper S, Khatib F, Shen BW, et al. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. Nat Biotechnol. 2012;30(2):190-192. [FREE Full text] [doi: 10.1038/nbt.2109] [Medline: 22267011]

25. Kleffner R, Flatten J, Leaver-Fay A, Baker D, Siegel JB, Khatib F, et al. Foldit standalone: a video game-derived protein structure manipulation interface using Rosetta. Bioinformatics. 2017;33(17):2765-2767. [FREE Full text] [doi: 10.1093/bioinformatics/btx283] [Medline: 28481970]

26. Andreasson JOL, Gotrik MR, Wu MJ, Wayment-Steele HK, Kladwang W, Portela F, et al. Crowdsourced RNA design discovers diverse, reversible, efficient, self-contained molecular switches. Proc Natl Acad Sci U S A. 2022;119(18):e2112979119. [FREE Full text] [doi: 10.1073/pnas.2112979119] [Medline: 35471911]

27. Krüger A, Watkins AM, Wellington-Oguri R, Romano J, Kofman C, DeFoe A, et al. Community science designed ribosomes with beneficial phenotypes. Nat Commun. 2023;14(1):961. [FREE Full text] [doi: 10.1038/s41467-023-35827-3] [Medline: 36810740]

28. Lee J, Kladwang W, Lee M, Cantu D, Azizyan M, Kim H, et al. RNA design rules from a massive open laboratory. Proc Natl Acad Sci U S A. 2014;111(6):2122-2127. [FREE Full text] [doi: 10.1073/pnas.1313039111] [Medline: 24469816]

29. Shi J, Das R, Pande VS. SentRNA: improving computational RNA design by incorporating a prior of human design strategies. arXiv. Preprint posted online Mar 06, 2019. [doi: 10.48550/arXiv.1803.03146]

30. Treuille A, Das R. Scientific rigor through videogames. Trends Biochem Sci. 2014;39(11):507-509. [FREE Full text] [doi: 10.1016/j.tibs.2014.08.005] [Medline: 25300714]

31. Wayment-Steele HK, Kladwang W, Strom AI, Lee J, Treuille A, Becka A, et al. RNA secondary structure packages evaluated and improved by high-throughput experiments. Nat Methods. 2022;19(10):1234-1242. [FREE Full text] [doi: 10.1038/s41592-022-01605-0] [Medline: 36192461]

32. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. Lancet Respir Med. 2016;4(3):213-224. [FREE Full text] [doi: 10.1016/S2213-2600(16)00048-5] [Medline: 26907218]

33. Kalantarian H, Jedoui K, Washington P, Wall DP. A mobile game for automatic emotion-labeling of images. IEEE Trans Games. 2020;12(2):213-218. [FREE Full text] [doi: 10.1109/tg.2018.2877325] [Medline: 32551410]

34. Kalantarian H, Washington P, Schwartz J, Daniels J, Haber N, Wall D. A gamified mobile system for crowdsourcing video for autism research. 2018. Presented at: 2018 IEEE International Conference on Healthcare Informatics (ICHI); June 4-7, 2018;350-352; New York, NY. [doi: 10.1109/ICHI.2018.00052]

35. Penev Y, Dunlap K, Husic A, Hou C, Washington P, Leblanc E, et al. A mobile game platform for improving social communication in children with autism: a feasibility study. Appl Clin Inform. 2021;12(5):1030-1040. [FREE Full text] [doi: 10.1055/s-0041-1736626] [Medline: 34788890]

36. Kalantarian H, Washington P, Schwartz J, Daniels J, Haber N, Wall DP. Guess what? Towards understanding autism from structured video using facial affect. J Healthc Inform Res. 2019;3(1):43-66. [FREE Full text] [doi: 10.1007/s41666-018-0034-9] [Medline: 33313475]

37. Kalantarian H, Jedoui K, Washington P, Tariq Q, Dunlap K, Schwartz J, et al. Labeling images with facial emotion and the potential for pediatric healthcare. Artif Intell Med. 2019;98:77-86. [FREE Full text] [doi: 10.1016/j.artmed.2019.06.004] [Medline: 31521254]

38. Hou C, Kalantarian H, Washington P, Dunlap K, Wall DP. Leveraging video data from a digital smartphone autism therapy to train an emotion detection classifier. medRxiv. Preprint posted online Aug 01, 2021. [doi: 10.1101/2021.07.28.21260646]

39. Kalantarian H, Jedoui K, Dunlap K, Schwartz J, Washington P, Husic A, et al. The performance of emotion classifiers for children with parent-reported autism: quantitative feasibility study. JMIR Ment Health. 2020;7(4):e13174. [FREE Full text] [doi: 10.2196/13174] [Medline: 32234701]

40. Washington P, Kalantarian H, Kent J, Husic A, Kline A, Leblanc E, et al. Improved digital therapy for developmental pediatrics using domain-specific artificial intelligence: machine learning study. JMIR Pediatr Parent. 2022;5(2):e26760. [FREE Full text] [doi: 10.2196/26760] [Medline: 35394438]

41. Washington P, Kalantarian H, Kent J, Husic A, Kline A, Leblanc E, et al. Training an emotion detection classifier using frames from a mobile therapeutic game for children with developmental disorders. arXiv. Preprint posted online Dec 16, 2020 . [doi: 10.48550/arXiv.2012.08678]

42. Varma M, Washington P, Chrisman B, Kline A, Leblanc E, Paskov K, et al. Identification of social engagement indicators associated with autism spectrum disorder using a game-based mobile app: comparative study of gaze fixation and visual scanning methods. J Med Internet Res. 2022;24(2):e31830. [FREE Full text] [doi: 10.2196/31830] [Medline: 35166683]

43. Chi NA, Washington P, Kline A, Husic A, Hou C, He C, et al. Classifying autism from crowdsourced semistructured speech recordings: machine learning model comparison study. JMIR Pediatr Parent. 2022;5(2):e35406. [FREE Full text] [doi: 10.2196/35406] [Medline: 35436234]

44. Lakkapragada A, Kline A, Mutlu OC, Paskov K, Chrisman B, Stockham N, et al. The classification of abnormal hand movement to aid in autism detection: machine learning study. JMIR Biomed Eng. 2022;7(1):e33771. [FREE Full text] [doi: 10.2196/33771]

45. Washington P, Kline A, Mutlu OC, Leblanc E, Hou C, Stockham N, et al. Activity recognition with moving cameras and few training examples: applications for detection of autism-related headbanging. 2021. Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8-13, 2021;1-7; Yokohama, Japan. [doi: 10.1145/3411763.3451701]

46. Paolacci G, Chandler J, Ipeirotis PG. Running experiments on Amazon Mechanical Turk. Judgm Decis Mak. 2010;5(5):411-419. [FREE Full text] [doi: 10.1017/s1930297500002205]

47. Ipeirotis PG, Provost F, Wang J. Quality management on Amazon Mechanical Turk. 2010. Presented at: HCOMP '10: Proceedings of the ACM SIGKDD Workshop on Human Computation; July 25, 2010;64-67; Washington DC. [doi: 10.1145/1837885.1837906]

48. Lee YJ, Arida JA, Donovan HS. The application of crowdsourcing approaches to cancer research: a systematic review. Cancer Med. 2017;6(11):2595-2605. [FREE Full text] [doi: 10.1002/cam4.1165] [Medline: 28960834]

49. Ross J, Zaldivar A, Irani L, Tomlinson B. Who are the turkers? Worker demographics in Amazon Mechanical Turk. ResearchGate. 2009. URL: https://www.researchgate.net/publication/268427703_Who_are_the_Turkers_Worker_Demographics_in_Amazon_Mechanical_Turk [accessed 2024-03-22]

50. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis. 2015;115:211-252. [doi: 10.1007/s11263-015-0816-y]

51. Sheng VS, Zhang J. Machine learning with crowdsourcing: a brief summary of the past research and future directions. Proc AAAI Conf Artif Intell. 2019;33(01):9837-9843. [doi: 10.1609/aaai.v33i01.33019837]

52. Sorokin A, Forsyth D. Utility data annotation with Amazon Mechanical Turk. 2008. Presented at: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops; June 23-28, 2008;1-8; Anchorage, AK. [doi: 10.1109/cvprw.2008.4562953]

53. Xintong G, Hongzhi W, Song Y, Hong G. Brief survey of crowdsourcing for data mining. Expert Syst Appl. 2014;41(17):7987-7994. [doi: 10.1016/j.eswa.2014.06.044]

54. Irani LC, Silberman MS. Turkopticon: interrupting worker invisibility in Amazon Mechanical Turk. 2013. Presented at: CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 27, 2013 - May 2, 2013;611-620; Paris, France. [doi: 10.1145/2470654.2470742]

55. Irani L, Silberman MS. From critical design to critical infrastructure: lessons from turkopticon. Interactions. 2014;21(4):32-35. [doi: 10.1145/2627392]

56. Kummerfeld JK. Quantifying and avoiding unfair qualification labour in crowdsourcing. arXiv. Preprint posted online May 26, 2021. [doi: 10.48550/arXiv.2105.12762]

57. Hansson K, Ludwig T. Crowd dynamics: conflicts, contradictions, and community in crowdsourcing. Comput Support Coop Work. 2019;28:791-794. [FREE Full text] [doi: 10.1007/s10606-018-9343-z]

58. Shen XL, Lee MKO, Cheung CMK. Exploring online social behavior in crowdsourcing communities: a relationship management perspective. Comput Human Behav. 2014;40:144-151. [doi: 10.1016/j.chb.2014.08.006]

59. Wu W, Gong X. Motivation and sustained participation in the online crowdsourcing community: the moderating role of community commitment. Internet Res. 2021;31(1):287-314. [doi: 10.1108/intr-01-2020-0008]

60. Brady CJ, Villanti AC, Pearson JL, Kirchner TR, Gupta OP, Shah C. Rapid grading of fundus photographs for diabetic retinopathy using crowdsourcing. J Med Internet Res. Oct 30, 2014;16(10):e233. [FREE Full text] [doi: 10.2196/jmir.3807] [Medline: 25356929]

61. Mitry D, Peto T, Hayat S, Blows P, Morgan J, Khaw KT, et al. Crowdsourcing as a screening tool to detect clinical features of glaucomatous optic neuropathy from digital photography. PLoS One. 2015;10(2):e0117401. [FREE Full text] [doi: 10.1371/journal.pone.0117401] [Medline: 25692287]

62. Cheng J, Manoharan M, Zhang Y, Lease M. Is there a doctor in the crowd? Diagnosis needed! (For less than $5). iConference 2015 Proceedings. 2015. URL: https://www.ideals.illinois.edu/items/73844 [accessed 2024-03-22]

63. Leblanc E, Washington P, Varma M, Dunlap K, Penev Y, Kline A, et al. Feature replacement methods enable reliable home video analysis for machine learning detection of autism. Sci Rep. 2020;10(1):21245. [FREE Full text] [doi: 10.1038/s41598-020-76874-w] [Medline: 33277527]

64. Tariq Q, Daniels J, Schwartz JN, Washington P, Kalantarian H, Wall DP. Mobile detection of autism through machine learning on home video: a development and prospective validation study. PLoS Med. 2018;15(11):e1002705. [FREE Full text] [doi: 10.1371/journal.pmed.1002705] [Medline: 30481180]

65. Tariq Q, Fleming SL, Schwartz JN, Dunlap K, Corbin C, Washington P, et al. Detecting developmental delay and autism through machine learning models using home videos of Bangladeshi children: development and validation study. J Med Internet Res. 2019;21(4):e13822. [FREE Full text] [doi: 10.2196/13822] [Medline: 31017583]

66. Washington P, Tariq Q, Leblanc E, Chrisman B, Dunlap K, Kline A, et al. Crowdsourced privacy-preserved feature tagging of short home videos for machine learning ASD detection. Sci Rep. 2021;11(1):7620. [FREE Full text] [doi: 10.1038/s41598-021-87059-4] [Medline: 33828118]

67. Washington P, Leblanc E, Dunlap K, Penev Y, Kline A, Paskov K, et al. Precision telemedicine through crowdsourced machine learning: testing variability of crowd workers for video-based autism feature recognition. J Pers Med. 2020;10(3):86. [FREE Full text] [doi: 10.3390/jpm10030086] [Medline: 32823538]

68. Washington P, Leblanc E, Dunlap K, Penev Y, Varma M, Jung JY, et al. Selection of trustworthy crowd workers for telemedical diagnosis of pediatric autism spectrum disorder. Pac Symp Biocomput. 2021;26:14-25. [FREE Full text] [Medline: 33691000]

69. Washington P, Kalantarian H, Tariq Q, Schwartz J, Dunlap K, Chrisman B, et al. Validity of online screening for autism: crowdsourcing study comparing paid and unpaid diagnostic tasks. J Med Internet Res. 2019;21(5):e13668. [FREE Full text] [doi: 10.2196/13668] [Medline: 31124463]

70. Washington P, Tariq Q, Leblanc E, Chrisman B, Dunlap K, Kline A, et al. Crowdsourced privacy-preserved feature tagging of short home videos for machine learning ASD detection. Sci Rep. 2021;11(1):7620. [FREE Full text] [doi: 10.1038/s41598-021-87059-4] [Medline: 33828118]

71. Washington P, Chrisman B, Leblanc E, Dunlap K, Kline A, Mutlu C, et al. Crowd annotations can approximate clinical autism impressions from short home videos with privacy protections. Intell Based Med. 2022;6:100056. [FREE Full text] [doi: 10.1016/j.ibmed.2022.100056] [Medline: 35634270]

72. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA. 2017;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]

73. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med. 2019;2:88. [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]

74. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. Neuropsychopharmacology. 2021;46(1):45-54. [FREE Full text] [doi: 10.1038/s41386-020-0771-3] [Medline: 32679583]

75. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology. 2016;41(7):1691-1696. [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]

76. Washington P, Mutlu CO, Kline A, Paskov K, Stockham NT, Chrisman B, et al. Challenges and opportunities for machine learning classification of behavior and mental state from images. arXiv. Preprint posted online Jan 26, 2022. [doi: 10.48550/arXiv.2201.11197]

77. Baumeister H, Montag C, editors. Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics. Cham, Switzerland. Springer International Publishing; 2019.

78. Laport-López F, Serrano E, Bajo J, Campbell AT. A review of mobile sensing systems, applications, and opportunities. Knowl Inf Syst. 2020;62(1):145-174. [doi: 10.1007/s10115-019-01346-1]

79. Macias E, Suarez A, Lloret J. Mobile sensing systems. Sensors (Basel). 2013;13(12):17292-17321. [FREE Full text] [doi: 10.3390/s131217292] [Medline: 24351637]

80. Nazir S, Ali Y, Ullah N, García-Magariño I. Internet of things for healthcare using effects of mobile computing: a systematic literature review. Wirel Commun Mob Comput. 2019;2019:1-20. [FREE Full text] [doi: 10.1155/2019/5931315]

81. Silva BMC, Rodrigues JJPC, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform. 2015;56:265-272. [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]

82. Sim I. Mobile devices and health. N Engl J Med. 2019;381(10):956-968. [doi: 10.1056/nejmra1806949]

83. Yürür O, Liu CH, Sheng Z, Leung VCM, Moreno W, Leung KK. Context-awareness for mobile sensing: a survey and future directions. IEEE Commun Surv Tutor. 2014;18(1):68-93. [doi: 10.1109/comst.2014.2381246]

84. Picard RW. Affective Computing. Cambridge, MA. MIT Press; 2000.

85. Picard RW. Affective computing: challenges. Int J Hum-Comput Stud. 2003;59(1-2):55-64. [doi: 10.1016/s1071-5819(03)00052-1]

86. Poria S, Cambria E, Bajpai R, Hussain A. A review of affective computing: from unimodal analysis to multimodal fusion. Inf Fusion. 2017;37:98-125. [doi: 10.1016/j.inffus.2017.02.003]

87. Scherer KR, Bänziger T, Roesch E, editors. A Blueprint for Affective Computing: a Sourcebook and Manual. Oxford, United Kingdom. Oxford University Press; 2010.

88. Tao J, Tan T. Affective computing: a review. In: Affective Computing and Intelligent Interaction. Berlin, Heidelberg. Springer; 2005. Presented at: First International Conference, ACII 2005; October 22-24, 2005;981-995; Beijing, China. [doi: 10.1007/11573548_125]

89. Wang Y, Song W, Tao W, Liotta A, Yang D, Li X, et al. A systematic review on affective computing: emotion models, databases, and recent advances. Inf Fusion. 2022;83-84:19-52. [doi: 10.1016/j.inffus.2022.03.009]

90. Zhao S, Wang S, Soleymani M, Joshi D, Ji Q. Affective computing for large-scale heterogeneous multimedia data: a survey. ACM Trans Multimed Comput Commun Appl. 2019;15(3s):1-32. [doi: 10.1145/3363560]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
FDA: Food and Drug Administration
MTurk: Amazon Mechanical Turk

XSL•FO
RenderX