

Cite this: *Lab Chip*, 2012, 12, 4102–4106www.rsc.org/loc

PAPER

Crowd-sourced BioGames: managing the big data problem for next-generation lab-on-a-chip platforms†

Sam Mavandadi,^a Stoyan Dimitrov,^{‡a} Steve Feng,^{‡a} Frank Yu,^{‡a} Richard Yu,^{‡a} Uzair Sikora^a and Aydogan Ozcan^{§*abcd}

Received 28th May 2012, Accepted 8th August 2012

DOI: 10.1039/c2lc40614d

We describe a crowd-sourcing based solution for handling large quantities of data that are created by *e.g.*, emerging digital imaging and sensing devices, including next generation lab-on-a-chip platforms. We show that in cases where the diagnosis is a binary decision (*e.g.*, positive *vs.* negative, or infected *vs.* uninfected), it is possible to make accurate diagnosis by crowd-sourcing the raw data (*e.g.*, microscopic images of specimens/cells) using entertaining digital games (*i.e.*, BioGames) that are played on PCs, tablets or mobile phones. We report the results and the analysis of a large-scale public BioGames experiment toward diagnosis of malaria infected human red blood cells (RBCs), where binary responses from approximately 1000 untrained individuals from more than 60 different countries are combined together (corresponding to more than 1 million cell diagnoses), resulting in an accuracy level that is comparable to those of expert medical professionals. This BioGames platform holds promise toward cost-effective and accurate tele-pathology, improved training of medical personnel, and can also be used to manage the “Big Data” problem that is emerging through next generation digital lab-on-a-chip devices.

The internet revolution can be traced back to the invention of the transistor. It was the integration of transistors into electronic chips, and subsequently into computational systems that led to the development of computers and the internet. For over three decades, the number of transistors in integrated circuits has been doubling approximately every 18 months, a trend that is dubbed as the Moore's Law. Quite interestingly, a similar trend seems to exist for the number of pixels installed on mobile phone cameras. In other words the megapixel count of mobile phones has been following Moore's Law since the wide-scale introduction of camera phones in 2002 with 0.3 megapixels, reaching 41 megapixels in 2012.¹ The massive volume of these consumer electronics components, reaching billions, has brought a phenomenal reduction in cost and unprecedented levels of access to such advanced digital devices despite their sophisticated hardware and software capabilities.

As a result of this revolution in digital electronics, we are now seeing a paradigm shift in imaging and sensing technologies with ultra-portable, cost-effective and high-throughput lab-on-a-chip

platforms providing innovative solutions for *e.g.*, point-of-care and telemedicine applications, among others.^{2–18} Cost-effectiveness and portability of these emerging digital imaging and sensing technologies through lab-on-a-chip platforms will allow wide-scale generation of large amounts of biomedical and environmental data even in traditionally resource-limited and remote parts of the world. The richness of this data will surely provide new opportunities for better understanding various phenomena such as global spatio-temporal patterns of various diseases and health conditions, helping us relate such patterns to their causes, potentially also influencing our global health and environmental policies. On the other hand, such massive amounts of data that will be continuously created by the next generation digital imaging, sensing, and lab-on-a-chip technologies will also add to our “Big Data” problem, which can fundamentally be related to the fact that the performance, training, and the number of available human experts (*e.g.*, medical personnel, data analysts) do not scale as fast as Moore's Law.

In this work, we introduce a platform that aims to tackle this emerging Big Data problem through entertaining games that crowd-source the digital data of interest (*e.g.*, microscopic images of specimens) to minimally trained humans/gamers for diagnostic decision making or labelling. These digital games (termed BioGames – see Fig. 1) can be played all around the world on PCs, tablets or mobile phones. The individual responses of the gamers are collected using secure servers for analysis and decoding of the final decisions and the labelling of the original data.

^aElectrical Engineering Department, University of California, Los Angeles, CA, USA. E-mail: ozcan@ucla.edu

^bBioengineering Department, University of California, Los Angeles, CA, USA

^cCalifornia NanoSystems Institute, University of California, Los Angeles, CA, USA

^dDepartment of Surgery, School of Medicine, University of California, Los Angeles, CA, USA

† <http://biogames.ee.ucla.edu>

‡ These authors contributed equally to this work.

§ <http://innovate.ee.ucla.edu>

At the centre of the BioGames platform lies crowd-sourcing, which has recently emerged as a powerful strategy for tackling computationally difficult scientific problems.^{19–24} The basic idea behind a crowd-sourcing platform is to first break the task of interest into smaller pieces that can each be completed in a relatively short amount of time; and second to distribute these pieces of the problem to individual humans through a convenient medium, such as the internet; and finally to combine the collective responses of the individual participants to yield an optimal solution. A promising approach to distributing the data and finding human volunteers has been the use of entertaining computer games.^{21,25,26} In this approach the scientific or computational problem is embedded into a digital game and then distributed to gamers. The individual gamers may then co-operate, compete, or play independently to solve parts or the entire problem of interest.

Along the same lines, we have recently proposed a crowd-sourcing-based solution for diagnosing or labelling bio-medical images.²⁴ In this small scale experiment (involving ~30 gamers), we showed that it is conceivable to achieve accurate diagnosis results by asking multiple individuals to label human RBC images that are potentially infected with the malaria parasite *Plasmodium falciparum*. In this manuscript, we expand on our earlier work and report the results and the related analysis of a large-scale public BioGames experiment toward diagnosis of malaria infected human RBCs, where binary responses from approximately 1000 minimally trained individuals from 6 different continents that have collectively diagnosed more than 1 million RBCs are combined, resulting in a diagnostic accuracy level that is comparable to those of expert medical professionals examining the same set of images (See Fig. 1).

We chose malaria for this public BioGames experiment since it is a disease that still afflicts a large number of people around the globe, and is most prevalent in impoverished and remote locations of the world. Furthermore, the binary nature of malaria diagnosis makes it suitable for the proposed BioGames approach. According to the World Health Organization, there were an estimated 174 million cases of malaria in 2010 resulting in 655 000 deaths, where 81% of the cases and 91% of the deaths

occurred in Africa.²⁷ Furthermore, approximately 86% of the global death toll has been in children under the age of 5.

Toward diagnosis of malaria using BioGames, we have designed a computer game in which a gamer is presented with multiple game *frames* containing a grid of RBC images (see Fig. 1). These images are taken from *Giemsa stained thin blood smear* samples using traditional bright-field microscopes with 100× objective-lenses. The Giemsa stain typically causes the malaria parasite to appear with a bluish colour, helping its diagnosis. Each gamer is given a brief online tutorial of what malaria infection looks like and is asked to *kill* or *bank* what s/he thinks are infected and healthy cells respectively in the frames that are presented throughout the game. Within each frame, we also include a set of *control* cells, *i.e.*, RBC images whose labels (infected or healthy) are known to the game. The purpose of these control images is to allow us to determine the performance and error rates of each individual gamer as s/he is going through the frames. We use these performance numbers when combining and weighing the responses coming from multiple gamers. Stated differently, those gamers who are found to be highly accurate through these control images will be given more weight in the decision process compared to other gamers who perform poorly during the same games.

In our decoding strategy, we take a tele-communications analogy and start by assigning the labels 0 and 1 to correspond to healthy and infected cells, respectively. Using the control images, we compute the error probabilities associated with each gamer as $p(y_i^k | x_i = l)$ where y_i^k is the label that gamer k has provided for the i^{th} cell image, x_i is the true label of the image, and $l \in \{0,1\}$. Assuming we have M gamers and a prior $\phi_l \triangleq p(x_i = l)$, then we can compute the *maximum a posteriori probability* (MAP)²⁴ estimate z_i for the i^{th} RBC as:

$$z_i = \arg \max_{l \in \{0,1\}} \left[\log \phi_l + \sum_{j=0}^M \log p(y_i^j | x_i = l) \right] \quad (1)$$

Given that we do not necessarily know the parasitemia level of the RBCs presented to the gamers, we assume equal prior probabilities for the infected and healthy cells, thus making the



Fig. 1 *Left:* The designed game can be played on multiple platforms. The user is asked to *kill* or *bank* infected and healthy cells, respectively. *Right:* Geographic locations of gamers that have generated the diagnoses so far; each balloon indicates an individual gamer. A blue balloon indicates a gamer with more than 100 submitted cell diagnoses, and a red balloon indicates a gamer with less than 100 submitted cell diagnoses. Since its public launch in May 2012, we have had more than 2,150 gamers from 77 countries, who registered on our servers generating more than 1.5 million individual cell diagnoses.

first term in the maximisation argument in eqn (1) irrelevant. Based on this decoding scheme, each gamer is essentially treated as a “repeater” in a noisy tele-communication channel, and using a MAP scheme we estimate the correct label for each unknown image in the game.²⁴

We made our BioGames interface available to the public on May 3, 2012, and as of August 4, 2012 we have had more than 2150 gamers from 77 countries, who registered on our servers generating more than 1.5 million individual cell diagnoses. Our image database contained approximately 8500 individual RBC images taken with different optical microscopes under various imaging conditions. Since labelling the full dataset (*i.e.*, 8500 images) takes more time than what most gamers are willing to commit on average, it was expected that not all the gamers will diagnose the full set of images. For the purposes of data analysis and diagnosis of unknown RBC images, we only used the data collected from the gamers who managed to at least label 100 cells. This threshold is mainly to have sufficient statistics for estimating the gamer performance levels, but it also helps us to identify and select the “committed” gamers, yielding a total of 989 committed gamers (from 63 countries) with varying accuracy levels as shown in Fig. 2.

The ground truth labels for the images used in BioGames were obtained by asking a set of 9 medical experts to digitally label them and then taking their consensus responses. Based on this, we were able to quantify the individual performances of our gamers as well as their collective MAP decision estimates. We monitored the performance of the individual gamers using every 25th cell in this database as part of our control images. Fig. 3 shows the deviation of these accuracy estimates (based on control images) from the “true” accuracies of the gamers (based on all 8500 images). The distribution of these deviations is super-Gaussian and is concentrated close to zero, which implies that our control images (even though they are quite rare within the entire image set) provide sufficiently accurate sampling to represent each gamer’s overall error probability.

Combining the responses of these 989 gamers using MAP estimation,²⁴ we were able to achieve an accuracy of 98.13% when compared to the ground truth data (generated by the

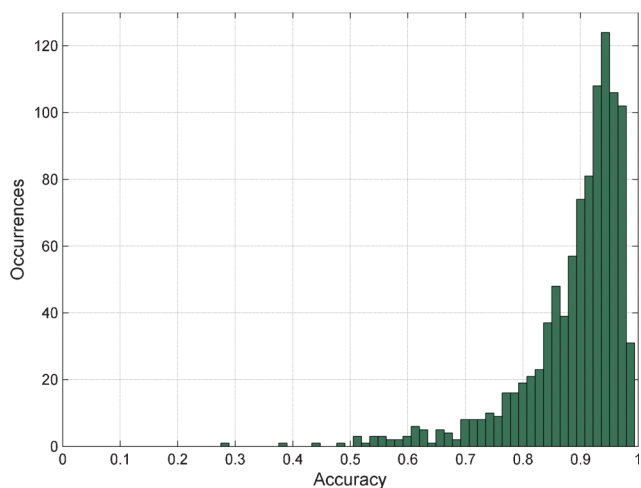


Fig. 2 Histogram of the overall accuracy levels of the 989 individual gamers who managed to label more than 100 cell images.

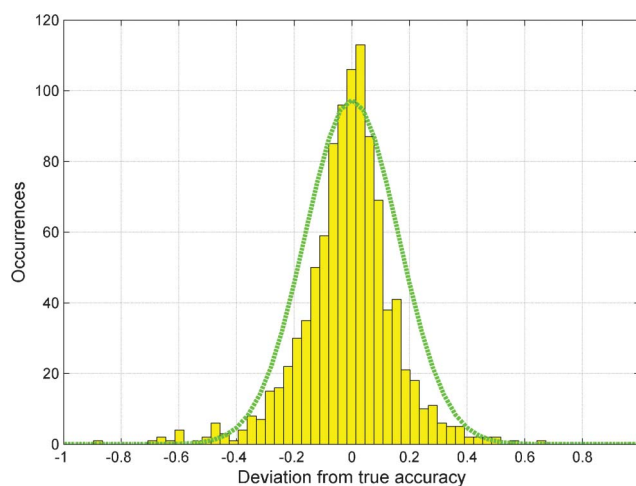


Fig. 3 Distribution of the deviation from the true accuracies of those estimated using the control images; the mean is at -0.014 and the variance is 0.026 . The green curve is that of a Normal distribution with the same variance as the deviations and a mean of 0 .

consensus of 9 medical experts). This number on its own, however, can be misleading due the fact that there are significantly more uninfected cells in the database as compared to infected ones (which is naturally expected since typical parasitemia for a malaria infected patient is less than 5%). Some other important metrics include: the *Positive Predictive Value* (PPV), the *Negative Predictive Value* (NPV), and the *False Positive Rate* (FPR) (see Fig. 4 for definitions). In our BioGames experiments, the PPV was 76.85%, meaning that more than three quarters of the cells that were labelled as infected, were indeed infected. We also achieved an NPV of 98.78%, such that almost all of the cells labelled as negative are correctly labelled as such.

In Fig. 4 we also report the change of the above discussed performance metrics as a function of the *effective crowd size*, which we define as the minimum number of times that *all* of the cells in the dataset have been diagnosed by the gamers. Since not all of the gamers diagnosed the complete set of images, the maximum effective crowd size that was achieved in our public BioGames experiment was 51 with some cells being diagnosed up to 450 times due to the overlap that existed among the ranges of cells diagnosed by different gamers. It is important to note that both the overall accuracy and the PPV are increasing as the effective size of the crowd increases. Furthermore, as desired, the FPR is consistently decreasing with larger crowd sizes (see Fig. 4). These experimental results validate the efficacy of our BioGames approach toward obtaining highly accurate diagnostics results and image labels with large-enough crowds of minimally trained individuals, that are physically scattered across several continents.

We should also emphasise that this distributed BioGames platform can be extended to scenarios where a binary diagnosis is no longer the case. It is possible to systematically combine results of *n*-ary decisions in an optimal fashion to produce a more accurate overall result/decision. However, a possible issue in such a more complicated *n*-ary scenario is the increased number of crowd decisions that may be required for accurate

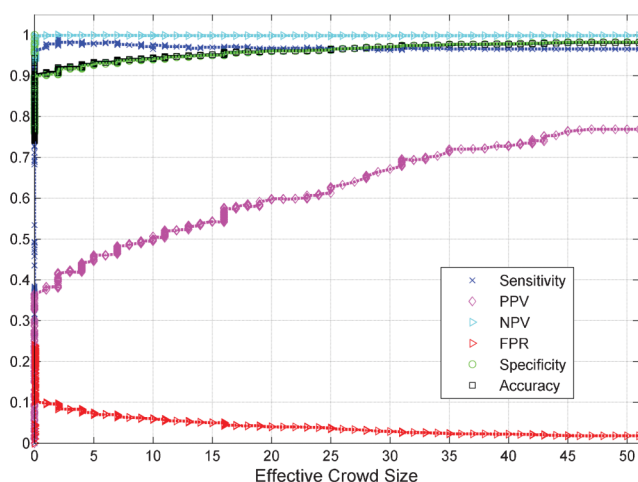


Fig. 4 The performance results obtained through combining responses of 989 gamers. The “effective crowd size” is the minimum number of times that every cell in the dataset has been diagnosed. Accuracy \triangleq (TP + TN)/(TP + TN + FP + FN), Sensitivity \triangleq TP/(TP + FN), Specificity \triangleq TN/(TN + FP), PPV \triangleq TP/(TP + FP), NPV \triangleq TN/(TN + FN), FPR \triangleq FP/(TN + FP), where TP, TN, FP, and FN correspond to the number of true positive, true negative, false positive, and false negative labels respectively.

diagnosis, making certain tasks (especially the ones that demand real-time decisions) impractical for crowd-sourcing purposes.

Looking forward, the BioGames concept could help us better manage the Big Data problem emerging with the introduction of next generation imaging, sensing and lab-on-a-chip devices which all benefit from ubiquitous digital communications technologies (e.g., mobile phones, tablet PCs, etc) for wide-scale generation of massive amounts of biomedical and environmental data, even in resource poor settings and remote locations. Through intelligent crowd-sourcing and digital gaming strategies, we can potentially harness the power of human crowds and their innate visual pattern recognition and learning abilities to better sort, classify, diagnose, and manage this emerging Big Data.

Our grand vision is to further develop this BioGames interface into a generic telemedicine and telepathology platform, further extending it to other medical diagnostics tasks. In addition to generating remote biomedical diagnosis through engaging games, the presented platform can serve as an information hub for the global healthcare and environmental research community. This digital hub will allow for the creation of very large databases of microscopic images that can be used for the purposes of training experts and automated computer vision algorithms. It can also serve as an analysis tool for health-care and environment related policy makers toward better management and/or prevention of epidemics, pandemics and disasters.

In conclusion, here we described how the task of binary image-based diagnosis can be crowd-sourced to minimally trained individuals and yet yield accurate results. We reckon that under many circumstances this methodology is more practical than the use of automated computer algorithms for the same purpose since the human visual system offers a very low-cost and highly superior pattern recognition platform

for such *image understanding* tasks. We demonstrated our BioGames results for images taken with traditional bright-field microscopes, but given that the task of diagnosis is independent of the imaging modality, as long as enough detail and resolution is present in the image, the proposed approach is applicable to other imaging modalities. This makes it a viable approach for managing the large quantities of data that will be created by our next generation lab-on-a-chip imaging and sensing devices that are rapidly emerging for especially point-of-care diagnostics and telemedicine applications through ubiquitous digital communications platforms such as mobile phones, tablet PCs etc.

Acknowledgements

A. Ozcan gratefully acknowledges the support of the Presidential Early Career Award for Scientists and Engineers (PECASE), ARO Young Investigator Award, NSF CAREER Award, the ONR Young Investigator Award 2009 and the NIH Director's New Innovator Award DP2OD006427 from the Office of The Director, NIH.

References

- 1 Nokia, Nokia 808 Pureview, 2012, <http://europe.nokia.com/find-products/devices/nokia-808-pureview>.
- 2 D. N. Breslauer, R. N. Maamari, N. A. Switz, W. A. Lam and D. A. Fletcher, *PLoS One*, 2009, **4**, e6320.
- 3 D. Tseng, O. Mudanyali, C. Oztoprak, S. O. Isikman, I. Sencan, O. Yaglidere and A. Ozcan, *Lab Chip*, 2010, **10**, 1787–1792.
- 4 W. Bishara, T.-W. Su, A. F. Coskun and A. Ozcan, *Opt. Express*, 2010, **18**, 11181–11191.
- 5 A. R. Miller, G. L. Davis, Z. M. Oden, M. R. Razavi, A. Fateh, M. Ghazanfari, F. Abdolrahimi, S. Poorazar, F. Sakhaie and R. J. Olsen, *et al.*, *PLoS One*, 2010, **5**, e11890.
- 6 Z. J. Smith, K. Chu, A. R. Espenson, M. Rahimzadeh, A. Gryshuk, M. Molinaro, D. M. Dwyre, S. Lane, D. Matthews and S. Wachsmann-Hogiu, *PLoS ONE*, 2011, **6**, 11.
- 7 S. O. Isikman, W. Bishara, U. Sikora, O. Yaglidere, J. Yeah and A. Ozcan, *Lab Chip*, 2011, **11**, 2222–2230.
- 8 H. Zhu, O. Yaglidere, T.-W. Su, D. Tseng and A. Ozcan, *Lab Chip*, 2011, **11**, 315–322.
- 9 H. Zhu, S. Mavandadi, A. F. Coskun, O. Yaglidere and A. Ozcan, *Anal. Chem.*, 2011, **83**, 6641–6647.
- 10 A. Greenbaum, U. Sikora and A. Ozcan, *Lab Chip*, 2012, **12**, 1242–1245.
- 11 O. Mudanyali, S. Dimitrov, U. Sikora, S. Padmanabhan, I. Navruz and A. Ozcan, *Lab Chip*, 2012, **12**, 2678, DOI: 10.1039/c2lc40235a.
- 12 P. Yager, T. Edwards, E. Fu, K. Helton, K. Nelson, M. R. Tam and B. H. Weigl, *Nature*, 2006, **442**, 412–418.
- 13 S. K. Sia, V. Linder, B. A. Parviz, A. Siegel and G. M. Whitesides, *Angew. Chem., Int. Ed.*, 2004, **43**, 498–502.
- 14 A. W. Martinez, S. T. Phillips, E. Carrilho, S. W. Thomas, H. Sindi and G. M. Whitesides, *Anal. Chem.*, 2008, **80**, 3699–3707.
- 15 B. Schatz, R. Berlin and J. RichardB. Berlin, *Healthcare Infrastructure: Health Systems for Individuals and Populations*, Springer, 2011.
- 16 G. McKernan, *Searcher: The Magazine for Database Professionals*, 2010, **18**, 48.
- 17 R. Adler, *California Healthcare Foundation*, 2007.
- 18 R. Hurling, M. Catt, M. D. Boni, B. W. Fairley, T. Hurst, P. Murray, A. Richardson and J. S. Sodhi, *J. Med. Internet Res.*, 2007, **9**, e7.
- 19 L. Von Ahn, B. Maurer, C. McMillen, D. Abraham and M. Blum, *Science*, 2008, **321**, 1465–1468.
- 20 D. Roman, *Commun. ACM*, 2009, **52**, 12–12.
- 21 S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popovic and F. Players, *Nature*, 2010, **466**, 756–760.

-
- 22 S. Cooper, D. Baker, Z. Popovic, A. Treuille, J. Barbero, A. Leaver-Fay, K. Tuite, F. Khatib, A. C. Snyder and M. Beenen, *et al.*, *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10*, 2010, 40–47.
- 23 F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popovic, D. Baker and F. Players, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 18949–18953.
- 24 S. Mavandadi, S. Dimitrov, S. Feng, F. Yu, U. Sikora, O. Yaglidere, S. Padmanabhan, K. Nielsen and A. Ozcan, *PLoS One*, 2012, **7**, e37245.
- 25 L. von Ahn and L. Dabbish, *Commun. ACM*, 2008, **51**, 58–67.
- 26 EteRNA, *EteRNA—Played by Humans, Scored by Nature*, 2011 <http://eterna.cmu.edu/content/EteRNA>.
- 27 W. H. O. Who, *World*, 2011, 278.