Case Study on a User-Centered Annotation Tool for Digital Pathology Education

Dariia Drobna*
Supervised by: Ing. Miroslav Laco, PhD.†

Faculty of Informatics and Information Technologies,
Slovak University of Technology Bratislava / Slovak Republic

Abstract

Digital pathology depends on creating structured annotations on whole-slide images (WSIs) using specialized software with medical image annotation tools. Existing solutions vary from open-source to proprietary, offering a wide range of annotation creation tools, integrated deep learning pipelines, and scripting options. In a medical education setting, the clinical and research tools are often too complex. In this paper, we propose a modification to the Human-Centered Design (HCD) methodology, involving the concepts of user experience (UX) design, usability, and user testing. The proposed modification focuses on user interface (UI) design in the specific medical education context and was applied during a case study to design a prototype for integrating a medical image annotation module into an existing educational tool for digital pathology. The existing tool already enables users to study and browse course materials and simulate microscopy to view digitized histological slides in WSI format. User research, such as contextual inquiry, surveys, interviews, and usability testing, was conducted to formalize users' motivations, objectives, and pain points. Based on these findings, a functional high-fidelity prototype was designed and user-tested by five participants using standard usability testing procedures. Quantitative results of the testing are combined with quantitative metrics to identify and discuss usability issues, thereby enhancing the proposed annotation tool design in subsequent iterations.

Keywords: User Experience, Human-Centered Design, Usability, User Interface, Medical Image Annotation Tools

1 Introduction

In digital pathology, annotations are created in order to differentiate and label morphological structures in tissue samples. Accurate annotations determine the clinical interpretation of the sample and, in an educational setting, provide a necessary foundation for developing appropriate expertise [19]. Nowadays, pathologists can leverage digital pathology solutions, where it is possible to create structured annotations on whole-slide images (WSIs) – digitized slides in high resolution. The variety of tools and functionalities, such as machine learning pipelines and scripting, implemented in the vast majority of solutions can prove overwhelming to end-users, especially pathology students. This paper explores existing systems in the domain of digital pathology [18, 8, 12], with a focus on the image annotation creation process. While numerous digital pathology applications already offer annotation capabilities, they are primarily designed for covering the needs of clinical practice or computational research. Transfer of these annotation workflows to the educational setting is not feasible since the end-users and their motivations, together with the user scenarios, differ considerably: students, unlike experienced pathologists, require simplified interactions with a reduced cognitive overhead to focus on the subject matter rather than on the software itself. Thus, the main objective is to integrate a medical image education module into the existing histopathology education platform, HIPA [18] through a human-centered design (HCD) process. HIPA is a web-based application developed for digital pathology education based on the domain expertise from LF UK to support access to study materials and facilitate studying using digitized slides rather than physical microscope setups, in a way that is more intuitive compared to existing histopathology software. By grounding the design in HCD principles, usability heuristics, and best practices in user interface (UI) design [11], we aim to achieve greater user satisfaction and utility compared to reusing existing tools not originally intended for education. The data obtained from user research was analyzed and used to iteratively develop prototypes of the annotation module, informed by evaluations and testing of similar solutions described in Section 2.

2 Related work

In this section, we explore state-of-the-art concepts regarding medical image annotation principles and current existing solutions incorporating medical image annotation modules in digital pathology. This state-of-the-art forms a foundation for our research method proposal to design a

*xdrobnad1@stuba.sk
†miroslav.laco@stuba.sk

medical image annotation module for the HIPA education tool. Consistent annotation of WSI is important for capturing changes in tissue samples. Thus, following the annotation guidelines, such as specifying ROIs using precise annotation tools and magnification, together with providing classification using conventional labeling, is important for yielding accurate results [19].

The findings from an international survey on annotation practices conducted by Montezuma, Oliveira et al. [10] under the European Society of Digital and Integrative Pathology (ESDIP), with 137 respondents, of whom 23.4% were medical students and 68.6% of the respondents were based in Europe, indicate that Open-source software QuPath[1] is one of the most widely used medical image annotation software programs, scoring 59.3%; the second most used software is Leica's Biosystems ImageScope[2], which is proprietary freeware with no active development and is suited for research purposes only. 97.8% of users prefer using a mouse in order to create annotations. Half of the participants did not use automation for annotations, and only 11% of the participants used QuPath-based algorithms. The review of Korzynska et al. comparing multiple annotation tools highlights the need for an intuitive data extraction workflow [9]. Medical image annotation software, such as RIL-Contour [12], was designed with deep learning integration, enabling the visualization of descriptive statistics computed over ROIs. Therefore, narrowing the variety of options for staining, data formats, and input methods keeps the education tool focused on domain problems rather than technical complexity. QuPath has been widely adopted by professionals and the research community as an effective tool for such tasks as cell detection and object classification [1]. Based on this fact, we should consider a UX audit of the tool to extract the most useful features, as there is no evidence of such information so far.

To place medical image annotation within an educational context, user scenarios should be expanded to precisely define the intended purpose of the final product. This can be achieved through contextual inquiry [3] with domain experts and by deriving quantitative and qualitative findings from the evaluation. Proof of this can be found in the prior work of Váczlavová and Laco [18], where the following user scenarios were examined and implemented for the HIPA project: studying from medical image sets, uploading medical images, and accessing supporting study materials. One of the design decisions was to split the personas of end-users into two — a Lecturer and a Student — to match different points of view on the educational process. The following user scenarios are supported in HIPA: studying using the available course materials, which are organized by body system or organs (Student); browsing the materials (Student); and uploading slides into the system (Lecturer). Users can closely inspect WSI images of tissue samples using zoom and pan tools, which simulate microscopy. Additionally, the application supports the addition of notes and the exploration of anatomical structures by browsing keywords and toggling the visibility of the structure names. As of now, HIPA does not support the creation of annotations, although this user flow has already been proposed and designed. The role of artificial intelligence (AI) in medical image annotation was also addressed in the work. The AI-based enhancements of the tool were differentiated into automation — as an overlay of AI outputs on WSI — and augmentation — as hints of ROIs locations and the gradual appearance of annotations for users to approve or decline. In this paper, a comparable methodological approach was adopted, with the exception of excluding AI from medical image annotation. Such a decision was made because AI is capable of overturning correct decisions made by professionals [13]. In the case of radiology, the probability of inexperienced radiologists following incorrect suggestions produced by AI was significantly higher: accuracy dropped from 79.9% to 19.8% when provided with incorrect Breast Imaging Reporting and Data System (BI-RADS) predictions [6]. Since the primary focus is on designing the tool for educational purposes, reliance on AI could undermine diagnostic skills that are still being developed in students.

3 Design Methodology Proposal

Since our aim is to extend the existing education platform HIPA with a medical image annotation module while maintaining consistency with conventional histopathology software, we analyzed existing solutions and compared them to define the scope of work. The separation of the Student and Lecturer views of the application is also considered, resulting in separate information architectures (IAs) based on their respective roles. Based on the results of the observation phase (see Subsection 4.2), it can be concluded that end-users currently rely on clinical-grade software that provides an overwhelming set of tools, scripting options, and application logs, with no intuitive layout or user flows. This can be cumbersome for users, as instead of focusing on the subject matter, they have to first learn how to operate the software. Therefore, implementing support for the annotation-creation user flow constitutes an important part of HIPA. We propose an extension of the standard **HCD** process to create UI/UX design of the *Annotations* module. The extended methodology establishes an initial step, preceding the observation phase — an exploration phase (see Subsection 4.1) — and introduces interconnections between the exploration and prototyping phases (see Fig. 1). This methodology facilitates iterative feedback beyond the standard HCD flow. In highly specialized domains such as digital histopathology, direct observation without prior domain familiarization is

[1] *QuPath. Open Software for Bioimage Analysis*, version 0.6.0 (QuPath: Open source software for digital pathology image analysis, 2025). Available online on 07.04.2026 at qupath.github.io.

[2] *Aperio ImageScope - Pathology Slide Viewing Software*, versions 12.3.3 and 12.4.6 (Leica Biosystems Imaging, Inc., 2025). Available online on 07.04.2026 at leicabiosystems.com.

unlikely to yield meaningful insights. Therefore, we propose an **exploration** phase to establish the foundational domain knowledge necessary for all subsequent stages of the modified HCD methodology. This phase comprises a heuristic evaluation and exploratory usability testing involving a minimum of five participants, aimed at calibrating the author's mental model of the domain and identifying consistency patterns and anti-patterns for subsequent analysis. In the **observation** phase, we offer to use contextual inquiry to obtain data such as motivations and frustrations (see Subsection 4.2). Additionally, we propose to employ the *domain transfer* method [16] to obtain valuable insights from the domain specialists' perspectives, specifically from the fields of visual graphics and information technology (IT). During the **ideation** phase, we propose to create personas, along with user scenarios, to unify and formalize the typical users and their requirements (see Subsection 4.3). Also, we suggest the IA to be based on existing clinical software widely used by both students and professionals. We divide the **prototyping** phase into two steps: *low-fidelity prototypes* and *high-fidelity prototypes*. Low-fidelity prototypes take place during the *exploration phase*, and are created prior to and during user interviews to collect feedback to determine which layout feels more natural to the participants. The first iteration of a high-fidelity prototype is designed in the *prototyping phase* informed by the aforementioned research. The prototype is then evaluated in the **testing** phase via usability testing with a minimum of five participants (see Subsection 4.5). The sample size is consistent with established usability practices for first-iteration evaluation, where approximately 85% of major usability issues are uncovered [11]. The estimate is most reliable when the population is homogeneous and the task scope is narrow [20] – both conditions met in this case study, as the annotation module is focused and the participants were the first adopters, representative of the end-users. Prior iterative medical tool evaluations further confirm that feedback saturation can be reached within a small number of sessions when the task domain is well-defined [17]. Using the thinking aloud method [11] during testing we plan to gather qualitative verbal data from the participants to formalize mental models of the users together with IA issues, supported by quantitative metrics, such as the Task Completion Rate (TCR), the Error Rate (ER), the Task Completion Time (TCT), the System Usability Scale (SUS) [4], and the Single Ease Question (SEQ) [14], which are analyzed to identify usability issues together with the usability score, based on the perceived ease-of-use. The content and goals of the next iteration of the prototype rely on the findings of the latter phase.

4 Case Study

Preliminary validation of the proposed research approach to the design process of the medical image annotation tool

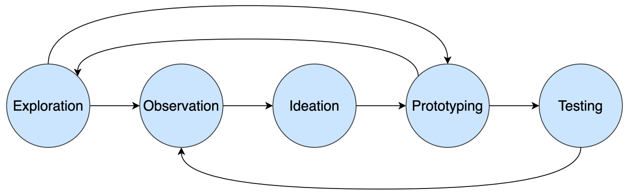

Figure 1: Modified human-centered design (HCD) methodology with additional initial exploration phase and refined interconnections between phases.

was conducted for a case study on UX and UI design of an annotation module extending the HIPA application described in Section 2. In the following subsections, the adoption of the modified HCD methodology is described in detail.

4.1 Exploration

Based on the findings in Section 2, **QuPath** was selected for preliminary user testing as a clinical-grade software to collect data on usability issues, layout, annotation creation tools, and the overall structure of the existing solution. This decision provided a simpler introduction to the domain compared to interviews with domain experts, where establishing a common language between medicine and IT would have posed a challenge. QuPath was chosen because of its availability and prevalence among pathologists, as mentioned in Section 2 [10]. Consequently, the goal was to maintain consistency with this particular software. The exploration phase comprised two parts: a heuristic evaluation based on Nielsen's heuristics [11], conducted by the author, and exploratory usability testing conducted with six participants (average age 30 years; female to male ratio 5:1). During the heuristic evaluation, the core workflows were reviewed — project creation, slide navigation, annotation creation, editing, deletion, and export — and documented in accordance with each heuristic. For the usability testing, one participant was a medical domain expert (a medical doctor), while the others were experienced in IT, one of whom had experience in visual graphics. Each session began with a semi-structured interview covering the participants' existing workflows, tool preferences, and pain points related to image viewing and editing software. The questions covered the following areas: familiarity with image editing applications and frequency of use, experience with professional tools, sources of frustration in current software, usage of keyboard and mouse shortcuts, preferences for which functions should be visible and which should be hidden, and preferred annotation tools (e.g., freehand pen, polyline, geometric shapes, wand tool). Additionally, participants were asked about specific interaction preferences, including pen and eraser width and color settings, bulk annotation deletion, and export capabilities. Following the interview, the participants were introduced to the basic principles of annotation creation in digital pathology; thus, there was no need for task adaptation for non-experts. Also, the participants were encouraged to sketch a low-fidelity prototype at any time during the session to convey their mental model. The session proceeded with seven tasks that included locating a ROI, creating three annotations using different tools, editing an annotation, and exporting the results. The sessions were recorded via video and audio, capturing participants' behavior and actions while completing the tasks. The recordings were then manually analyzed to obtain the TCT, TCR, and ER metrics. Additionally, three low-fidelity prototypes with different layout schemes were presented to the participants. The layout resembling existing graphics software was preferred, as the respondents indicated that there was no need to change the existing conventions. Based on the results, the following conclusions could be drawn:

- QuPath provides a wide range of functionalities for advanced users, including automation, a logger, and support for Python scripting.
- Several tools that could be useful for users are missing, including an eraser tool, undo and redo buttons, a zoom slider, and pen tool thickness and color settings. The participants would like to have those functionalities, even though one respondent (medical doctor) did not find it necessary to change the color of any tools because yellow is contrasting on a purple-tinted Hematoxylin & Eosin (H&E) slide, as the two colors are complementary.
- Users would like to have a button to delete all previously created annotations.
- Users could not find how to export the annotations and measurements in a text format. The participants would like to have a simple save-like button to create an export.
- QuPath reveals usability issues mapped to Nielsen's heuristics [11], such as user control and freedom (no explicit undo and redo buttons), helping users with errors (the design does not offer straightforward recovery for novice users), match between the system and the real world (the UI terminology and icons are unfamiliar and unclear to end-users), and recognition rather than recall (unclear export options, which users have to memorize). Additionally, the *blank slate* [5] could complicate first-time use of the program.

4.2 Observation

To gain a preliminary understanding of the natural annotation workflow, sufficient at this formative stage given the narrow task scope, a contextual inquiry [3, 7] was conducted with a histopathology student, during which the participant performed a typical annotation task in QuPath using a stylus and keyboard. The participant demonstrated the annotation of H&E-stained slides, a process that usually required 1.5 to 2 hours per specimen. The respondent relied primarily on the freehand drawing tool and identified several pain points: the absence of pen width settings, the lack of an eraser tool, and the need to frequently switch between the drawing and move tools to navigate the slide due to the absence of tool width settings. The participant also expressed a preference for a thin-line drawing mode with automatic region fill upon closure, as well as automatic viewport panning when the stylus reaches the edge of the screen. These observations provided insight into the inefficiencies of the current workflow and confirmed the need for a simplified alternative. These findings, together with the usability issues identified during the exploration phase, informed the creation of personas, user scenarios, and the IA proposed in Subsection 4.3.

4.3 Ideation

During the ideation phase of our proposed modified HCD methodology, we identified two personas, five user scenarios, developed and verified the IA.

4.3.1 Personas and user scenarios

Based on the results of the user research, **personas** and **persona scenarios** were created: a student, Sveta, and a histopathology lecturer, Gene. The persona Sveta was primarily derived from the contextual inquiry with a histopathology student, as the respondent's motivations and challenges were specific to the current educational context. Sveta seeks to study efficiently and is frustrated by the amount of time spent on inefficient manual annotation creation. Gene was constructed based on the usability testing with the medical doctor participant, as this individual has experience in pedagogy and understands the motivations and frustrations of medical students. This persona would like to have a unified system for delivering knowledge to students and is averse to overly complicated and feature-saturated software. For the Sveta persona, two main user scenarios were derived based on the user interviews mentioned in Subsection 4.1:

1. *Exam preparation*. Sveta needs to organize and gain knowledge more conveniently and efficiently using the provided software and her setup to create annotations. Sveta would like to have user-friendly software where she can create annotations on digitized and already categorized preparations provided by the faculty.
2. *Annotation creation*. Sveta needs to examine some slides to practice after the lecture. She would like to open a specific topic or body system and inspect the slides to create annotations using a mouse or a wireless pen to outline ROIs.

Gene persona has three user scenarios:

1. *Grade submissions*. Gene wants to implement recall techniques in lectures and seminars. Gene explains

the topic along with the manual annotation process on paper during the lecture. He would like to demonstrate the annotation process and review students' annotations using user-friendly software.

2. *Annotation creation.* Gene wants to create detailed annotations with comments and demonstrate the process during the lecture.
3. *Annotation export.* Gene would like to share the annotations after the lecture so that the students can revise the material.

4.3.2 Information Architecture

An **information architecture** model was created to visualize the tool layout in QuPath. The structure was then analyzed, and a significant number of elements were removed, such as wand, polyline, logging, and scripting tools, along with the combine elements and hierarchy options, as the results of the user research did not indicate the need to retain them. The following elements were added: explicit undo, redo, and save buttons, an eraser tool with partial or complete deletion, tool width and color settings, a delete all elements button, and a zoom slider. Toolbox and the workspace layout along with the move tool were kept unchanged, as this is standard in histopathology software, reviewed in Section 2. The icon for move tool was changed to resemble a hand based on user research. The IA was then validated using the closed card sorting method, in which the categories were predefined [2] (see Fig. 2). The results confirmed that the proposed structure was clear to the respondents.

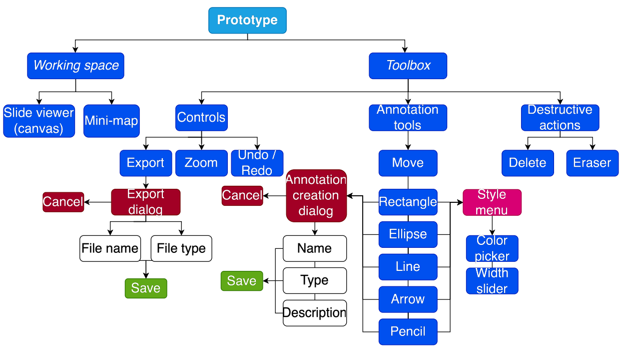

Figure 2: Information architecture (IA) diagram of the first iteration high-fidelity prototype.

4.4 Prototyping

As mentioned in Subsection 4.1, low-fidelity prototypes were prepared in advance and presented during user interviews, allowing participants to compare the options and indicate which layout they found more suitable. Participants also produced their own sketches to communicate the ideas more effectively. This approach facilitated prompt feedback collection, thus mitigating possible shared language barriers. Furthermore, based on the survey study [10] mentioned in Section 2, the most commonly used staining method is H&E, scoring 94.9% among respondents; therefore, an H&E-stained image was chosen for the high-fidelity prototype. Based on the results of the user research combined with low-fidelity prototyping evaluation and IA proposal, the first iteration **high-fidelity prototype** was designed in *Axure 11 RP*[3], consistent with HIPA's design (see Fig. 3). Axure RP 11 was chosen due to its capability of integrating web-based functionalities by embedding custom code within prototypes. This allowed the live ROI definition step of the annotation creation user scenario to be performed directly within the prototype, eliminating the need for a Wizard of Oz method [11]. The following user scenarios, defined in Subsection 4.3 were implemented:

1. *Annotation creation* – persistent annotations using rectangle, ellipse, line, arrow, and pencil tools.
2. *Annotation editing* – reversal of action using undo, redo, and delete buttons.
3. *Export* – exporting the project in a chosen format.

The prototype was designed as a modular system in which the user interface and the drawing engine operate as interconnected components. The architecture of the prototype consists of two modules: a UI layer and a drawing engine. The UI layer, represented by the prototype in Axure 11 RP, functions as an interface with which users interact with the system according to the defined user scenarios. User interactions are communicated as action calls to the embedded drawing engine located within the workspace of the prototype. Each action is dispatched via an "Open Link / javascript:..." POST message to the *iframe* element, which loads the canvas.html file. Within the iframe, shapes are rendered on the canvas element and persisted in JavaScript arrays.

4.5 Testing

The high-fidelity prototype was tested via usability testing with five participants. The average age of the participants was 29.4 years, and the medical doctor from the pilot testing also participated. There were nine tasks divided into groups based on the difficulty level of completion:

1. [Impression task] Describe the initial impression of the interface and its perceived purpose.
2. [Motivational task] Interpret the meaning of the icons present in the interface.
3. Create an annotation using the first available tool.
4. Create an annotation using an alternative tool.
5. Change the color of the annotation tool before creating an annotation.
6. Find the details of the annotation.
7. Undo functionality – correct a mistake.

[3] *Axure RP 11*, Team edition 11.0.0.4134 (Axure Software Solutions, Inc.). Available online on 07.04.2026 at axure.com.

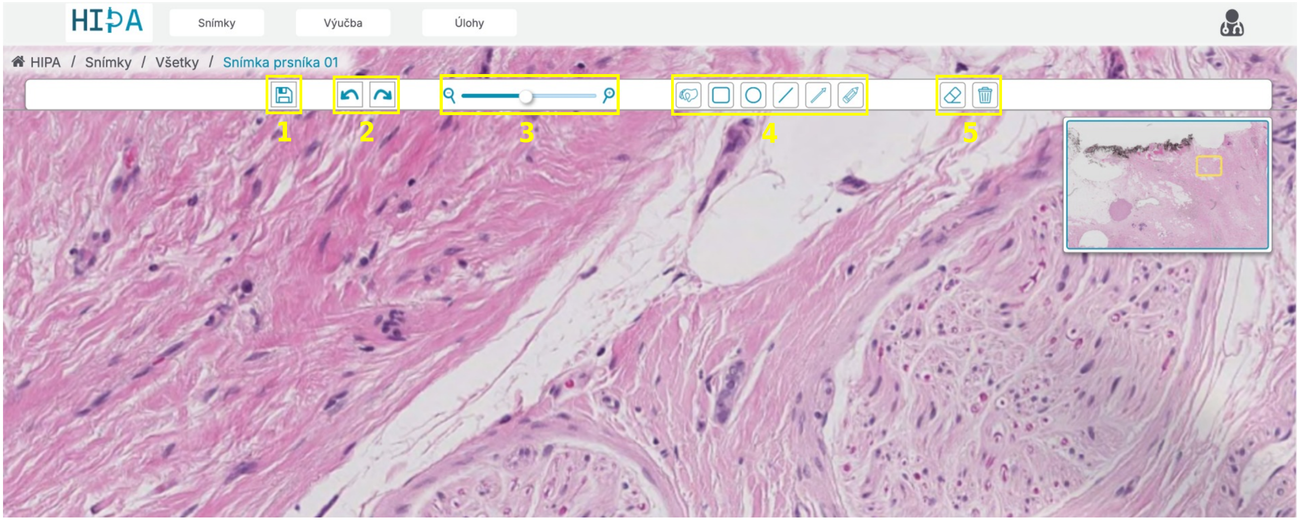

Figure 3: First iteration high-fidelity prototype in Axure RP 11, consistent with HIPA's design and is integrated into the existing microscopy simulation window. Labeled functionalities: **(1)** export; **(2)** undo/redo; **(3)** zoom; **(4)** move and drawing tools; **(5)** erase and delete.

8. Delete all functionality – start from the beginning without leaving the window.
9. Perform an export of the annotated image.

The testing sessions were in the 1:1 format, where the participants' actions and comments were recorded on video to later examine and derive metrics. One session was roughly 30 to 60 minutes, with the average time being 42 minutes. The average TCT per task was 78.0 seconds, with a standard deviation of 66.0 seconds. The recordings were then post-processed for subsequent analysis and metric derivation, as described in Section 3. Finally, qualitative feedback was collected to identify possible improvements to the UI design. Based on the feedback, the addition of the selection of the created annotation or object is required in order to edit the shape using individual points.

The **SEQ** metric (Fig. 4) indicated that the fifth task was the most difficult, with 40% of respondents rating this task as difficult. The participants could not find a way to

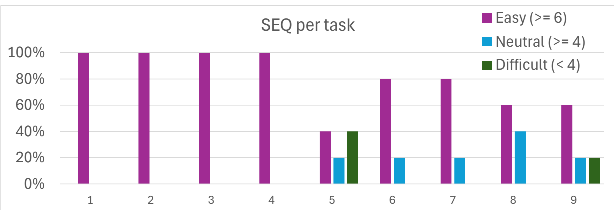

Figure 4: The Single Ease Question (SEQ) metrics for each task

change a tool's color before creating an annotation, which was caused by the workflow where the color is set after the ROI is outlined, in a separate properties window. This task required the color to be chosen before the annotation creation through a context menu triggered by a right click on the tool button. Some participants indicated the tool icon should include an additional visual element to indicate the availability of a context menu. Other participants noted that the color settings should be separated from the tools altogether and represented by a separate button that would expand into a color selection menu on click. In the ninth task, the export button was not intuitive for the users, with 20% of participants labeling this task as difficult. Continuous design improvement was applied, and the icon was replaced with a floppy disk save icon based on participants' feedback from the completed sessions. The SEQ metrics indicate that the tasks were generally found to be easy, with a mean of 6.24 on a 7-point scale. At least half of the participants rated the tasks the highest score; however, based on the standard deviation of 1.40, some respondents found the tasks to be more difficult. The average **ER** across the tasks was 0.73 with a standard deviation of 1.30, suggesting that while errors were not frequent, their occurrence varied considerably among participants. The statistical metrics of the **SUS** scores (Table 1) indicate a high level of perceived usability of the prototype, as the mean score of 91.0 is greater than the benchmark of 68.0 for average usability [15]. The range of scores did not vary substantially, with a minimum of 82.5, a maximum of 95.0, and a standard deviation of 5.18, suggesting that all participants rated the prototype consistently highly. The **TCR** across all participants was 87.78% with a standard deviation of 17.87% (Table 2).

Table 1: SUS statistical metrics

Statistical metric	Value
Mean	91.0
Median	92.5
Range (min–max)	82.5–95.0
Standard deviation	5.18

Table 2: TCR statistical metrics

Task number	**% Value**
1	100.0
2	100.0
3	100.0
4	100.0
5	50.0
6	90.0
7	100.0
8	80.0
9	70.0
Statistical metric	**% Value**
Mean	87.78
Range (min–max)	50.0–100.0
Standard deviation	17.87

5 Discussion

Overall, the results of the testing suggest that the prototype performed well, achieving a high level of usability, thus validating the design decisions based on the modified HCD methodology. The iterative approach, with the initial step of exploration and interconnected exploration and prototyping phases, together with evaluated findings from Section 2, contributed to addressing the usability issues identified in Subsection 4.1.

The TCR of 87.78% (Table 2) indicates that the majority of the tasks were accomplished successfully, with a standard deviation of 17.87, largely driven by the sixth task, which scored 50.0%. Usability was perceived to be high based on the results of the SUS score of 91.0 (Table 1); however, the metrics should be interpreted with caution because the participant sample included one domain expert. While this limits the applicability of the findings to the target users, it is worth noting that the domain expert rated the prototype the highest (95.0), which might suggest that, in this case, domain expertise does not negatively affect the perceived usability. This limitation could be addressed by recruiting domain experts as respondents in future usability testing iterations. The SEQ assessment (Figure 4) shows that the fifth and ninth tasks were reported as being difficult, as mentioned in Subsection 4.5. The difficulty in both tasks stems from the IA issues, as the design and layout of the UI did not match the users' mental model. The seventh task was reported as neutral by 20% of participants, as they attempted to directly edit the outline of the annotation. Similarly, the eighth task received a neutral rating from 40% of participants, who attempted to delete all annotations via direct selection and deletion. Neither of these functionalities was designed in the current iteration of the prototype, as the primary objective was to evaluate the overall layout and annotation tools within the given time constraints. Therefore, participants were required to find an alternative approach to complete the task.

These reported difficulties provide support for improvement in future iterations. Despite the absence of a dedicated help icon [11], no participant required assistance during evaluation, and the intentional simplicity of the annotation tools was confirmed by the domain expert as beneficial for reducing students' cognitive load.

6 Conclusion

In this paper, we proposed a modified HCD methodology alongside the method proposal to design the UX/UI of an annotation module for the existing digital pathology education tool HIPA [18]. The standard HCD process was extended with an initial exploration phase and an additional feedback loop between the exploration and prototyping phases due to the specifics of UX/UI design in medical education and the novel field of digital pathology, which joins the domains of applied informatics and medicine. The proposed methodology was validated through a case study on the design of the annotation module for HIPA. The exploration phase helped identify usability issues in existing clinical-grade software and provided a foundation for understanding the domain context. The observation and ideation phases informed two personas (Student and Lecturer), five user scenarios, and a validated IA. Based on these results, a first-iteration high-fidelity prototype of the annotation creation module was designed in Axure RP 11, consistent with HIPA. We conducted user testing with applied thinking-aloud methodology with five participants. The evaluation combined the following quantitative metrics – TCR (87.78%), SUS (91.0 on a 100-point scale), SEQ (6.24 on a 7-point scale), ER (0.73), and TCT (78.0 seconds) – with qualitative feedback. From these results, we assume that the preliminary solution received a high level of perceived usability, suggesting that the modified HCD process is effective in facilitating iterative feedback and addressing usability concerns. We consider that the resulting UX/UI design may serve as a specification for implementing the annotation module in HIPA after further iteration of the high-fidelity prototype, incorporating recommendations derived from the usability issues identified during user testing, such as improved IA for tool settings and direct annotation object editing support. Moreover, the preliminary validated modified HCD methodology may serve as a basis for future UX/UI design of medical education applications, where tailoring professional software to the needs of domain adepts is important. Future work involves refining the subsequent prototype iterations, recruiting a larger sample of domain experts, and integrating the annotation module into the HIPA project for deployment in a real educational setting.

References

[1] Peter Bankhead, Maurice B. Loughrey, José A. Fernández, Yvonne Dombrowski, Darragh G. McArt, Philip D. Dunne, Stephen McQuaid, Ro-

nan T. Gray, Liam J. Murray, Helen G. Coleman, Jacqueline A. James, Manuel Salto-Tellez, and Peter W. Hamilton. Qupath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1), 2017.

[2] David Benyon. *Designing User Experience: A Guide to HCI, UX and Interaction Design*. Pearson, 4 edition, 2019.

[3] H. Beyer and K. Holtzblatt. *Contextual Design: Defining Customer-Centered Systems*. Interactive Technologies. Elsevier Science, 1998.

[4] John Brooke. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 11 1995.

[5] Alan Cooper, Robert Reimann, David Cronin, and Christopher Noessel. *About Face: The Essentials of Interaction Design*. Wiley, 4 edition, 2014.

[6] Thomas Dratsch, Xue Chen, Mohammad Rezazade Mehrizi, Roman Kloeckner, Aline Mähringer-Kunz, Michael Püsken, Bettina Baeßler, Stephanie Sauer, David Maintz, and Daniel Pinto dos Santos. Automation bias in mammography: The impact of artificial intelligence bi-rads suggestions on reader performance. *Radiology*, 307(4):e222176, 2023. PMID: 37129490.

[7] Karen Holtzblatt, Jessamyn Burns Wendell, and Shelley Wood. *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. Morgan Kaufmann, San Francisco, 2005.

[8] M.P. Humphries, P. Maxwell, and M. Salto-Tellez. Qupath: The global impact of an open source digital pathology system. *Computational and Structural Biotechnology Journal*, 19:852–859, 2021.

[9] Anna Korzynska, Lukasz Roszkowiak, Jakub Zak, and Krzysztof Siemion. A review of current systems for annotation of cell and tissue images in digital pathology. *Biocybernetics and Biomedical Engineering*, 41(4):1436–1453, 2021.

[10] Diana Montezumaa, Sara P. Oliveiraa, Yuri Tolkacha, Peter Boora, Alex Haragana, Rita Carvalhoa, Vincenzo Della Meaa, Tim-Rasmus Kiehla, Sabine Leha, Mustafa Yousifa, David Ameisena, Mircea-Sebastian Șerbanescua, Norman Zerbea, and Vincenzo L'Imperio. Annotation practices in computational pathology: A european society of digital and integrative pathology (esdip) survey study. *Laboratory Investigation*, 2024.

[11] Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann, 1993.

[12] K.A. Philbrick, A.D. Weston, Z. Akkus, et al. Rilcontour: a medical imaging dataset annotation tool for and with deep learning. *J Digit Imaging 32, 571–581*, 2019.

[13] Emely Rosbach, Jonathan Ganz, Jonas Ammeling, Andreas Riener, and Marc Aubreville. Automation bias in ai-assisted medical decision-making under time pressure in computational pathology. In Christoph Palm, Katharina Breininger, Thomas Deserno, Heinz Handels, Andreas Maier, Klaus H. Maier-Hein, and Thomas M. Tolxdorff, editors, *Bildverarbeitung für die Medizin 2025*, pages 129–134, Wiesbaden, 2025. Springer Fachmedien Wiesbaden.

[14] Jeff Sauro and Joseph Dumas. Comparison of three one-question, post-task usability questionnaires. *Conference on Human Factors in Computing Systems - Proceedings*, pages 1599–1608, 04 2009.

[15] Jeff Sauro and James R. Lewis. Chapter 8 - standardized usability questionnaires. In Jeff Sauro and James R. Lewis, editors, *Quantifying the User Experience (Second Edition)*, pages 185–248. Morgan Kaufmann, Boston, second edition edition, 2016.

[16] Christian Stary and Stefan Cronholm. Method transfer across domains and disciplines: enriching universal access development. *Universal Access in the Information Society*, 14(2):145–150, Jun 2015.

[17] Katy E Trinkley, Miranda E Kroehl, Michael G Kahn, Larry A Allen, Tellen D Bennett, Gary Hale, Heather Haugen, Simeon Heckman, David P Kao, Janet Kim, Daniel M Matlock, Daniel C Malone, Robert L Page 2nd, Jessica Stine, Krithika Suresh, Lauren Wells, and Chen-Tan Lin. Applying clinical decision support design best practices with the practical robust implementation and sustainability model versus reliance on commercially available clinical decision support tools: Randomized controlled trial. *JMIR Medical Informatics*, 9(3), 2021.

[18] Erika Václavová and Miroslav Laco. Medical-domain-expert-centered ai-assisted digitized histopathology education tool. In Marta Kristín Lárusdóttir, Bilal Naqvi, Regina Bernhaupt, Carmelo Ardito, and Stefan Sauer, editors, *Human-Centered Software Engineering*, pages 281–290, Cham, 2024. Springer Nature Switzerland.

[19] Noorul et al. Wahab. Semantic annotation for computational pathology: multidisciplinary experience and best practice recommendations. *The journal of pathology. Clinical research vol. 8,2. 116-128. doi:10.1002/cjp2.256*, 2022.

[20] Alan Woolrych and Gilbert Cockton. Why and when five test users aren't enough. In *Proceedings of IHM-HCI 2001 Conference*, volume 2, pages 105–108. Cépaduès-Éditions, 2001.