Research Article

Use of Mechanical Turk as a MapReduce Framework for Macular OCT Segmentation

Aaron Y. Lee,[1,2] Cecilia S. Lee,[1,2] Pearse A. Keane,[2,3,4] and Adnan Tufail[2,3,4]

[1]*Department of Ophthalmology, University of Washington, Seattle, WA 98104, USA*
[2]*Medical Retina Service, Moorfields Eye Hospital NHS Foundation Trust, London EC1V 2PD, UK*
[3]*Institute of Ophthalmology, University College London, London WC1E 6BT, UK*
[4]*National Institute for Health Research Biomedical Research Centre for Ophthalmology,*
 Moorfields Eye Hospital NHS Foundation Trust, London SE1 4TT, UK

Correspondence should be addressed to Adnan Tufail; adnan.tufail@moorfields.nhs.uk

Purpose. To evaluate the feasibility of using Mechanical Turk as a massively parallel platform to perform manual segmentations of macular spectral domain optical coherence tomography (SD-OCT) images using a MapReduce framework. *Methods*. A macular SD-OCT volume of 61 slice images was map-distributed to Amazon Mechanical Turk. Each Human Intelligence Task was set to $0.01 and required the user to draw five lines to outline the sublayers of the retinal OCT image after being shown example images. Each image was submitted twice for segmentation, and interrater reliability was calculated. The interface was created using custom HTML5 and JavaScript code, and data analysis was performed using R. An automated pipeline was developed to handle the map and reduce steps of the framework. *Results*. More than 93,500 data points were collected using this framework for the 61 images submitted. Pearson's correlation of interrater reliability was 0.995 ($p < 0.0001$) and coefficient of determination was 0.991. The cost of segmenting the macular volume was $1.21. A total of 22 individual Mechanical Turk users provided segmentations, each completing an average of 5.5 HITs. Each HIT was completed in an average of 4.43 minutes. *Conclusions*. Amazon Mechanical Turk provides a cost-effective, scalable, high-availability infrastructure for manual segmentation of OCT images.

1. Introduction

Crowdsourcing is a relatively novel technique involving the distribution of work to a large group of people, typically through online frameworks [1]. It allows the subdivision of tedious tasks into discrete tasks that can be completed individually. Amazon Mechanical Turk is the largest and most popular of the online crowdsourcing systems [2]. In this system, simple Human Intelligence Tasks (HITs) are submitted to online untrained users for a small compensation. Recently in computer science, the MapReduce programming model has caused a paradigm shift in the way that large data sets are distributed in parallel within a computing cluster [3]. Notably Google used the MapReduce framework to regenerate their index of the Internet, and the MapReduce framework has become popularized as a generic framework to solve big data

problems in multicore cluster systems. In this study, our goal was to utilize human intelligence as a MapReduce framework for the segmentation of a macular optical coherence tomography (OCT) volume [4–6].

OCT is an important noninvasive diagnostic tool in the field of ophthalmology [6] and in the management of age-related macular degeneration (AMD), the commonest cause of blindness in the developed world [7, 8]. OCT measurements such as retinal thickness, subretinal fluid, and pigment epithelial detachment are important parameters in the diagnosis and monitoring of various retinal diseases [9, 10] and are thus integral in both large-scale clinical trials and routine clinical practice [11]. However, automated measurements provided by the OCT software result in frequent errors in quantifying critical parameters such as macular thickness and volume [12, 13]. Errors of retinal boundary detection and
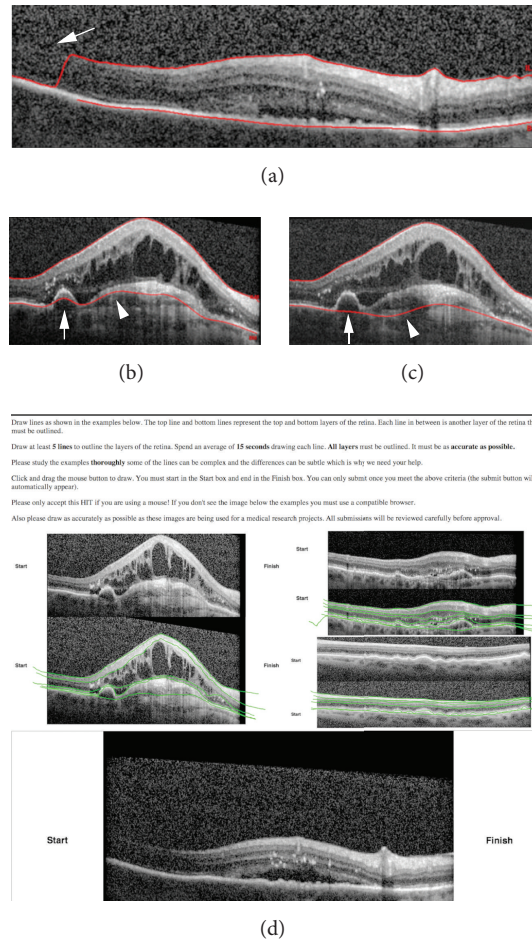
(a)

(b)

(c)

(d)

FIGURE 1: Examples of incorrect segmentations by automated software and user-interface for Mechanical Turk. Panel (a) is an example of a macula SD-OCT image with missing information (arrow) causing a sudden jump in the identification of the Internal-Limiting Membrane (ILM) by automated software included with Heidelberg Spectralis. Panels (b and c) show two similar macular OCT images with different automated segmentations caused by pigment epithelial detachment (arrow) and subretinal fibrosis (arrowhead). Panel (d) is a screenshot of web-based user-interface submitted to Amazon Mechanical Turk for manual segmentations.

thickness measurements have been reported as high as 92% in segmentation performed by the Stratus OCT system (Carl Zeiss Meditec, Germany) [14]. Although spectral domain OCT (SD-OCT) is expected to produce more accurate measurements with higher resolution and less artifacts, the segmentation errors continue to be a significant problem in measuring macular thickness, particularly in eyes with pathology [15–17].

There has been increasing interest in ways to overcome automated segmentation errors. Publicly available image analysis software, OCTOR (Doheny Image Reading Center, Los Angeles), quantifies OCT-derived parameters after a trained OCT grader delineates the retinal boundaries of interest manually. The software calculates the distance in pixels between two manually drawn layers. Then using the dimensions of the B-scan image, the data is converted into a thickness measurement [18]. Even though OCTOR is less subject to segmentation errors, it is time-consuming and impractical for use in large-scale clinical trials.

Automated segmentations have been attempted using dual-scale gradient or intensity information. Then the edges of the boundary were optimized using a shortest path search method [19]. Statistical models have been utilized for a more reliable automatic segmentation system [20]. Retinal layers have been segmented using seven features extracted from the OCT data with a random forest classifier [21]. Despite these achievements in the field of automated segmentations, macular OCT images with complex subretinal pathology, intraretinal/subretinal fluid, or low signal to noise ratio continue to pose a challenge for computer vision (Figures 1(a)–1(c)).

Amazon Mechanical Turk and other modalities of crowd-sourcing have been previously used in medical applications and demonstrated high level of accuracy in diagnostic accuracy [22–24]. In ophthalmology, retinal fundus photographs have been recently analyzed and showed an accuracy level at least comparable to automated programs and some trained graders [25]. To our knowledge, it has not been used to

attain segmentations in macular OCT images with complex pathology. We sought to achieve highly reliable segmentation by designing a system for distributed OCT segmentation over a scalable, human based infrastructure and to show proof of concept results.

2. Materials and Methods

Patient identifiers were stripped out completely and pseudoanonymized, and on this basis and for retrospective use of anonymized data in the UK formal ethics committee review is not required. However, consent was still obtained from all patients in this study to use their OCT images for research. This study was conducted in accordance with the Declaration of Helsinki and the United Kingdom's Data Protection Act.

A total of 61 individual macular SD-OCT images were taken using a commercially available SD-OCT device (Spectralis, Heidelberg Engineering, Heidelberg, Germany) as part of routine medical care for AMD. The images were extracted using commercially provided software (Heyex DICOM Interface, Heidelberg Engineering, Heidelberg, Germany), and no image manipulation was performed.

A custom web-based user-interface was created with Hyper Text Markup Language 5 (HTML5) and JavaScript to allow Mechanical Turk users to directly draw on the images through their web browser (Figure 1(d)). This interface gave each user an example image of segmentation by an expert retina trained physician. The JavaScript interface allowed capture of the mouse input to draw segmentation lines on the provided OCT image and captured timing data as the user drew segmentation lines. Each user was instructed to draw 5 lines to segment the provided image and they were required to spend at least an average of 15 seconds per line before they were allowed to submit their work. No image enlargement or zoom was allowed and users were given 3 example segmentations provided by a trained OCT grader using the same system. Each image was created as a separate HIT and the reward was set to \$0.01 (USD). Mechanical Turk users were required to have a prior approval rate of 80% before being allowed to participate in these HITs. In addition to the lines drawn, data was collected on the time spent drawing each line segment and time to completion of segmentation, and each image was submitted twice for segmentation.

After all segmentations were performed, the data was collected and image processing was performed to enhance the accuracy of the manual segmentations. This automated analysis pipeline used adjustments based on finding the consistently highest contrast value within 5 pixels of where the segmentation line was drawn. If there were no improved changes detected, then data from the original manual segmentation was used. Automated quality control heuristics were implemented to ensure that no two segmentations from the same user of the same image crossed paths. The reduction step of combining consensus segmentations of the same image was classified using a linear correlation heuristic, and these segmentation data were used to calculate interrater reliability. The final reduction step was utilized to recreate a three-dimensional segmented model of the retina. Custom Ruby and R code was created to automate the creation,

submission, collection, image processing, and data analysis. All custom software is available upon request.

3. Results

The automated analysis pipeline, using a MapReduce framework, was able to create, submit, collect, collate, process, and analyze a total of over 92,500 data points from the 61 macular OCT images that were manually segmented twice over Amazon Mechanical Turk. Time of submission of the 122 HITs to completion of all tasks was 3 days with greater than 75% of HITs finished within the first 24 hours. A total of 22 individual Mechanical Turk users provided segmentations each completing an average of 5.5 HITs.

Each HIT was completed on average of 4.43 minutes (range: 1.83–24.45 minutes) with each segmentation line completed on average of 20.40 seconds (Figure 2(a)). In a subset of users who had segmented four or more HITs, we noted that there was a trend in decreasing time to completion of the task (Figure 2(b)). A total of 646 segmentations were collected, and an average of 5.30 segmentations per macular OCT was provided (range: 5 to 7). The total cost of segmentations of all images was \$1.22 (USD).

Representative segmentations with the associated image processing are shown in Figure 3. All slices from both manual segmentation and the combined final segmentations are shown in the Supplementary Materials (see Supplementary Material available online at http://dx.doi.org/10.1155/2016/6571547). Pearson's correlation of interrater reliability was 0.995 ($p < 0.0001$) and coefficient of determination was 0.991. A Bland-Altman plot was calculated to estimate interrater agreement based on the consensus segmentation lines (Figure 4).

4. Discussion

OCT is a critical tool in clinical practice for ophthalmology, and objective, quantitative OCT parameters have the potential of guiding clinical practice and establishing new endpoints for clinical trials. Automated segmentation approaches have traditionally suffered in the setting of complex retinal pathology such as pigment epithelial detachments, subretinal fibrosis, or intraretinal and subretinal fluid. Indeed the automated segmentation that is provided with the commercial device used in this study failed in many situations (Figures 1(a)–1(c)). With the advent of Mechanical Turk and programming APIs, automating simple human vision tasks through a MapReduce framework has become not only feasible but also cost-effective. The advantages of utilizing manual segmentations using human vision include the ability to complete areas of macular OCT where there is poor signal to noise ratio (Figure 3(a)) or complex pathology (Figure 3(c)).

Next steps of this study would be to compare the accuracy of the Mechanical Turk based segmentation to the ones performed by trained experts. Using the segmentation lines performed by trained experts as the gold standard, we will plan to evaluate the correlation between the accuracy and the time spent by the users, previous experience of the users,
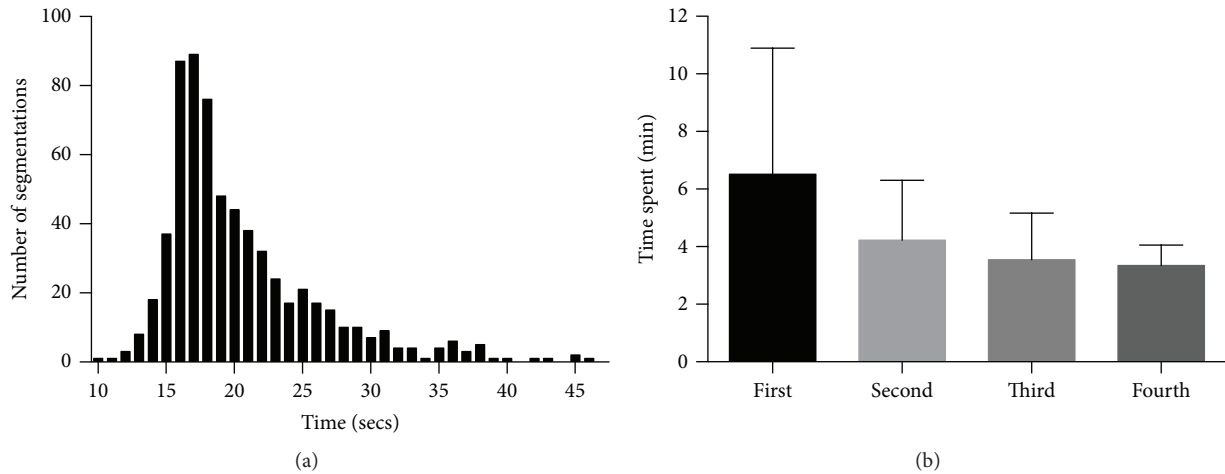
(a)

(b)

FIGURE 2: Temporal data of segmentations. Panel (a) shows a histogram of the time spent for each segmentation line. Average was 20.40 seconds with a range of 10.01 to 46.22 seconds. Panel (b) shows the decreasing trend in total time spent in minutes segmenting one SD-OCT image in a subset of users who segmented 4 or more times (7 out of 22 users). Error bars are standard deviation.

(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)
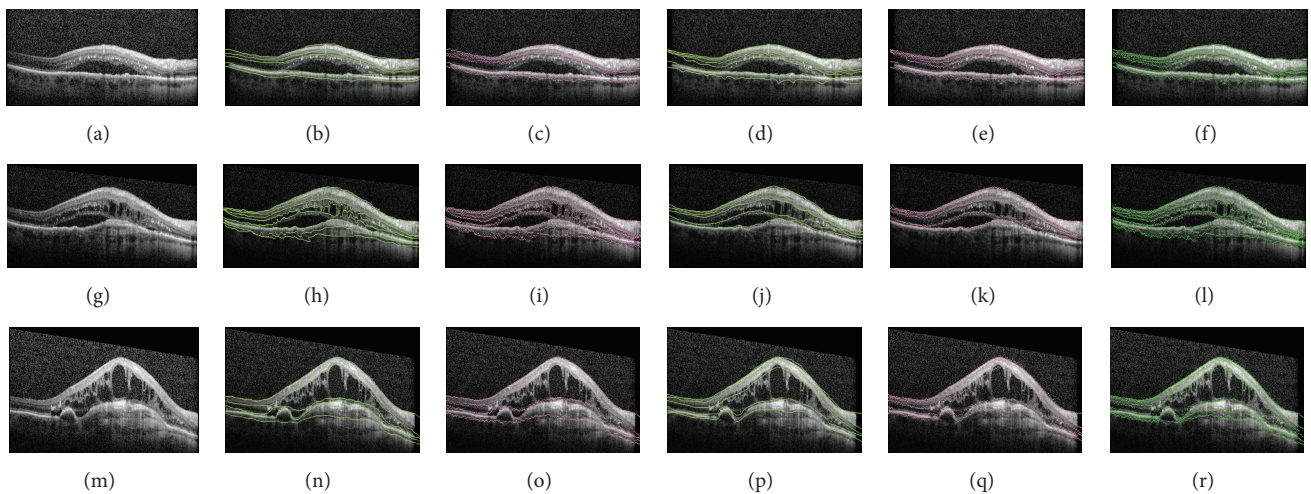
(m) (n) (o) (p) (q) (r)

FIGURE 3: Representative segmentations by Mechanical Turk based manual segmentations with contrast based enhancements. Panels (a, g, m) show three raw SD-OCT macular images. Panels (b, d, h, j, n, p) demonstrate each image segmented by two different people on Mechanical Turk. Manual segmentations are shown as green lines. Panels (c, e, i, k, o, q) show local contrast based enhancement of manual segmentations as magenta lines. Panels (f, l, r) are the final consensus segmentations (green lines) after combining segmentations.

and any learning effect by repetitive performance of the same users.

Limitations of this approach stem from the lack of professionally trained OCT readers and the lack of knowledge or training of the Mechanical Turk users. Future analysis pipelines may include an expert validation step, which reviews the consensus segmentations and decides whether to accept or to reject the submitted segmentations, which then could be resubmitted for another round of segmentation. In addition, future, large-scale studies will be necessary to assess the external validity of this system by submitting macular OCT images for segmentation by expert graders versus Mechanical Turk.

Large data sets are becoming increasingly common with today's clinical studies and multicenter trials. Rapid, reliable, cost-effective methods of interpreting large data will be crucial in the future. Crowdsourcing in OCT segmentation has the potential of minimizing the errors seen in automated segmentation system with less time and cost than manual segmentations performed in reading centers. Additional ways to improve this tool such as more effective training of the users, preselection of qualified users, or creating an automated system based on users' initial segmentation would be important areas to be investigated. Implementation of our current method in the RISE/RIDE study, for example, where 759 patients received monthly OCT imaging, would
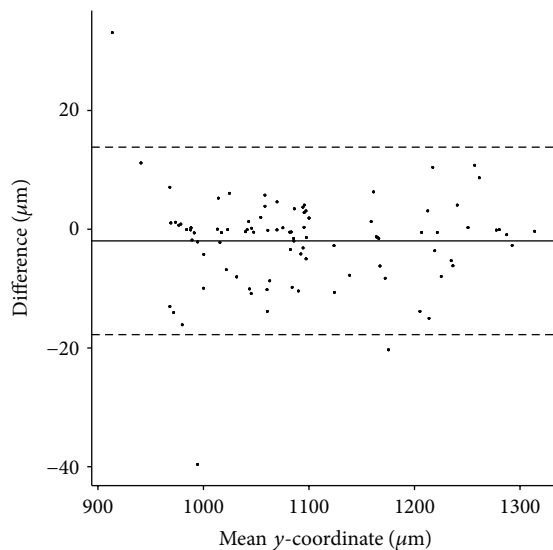
FIGURE 4: Bland-Altman plot showing agreement between segmentations. Consensus segmentations of the same image between two independent Mechanical Turk users were used to determine interrater reliability. The average y-coordinate value in microns for each consensus line was used and the Bland-Altman plot was created.

cost approximately \$273.24 per study month for a standard 18-slice macular OCT.

This study has applied a novel proof of concept of applying manual segmentation of OCT images in a distributed way to nonexpert graders. The retinas with various pathologies provide challenge to currently available automated segmentation systems. Mechanical Turk provides a cost-effective, scalable, high-availability infrastructure for manual segmentation of OCT images of the type which are difficult for automated algorithms to handle. The resulting images can be recombined for high-resolution 3D analysis. This approach may be applied to the analysis of high volumes of OCT images in clinical studies or training of future automated segmentation algorithms.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] D. C. Brabham, K. M. Ribisl, T. R. Kirchner, and J. M. Bernhardt, "Crowdsourcing applications for public health," *American Journal of Preventive Medicine*, vol. 46, no. 2, pp. 179–187, 2014.

[2] M. J. C. Crump, J. V. McDonnell, and T. M. Gureckis, "Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research," *PLoS ONE*, vol. 8, no. 3, Article ID e57410, 2013.

[3] R. Bellazzi, "Big data and biomedical informatics: a challenging opportunity," *Yearbook of Medical Informatics*, vol. 9, pp. 8–13, 2014.

[4] R. C. Taylor, "An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics," *BMC Bioinformatics*, vol. 11, supplement 12, article S1, 2010.

[5] D. Huang, E. A. Swanson, C. P. Lin et al., "Optical coherence tomography," *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[6] I. Voo, E. C. Mavrofrides, and C. A. Puliafito, "Clinical applications of optical coherence tomography for the diagnosis and management of macular diseases," *Ophthalmology Clinics of North America*, vol. 17, no. 1, pp. 21–31, 2004.

[7] R. Klein, T. Peto, A. Bird, and M. R. Vannewkirk, "The epidemiology of age-related macular degeneration," *American Journal of Ophthalmology*, vol. 137, no. 3, pp. 486–495, 2004.

[8] N. Congdon, B. O'Colmain, C. C. W. Klaver et al., "Causes and prevalence of visual impairment among adults in the United States," *Archives of Ophthalmology*, vol. 122, no. 4, pp. 477–485, 2004.

[9] D. M. Brown, P. K. Kaiser, M. Michels et al., "Ranibizumab versus verteporfin for neovascular age-related macular degeneration," *The New England Journal of Medicine*, vol. 355, no. 14, pp. 1432–1444, 2006.

[10] P. J. Rosenfeld, D. M. Brown, J. S. Heier et al., "Ranibizumab for neovascular age-related macular degeneration," *The New England Journal of Medicine*, vol. 355, no. 14, pp. 1419–1431, 2006.

[11] D. M. Brown, Q. D. Nguyen, D. M. Marcus et al., "Long-term outcomes of ranibizumab therapy for diabetic macular edema: the 36-month results from two phase III trials: RISE and RIDE," *Ophthalmology*, vol. 120, no. 10, pp. 2013–2022, 2013.

[12] P. J. Patel, F. K. Chen, L. da Cruz, and A. Tufail, "Segmentation error in Stratus optical coherence tomography for neovascular age-related macular degeneration," *Investigative Ophthalmology and Visual Science*, vol. 50, no. 1, pp. 399–404, 2009.

[13] P. A. Keane, P. J. Patel, S. Liakopoulos, F. M. Heussen, S. R. Sadda, and A. Tufail, "Evaluation of age-related macular degeneration with optical coherence tomography," *Survey of Ophthalmology*, vol. 57, no. 5, pp. 389–414, 2012.

[14] P. A. Keane, S. Liakopoulos, R. V. Jivrajka et al., "Evaluation of optical coherence tomography retinal thickness parameters for use in clinical trials for neovascular age-related macular degeneration," *Investigative Ophthalmology and Visual Science*, vol. 50, no. 7, pp. 3378–3385, 2009.

[15] Y. Song, B. R. Lee, Y. W. Shin, and Y. J. Lee, "Overcoming segmentation errors in measurements of macular thickness made by spectral-domain optical coherence tomography," *Retina (Philadelphia, Pa)*, vol. 32, no. 3, pp. 569–580, 2012.

[16] M. Kim, S. Lee, J. Han, S.-Y. Yu, and H. Kwak, "Segmentation error and macular thickness measurements obtained with spectral-domain optical coherence tomography devices in neovascular age-related macular degeneration," *Indian Journal of Ophthalmology*, vol. 61, no. 5, pp. 213–217, 2013.

[17] P. A. Keane, P. S. Mand, S. Liakopoulos, A. C. Walsh, and S. R. Sadda, "Accuracy of retinal thickness measurements obtained with Cirrus optical coherence tomography," *British Journal of Ophthalmology*, vol. 93, no. 11, pp. 1461–1467, 2009.

[18] S. R. Sadda, S. Joeres, Z. Wu et al., "Error correction and quantitative subanalysis of optical coherence tomography data using computer-assisted grading," *Investigative Ophthalmology & Visual Science*, vol. 48, no. 2, pp. 839–848, 2007.

[19] Q. Yang, C. A. Reisman, Z. Wang et al., "Automated layer segmentation of macular OCT images using dual-scale gradient information," *Optics Express*, vol. 18, no. 20, pp. 21293–21307, 2010.

[20] V. Kajić, B. Považay, B. Hermann et al., "Robust segmentation of intraretinal layers in the normal human fovea using a novel statistical model based on texture and shape analysis," *Optics Express*, vol. 18, no. 14, pp. 14730–14744, 2010.

[21] A. Lang, A. Carass, E. Sotirchos, P. Calabresi, and J. L. Prince, "Segmentation of retinal OCT images using a random forest classifier," in *Proceedings of the Medical Imaging: Image Processing*, vol. 8669 of *Proceedings of SPIE*, Lake Buena Vista, Fla, USA, February 2013.

[22] R. R. Carter, A. DiFeo, K. Bogie, G.-Q. Zhang, and J. Sun, "Crowdsourcing awareness: exploration of the ovarian cancer knowledge gap through Amazon Mechanical Turk," *PLoS ONE*, vol. 9, no. 1, Article ID e85508, 2014.

[23] S. Mavandadi, S. Dimitrov, S. Feng et al., "Distributed medical image analysis and diagnosis through crowd-sourced games: a malaria case study," *PLoS ONE*, vol. 7, no. 5, Article ID e37245, 2012.

[24] T. B. Nguyen, S. Wang, V. Anugu et al., "Distributed human intelligence for colonic polyp classification in computer-aided detection for CT colonography," *Radiology*, vol. 262, no. 3, pp. 824–833, 2012.

[25] D. Mitry, T. Peto, S. Hayat, J. E. Morgan, K.-T. Khaw, and P. J. Foster, "Crowdsourcing as a novel technique for retinal fundus photography classification: analysis of images in the EPIC norfolk cohort on behalf of the UKBiobank eye and vision consortium," *PLoS ONE*, vol. 8, no. 8, Article ID e71154, 2013.