Intelligent Health Systems – From Technology to Data and Knowledge
E. Andrikopoulou et al. (Eds.)

doi:10.3233/SHTI250330

Formative Usability Testing of Artificial Intelligence in Pathology: The Challenge of Assessing Acceptability

Natasha ALVARADO[a,b,1], Derek MAGEE[c], Darren TREANOR[c,d,e], Peter ELLERY [e], Helen MILLOT[f], and Rebecca RANDELL[a,b]
[a] *Centre for Digital Innovation in Health and Social Care, University of Bradford, Bradford, UK*
[b] *Wolfson Centre for Applied Health Research, Bradford, UK*
[c] *University of Leeds, Leeds, UK*
[d] *Linköping University Linköping, Sweden*
[e] *Leeds Teaching Hospital NHS Trust, Leeds, UK*
[f] *Faculty of Health Studies, University of Bradford, Bradford, UK*
ORCiD ID: Natasha Alvarado https://orcid.org/ 0000-0001-9422-4483

Abstract. Digital Pathology has provided a platform to use Artificial Intelligence (AI) to assist pathologists with diagnosis and reporting. An AI tool is being developed that analyzes digital Hematoxylin and Eosin (stained tissue) images associated with a skin cancer case and pre-populates a report with required parameters. The aim of this AI pathology assistant is to save pathologist time and increase reporting efficiency. This study assessed ease of use and acceptability of a first iteration of the AI tool. Twelve pathologists were recruited across seven UK hospitals and participated in a think-aloud evaluation, completing a pathology report using the novel tool, after which they participated in a brief interview. The think-aloud identified several issues that can inform tool development to improve ease of use. AI performance (inaccuracy populating report items) constrained assessment of tool acceptability and added tasks to the reporting process. This finding emphasizes the importance of AI accuracy (1) for assessing if and how such tools can be integrated into clinician's workflow to increase efficiency, and (2) for cultivating clinician trust in tool performance to support adoption in practice.

Keywords. Artificial Intelligence, pathology reporting, usability testing

1. Introduction

Traditionally pathology reports for primary diagnosis have been produced using glass slides and a microscope. The introduction of Digital Pathology (DP) has transformed the reporting process e.g., enabling remote diagnostic reporting and facilitating second and expert opinion reporting [1]. The roll out of DP across hospitals is providing increasing numbers of pathologists with the choice of viewing case images *via* microscope or digitally, while populating the clinical report using their usual systems e.g., Word document/ dictation devices. DP has also provided a platform to use Artificial Intelligence (AI) with potential to further transform pathologist workflow. An innovative

[1] Corresponding Author: Natasha Alvarado; E-mail: n.alvarado@bradford.ac.uk.

AI tool – named the AI pathology assistant - is being developed to assist pathologists report Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC), the most common types of skin cancer cases [2]. The aim is to produce the same report as would be produced manually by a clinician but to save time and increase workflow efficiency. Drawing on the principles of Human Centered Design (HCD) - an iterative, collaborative approach to technology development [3] - the objectives of this study were to inform development of the AI pathology assistant in ways that better meet user needs and preferences by assessing ease of use and acceptability of the first iteration of the tool.

2. The AI Pathology Assistant

The AI pathology assistant was developed with input from pathologists in one hospital, who were experienced using DP. The tool integrates an image and report viewer that are used simultaneously, ideally using two screens. The report viewer displays the minimum dataset for reporting BCC or SCC cases and the image viewer displays the digital Hematoxylin and Eosin (H&E) images. The AI analyzes the H&E images, pre-populates dataset items with required parameters, and overlays the digital images with annotations e.g., measurement lines and circled areas that indicate nerves or lymph nodes for checking for Perineural Invasion (PNI) and Lymphovascular Invasion (LI) – invasion of cancer in the space surrounding the nerve of lymph nodes respectively. Pathologists review report items, cross-reference them with the image annotations, and then confirm or reject and amend them as necessary using the tool functions. The image viewer has functions to manually measure tumor regions to amend measurement items and all items in the report viewer can be over-typed by pathologists.

3. Methods

Pathologists (consultant and trainees) were recruited from seven UK hospitals. The aim was to include specialists in skin cancer reporting and generalists who report a range of cases. Methods included (1) **think-aloud;** this is an inspection method that identifies issues experienced when participants use technology to complete tasks [4]. Pathologists were asked to use the AI pathology assistant to report an SCC case consisting of five H&E images and to articulate their thought process as they did so. Before completing the report, the researcher demonstrated how to use the tool functions. The think-aloud was video recorded. (2) **Short interviews**: Short semi-structured interviews were conducted post-think aloud to probe for further detail about user experiences. Interviews were digitally recorded.

3.1. Data Analysis

Data from the think-aloud were extracted using a proforma to capture experiences including any challenges experienced by item. Based on previous experience of usability testing [5] interview data were categorized using: (1) a task-related item, capturing ease of use when completing reporting tasks, (2) a visualization-related item, capturing feedback on the choice of visualizations, (3) a data-related category, capturing comments relating to the data provided by the AI, and (4) an 'other' category, capturing any further comments including benefits and suggestions for improvements.

3.2. Ethical Considerations

Following a protocol reviewed favorably by the West Midlands – Edgbaston Research Ethics Committee Ref: 23/WM/0159 and the Humanities, Social and Health Sciences Research Ethics Panel at the University of Bradford, Ref: E2223/00047 informed consent was taken from participants to conduct the think-aloud and interview.

4. Results

In total, 12 pathologists were recruited, see Table 1 for recruitment summary by role.

Table 1 Recruitment summary by role

Role	No. recruited
Consultant/ skin specialist	1
Consultant/ Generalist	8
Trainee	3
Total	12

4.1. Ease of Use

Pathologists used different approaches to complete the report; some went through the report item by item, referring to the image viewer to check pre-populated items as they did so, others began by viewing case images first to identify tumor regions and then navigating between the report and the image viewer to check and complete report items.

Frequently pathologists opted to look for areas of interest (PNI and LI) manually rather than using AI annotations (discussed further in 4.2. Acceptability) and commented that the number of annotations were distracting to the task. Participants could reduce the number of annotations displayed, but often needed reminding of this function. Participants also required reminders of how to draw and edit lines in the image viewer, suggesting more time was needed for them to familiarize themselves with tool functions. Pathologists experienced with DP also discussed image viewing systems they used in practice with features - not available in the AI pathology assistant - that facilitated the reporting process e.g., the ability to mark areas of interest whilst scanning images, the ability to rotate images to facilitate drawing line measurements and the ability to view more than one image at the same time.

Commonly observed challenges using the report viewer included spontaneous cursor movement that disrupted pathologists when they manually typed information, manually typed items that reverted to the original AI entry, and pop-up information that displayed repeatedly and required a response before the pathologist could proceed. Some participants commented that tool features were not as intuitive as they could be e.g., referring to clicking an 'X' to confirm an AI populated item, a pathologist commented, '*if you see a cross you think it means get rid of or delete rather than confirm*'. Additionally, pathologists were only able to draw lines when certain items were selected in the report viewer i.e., the image viewer functions changed according to the report item selected; this restricted how pathologists completed some tasks.

4.2. Acceptability

All study participants currently used separate image viewing and reporting systems and integration of these systems was considered acceptable and facilitative to the reporting process. The benefit of pre-populated items was also noted; a pathologist commented:

> *Even with digital, you have to actually put down the actual measurements and […] there are certain things that I have to delete, add, delete, add continuously […] but this one is actually what I want to put in the report comes up in the thing..* **Pathologist 8, Post Think-Aloud interview**

At this site, the pathologist used a Word document (proforma) to complete reports that contained information that needed deleting and adding to as appropriate for each case. While potential benefits of the AI to pathologist workflow were recognized, the think-aloud highlighted that assessing acceptability of the current tool was constrained by AI performance i.e., accuracy measuring tumor margins and identifying areas of interest (PNI and LI). A pathologist discussed:

> *The AI has highlighted lots of things that wouldn't make me look twice. So, yeah, maybe it sort of almost slightly slows you down because if you were to actually just click through all the things that it had highlighted as possible PNI, you'd, be looking at things that you wouldn't have bothered looking at before.* **Pathologist 5, Post Think-Aloud Interview**

In this example, the AI had reported no PNI in the case; the pathologist explained that they would want to check this decision '*closely'*. However, many of the AI annotations to support identification of PNI did not highlight nerves, therefore pathologists were directed to areas not relevant to the task.

The process of amending AI measurements was also discussed; pathologists quickly identified that the Peripheral Margin (side edges of the biopsy) measurement was incorrect (it was in an image where there was no cancer). Therefore, they sought to draw the required line measurement manually using the functions in the image viewer, but a pathologist explained:

> *I wouldn't normally extend a line* [a border line] *[…] it's a step for the sake of the algorithm, not for the sake of me and the speed of my report.* **Pathologist 3, Post Think-Aloud Interview**

The pathologist commented that in current practice they 'eyeball' case images and draw a line based on their judgement. Therefore, editing borders to measure margins and drawing multiple lines are not necessary. The process of checking and then amending pre-populated items extended the time it took to complete reporting tasks, especially for pathologists experienced using DP systems, a pathologist commented:

> *I don't anticipate this case taking more than five minutes actually, even with perineural invasion but all those lines, inability to rotate the image and then that pre-filled thing where you've to confirm, refuse, revise… I think it just made a very simple task for an experienced or an expert pathologist and making it difficult* **Pathologist 4, Post Think-Aloud interview**

5. Discussion

The think-aloud technique identified several issues that can inform tool development and future usability testing of AI technologies. Issues that impacted ease of use such as lack of familiarity with the image viewer functions and unwanted cursor movement in the report viewer can be used to improve the tool interface and development of training materials to support pathologist use of tool functions.

While integration of the report and image viewer was acceptable, and desirable, to participants, assessing acceptability of the AI itself was constrained by its performance measuring required margins and identifying areas of interest (PNI, LI). Participant

feedback focused on AI performance in the task at hand, rather than the utility of annotated images and pre-populated dataset items more generally and how they could be integrated to benefit pathologists' workflow.

The findings suggest that AI performance influences assessment of tool acceptability and points to potential issues with adoption of such tools - clinicians need to have confidence that AI can perform required tasks correctly to build trust in its performance before using it to impact their workflow. This resonates with a growing body of literature that discusses trust as a key mechanism for AI adoption [6-8].

6. Conclusions

AI tools offer potential to improve workflow efficiency in the context of pathology reporting and other areas of clinical decision making. Usability testing of such tools *via* the think-aloud technique, as part of UCD, help pinpoint interface issues that can improve ease of use. To optimize learning about acceptability of AI tools in the clinical workflow, consideration should be given to AI performance as this has the potential to constrain clinicians' assessment of this area.

Acknowledgements

We would like to thank Chloe Kirby, Gillian Mathews, Ruben Mujica-Mota and the remaining project team for their support and guidance throughout this study and the pathologists who generously gave up their time to participate.

References

[1] Rakha EA, Toss M, Shiino S, Gamble P, Jaroensri R, Mermel CH, Chen PHC. Current and future applications of artificial intelligence in pathology: a clinical perspective. J Clin Pathol. 2021;74(7):409–14.
[2] Firnhaber JM. Basal cell and cutaneous squamous cell carcinomas: diagnosis and treatment. Am Fam Physician. 2020;102(6):339–46.
[3] Giacomin J. What is human centred design? Des J. 2014;17(4):606–23.
[4] Van Someren M, Barnard YF, Sandberg J. The think aloud method: a practical approach to modelling cognitive. London: Academic Press; 1994.
[5] Elshehaly M, Randell R, Brehmer M, McVey L, Alvarado N, Gale CP, Ruddle RA. QualDash: adaptable generation of visualisation dashboards for healthcare quality improvement. IEEE Trans Vis Comput Graph. 2021;27(2):689–99.
[6] Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res. 2020;22(6):e15154.
[7] Rojas JC, Teran M, Umscheid CA. Clinician trust in artificial intelligence: what is known and how trust can be facilitated. Crit Care Clin. 2023;39(4):769–82.
[8] Shamszare H, Choudhury A. Clinicians' perceptions of artificial intelligence: focus on workload, risk, trust, clinical decision making, and clinical integration. Healthcare (Basel). 2023.