MPHY0047 – Surgical Data Science

Coursework 1

Release Date:	26 January 2026
Submission Deadline:	16 February 2026
Estimated Coursework Return:	Three weeks after submission
Topics Covered:	Topics 1 – 3
Expected Time on Task:	8 hours

Guidance for Submissions: *Failure to follow this guidance might result in a penalty of up to 10% on your marks.*

I. Submit a single Word//PDF document with questions in ascending order. Explain in detail your reasoning for every mathematical step taken. Alternatively this can be integrated in a PYTHON notebook together with your source code.

II. Insert relevant output (e.g calculations, graphs or figures), and describe them in your document. All calculations, figures and tables must be labelled, showing relevant parameters and units.

III. You will need PYTHON coding to solve the questions. Include **all code** as separate PYTHON source files or as part of a PYTHON notebook that can be also used to produce the report of your submission. Remember to comment on your code, explaining your steps. The submitted source code will be take into account during assessment.

This coursework counts towards 25% of your final MPHY0047 grade and comprises five questions. Question 1 will make up 15% of your CW1 grade, Questions 2, 3, 4 will each make up 25% of your CW1 grade and Question 5 will make up 10% of your CW1 grade.

On Academic Integrity (Read more about it here)

Academic integrity means being transparent about our work.

- **Research:** You are encouraged to research books and the internet. You can also include and paraphrase any solution steps accessible in the literature and online content if you reference them.
- **Acknowledge others:** We are happy when you acknowledge someone else's work. You are encouraged to point out if you found inspiration or part of your answers in a book, article or teaching resource. Read more about how to reference someone else's work here and how to avoid plagiarism here.
- **Ask *good* questions to your peers:** These include but are not limited to questions like "What is the best mathematical method for this question?", "Should I review any books/materials/videos?", "Which PYTHON function did you use for this problem?", "How was the structure of your PYTHON code?".
- **Be *helpful* and *ethical* when answering questions from your peers:** These include "I think that it would be helpful to review X video, Y page of the slides/notes". "I found this good video online", "I used X PYTHON function, structured that way".
- **Do not share and do not copy:** We expect students **not to share and not to copy** assessment solutions or PYTHON code from their peers, even if partially.
- **Do not publish MPHY0047 assessment material:** We expect students **not to share** MPHY0047 assessment materials at external online forums, including tutoring or "homework" help websites.

Students found in misconduct can receive a 0 mark in that assessment component and have a record of misconduct in their UCL student register. In some extreme cases, academic misconduct will result in the *termination* of your student status at UCL.

Recent research focuses on recording and analysing the eye movements and points of gaze focus of surgeons while performing surgical tasks. Eye tracking for skills assessment has been used in [1], with anaesthesiologists of different level of expertise when performing an epidural block. Eye training significantly improves learning of knot tying compared to a traditional technical approach was shown in [2]. Another study found that expert neurosurgeons spend more time fixating on the region of interest before performing an action [3]. Blind spots, a key performance metric in real-time EGD were identified with the use of eye tracking [4]. From these studies it is established that eye tracking data can provide meaningful information about the operator's technical and cognitive skills and overall performance on surgical tasks. The locations of eye focus and spatial distribution of gaze fixations can be helpful for assessing the use of instruments and the sequence and quality in executing tasks.

You will work with a UCL eye-tracking dataset collected during an ophthalmic surgery skills course at the Royal College of Ophthalmologists. During the one-day course attendees (trainee ophthalmic surgeons) performed ophthalmic knot tying tasks while their eye movement was recorded with an eye tracking device.

Your main task is to structure performance measures derived by the eye tracking data and evaluate their ability to discriminate between novice and expert users.

Good luck !

References

[1] Capogna E., et al., "Novice and Expert Anesthesiologists' Eye-Tracking Metrics During Simulated Epidural Block: A Preliminary, Brief Observational Report", *Local Reg*; 2020; 2020(13): 105-109.

[2] Causer J., et al., "Quiet eye training improves surgical knot tying more than traditional technical training. A randomized, controlled study", *Am J Surg;* 2014; 208(2): 171-177.

[3] Chainey J., et al., "Eye-Hand Coordination of Neurosurgeons: Evidence of Action-Related Fixation in Microsuturing", *World Neurosurg.*; 2021; 155:196-202

[4] Lee A., et al., "Identification of gaze pattern and blind spots by upper gastrointestinal endoscopy using an eye-tracking technique", Surg. Endosc;2021.

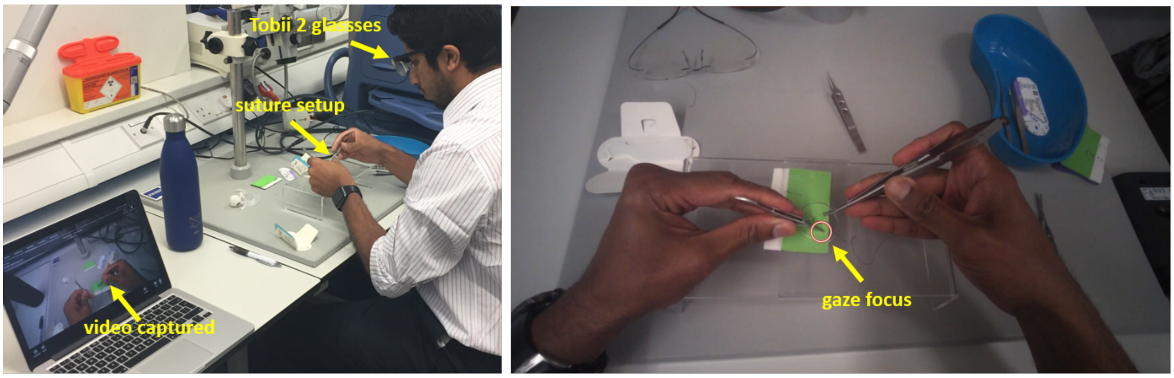
Problem :Assessing ophthalmic knot-tying performance using eye tracking data

Preliminaries

A study was conducted in September 2021 in the Royal College of Ophthalmologists (London, UK) during the training course “Introduction to Ophthalmic Surgery”. The course involves two practical sessions for practicing surgical knots under the guidance and evaluation of faculty members.

Prior to practicing, participants are educated through lectures, example videos, and live demonstrations. In the first session, attendees were instructed to practice the surgical knot on a standard suture training board, completing on average five knots.

During the practical sessions the Tobii Pro Glasses 2 (Stockholm, SWE) was used and deployed as shown in Fig. 1(a) for capturing the participants’ eye movements. The device uses near-infrared illumination to create reflection patterns on the cornea and pupil of the subject’s eye, while image sensors are used to capture images of the eyes with the reflection patterns as shown in Fig. 2. Image processing and a physiological 3D model of the eye are then used to estimate the point of gaze within the field of view (FoV) provided by a camera at the front side of the glasses. Time-stamped gaze focus points (pixel coordinates) are synchronized with the FoV video recording and overlayed (red circle) as shown in Fig.1(b).



(a) (b)

Figure 1: (a) Experimental setup, (b) Snapshot of a video showing the wearer’s field of view along with gaze focus point.

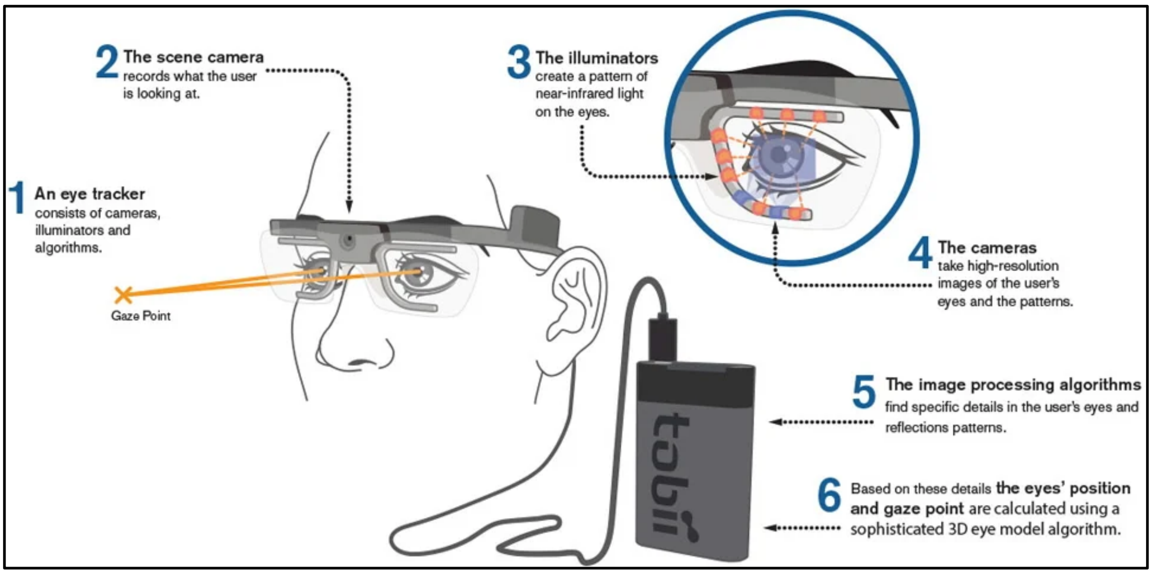


Figure 2: The Tobii Pro Glasses 2 and its operating principle

In total, **20** volunteers took part in the experiments, **11 novice trainees** (in the early stages of their residency programme) and **9 experts** (consultant ophthalmic surgeons and academic faculty members teaching the course). Wearing the "Tobii Pro Glasses 2", each participant was instructed to perform a single surgical knot on a training board. We recorded one video from each participant after they had a basic understanding of the activity, with calibration occurring before each recording.

Annotations

A surgical knot consists of two major subtasks, as shown in Fig. 3. These are **needle passing** and **knot tying**, where typically surgeons tie three knots to secure the suture. By reviewing the videos, we annotated the **start and end** moments of each task, as shown in Fig. 2. We then calculated **i)** the **total duration** of the entire suturing task, which commences with the needle passing and ends with suture cutting; **ii)** the **duration of the needle passing** subtask, as well as **iii)** the **duration of tying** the first knot. These time-based metrics will be further used for statistical analysis.

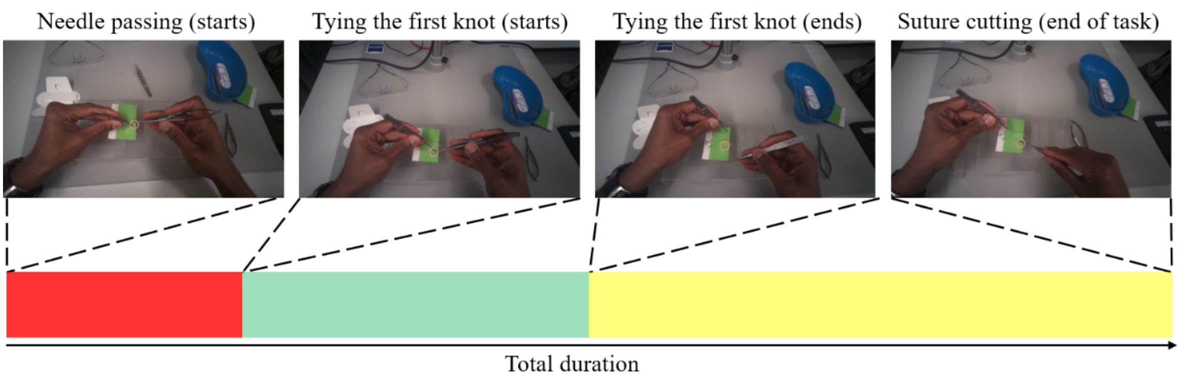


Figure 3: Annotation of a surgical suture task from our dataset.

Gaze tracking data allows us to investigate the spatial distribution of gaze samples and fixations during the suturing task. A **fixation** is detected if the velocity of directional shifts of the eye tracks is below 30 visual degrees per second. The distribution of **fixations** is illustrated as a color-coded heatmap plot (see Fig. 4) in which warmer colors signify a larger concentration of fixations observed on that point.

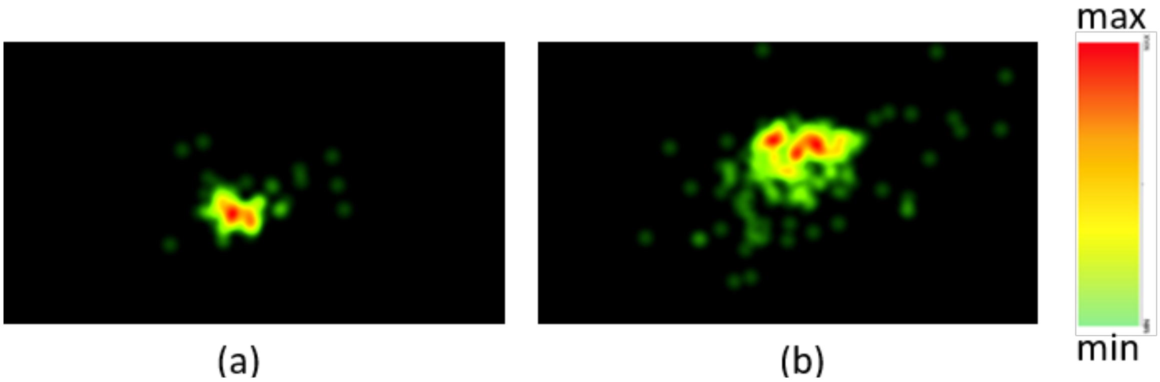


Figure 4. Example of the spatial distribution of fixations during knot tying (a) expert surgeon, (b) novice surgeon.

The task of surgical knot tying is composed of a sequence of subtasks. Considering the vocabulary in Table 1, the proper sequence for executing a knot tying task can be defined and used as reference. The eye tracking videos were annotated by our clinical partners and the sequence of the subtasks that each participant followed was noted. This allows us to count the deviations from the ideal sequence as an additional **error metric** of performance.

Table 1: Vocabulary of surgical subtasks in the knot tying task

surgical gesture	**ID**
pushing the needle through the tissue	S1
disentangling threads	S2
picking the appropriate instruments	S3
knot tying	S4
suture cutting	S5
Ideal Sequence	
1-3-4-4-(4)-5	

Error annotation took place with the following rules:

- The ideal sequence allows two (min) to three (max), knot tying gestures (S4).
- The existence of S2 (disentangling threads) in a sequence constitutes an error (+1)
- The absence of S3 (picking appropriate instruments), the absence of S5 “suture cutting”, or having less than two S4 (knot tying) gestures constitutes an error (each adds **+1** to the error metric)

Examples:

Compared to the ideal sequence, a sequence of “**1-2-4-5**” would constitute to an error metric of “3”. **S2** is present **(+1)**, **S3** is absent **(+1)**, only one **S4 (+1)** → **error metric** = 3

Compared to the ideal sequence, a sequence of “**1-3-4-4-4-5**” would constitute to an error metric of “0”. **S2** is absent, three **S4** and **S3/S5** present at the correct sequence → **error metric** = 0

Dataset

Every participant has its own unique ID number. The available dataset, that you will work with, contains for each participant (9 experts, 11 novices):

1) The three time parameters (sec): i) **total duration** of the entire task; **ii) duration of the needle passing** subtask, and **iii) duration of first knot tying** subtask

2) The annotated **subtask sequences** for each participant which allows for **error analysis**

3) Greyscale **fixation heatmaps**.

Data for 1) are provided in .csv files (time_experts.csv, time_novices.csv. These can be loaded directly to a PYTHON script using the “pandas” library.

https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html

Data for 2) are provided in .xlsx file (error_data.xlsx). These can be loaded directly to a PYTHON script using the “pandas” library.

https://pandas.pydata.org/docs/reference/api/pandas.read_excel.html

Data for 3) are given as greyscale bitmap images (.png). These can be loaded to PYTHON script using “opencv” or the “PIL” libraries.

https://www.geeksforgeeks.org/python-opencv-cv2-imread-method/

https://www.geeksforgeeks.org/python-pil-image-open-method/

Question 1 [15 marks]

Consider the three time parameters for the two groups (9 experts, 11 novices). Calculate analytically (without using any available Python library) the: i) mean, ii) median, iii) variance, iv) skewness and v) kurtosis and summarize them in a table.

Draw histograms and boxplots for each time parameter (for both groups). (12 figures in total). Identify possible outliers and together with the answer in Q1 comment on which of the three time parameters is more robust (i.e. presents the smaller dispersion).

Question 2 [25 marks]

Considering the three time parameters and the grouping of experts and novices, perform statistical testing and identify any parameters that present statistical significance between the two groups for the knot tying task. **Justify and explain** your selection of the method (test) chosen to evaluate statistical significance and **provide an interpretation** of the results of every test (parameter).

Question 3 [25 marks]

Following the error annotation process described earlier perform error analysis and **count the number of errors (error metric)** for each participant. Perform statistical testing and comment if the error metric presents statistical significance between the two groups for the knot tying task. **Justify and explain** your selection of the method (test) chosen to evaluate statistical significance and **provide an interpretation** of the results.

Question 4 [25 marks]

To quantify the spatial distribution of fixations during the knot tying phase, we introduce a metric called **fixation sparsity**, as the **ratio** of the **total number** of **non-white pixels** (locations that we obtain fixations) in the heatmap divided by the total **number of pixels** image resolution 1920x1080.

Perform statistical testing and comment if the fixation sparsity metric presents statistical significance between the two groups for the knot tying task. **Justify and explain** your selection of the method (test) chosen to evaluate statistical significance and **provide an interpretation** of the results.

Question 5 [10 marks]

“Which of these parameters appears to be able to better discriminate expert and novice ophthalmic surgeons in the task of knot tying ?”. According to your outcomes in Questions 2–4, provide a ranking of the five metrics (3 time parameters, 1 error metric, 1 fixation sparsity metric) on their potential/ability to capture the differences between the two groups of surgical experience (experts, novices). Justify your answer.