Medical image interpretation training with a low-cost eye tracking and feedback system: A preliminary study

Mathieson Tan Zui Quen | J. Mountstephens | Yong Guang Teh | J. Teo

1 Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia
2 Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia

Correspondence
J. Mountstephens, Faculty of Computing and Informatics, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia.
Email: james@ums.edu.my

Funding Information
The APC for this work was provided by PPPI, Universiti Malaysia Sabah.

Abstract
All medical students must learn to detect tumours in chest X-ray films. Eye tracking technology may be able to improve the training of this skill, with the potential benefit of saving thousands of lives each year. Eye trackers can record where students actually look in relation to known abnormalities in images, allowing the automated provision of corrective feedback during medical training. Research using dedicated eye tracking units suggests their potential as pedagogical tools but this hardware can be expensive, which prevents their use by medical students. To overcome this financial barrier, this project used eye tracking based on low-cost webcams to develop a system with the potential to allow all medical students to improve their interpretation skills. In self-study mode, students can practice and improve based on automated performance feedback in a browser-based system. Improvements in performance over time can be determined and mistakes can be played back and analysed for correction. Eight medical trainees used the system on a custom dataset of 60 chest X-ray images and were able to improve their decision times by self-study alone. The system was also rated highly for its ability to provide valuable objective information during subsequent one-on-one instruction sessions.

1 INTRODUCTION

Imaging techniques afford medical professionals a view of their patient’s internal state without the need for invasive surgery [1]. As part of their education, trainee doctors must develop the ability to correctly interpret medical images. This skill is part perceptual and part cognitive. Clearly, understanding what the features of images mean is important but learning where to look in the first place is an essential prerequisite [2]. Radiography is the oldest branch of medical imaging with plain film radiography (commonly known as X-ray) the most mature technique. Newer imaging modalities such as CT and MRI scans are available to most doctors but X-ray imaging still remains extremely important because it is inexpensive, readily accessible, and easy to perform [3]. For these reasons, it is “commonly the first-line imaging modality utilized by clinicians.” (ibid). Medical images must be interpreted in order to be useful and while advances have been made in computerised interpretation [4, 5], human interpretation still rules for the time being [6]. The correct interpretation of chest X-rays is of particular importance since radiography is most frequently used in initial investigations of the thorax [3]. Detecting tumours at an early stage during these initial examinations is a crucial function of chest X-rays, as illustrated by Figure 1.

To the experienced clinician, chest X-rays offer rich and detailed information about a patient’s condition, but it is well-known that they “can seem baffling and intimidating for junior doctors” [7]. And the consequences for error are serious. The correct interpretation of medical images saves lives but thousands die each year due to incorrect interpretation. Brady [2] estimates that in the US alone, up to 98,000 preventable deaths occur annually due to errors in interpreting radiographic images. Reporting for the US Institute of Medicine, Brady assigns only 20–40% of this error to cognitive defects (i.e. to misinterpreting what was seen, and its significance). The majority of errors are actually perceptual: the doctor failed to see the abnormality in the first place.

Improvements in medical education are considered to be a key way to rectify this problem [2], and the application of computing technologies to medical training offers great potential to...
achieve this goal [6, 8]. Medical image interpretation is taught in senior classes on radiography but is only developed as a practical skill later during internship as medical officers (MOs). Knowing where to direct and allocate gaze is crucial for success and this is learned based on one-to-one feedback from experienced doctors. However, the time available for personal guidance is limited, and teaching this particular skill is complicated because the instructor cannot directly know where the trainee is actually looking to offer the most precise feedback.

These problems could be addressed by modern eye-tracking technology [9, 10] (see Figure 2) since it can directly determine student gaze and offer the basis for automated feedback that does not require the presence of an experienced doctor or instructor. In addition to aiding self-study, records of exactly how students view training images can be used for more informed one-to-one feedback sessions with instructors. Gaze pattern data can be compared between experienced and trainee doctors and significant patterns leading to error or success could even be incorporated into medical training more formally.

The use of eye tracking has steadily grown within the medical field [eg. 10, 11]. The educational potential of eye tracking for medical training has been noted by Leveque et al [12], Bertram et al. [13], and Brunyé et al [14, 15] in their eye-tracking studies of the visual search patterns of medical professionals. With eye tracking, it can be clearly demonstrated that experienced doctors gaze at images very differently to trainees in terms of spread and decision time (ibid). It has been proposed that eye tracking can be used to determine where students actually look and that this data be used as the basis for objective performance feedback. However, a serious obstacle to realising this potential is that dedicated eye tracking units can be expensive, which severely limits their availability to students. For example, an entry level Tobii unit can still cost over USD 250 [16].

To overcome this financial barrier, the research presented here used low-cost webcams (found in all modern laptops) to develop a system with the potential to allow all medical students to improve their image interpretation skills. It was intended that students can independently practice and improve their skills based on automated performance feedback in a system that requires only a web browser to run. Improvements in performance over time are recorded and mistakes can be played back and analysed for correction. This automated feedback is free and always available.

For this project, the choice was made to use a webcam compatible eye tracking library that is specifically browser based, making it cross-platform and available for use anywhere there is an internet connection. If widely adopted, this would also offer future potential to aggregate gaze data internationally on a standardised image dataset which can yield analysis and improvements to medical education in this domain worldwide. However, it should be noted at the outset that webcam-based eye tracking cannot rival dedicated hardware for accuracy. The research question addressed here is whether it is adequate to improve the learning process for both student and instructor.

## 2 | LOW-COST EYE TRACKING AND FEEDBACK SYSTEM

The overall design of our Low-Cost Eye Tracking and Feedback System is shown in Figure 3. The hardware components are webcam and keyboard, which are found in all modern laptops. The software side was implemented using custom JavaScript for the browser-based interface, with MySQL for the database backend.

The eye-tracking software used was Webgazer.js, a free and open-source library that uses standard desktop and laptop webcams to infer the locations of user gaze on a webpage in real-time [17, 16, 18, 19]. Webgazer’s eye tracking model is calibrated using a nine-point method which is used to
determine a mapping between eye features and screen locations. WebGazer.js is written entirely in JavaScript and can be integrated into most webpages with minimal code. This library runs entirely in the client browser, precluding the need to send high bandwidth video data to a server. The primary components of WebGazer.js are the tracker module and the regression module. The tracker controls how eyes in a facial image are detected, while the regression module governs how a regression model is learned and how gaze locations are predicted based on the eye patches extracted from the tracker. The default tracker module in Webgazer is the Facemesh library by MediaPipe [20], a face geometry detection module that estimates 468 3D face landmarks in real-time and uses machine learning to infer the underlying 3D surface geometry. Facemesh requires only a single camera input without the need for a dedicated depth sensor. Webgazer’s regression module estimates a regression model of the form:

$$\text{minimise } \sum_{x_i \in x} \| D_{x_i} - f(x_i) \|^2 + \lambda \| w \|^2$$  \hspace{1cm} (1)

where $x$ is the predicted x-coordinate of user gaze (the equation is identical for the y-coordinate), $w$ is a weight vector, $D_{x_i}$ are the display coordinates of the nine calibration points, and $\lambda$ is a regularisation term. In practice, Webgazer.js provides three regression modules:

1. Ridge, a simple ridge regression model that maps pixels from the eye patches detected by Facemesh to ($x, y$) screen locations. This is the model used in the system developed in the current work.
2. Weighted Ridge, a weighted ridge regression model where the most recent user interactions contribute more to the model.
3. Threaded Ridge, an implementation of ridge regression employing threads to improve speed.

The basic functionality of the feedback system in use is shown in Figure 4 below. There are two basic modes: testing and feedback. During testing, a randomised set of chest X-ray images is shown sequentially in a slideshow to the user who must classify each image as either abnormal (containing tumours) or normal (clear) by clicking buttons for these two options at the bottom of the screen (Figure 5). Clicking will advance to the next image. The length of the image sequence can be set by the user and no time limit on the session is imposed.

During this phase of viewing and classification, the user’s eyes are tracked as they view each image and a sequence of time stamped gaze locations ($x, y, t$) are written to the database along with other relevant parameters for this session (ie user ID, image IDs, date and location). Calibration is conducted before testing begins to ensure the best possible accuracy of reported gaze locations.

When the sequence of images is complete, the system enters feedback mode and performance for the particular session is shown in a variety of ways. Initially, overall statistics are displayed for accuracy of classification, time taken, and percentage of gaze points on tumours (when they exist). As shown in Figure 6, The whole image sequence is listed as thumbnails, and individual images may be selected for detailed inspection, which is likely if errors were made on them.

During the detailed inspection of an image (Figure 7), the locations of tumours (if present) are highlighted and the student’s gaze data is superimposed onto the image, allowing analysis of its own characteristics (such as spread and coverage) but most importantly as compared to the locations of abnormalities. Gaze data can be viewed in three ways: as a simple set of points, as a heat map which reveals gaze point density more clearly, and as a sequence of points which can be advanced...
manually by the user. This sequential mode reveals the temporal progression of viewing in a way that the other two modes do not.

Since all session data are stored in the database, the system can also enter feedback mode for a previously conducted session for the purposes of self review or to allow a senior doctor to give feedback on the student’s performance with the aid of known gaze and tumour locations.

3 | EXPERIMENTS

The proposed system was tested on real trainee doctors in two experimental contexts. The first experiment tested the system in a self-study context and the second tested the system in the traditional setting of one-to-one feedback with a senior doctor.

The self-study experiment aimed to determine whether automated eye-tracking feedback would improve student interpretation performance on chest X-rays. The basic experimental design was to first divide participants into Treatment and Control groups. The Treatment group would use the system and receive eye tracker feedback from the system and the Control group would not. Both groups would be initially presented with a Test set of X-ray images in randomised order, for unlimited time. This was followed by a feedback session from the system of up to 10 min where the Treatment group was able to see gaze data but the Control group was not. Then both groups would classify a Retest set of X-ray images and objective differences in performance could be ascertained between those receiving automated feedback and those who did not. The two measures of performance considered here were classification accuracy and decision time. The overall design of the self study experiment is shown in Figure 8.

One week later, a one-on-one feedback session with a senior radiologist would offer feedback on each student’s performance using the system to display the gaze data collected during self-study. The efficacy of the system for this purpose would be determined subjectively by the opinion of both student and instructor.

Both experiments were carried out at the Sabah Woman and Children’s Hospital, Kota Kinabalu, East Malaysia (see Figure 9). Eight medical officers (MOs) currently undergoing internship at this facility agreed to participate in our study. The MOs each had differing levels of experience and were allocated to Treatment and Control groups so that each group had approximately the same total level of experience, as measured by months of internship so far (see Table 1).

60 Chest X-ray images (Figure 10) were collected by a senior radiologist at the Sabah Woman and Children’s Hospital, Kota Kinabalu, Sabah, Malaysia. These images consisted of 30 abnormal (containing one or more tumours) and 30 normal (containing no tumours) images and were marked up for abnormalities.
and rated for difficulty by the senior radiologist which allowed division into Test and Retest image sets of equal difficulty. Test and Retest individually contained 15 normal and 15 abnormal images.

4 RESULTS

The results for the automated feedback session are summarised in Tables 2 and 3 below. These table are divided into sections for the three stages of the experiment: Test, Feedback, and Retest. For Test and Retest, the number of accurately classified images (out of 30) is listed for each participant, followed by the time taken to classify the full 30 images. Time taken during the intermediate Feedback stage is shown in the centre column for both the Treatment group, who were able to view their gaze data, and the Control group, who were not. Averages per group for all these measures are shown beneath each table.

It can be seen that there was little difference in overall accuracy between Treatment and Control groups. There was a very small decrease in mean accuracy found after retest for both groups (~1% and 2%, respectively). Although an increase in accuracy would have been the desired outcome, this small reduction was not considered significant and could be attributed to fatigue and discomfort from remaining in a fixed viewing position for around 30 mins. It can also be seen that, with the exception of participant SHU in the Treatment group, there was little difference in the time chosen to study the feedback provided by either group.

Although overall accuracy did not improve after receiving feedback on gaze location, the Treatment group did demonstrate a significant average improvement in decision time over the Control group as a whole. This result suggested that even with similar feedback times, the quality of feedback based on eye tracking and actual gaze locations was more useful for improvement by self-study. It was remarked afterwards by the least-experienced members of the Treatment group that being able to see that they were performing well gave them more confidence in their performance, allowing them to work faster. This could be the main reason why the Control group did not perform much faster on Retest. And it is worth noting that confidence after feedback was an important factor in the only exception to improved decision time observed in the Treatment group. Unlike the two least-experienced MOs (NAD and SHU), who reduced their decision times by ~50%, the most experienced MO (TKY) actually took twice as long after feedback because knowing that mistakes had been made encouraged a more careful approach in the subsequent retest.

Two weeks later, our system and the recorded self-study sessions were used to explore how eye tracking performance could improve traditional one-to-one feedback from experienced radiologists. Ordinarily, instructional reference to X-ray films and images can be ad hoc and no standard database of images is available. And, even with a structured set of training images, no knowledge of where the student was actually looking has ever been available. Our system was able to supply both these elements, and it was hoped that they would be found useful by both instructor and student. Apart from a loose 15 min time limit, the content of the feedback session was not formally structured. The senior radiologist and student were free to discuss...
TABLE 3  Control group performance during self-study session

| ID   | Test accuracy (of 30) | Test time (min) | Feedback time (min) | Retest accuracy (of 30) | Retest time (min) |
|------|-----------------------|-----------------|---------------------|-------------------------|------------------|
| NIZ  | 27                    | 15.54           | 4                   | 26                      | 16.55            |
| LIM  | 26                    | 20.1            | 2.35                | 21                      | 14.38            |
| SIM  | 25                    | 17.53           | 2.32                | 22                      | 23.2             |
| ATI  | 18                    | 18              | 1.51                | 25                      | 11.03            |
| Mean | 24                    | 17.79           | 2.55                | 23.5                    | 16.29            |

FIGURE 11  Spread of gaze for MOs (top) and experienced (bottom) doctors

whichever images and issues they considered most salient with reference to specific images and actual gaze performance.

Both senior radiologist and all trainees rated the system positively. The instructor was pleased to be able to refer directly to student gaze in his comments. With this information he was able to identify deficiencies common to most students. For example, the difference in the spread of gaze between student and the experienced radiologist himself shown in Figure 11 below were found to be common. Please note that the bottom image obtained from the experienced doctor looks messier due to the limited accuracy of the webcam eye tracker. In fact, the greater spread of the experienced doctor is desirable to ensure no tumour is missed and the narrow focus of the student is something to be overcome by training. This pattern demonstrated by webcam eye tracking is consistent with previous differences in gaze pattern by experience determined by expensive dedicated eye trackers.

TABLE 4  Medical Officer comments on system-assisted human feedback session

| #  | ID | Comment                                                                 |
|----|----|--------------------------------------------------------------------------|
| 1  | TKY | “Need to focus on right side; be disciplined in reviewing the X-ray, cover the review areas” |
| 2  | HAL | “Need to recognise normal; cover review area especially apex, behind the heart; need to focus on left side” |
| 3  | NAD | “The application is useful for first introduction to X-ray; feedback session was great for us to know our scanning pattern” |
| 4  | SHU | “Need to be more disciplined in viewing the X-ray, more knowledge” |
| 5  | NIZ | “Good viewing pattern, recognizing problem” |
| 6  | LIM | “Good in teaching usage (to know how student answering questions during exam)” |
| 7  | SIM | “Great to know the scanning pattern” |
| 8  | ATI | “Good viewing pattern, need to improve knowledge” |

Each MO was also asked to provide short comments on the effectiveness of our system in this more traditional teaching context. As seen in Table 4 below, these comments are unanimously positive. And significantly, they describe the specific locations of issues that must be improved which could only be determined by knowing their gaze locations through eye tracking.

5  | CONCLUSION

This research is still very much a pilot study but the results found at this early stage are promising. It was possible to develop a functional training system using free webcam-based eye tracking. The system was able to achieve some objective performance improvements in terms of reduced decision time after use as a self-study aid. And the system’s potential for integration into more traditional one-on-one instruction is demonstrated by positive feedback from both instructor and students. The accuracy of webcam-based eye tracking is clearly not equal to dedicated hardware but it appears adequate to provide genuine benefit as a learning technology and to identify gaze patterns consistent with those found in more expensive yyyy studies.
Many improvements under future work are currently underway. First of all, it is clear that experimentally, larger sample sizes are required to ensure that the results found here are representative. Second and perhaps more importantly, the nature of the automated feedback should be improved. The current approach is simply to present information for the student to interpret independently. The next phase of system evolution will be to develop a more proactive style of feedback which is more akin to human guidance in correction and instruction. This is a major undertaking since no expert system for interpreting chest X-rays is currently known to exist. Constructing the rules governing a more active feedback would require the synthesis of known best practices in the medical literature and the less easy to define intuition possessed by experienced doctors—a task that will require the input of medical practitioners and instructors from around the world. However, considering the benefits in terms of thousands of potential lives saved, it would be more than worth it.

**FUNDING**

The APC for this work was provided by PPPI, Universiti Malaysia Sabah.

**REFERENCES**

1. Bercovich, E., Javitt, M.C.: Medical imaging: from roentgen to the digital revolution, and beyond. Rambam Maimonides Med. J. 9(4), 0–34 (2018)
2. Brady, A.P.: Error and discrepancy in radiology: inevitable or avoidable?. Insights Imaging 8(1), 171–182 (2017)
3. Darby, M., Barron, D., Hyland, R.: Oxford Handbook of Medical Imaging. Oxford University Press, Oxford (2012)
4. Farahani, N., et al. Three-dimensional imaging and scanning: Current and future applications for pathology. J. Pathol. Inform. 8, 36 (2017)
5. Foran, D.J., Chen, W., Yang, L.: Automated image interpretation computer-assisted diagnostics. Anal. Cell Pathol. 34(6), 279–300 (2011)
6. Chan, K.S., Zary, N.: Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMRI Med. Edu. 5(1), e13930 (2019)
7. Eng, P., Cheah, F.K.: Interpreting Chest X-Rays: Illustrated with 100 Cases. Cambridge University Press, Cambridge, UK (2005)
8. Taveira-Gomes, T., et al.: What are we looking for in computer-based learning interventions in medical education? A systematic review. J. Med. Internet Res. 18(8), e204 (2016)
9. Singh, H., Singh, J.: Human eye-tracking and related issues: A review. Int. J. Sci. Res. Publ. 2, 1–9 (2012)
10. Wong, B.S.F., Ho, G.T.S., Tsui, E.: Development of an intelligent e-healthcare system for the domestic care industry. Ind. Manag. Data Syst. 117, 1426–1445 (2017)
11. Wang, Y., Li, Z., Zheng, Y.: Automatic emotion perception using eye movement information for E-healthcare systems. Sensors 18(9), 2826 (2018)
12. Lévêque, L., et al.: State of the art: Eye-tracking studies in medical imaging. IEEE Access 6, 37023–37034 (2018)
13. Bertram, R., et al.: Eye movements of radiologists reflect expertise in CT study interpretation: A potential tool to measure resident development. Radiology 281(3), 805–815 (2016)
14. Brunyé, T.T., et al.: A review of eye tracking for understanding and improving diagnostic interpretation. Cogn. Research 4, 7 (2019)
15. Brunyé, T.T., Nallamothu, B.K., Elmore, J.G.: Eye-tracking for assessing medical image interpretation: A pilot feasibility study comparing novice vs expert cardiologists. Perspect Med Educ 8, 65–73 (2019)
16. Papoutsaki, A., et al.: Scalable webcam eye tracking using user interactions. In Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI) 2016, New York, pp. 3839–3845 (2016)
17. Webgazer: https://webgazer.cs.brown.edu/, Accessed 22 March 2021
18. Papoutsaki A., Laskey, J., Huang, J.: SearchGazer: Webcam eye tracking for remote studies of web search. In Proceedings of the ACM SIGIR Conference on Human Information Interaction & Retrieval (CHIIR), Oslo, Norway (2017)
19. Papoutsaki A., et al.: The eye of the typer: a benchmark and analysis of gaze behavior during typing. In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications, Warsaw, Poland (2018)
20. Facemesh: https://google.github.io/mediapipe/solutions/face_mesh.html, Accessed 22 March 2021
21. Tobii: https://www.tobii.pro/product-listing/nano/, Accessed 22 March 2021.
22. Jacob, R.J., Kern, K.S.: Eye tracking in human-computer interaction and usability research. The Mind’s Eye, Elsevier BV, Amsterdam, Netherlands, pp. 573–605 (2003)

**How to cite this article:** Quen, M. T. Z., et al.: Medical image interpretation training with a low-cost eye tracking and feedback system: A preliminary study. Healthc. Technol. Lett. 8, 97–103 (2021). https://doi.org/10.1049/htl2.12014