medRxiv preprint doi: https://doi.org/10.1101/2020.10.19.20215087; this version posted October 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Evaluation of Eye-tracking Methods for a Decision Support Application

Shyam Visweswaran[1,2], Andrew J King[1,3], Mohammadamin Tajgardoon[2], Luca Calzoni[1], Gilles Clermont[3], Harry Hochheiser[1,2], Gregory F Cooper[1,2]

[1]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA
[2]Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA, USA
[3]Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

Eye-tracking is used widely to investigate visual and cognitive processes in the context of electronic medical record systems. We investigated a novel application of eye tracking to collect training data for machine learning-based clinical decision support. Specifically, we recorded the information-seeking behavior of physicians while they used electronic medical records in the context of a specific clinical task. Using data captured by a low-cost eye tracking device, we evaluated the performance of several methods for processing gaze points that were recorded using the device. Our results support the use of a low-cost eye tracking device and relatively simple methods for processing gaze points to record the information-seeking behavior of physicians. The eye-tracking methods and scripts that we developed offer a first step in developing novel uses for eye-tracking for clinical decision support.

Keywords: Eye-tracking; Electronic medical record system; Information-seeking behavior

Lay summary

In the context of electronic medical record systems, eye-tracking is used extensively to explore visual and cognitive processes. We investigated a novel application of eye tracking to collect training data for machine learning-based clinical decision support. We developed and evaluated several methods for processing gaze points that were recorded using a low-cost eye tracking device. The eye-tracking methods and scripts that we developed offer a first step in developing novel uses for eye-tracking for clinical decision support.

1 Background and significance

Eye tracking is a sensor technology that detects where a person is looking and is widely used to record and study human visual and cognitive processes. With the advent of affordable eye tracking devices, there is growing interest in leveraging eye-tracking to provide input and to control computer applications for people with disabilities.

In the context of electronic medical record (EMR) systems, eye tracking has been used to study information search patterns [1], evaluate usability [2], understand clinical reasoning [3], measure time use [4], and quantify cognitive loads while performing tasks in the EMR system [5]. An intriguing application of eye-tracking in EMR systems is to enable clinical decision support applications. For example, we have used eye-tracking to collect training data for machine learning models of information-seeking behavior in an EMR system to identify and highlight data in the EMR that are likely to be relevant to the user [6, 7, 8]. The availability of inexpensive eye-tracking devices makes the broad deployment of eye-tracking enabled systems feasible.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.19.20215087; this version posted October 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Eye tracking devices record a sequence of gaze points (the onscreen locations where a person is looking) with a regular sampling rate (e.g., a sampling rate of 60 Hz produces a sequence of 60 gaze points per second). Based on gaze points, eye movements are broadly classified into fixations and saccades. A series of gaze points that cluster closely in space constitutes a fixation. Fixations are periods when the eyes are more or less still and are focused on a singular region of interest. Gaze points that occur between fixations and are widely dispersed in space constitute saccades. Saccades are fast movements of the eyes from one region of interest to another.

Visual attention and cognitive processing are assumed to occur only during fixations and thus methods have been developed to identify fixation durations and locations from a raw sequence of gaze points. These methods are often used to measure the duration of visual attention directed to a region of interest and to infer that if the duration is above a certain threshold then the information in that region has been cognitively processed. Simple gaze point methods count the number of gaze points that fall in each region of interest during the study period. More sophisticated methods count the number of gaze points that fall in each region of interest during fixations. The proportion of all gaze points, or gaze points during fixations, that are directed toward one region relative to others indicates the extent of the attention paid to that region.

For clinical decision support, eye tracking can be effective in several ways: (a) in the collection of training data for machine learning models of information-seeking behavior, (b) in the generation of personalized alerts based on what the user has seen and cognitively processed –thus decreasing alert fatigue–, and (c) in investigating visual and cognitive processes –the traditional application of eye tracking– in the context of using decision support.

2 Objective

We investigated the suitability of eye tracking for collecting training data for machine learning-based clinical decision support. Specifically, we record the information-seeking behavior of physicians while they use electronic medical records in the context of a specific clinical task. We evaluated the accuracy of several methods for processing gaze points that were recorded using a low-cost eye tracking device.

3 Methods

In this section, we introduce the learning electronic medical record (LEMR) system and the eye-tracking device. We then describe the data we collected and the gaze point processing methods we evaluated.

3.1 The learning electronic medical record (LEMR) system

We developed the LEMR clinical information display to use machine learning models to prioritize patient information that is relevant in the context of a clinical task [9, 10]. For example, the task of summarizing a patient's clinical status at morning rounds in the intensive care unit (ICU) typically requires the presenter to first review and identify relevant patient information in the EMR system. This "pre-rounding" task is time-consuming and laborious, and the LEMR system was designed to assist in this task.

The LEMR interface (see Figure 1) enables the collection of physician information-seeking behavior in two ways: (a) manual selection, when the user annotates relevant information by clicking on checkboxes, and (b) gaze-derived, when an eye-tracking device records gaze points while the user is reviewing the patient's record. Both methods produce a list of clinical data for a patient that were considered relevant by the user and are used to train the LEMR machine learning models. These models are then applied to new patients to identify and highlight relevant information that the user is predicted to view.

3.2 Eye-tracking device

We used Tobii EyeX, an inexpensive portable eye-tracking device and software package, which is primarily marketed for developing computer gaming and virtual reality applications [11]. The

medRxiv preprint doi: https://doi.org/10.1101/2020.10.19.20215087; this version posted October 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1: A computer monitor displaying the LEMR system with an eye-tracking device mounted at the bottom. From left to right, the system displays patient information on vital signs, ventilator settings, and intake and output, medication administrations, laboratory test results, and free-text notes and reports.

hardware component, the Tobii EyeX Controller, is mounted at the bottom edge of a computer monitor and samples eye gaze point coordinates at approximately 60 Hz. The software component, the Tobii EyeX Software Development Kit, records and outputs x-axis and y-axis gaze point coordinates for each eye.

3.3 Data collection

We randomly selected 178 ICU patients who were admitted to an ICU between June 2010 and May 2012 at the University of Pittsburgh Medical Center, and had a diagnosis of either acute kidney failure (AKF; ICD-9 584.9 or 584.5; 93) or acute respiratory failure (ARF; ICD-9 518.81; 85). Eleven critical care physicians reviewed the EMRs of the selected patients in the LEMR system.

Each patient record was loaded into the LEMR system as shown in Figure 1. The physicians reviewed a record by completing three tasks. In the familiarization task, the physician was shown patient information from the time of ICU admission up to 8:00 AM on a random ICU day between day two and the day before discharge from the ICU (inclusive). The physician was asked to understand the clinical course of the patient. In the preparation task, the physician was shown an additional 24 hours of patient information and was instructed to review the information for the task of summarizing the patient record for presentation at morning rounds. During this task, eye-tracking was used to record the physician's gaze points as they reviewed the latest patient information. In the selection task, the physician indicated which patient information was relevant by clicking on checkboxes.

From gaze points recorded during the preparation task, we estimated which patient information was considered to be relevant by the physician; from the manual checkbox selections recorded during the selection task, we derived a reference standard of patient information that was specified as relevant by the same physician. We evaluated the performance of the gaze-derived relevant patient information against the manual selection reference standard.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.19.20215087; this version posted October 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

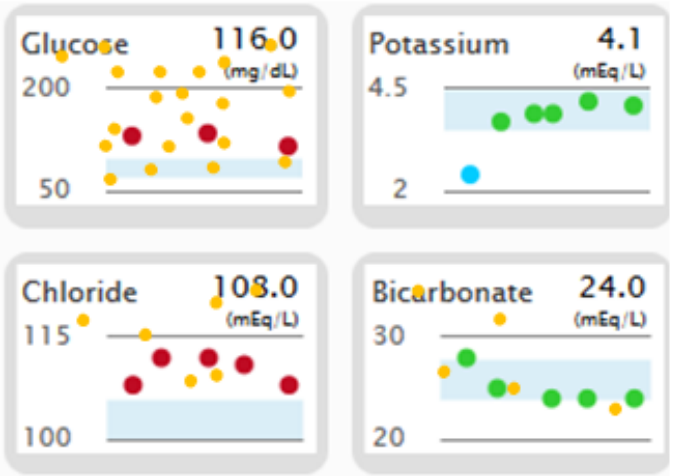
3.4 Eye-tracking methods

We evaluated four eye-tracking methods to infer relevant patient information from the gaze points recorded from the 11 physician reviewers. We selected two fixation identification algorithms that are widely used and include the dispersion-threshold identification (I-DT) and area-of-interest identification (I-AOI) methods. In addition, we developed two simple gaze point algorithms called the gaze point (GP) and distributed gaze point (DGP) methods.

The I-DT method identifies fixations from a series of gaze points collected during a study period. It categorizes a group of consecutive gaze points as a fixation if they are within a maximum distance of one another and within a period of time exceeding some minimum length (generally 100 milliseconds). Thus, the I-DT requires two input parameters, the dispersion threshold and the duration threshold.

In contrast to the I-DT that identifies fixations in any location in the visual field, the I-AOI identifies fixations that occur within one or more specified target areas. Typically, a target area is a rectangular region of interest that represents a unit of information in the visual field. Similar to I-DT, I-AOI utilizes a duration threshold to distinguish fixations in the target area from saccades crossing that area. Thus, the I-AOI requires a duration threshold as an input parameter.

We developed two simple and efficient gaze point methods that do not rely on fixation identification. The GP method maps all gaze points to specified target areas without classifying the points as part of a fixation or a saccade. In our application, a target area is a rectangular region that displays one type of patient information, e.g., a laboratory test time series or a medication order time series. A higher proportion of gaze points that map to a target area (compared to all gaze points that are recorded for the preparation task) indicate that more visual attention has been directed there (see Figure 2). Thus, the GP method does not require inputs like the dispersion threshold and the duration threshold.



Target area	Value
Glucose	20
Potassium	0
Chloride	6
Bicarbonate	5

Figure 2: The panel on the left shows four illustrative laboratory test results that are four target areas. The orange points denote gaze points. The GP method computes the value assigned to each target area as the number of gaze points that fall within each target area. For example, the value assigned by the GP method to the bicarbonate laboratory test target area is 5 as shown in the right had panel.

The DGP method is a probabilistic refinement of the GP method in which each gaze point contributes to several adjacent target areas in probability fashion (see Figure 3). The fractional contribution of a gaze point to a target area is equal to the density of a bivariate normal distribution. The means of the distribution are located at the center of the gaze point and the variances are derived from the average error of the eye-tracking device in the horizontal and vertical directions that we estimated in a prior study [7].

In our experiments, we explored a range of values for the input parameters for I-DT and I-AOI. For the I-DT algorithm, dispersion thresholds values were picked from values [50, 80, 100, 150, 200 pixels] and the duration threshold was selected from values [10, 20, 30, 40 data points]. For I-AOI, values [10, 20, 30, 40, 50, 100, 150, 200 data points] were assigned to the duration threshold. Since

medRxiv preprint doi: https://doi.org/10.1101/2020.10.19.20215087; this version posted October 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

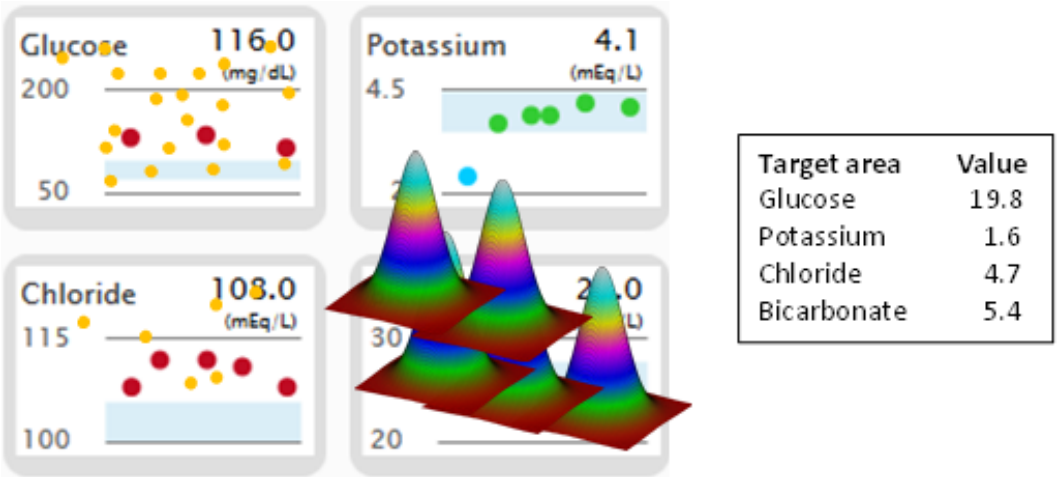


Figure 3: The panel on the left shows the same four illustrative laboratory test results with gaze points as in Figure 2. The DGP method computes the fractional contribution of a gaze point to a target area as equal to the density of a bivariate normal distribution. For illustration, five bivariate normal distributions are shown overlaid on five gaze points on the bicarbonate laboratory test target area. For example, the value assigned by the DGP method to the bicarbonate laboratory test target area is 5.4 as shown in the right had panel.

the sampling frequency is 60 Hz, the interval between consecutive data points is 16.7 milliseconds; thus 10 data points for the duration threshold translates to a duration of 167 milliseconds.

We evaluated the performance of the methods on several measures including accuracy, precision, recall. Accuracy is the percent of target areas, both relevant and irrelevant, that were correctly identified by an eye-tracking method. Precision is the fraction of relevant target areas among all target areas identified as relevant and recall is the fraction of relevant target areas among all target areas that are relevant.

4 Results

Overall, GP had the highest accuracy and precision of 69% and 53% and DGP has the highest recall at 48% (see Table 1). The two fixation methods, I-DT and I-AOI, had lower values on precision and recall and slightly lower accuracy than GP and DGP.

The scripts for the four eye-tracking methods with accompanying documentation are freely available on GitHub at https://github.com/ajk77/EyeBrowserPy.

Table 1: Accuracy, precision, and recall values with standard error of four eye-tracking methods. The highest values for each performance measure are in bold font.

Method	Accuracy %	Precision %	Recall %
I-DT	67 ± 0.03	46 ± 0.05	26 ± 0.23
I-AOI	68 ± 0.03	49 ± 0.09	31 ± 0.26
GP	$\mathbf{69 \pm 0.04}$	$\mathbf{53 \pm 0.10}$	38 ± 0.28
DGP	67 ± 0.05	50 ± 0.08	$\mathbf{48 \pm 0.25}$

5 Discussion

We evaluated several eye-tracking methods to collect data for training machine learning models that underlie a clinical decision support application. Specifically, we evaluated four methods including

medRxiv preprint doi: https://doi.org/10.1101/2020.10.19.20215087; this version posted October 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

two that we developed for processing gaze points collected while a physician was using a custom-built EMR system in the clinical context of pre-rounding in the ICU. Our results support the use of a low-cost eye tracking device and relatively simple methods for processing gaze points to record the information-seeking behavior of physicians. Next, we applied our methods to collect training data and derived machine learning models to predict the relevance of laboratory tests, medications, vital sign measurements, ventilator settings and fluid intake and output for the task of pre-rounding in the ICU. The models that were derived using eye-tracking data performed as well as models that were derived using manual annotations made by physicians to indicate relevant patient information using the LEMR system [6].

The eye tracking device, Tobii EyeX, was not developed explicitly for research applications and the device's modest temporal and spatial resolution, moderate precision, and low sampling frequency may be inadequate for many research applications, especially those that investigate human visual and cognitive processes [12]. However, it is adequate for applications that only require more than monitoring of simple eye movements. In an earlier evaluation, we found that the EyeX is adequate for collection of information-seeking behavior data for the LEMR system, and its accuracy—for our task—was comparable to that of a more expensive eye-tracking device [7].

There are several limitations to our study. One is the use of a custom-built LEMR system whose interface is significantly different from the vendor EMR systems currently used in clinical care. Further studies are needed to assess the adequacy of low-cost eye tracking devices when used in conjunction with vendor EMR systems. A second limitation is that the eye-tracking device currently for sale from the same company is a newer model (Tobii Eye Tracker 5) than the eye tracking device used in this study (Tobii EyeX), and has a higher sampling frequency. A third limitation is that the eye tracking device used in this study could not track head movements, which restricts its use to studies in which head movement is restrained. However, this limitation may be mitigated with newer devices like Tobii Eye Tracker 5 that are capable of tracking both head and eye movements offer the ability to robustly estimate the coordinates of eye-gaze even if the position of the head changes [13].

6 Conclusion

Eye-tracking provides an automated and unobtrusive method to capture the information-seeking behavior of physicians. Data obtained from an inexpensive eye-tracking device are suitable for building machine learning models that identify and highlight relevant patient information. The eye-tracking methods and scripts that we developed offer a first step in developing novel uses of eye-tracking in clinical decision support.

7 Acknowledgments

The research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health under award numbers R01 LM012095 and T15 LM007059. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors thank Ms. Yu Zhang for assistance with data preprocessing.

References

[1] Julie W Doberne, Ze He, Vishnu Mohan, Jeffrey A Gold, Jenna Marquard, and Michael F Chiang. Using high-fidelity simulation and eye tracking to characterize ehr workflow patterns among hospital physicians. In *AMIA Annual Symposium Proceedings*, volume 2015, page 1881. American Medical Informatics Association, 2015.

[2] Onur Asan and Yushi Yang. Using eye trackers for usability evaluation of health information technology: a systematic literature review. *JMIR Human Factors*, 2(1):e5, 2015.

[3] Katherine S Blondon, Rolf Wipfli, and Christian Lovis. Use of eye-tracking technology in clinical reasoning: a systematic review. In *MIE*, pages 90–94, 2015.

[4] Steven Rick, Alan Calvitti, Zia Agha, and Nadir Weibel. Eyes on the clinic: accelerating meaningful interface analysis through unobtrusive eye tracking. In *2015 9th International*

medRxiv preprint doi: https://doi.org/10.1101/2020.10.19.20215087; this version posted October 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), pages 213–216. IEEE, 2015.

[5] A Calvitti, N Weibel, L Liu, H Hochheiser, K Zheng, C Weir, S Ashfaq, S Rick, Z Agha, B Gray, et al. Can eye tracking and electronic health record mouse activity tell us when clinicians are overloaded. *Hum Factors Quart. Hum Factors Q*, 8:1, 2014.

[6] Andrew J King, Gregory F Cooper, Gilles Clermont, Harry Hochheiser, Milos Hauskrecht, Dean F Sittig, and Shyam Visweswaran. Leveraging eye tracking to prioritize relevant medical record data: Comparative machine learning study. *Journal of Medical Internet Research*, 22(4):e15876, 2020.

[7] Andrew J King, Harry Hochheiser, Shyam Visweswaran, Gilles Clermont, and Gregory F Cooper. Eye-tracking for clinical decision support: a method to capture automatically what physicians are viewing in the emr. *AMIA Summits on Translational Science Proceedings*, 2017:512, 2017.

[8] Mohammadamin Tajgardoon, Gregory F Cooper, Andrew J King, Gilles Clermont, Harry Hochheiser, Milos Hauskrecht, Dean F Sittig, and Shyam Visweswaran. Modeling physician variability to prioritize relevant medical record information. *medRxiv*, 2020.

[9] Andrew J King, Gregory F Cooper, Gilles Clermont, Harry Hochheiser, Milos Hauskrecht, Dean F Sittig, and Shyam Visweswaran. Using machine learning to selectively highlight patient information. *Journal of Biomedical Informatics*, 100:103327, 2019.

[10] Andrew J King, Gregory F Cooper, Harry Hochheiser, Gilles Clermont, Milos Hauskrecht, and Shyam Visweswaran. Using machine learning to predict the information seeking behavior of clinicians using an electronic medical record system. In *AMIA Annual Symposium Proceedings*, volume 2018, page 673. American Medical Informatics Association, 2018.

[11] *Tobii. An introduction to Tobii EyeX 2014*. `https://developer.tobii.com/an-introduction-to-the-tobii-eyex-sdk/`.

[12] Agostino Gibaldi, Mauricio Vanegas, Peter J Bex, and Guido Maiello. Evaluation of the tobii eyex eye tracking controller and matlab toolkit for research. *Behavior research methods*, 49(3):923–946, 2017.

[13] *Tobii. Tobii Eye Tracker 5 2016*. `https://gaming.tobii.com/product/eye-tracker-5/`.