# Crowdsourcing for Rapid Fundus Photograph Interpretation

Third-party interpretations of fundus photographs could lighten the burden on retina specialists.

## **BY CHRISTOPHER J. BRADY, MD**

Creening for diabetic retinopathy (DR) with retinal fundus examination is effective and cost-effective in preventing vision loss, but, despite this, screening rates remain low. Screening for retinopathy is particularly human resource–intensive, and telemedicine has been proposed as a way to make retinal screening more accessible to all individuals with diabetes. Current methods of telemedicine screening require a skilled interpretation of retinal fundus images, adding to the human resource burden, so new ways of processing this image data are needed. Crowdsourcing is a novel way of data processing that leverages human intelligence and pattern recognition to categorize images.

#### **TELEHEALTH PROGRAMS**

Telehealth programs using nonmydriatic fundus photography and remote interpretation are expanding, especially in rural and remote settings, and have become a method of increasing adherence to DR screening recommendations.<sup>1-3</sup> In addition to improving screening uptake, telehealth may provide ways to reduce provider, payer, and societal costs.<sup>4-6</sup> The cost of fundus photograph interpretation for DR screening can be high given labor-intensive interpretation protocols and the need to interpret multiple images per patient. Computerized, semiautomated image analysis techniques have been developed that will likely be able to reduce physician workload and screening costs in the near future<sup>7-9</sup>; however, at this time, these methods are neither approved by the US Food and Drug Administration nor widely used clinically. As telehealth expansion continues, novel, low-cost methods will be needed to interpret the large volume of fundus images expected due to the rising incidence of diabetes.

"As telehealth expansion continues, novel, low-cost methods will be needed to interpret the large volume of fundus images expected due to the rising incidence of diabetes."

#### WHAT IS CROWDSOURCING?

Darren C. Brabham, PhD, defines crowdsourcing as "an online, distributed problem-solving and production model that leverages the collective intelligence of online communities to serve specific organizational goals."<sup>10</sup> A subset of crowdsourcing, which Dr. Brabham terms "distributed human intelligence tasking," can involve subdividing larger tasks into small portions and then recruiting a group of individuals to complete each of these small portions, thereby completing the entire task.

The use of crowdsourcing in research initially blossomed in the behavioral sciences, and several biomedical research groups have adopted these methods for public health research.<sup>11</sup> Crowdsourcing can also be used to interpret medical imaging. For example, malaria researchers have created a web-based game to recruit untrained internet users to identify malaria parasites on images of thick blood smears.<sup>12</sup> The investigators were able to achieve accuracy rates similar to those of expert microscopists by combining the analyses of several users. Crowdsourcing has also been used to categorize fundus photographs with a variety of diagnoses as normal or



All HITs   HITs Available To You   HITs Assigned To You				
Find HITS <b>v</b> containing		that pay at least \$ 0.00	<ul> <li>for which you are qualified</li> <li>require Master Qualification</li> </ul>	60
Timer: 00:00:00 of 15 minutes	Want to work on this HIT? Wart to work on this HIT? Wa	ant to see other HITs?	Tota Total HITs S	l Earned: Unavailable ubmitted: 0
Fundus Photo Grading				
Requester: Christopher Brady Qualifications Required: None		Reward: \$0.	10 per HIT HITs Available: 100	Duration: 15 minutes
This is a photo of the inside of the eye. We are looking to label eyes as healthy or unhealthy with respect to diabetes as part of a reseach study. This task is difficult, but all good faith efforts will be accepted. Please answer both questions even if you are not sure. To close pop-out, move cursor left, or <u>tap here</u> on iPad.				
Click for magnified image	Hover for examples: Normal Normal Ex. 2 Abnormal: Mild Abnormal: Moderate Abnormal: Severe Abnormal: Severe Ex. 2 Abnormal: Severe Ex. 3	<ol> <li>Label this eye: NOTE: The color/tint of the</li> <li>Normal (Healthy)</li> <li>Abnormal (mild, moderate</li> <li>Is this photo good enough</li> <li>Yes</li> <li>No</li> </ol>	e photo is not a marker of diseas e, severe disease) a quality to grade?	e.

Figure. Turkers use this interface to determine if a particular fundus photograph shows evidence of disease.

abnormal.<sup>13</sup> In a trial conducted in the United Kingdom using untrained graders, the sensitivity was 96% or more for severely abnormal findings and between 61% and 79% for mildly abnormal ones.<sup>13</sup>

### **CROWDSOURCING WEBSITES**

There are several standalone websites dedicated to crowdsourcing. One such website, Zooniverse.org, allows users to participate in "virtual citizen-science" on a volunteer basis.<sup>14</sup> Amazon.com has developed a well-known fee-based site for crowdsourcing: Amazon Mechanical Turk (AMT). AMT is an online distributed human intelligence market that allows access to thousands of registered users who can quickly accomplish small, discrete tasks for small amounts of money. Typical AMT tasks include categorizing photos, providing translations, or writing very short articles for websites. AMT has its own vocabulary used by workers (called *turkers*) and task administrators (called *requestors*). A *human intelligence task* (HIT) is a small job that may be performed in a matter of seconds or minutes and, once approved by the requestor, may pay \$0.01 to \$0.25 or more per task, depending on the complexity of the HIT. A group of similar HITs is termed a *batch*. Depending on the complexity of the task and the payment offered by the requestor, a batch is often completed within minutes or hours of posting.

Given the small remuneration for each individual HIT, it is interesting to consider why a person might choose to perform these tasks conscientiously. AMT is a reputation-based economy in which turkers may only access the most desirable (ie, most interesting and most highly paid) HITs once they have a sufficient track record of previously accepted work.<sup>15</sup> Indeed, a turker's reputation will suffer following rejection of even a small number of HITs, and for this reason high-quality turkers may avoid a new requestor's HITs until the requestor has demonstrated his or her own fairness in approving and rejecting work. AMT is a complex ecosystem in which both high quality work on the part of the turkers and fairness on the part of requestors are rewarded.

Turkers perform their work anonymously, but demographic studies have been conducted. In a survey of 1000 turkers, Ipeirotis found that 46.8% of turkers were located in the United States, 34% were in India, and the remaining 19.2% were from 64 other countries.<sup>16</sup> In the United States, the majority of workers were women, most of whom reported AMT as a source of supplemental income, whereas most workers in India were men and reported AMT as their primary source of income. Across nations, turkers were younger and better educated than the general population of Internet users.<sup>16</sup>

In a proof-of-concept study, we demonstrated that turkers can rapidly and accurately identify fundus photographs of patients with DR.<sup>17</sup> For this study, we created a custom interface (Figure) that allowed turkers to review several training images describing a normal fundus and the various pathologic features of DR. On the same page, turkers were presented with one of 19 teaching images and asked to grade the fundus as normal or abnormal. Turkers spent an average of 25 seconds, including the time spent viewing the training images, and were correct 81.3% of the time. Using feedback from our initial batches, the web interface was improved, and turker accuracy likewise improved. A separate experiment confirmed that requesting 10 independent interpretations and using the average for the final "grade" was an appropriate strategy for correctly identifying DR in patients. When asked to grade images from a much larger public dataset with more subtle disease,<sup>18,19</sup> turkers successfully identified images as abnormal when moderate to severe disease was present but were less successful at identifying very mild disease (ie,  $\leq$ 5 microaneurysms) as abnormal.<sup>20</sup>

#### **BENEFITS OF CROWDSOURCING**

There are several potential benefits in the use of crowdsourcing for the interpretation of visual data in ophthalmology. First, an inexpensive, rapid, and accurate system to reduce the number of images requiring skilled human grading in large public health screenings is needed. An approach that accurately identifies normal (or very mild) disease on fundus photographs would be of great value and could reduce the skilled grader burden by 26% to 38% or more, according to some investigators using artificial intelligence (AI) programs.<sup>9</sup> A first pass to remove normal images is currently being done with an AI solution in Scotland's national screening program.<sup>21</sup> If appropriately validated, crowdsourcing could provide a similar service at lower cost and with less infrastructure in resource-poor settings. More abstractly, crowdsourcing might provide a complement to AI development efforts and computer vision technologies. There is a need for "ground truth" data in the development of computer vision algorithms, and crowdsourcing could be used to categorize fundus photographs and possibly other imaging outputs, which could then be used to test and improve AI algorithms.

An additional, unanticipated benefit of crowdsourcing biomedical image analysis is that it might raise awareness of the disease in question. Because our HITs allowed turkers to leave feedback, we were able to capture comments such as, "I have learn [sic] about diabetes little bit," "I really liked seeing the pics of the eye, very interesting," and "This HIT was very good and a nice break from all of the bubbling surveys. Thank you!" suggesting an interest in the subject matter perhaps beyond nonmedical or nonscientific HITs.

Christopher J. Brady, MD, is an assistant professor of ophthalmology at the Wilmer Eye Institute at the Johns Hopkins School of Medicine. Dr. Brady may be reached at cbrady5@jhmi.edu.



1. Sharp PF, Olson J, Strachan F, et al. The value of digital imaging in diabetic retinopathy. *Health Technol Assess*. 2003;7(30):1-119.

 Scanlon PH. The English national screening programme for sight-threatening diabetic retinopathy. J Med Screen. 2008;15(1):1-4.

3. Ng M, Nathoo N, Rudnisky CJ, Tennant MT. Improving access to eye care: teleophthalmology in Alberta, Canada. J Diabetes Sci Technol. 2009;3(2):289–296.

 Brady CJ, Villanti AC, Gupta OP, Graham MG, Sergott RC. Tele-ophthalmology screening for proliferative diabetic retinopathy in urban primary care offices: an economic analysis. *Ophthalmic Surg Lasers Imaging Retina*. 2014;45(6):556-561.

Au A, Gupta O. The economics of telemedicine for vitreoretinal diseases. *Curr Opin Ophthalmol*. 2011;22(3):194-198.
 Rein DB, Wittenborn JS, Zhang X, et al. The cost-effectiveness of three screening alternatives for people with diabetes with no or early diabetic retinopathy. *Health Serv Res*. 2011;46(5):1534-1561.

 Abramoff MD, Folk JĆ, Han DP, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmol. 2013;131(3):351–357.

 Trucco E, Ruggeri A, Karnowski T, et al. Validating retinal fundus image analysis algorithms: issues and a proposal. *Invest Ophthalmol Vis Sci.* 2013;54(5):3546–3559.

9. Goatman K, Charnley A, Webster L, Nussey S. Assessment of automated disease detection in diabetic retinopathy screening using two-field photography. *PloS One*. 2011;6(12):e27524.

10. Brabham DC. Crowdsourcing. Cambridge, MA: MIT Press; 2013.

11. Brabham DC, Ribisl KM, Kirchner TR, Bernhardt JM. Crowdsourcing applications for public health. Am J Prev Med. 2014;46(2):179-187.

12. 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.

 Mitry D, Peto T, Hayat S, et al. Crowdsourcing as a novel technique for retinal fundus photography classification: analysis of images in the EPIC Norfolk cohort on behalf of the UK Biobank Eye and Vision Consortium. *PIoS One*. 2013;8(8):e71154.

14. Reed J, Raddick MJ, Lardner A, Carney K. An exploratory factor analysis of motivations for participating in Zooniverse, a collection of virtual citizen science projects. Paper presented at: Hawaii International Conference on System Sciences; January 7-10, 2013; Wailea, HI.

 Martin D, Hanrahan BV, O'Neill J, Gupta N. Being a turker. Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing; February 15-19, 2014; Baltimore, MD.

Ipeirotis PG. Demographics of mechanical turk. CeDER Working Paper-10-01; New York University; 2010.
 Brady CJ, Villanti AC, Pearson JL, et al. Rapid grading of fundus photographs for diabetic retinopathy using crowdsourcing. J Med Internet Res. 2014;16(10):e233.

18. Sánchez CI, Niemeijer M, Dumitrescu AV, et al. Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data. *Invest Ophthalmol Vis Sci.* 2011;52(7): 4866-4871.

19. MESSIDOR: Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology. Accessed March 16, 2015. http://messidor.crihan.fr/index-en.php

 Brady CJ, Villanti AC, Pearson JL, et al. Rapid grading of fundus photos for diabetic retinopathy using crowdsourcing: Refinement and External Validation. Paper presented at: Retina Society 47th Scientific Meeting; September 11–13, 2014; Philadelphia, PA.

21. Medalytix Retinal Screening. Accessed December 3, 2014. www.medalytix.com.