Design and Development of a Training and Immediate Feedback Tool to Support Healthcare Apprenticeship

Conference Paper · April 2023

DOI: 10.1145/3544549.3585894

Borui Wang
University of Washington

Chen Chen
Florida International University

Larry Hernandez
University of California, San Diego

James D Murphy
University of California, San Diego

Design and Development of a Training and Immediate Feedback Tool to Support Healthcare Apprenticeship

Matin Yarmand
UCSD The Design Lab
La Jolla, CA, United States
myarmand@ucsd.edu

Borui Wang
UW HCDE
Seattle, WA, United States
theworld@uw.edu

Chen Chen
UCSD The Design Lab
La Jolla, CA, United States
chenchen@ucsd.edu

Michael Sherer
UCSD Health
La Jolla, CA, United States
msherer@ucsd.edu

Larry Hernandez
UCSD Health
La Jolla, CA, United States
l2hernandez@ucsd.edu

James D. Murphy
UCSD Health
La Jolla, CA, United States
j2murphy@ucsd.edu

Nadir Weibel
UCSD The Design Lab
La Jolla, CA, United States
weibel@ucsd.edu

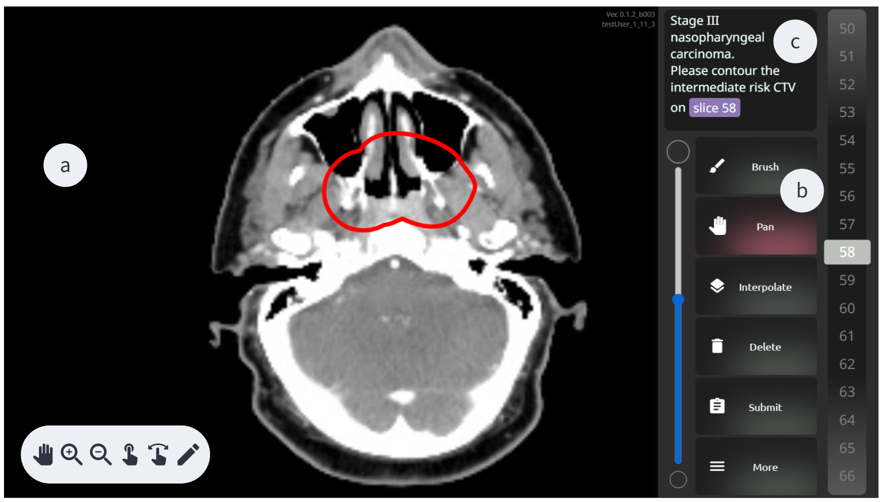

Figure 1: A contouring learning platform that (a) enables visualizing and delineating medical images via pan, zoom, tap, and swipe, (b) provides essential contouring features, such as interpolating, erasing contours, and adjusting brush size, and (c) incorporates an information panel in-situ of the main canvas that presents patient case and contouring feedback.

ABSTRACT

The apprenticeship model of learning, while valuable in facilitating direct expert supervision, lacks flexibility, timeliness, and feedback diversity, especially in high-stakes healthcare training (*e.g.*, residency programs) where teaching resources are limited. An example healthcare domain is radiation oncology, where residents learn from attending physicians how to contour tumors that require radiotherapy treatment. This paper explores the current apprenticeship strategies in radiation oncology, proposes designs, and develops a prototype to enhance the transfer of knowledge from experts to residents. We introduce three feedback mechanisms comprising visual and text-based elements that outline the degree of overlap with expert contour, specific guidance on over- and under-contoured regions, and long-term toxicity for tumors and nearby organs at risk. The design strategies of this work can inform the design of other learning platforms (in healthcare and beyond) to improve the delivery and access of expert feedback in apprenticeship models.

CCS CONCEPTS

• **Human-centered computing** → *Web-based interaction.*

KEYWORDS

Contouring, Healthcare Training, Apprenticeship Learning Models

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585894

ACM Reference Format:
Matin Yarmand, Borui Wang, Chen Chen, Michael Sherer, Larry Hernandez, James D. Murphy, and Nadir Weibel. 2023. Design and Development of a Training and Immediate Feedback Tool to Support Healthcare Apprenticeship. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585894

1 INTRODUCTION

Apprenticeship is a training model that incorporates direct supervision as the learner gains relevant skills and knowledge by closely monitoring the expert. Residency programs—a common curriculum structure for highly specialized healthcare domains—is a high-stakes learning scenario that relies on apprenticeship models, in which residents learn medical skills by observing the assigned faculty's general workflow, and then re-create their processes. Despite the existing medical reference aids, receiving direct feedback from faculty (in an apprenticeship model [31]) remains the main method of training in residency programs [23, 26].

This learning model, while a potentially powerful method of mentorship, poses important challenges due to a lack of timely and flexible learning opportunities, as well as limited diversity in feedback. Attending faculty take on a dual role of clinician and teacher, and when time is in short supply, patient care takes absolute priority over teaching [28]. As such, many residents report that the limited time and flexibility of supervision can hinder gaining valuable learning skills [14, 18]. In addition, this model of training assigns residents to only a few sequential rotations with different faculty [24]. Given that medical fields are far from gold standards and rely on physicians' tendencies, expert disagreements often arise in many workflows, such as identifying brain signals [5], assessing retinal conditions [30], and segmenting image regions [21]. The limited diversity of feedback can hinder exposure to various medical tendencies and minimize robust learning. Addressing these challenges is a dire need, especially in high-stakes healthcare fields that can ultimately lead to life or death decision-making.

This work aims to inform the further development of apprenticeship models, and specifically improve healthcare residency programs by presenting the preliminary exploration, design, and development of a learning interface that is augmented with valuable expert knowledge to enhance feedback delivery and access. More specifically, this work explores the domain of radiation oncology training, a residency-based program aimed to teach the skills of identifying and outlining tumors (*a.k.a. contouring*). Following a participatory design protocol with four faculty and six residents spanning across two years, this paper succinctly reports on the design and development of a contouring education system with three types of feedback embedded directly in the contouring user interface: (i) percentage of overlap with expert contours, (ii) step-by-step guidance on over- and under-contoured regions, and (iii) long-term toxicity values for tumour and nearby healthy organs. An upcoming nation-wide trial study will determine the feasibility and effectiveness of the feedback strategies offered in this work. While focused specifically on radiation oncology, this paper also provides a blueprint for other apprenticeship training models. We believe that the insights from our participatory design and the implementation strategies provided in this paper are general and can be applied broadly to improve expert training in healthcare and beyond.

2 BACKGROUND AND RELATED WORK

This section provides a discussion of how prior HCI and learning science studies explored and designed for enhancing education by implementing tutorship, as well as a brief dive into radiation oncology and the importance of contouring during radiotherapy treatment planning.

2.1 Curricular and Technological Support for Facilitating Tutorship

For decades, learning scientists and practitioners have aimed to sustainably replicate the gold-standard of education, tutoring. In the early 90's, Bloom's "2 Sigma Problem" [8] shed light on the advantage of 1-on-1 mentorship compared to a conventional classroom in which the students under tutorship achieved two standard deviation higher learning outcomes. As Bloom pointed out, one-to-one tutoring is too costly for societies to bear, and calls for curricular and technological innovations to close the learning gap.

To achieve similar learning outcomes as mentorship, many HCI works offered structural and computer-supported solutions. Flipped classrooms [6] aimed to enhance teacher-student interaction by devoting class time primarily to discussions. Learning For Mastery [7] is a group-based educational strategy in which students mainly learn by cooperating with their classmates. This philosophy argues for a flexible and individualized learning strategy in which every student can achieve mastery by considering that their aptitude is a function of *time needed* to achieve a *certain level* [11]. In addition, technological innovations facilitated tutor-like capabilities that can be deployed at scale. Intelligent Tutoring Systems (ITS), especially successful in the field of mathematics, incorporated evaluation snippets to gauge students' expertise level before moving to more complex concepts [2], similar to how a good tutor responds to the students' needs. Affective computing in education is a more recent line of research, in which these systems attempt to identify the affective states of the learner, assess their real-time needs and motivation status, and tailor the educational components to prolong the learning session [34].

2.2 Contouring: the Backbone of Radiotherapy Treatment Planning

Radiotherapy is the medical practice of treating cancer by delivering high dose of radiation to the tumors. The current procedure contains a multi-disciplinary workflow of multiple clinicians and specialized tasks [3]: after cancer detection, a group of medical practitioners (*incl.* radiation oncologists, radiologists, and surgeons) discuss details of treatment and request acquiring specific medical images. Later, radiation oncologists delineate regions (a process known as *contouring*) and distinguish between malignant tumors and surrounding healthy structures (*e.g.,* organs at risk) [10]. Dosimetrists finally deliver the high energy radiation to the target tissues.

Contouring is a cognitively demanding procedure that calls for extreme precision and requires clinicians to continuously take into account three categories of medical factors [4]: treatment context (*e.g.,* clinical symptoms), tumor context (*e.g.,* size and growth

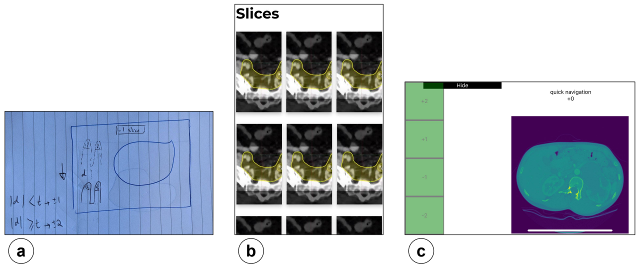

Figure 2: Three mechanisms of slice navigation prototyped during the focus group sessions in increasing fidelity. (a) A sketch demonstrating the concept of two-finger scrolling while the distance dragged determines speed of navigation. The experts indicated that this method can be time-consuming and imprecise. (b) A digital wireframe on Figma presenting a scrollable overview of all slices. While common in many photo applications, the physicians found this method infeasible to grasp the anatomical structure of organs. (c) A webpage with working navigation functionalities implemented via the four buttons on the left. The experts reported that this mechanism was intuitive and granular for contouring, yet preferred a slightly faster method especially for overviewing slices in succession.

direction of tumor), and tumorous areas (*e.g.,* satellite regions). In addition, contouring is considered the weakest link in radiation oncology treatment [27] due to substantial variability in providers' contours [17, 20, 35]. Clinical trials further reveal sobering insights into the high frequency of poor contouring: a study on anal cancer found that 81% of radiation plans had "incorrect contours" [19] and 70% of contours on brain cancer cases were "unacceptable" [16].

Importantly, radiation oncology, like many other medical specialties, is based on a resident model where clinicians learn how to become specialists in this field by observing the work of the more senior attending physicians. Due to highly specialized medical tasks, and a lack of learning technologies in this space, the traditional apprenticeship model is still the only widely used learning model. This is despite the limited availability of faculty, who also play the key role of clinicians treating their patients.

Designing effective tutoring systems like the ones existing in other fields, requires an in-depth understanding of the learning context, especially in high-stakes domains like healthcare residency programs. Following a participatory design protocol with physicians and principles of learning technologies in HCI, we designed and prototyped a feedback interface that improves the delivery and access of expert feedback, and showcases a blueprint for other residency- and mentorship-based programs.

3 PARTICIPATORY DESIGN PROTOCOLS

This work builds on our earlier effort [36] that guided physicians in designing their ideal contouring feedback interfaces. This section describes a follow-up participatory design protocol in a span of two years by engaging radiation oncology faculty and residents in two types of studies:

1) Think-aloud and Semi-structured Interviews — Four faculty and six residents (from UC San Diego Department of Radiation Medicine and Applied Sciences) participated in one-hour interviews, consisting of two main halves. The first part involved participants in think-aloud studies during their routine contouring tasks, concurrently (for faculty) and retrospectively (for residents). This difference of methods aimed to lessen the cognitive load of concurrently thinking out loud and completing difficult tasks (*i.e.,* contouring) for novices, which could have diminished their verbalization and task performance [33]. The research team video recorded the sessions and logged notable events, and later conducted video analysis of contouring behaviours and extracted common patterns. The second half comprised semi-structured interviews in which participants discussed their contouring practices (*e.g.,* the benefits and challenges of the their preferred tools and methods) and contouring feedback (*e.g.,* frequency and strategies of exchanging feedback). Inductive thematic analysis [9] revealed codes and themes: the first author open-coded the transcribed interviews and identified the main topics. Iterative discussions among the team merged these initial codes into preliminary, and then, final themes.

2) Bi-weekly Focus Groups — These sessions enabled rapid prototyping and feedback collection on contouring functionalities of the platform and the embedded expert support. Two participating radiation oncologists (male; 45 and 35 years old; associate professor of medicine, and fourth-year resident) provided insights and experiences from both expert and learner perspectives. The focus group sessions lasted 13 months. Figure 2 describes three examples of prototypes presented during the focus groups. We focus on the highly relevant pieces of insights resulting from our analysis (in Sections 4 and 5), and show how they directly motivate our final design decisions.

4 INTERFACE DESIGN AND CORE FEATURES

Based on several iterations and user feedback during the focus groups, we designed a platform that offers three main UI components: main canvas, sidebar, and an information panel, as described in Figure 1.

a) ***Canvas*** — The canvas is the largest section of the interface, and visualizes the selected medical scans from a stack of DICOM files (*i.e.,* the file standard for medical imaging data). The brush tool facilitates delineation via a circle centered around the user's pointer which can be dragged and lay out contours on the canvas. The brush tool, when not intersecting existing contours, creates a new structure. When overlapping with any existing contour, the tool can instead push the contour inwards or outward when dragging from the outside or inside of the contour, respectively. Figure 3 describes this process. The automatic switching between outlining and refining further expedites a common contouring workflow.

b) Control buttons and slider — We iteratively designed six control buttons in the right panel: *Brush*, *Pan*, *Interpolate*, *Delete*, *Submit*, and *More* which adds a *Help* and a *Fullscreen* button to the panel. Physicians use interpolation to automatically fill in empty slices given adjacent slices, yet only use this technique for slow-changing regions, as noted in the think-aloud study by one of the residents:

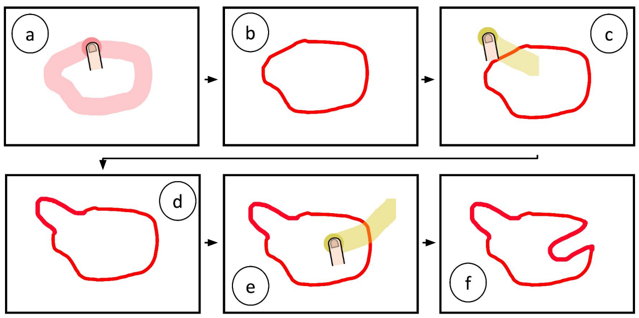

Figure 3: The process of outlining and refining on the contouring learning platform. a) The user's first stroke on an image constitutes as *outlining* and appears with low-opacity red (for the traced path) and a high-opacity red circle (for the position of the pointer). b) The resulting contour appears with a red opaque stroke. c) The user can *refine* the contour from inside to outside with a yellow low-opacity stroke describing the traced path. d) The contour has a part protruding out which marks the outline of the refined path. e) The user can also *refine* from outside to inside. f) The contour, now, has a part hollowed out.

> *"Because [the images] are changing so quickly, the interpolation probably won't work very well. So I'll just have to fix it manually anyway."* (male; 29 years old; year 2 resident)

A vertical slider for brush size appears to the left of the buttons that can also be controlled with the mouse wheel to enable easier adjustment. The proximity of the slider to the canvas addresses the frequent need of physicians to change the brush size from large (*i.e.,* coarse outlining) to small (*i.e.,* granular refining) on every slice, as revealed in the think-aloud studies. A slice scrolling selector (to the right of the buttons) provides an intuitive way of navigating between slices. This type of UI element not only facilitates granular navigation (*e.g.,* when selecting a specific slice), but it also enables quick sliding and comparing across many images, mentioned by physicians as a critical need in the focus groups.

c) Information Panel — The information panel appears on the top of the side panel, which provides patient information and text-based feedback descriptions. This design initially displays information about the patient, the contour type, and target slice. When the user submits their contour, the panel shows one of three types of feedback mechanisms, as described in more detail below, in Sect. 5.

5 FEEDBACK MECHANISMS

Our contouring learning platform contains three feedback mechanisms, separately implemented to enable independent evaluation during the follow-up user study.

5.1 Feedback I: Percentage of Overlap with Readily-available Expert Contour

The first method of feedback provides an immediate accuracy assessment via Dice Similarity Coefficient (DSC), a popular measure in radiation oncology training that gauges the similarity between two contours (learner and expert) with a score from 0 (no overlap) to 1 (perfect overlap) [15], and has been used as an evaluation heuristic in prior computer-supported contouring tools [12, 38]. Residents can also visually compare their contour to the expert's one as it appears on top after task completion (in blue, Fig. 4). To implement this accuracy assessment, we collected consensus expert data on contours from a group of faculty at UC San Diego Health, as described earlier in Sec. 3.

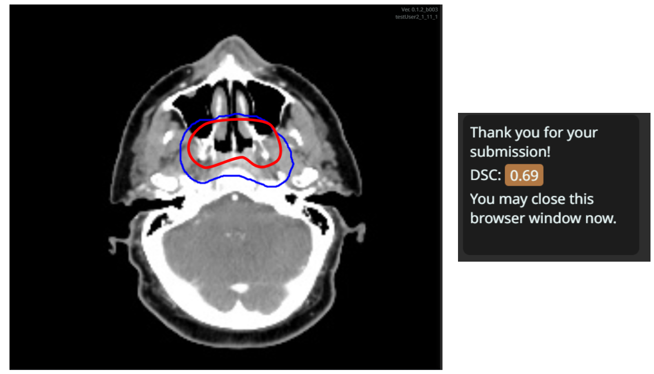

Figure 4: Screenshot of the canvas and the information panel for Feedback I. To the left of the figure is the medical image with a red novice contour and a blue expert contour. To the right of the figure is the information screen containing a calculated DSC based on the two contours.

This method enhances the existing feedback exchange protocol between faculty and residents. As identified in the interview sessions, a popular method of delivering feedback is to re-contour the entire case after the resident (which would take a considerable amount of time), and provide explanations via email, if needed. The residents are then expected to assess their task by visually comparing their contour to their faculty's, as reported by one of the participating experts during the semi-structured interviews:

> *"They get feedback in terms of looking at what I did versus what they did. [...] I think just over time you sort of develop a skill for looking at these differences and doing the proper windowing."* (male; 34 years old; assistant professor)

This method not only provides an additional DSC value for more numerical and (potentially) granular comparison, but it also improves the timeliness of feedback delivery, whereby the expert contour is immediately available for review. Figure 4 illustrates an example feedback which displays a resident contour (in red), an expert contour (in blue), and the calculated DSC value incorporated in the information panel.

5.2 Feedback II: Step-by-step Guidance on Over- and Under-contoured Regions

Informed by the nuances of detecting surrounding and granular regions, this feedback mechanism focuses on areas that could be subject to over-contouring (*e.g.,* mistakenly taking a nearby vessel for the tumour) or under-contouring (*e.g.,* conservative delineation and missing the outside tumor tissues). Once the system detects

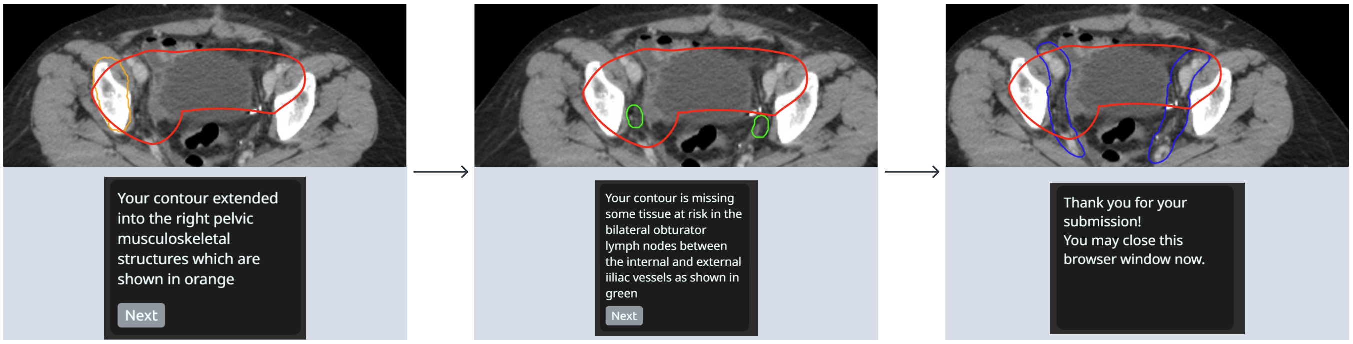

Figure 5: A series of contours and information panels describing example steps in Feedback II. Based on the user contour, three types of sequential screens inform the novice of over-contoured or under-contoured structures, as well as show the available expert contour. Structures are displayed one per screen to help prevent excessive information and improve readability. The difference of colors further assists distinguishing between over and under contouring.

conflicts with the critical areas, regions show up sequentially on the canvas with unique colors (*i.e.,* orange for over- and green for under-contouring) while the information panel displays the accompanying descriptions. The learner can navigate through each conflict and examine it before moving on to the next conflict. Figure 5 demonstrates how over- and under-contoured areas are displayed step by step in the system. One of the participating experts (male; 36 years old; assistant professor of radiation oncology) provided the regions and descriptions used in the system.

This feedback mechanism follows a long-standing HCI and Learning Sciences technique, most notably seen in many Intelligent Tutoring Systems (ITS), specifically a model tracing tutor which aims to solve a problem step-by-step while providing relevant assistance [13]. This technique provides individualized instruction and tailors the feedback according to the resident's understanding of the case. Instead of providing a general solution, this strategy provides extensive information about some areas (that the resident under- or over-contoured) while dismissing all the other regions with the aim of mitigating cognitive load of learning.

5.3 Feedback III: Long-term Toxicity for Tumour and Healthy Tissues

The third feedback strategy provides a deeper understanding behind contouring tasks and presents future consequences for the accompanying radiotherapy. This technique presents two sets of values, each type for both the resident and expert to facilitate learning by comparison [29]: long-term toxicity values for the tumorous region (*i.e.,* Tumor Control Probability or TCP), and Normal Tissue Complication Probability (NTCP), which measures potentially harmful radiation to the nearby organs at risk. These numbers contextualize the critical trade-off in radiotherapy treatment: while the ideal contour achieves high TCP values (*i.e.,* applies high radiation dose to the tumour), it must also minimize radiation to the surrounding healthy organs. Figure 6 displays an example novice contour, the expert contour, and the toxicity values in the information panel on the right. Residents can develop a better evaluation of their task by gauging discrepancies in TCP and NTCP numbers and connecting them to the visual differences in the contours.

Implementing long-term toxicity values followed a two stage pipeline: (i) predicting a spatial dose distribution (*a.k.a.* dose clouds), and (ii) utilizing toxicity models with existing parameters in prior radiation oncology studies. The dose prediction model uses a 2D Convolutional Neural Network (U-Net) to generate dose clouds from each set of contours. We used 101 unique treatment plans—containing different cases and ranging between 170 and 250 slices per case—to train our underlining machine learning algorithm. Specialists across the US, working alongside the editors of a popular contouring textbook [22], provided the expert contours (on tumours, organs at risk, and other notable regions) used in the training dataset. A toxicity model (*i.e.,* Lyman-Kutcher-Burman or LKB) [25] was then applied and produced TCP and NTCP values from the generated dose clouds. This model is an equation that utilizes the dose values and a few numerical parameters estimated from historical oncology studies.

6 DISCUSSION AND CONCLUSION

This work presents the design and development of a learning platform that enhance existing apprenticeships models (*i.e.,* residency) in radiation oncology. The proposed platform not only offers the core contouring functionalities, but also integrates three feedback mechanisms that are based on generalizable learning and engagement models, and apply them to the specific case of radiation oncology training.

The separation of the three feedback methods aims to assess the effectiveness of each technique independently, which will later inform careful combination of methods in order to maximize benefits (given unique needs) while lowering the cognitive load of contouring and learning.

To further mitigate the cognitive burden of healthcare training, we plan to also integrate principles of nudge theory [32] into the outlined generalizable models. We believe this can inform the design of learning interfaces that deliver short feedback snippets unobtrusively, especially after detecting learning uncertainties (*e.g.,* confusion and frustration). As revealed in the interview sessions, prolonged navigation through medical images might

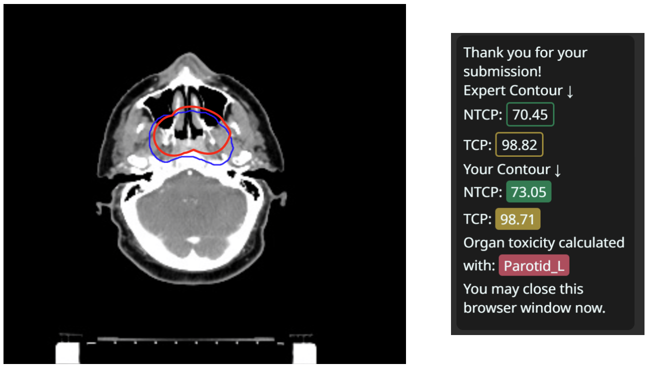

Figure 6: Screenshot combination of the canvas and the information panel for Feedback III. Two sets of values (Expert TCP & NTCP, Novice TCP & NTCP) are displayed to help compare contours. Values for the novice and expert are differentiated through their outlined and filled backgrounds, and background color. The name of the organ that informs NTCP calculation is also displayed.

indicate confusion about locating the tumour. We plan to exploit these moments and build on prior HCI works (such as [37] in the context of instructional videos) to incorporate elements of nudge and achieve higher learning and satisfaction outcomes.

A longitudinal user study can robustly determine specific benefits of each type of feedback, as well as holistically evaluate the learning platform. We plan to perform a randomized trial study and assign residents to four groups: no feedback (*i.e.,* control) and three treatment groups, one for each type of feedback. Participants will then complete similar tasks from the same disease site over a period of six weeks, and their progress (in terms of accuracy of contours and time spent contouring) will be measured throughout the study period. Ad-hoc and self-assessed surveys can further unveil affective states of learners as they interact with the interface. We also plan on collecting log data to capture the participants' unique usage of the platform. In addition, this trial study can further shape the design of each feedback mechanism: for instance, measuring the physicians' trust in the TCP and NTCP values can inform improving the *explainability* [1] of the machine learning algorithms in Feedback III (Sect. 5.3).

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their insightful feedback. This work was supported in part by the Agency for Healthcare Research and Quality (AHRQ).

REFERENCES

[1] Muhammad Aurangzeb Ahmad, Carly Eckert, and Ankur Teredesai. 2018. Interpretable machine learning in healthcare. In *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*. 559–560.

[2] John R Anderson, C Franklin Boyle, and Brian J Reiser. 1985. Intelligent tutoring systems. *Science* 228, 4698 (1985), 456–462.

[3] A Aselmaa, RHM Goossens, A Laprie, S Ken, T Fechter, A Ramkumar, and A Freudenthal. 2013. Workflow analysis report. *7th Framework Programme of the European Union, FP7-PEOPLE-2011-ITN, Summer Marie Curie Research Training Network (PITN-GA-2011-290148)* (2013).

[4] Anet Aselmaa, Richard HM Goossens, Ben Rowland, Anne Laprie, Yu Song, and Adinda Freudenthal. 2014. Medical factors of brain tumor delineation in radiotherapy for software design. In *5th International conference on applied human factors and ergonomics (AHFE)*. 4865–4875.

[5] Elham Bagheri, Justin Dauwels, Brian C Dean, Chad G Waters, M Brandon Westover, and Jonathan J Halford. 2017. Interictal epileptiform discharge characteristics underlying expert interrater agreement. *Clinical Neurophysiology* 128, 10 (2017), 1994–2005.

[6] Jacob Bishop and Matthew A Verleger. 2013. The flipped classroom: A survey of the research. In *2013 ASEE Annual Conference & Exposition*. 23–1200.

[7] James H Block and Robert B Burns. 1976. Mastery learning. *Review of research in education* 4 (1976), 3–49.

[8] Benjamin S Bloom. 1984. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher* 13, 6 (1984), 4–16.

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] Neil G Burnet, Simon J Thomas, Kate E Burton, and Sarah J Jefferies. 2004. Defining the tumour and target volumes for radiotherapy. *Cancer Imaging* 4, 2 (2004), 153.

[11] John B Carroll. 1963. A model of school learning. *Teachers college record* 64, 8 (1963), 1–9.

[12] Chen Chen, Matin Yarmand, Varun Singh, Michael V. Sherer, James D. Murphy, Yang Zhang, and Nadir Weibel. 2022. VRContour: Bringing Contour Delineations of Medical Structures Into Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 64–73. https://doi.org/10.1109/ismar55827.2022.00020

[13] Albert T Corbett, Kenneth R Koedinger, and John R Anderson. 1997. Intelligent tutoring systems. In *Handbook of human-computer interaction*. Elsevier, 849–874.

[14] Janet de Groot, Richard Tiberius, Joanne Sinai, Aileen Brunet, Peter Voore, David Sackin, Susan Lieff, and Susan Reddick. 2000. Psychiatric Residency. *Academic Psychiatry* 24, 3 (2000), 139–146.

[15] Frances Duane, Marianne C Aznar, Freddie Bartlett, David J Cutter, Sarah C Darby, Reshma Jagsi, Ebbe L Lorenzen, Orla McArdle, Paul McGale, Saul Myerson, et al. 2017. A cardiac contouring atlas for radiotherapy. *Radiotherapy and Oncology* 122, 3 (2017), 416–422.

[16] Alysa Fairchild, Damien C Weber, Raquel Bar-Deroma, Akos Gulyban, Paul A Fenton, Roger Stupp, and Brigitta G Baumert. 2012. Quality assurance in the EORTC 22033–26033/CE5 phase III randomized trial for low grade glioma: the digital individual case review. *Radiotherapy and Oncology* 103, 3 (2012), 287–292.

[17] Claudio Fiorino, Michele Reni, Angelo Bolognesi, Giovanni Mauro Cattaneo, and Riccardo Calandrino. 1998. Intra-and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiotherapy and oncology* 47, 3 (1998), 285–292.

[18] Chrystal Jaye, Tony Egan, Kelby Smith-Han, and Mark Thompson-Fawcett. 2009. Teaching and learning in the hospital ward. *The New Zealand Medical Journal (Online)* 122, 1304 (2009).

[19] Lisa A Kachnic, Kathryn Winter, Robert J Myerson, Michael D Goodyear, John Willins, Jacqueline Esthappan, Michael G Haddock, Marvin Rotman, Parag J Parikh, Howard Safran, et al. 2013. RTOG 0529: a phase 2 evaluation of dose-painted intensity modulated radiation therapy in combination with 5-fluorouracil and mitomycin-C for the reduction of acute morbidity in carcinoma of the anal canal. *International Journal of Radiation Oncology* Biology* Physics* 86, 1 (2013), 27–33.

[20] Robert Kosztyla, Robert Olson, Hannah Carolan, Susan Balkwill, Vitali Moiseenko, and Winkle Kwan. 2012. Evaluation of dosimetric consequences of seroma contour variability in accelerated partial breast irradiation using a constructed representative seroma contour. *International Journal of Radiation Oncology* Biology* Physics* 84, 2 (2012), 527–532.

[21] Elizabeth A Krupinski. 2000. The importance of perception research in medical imaging. *Radiation medicine* 18, 6 (2000), 329–334.

[22] Nancy Y Lee and Jiade J Lu. 2012. *Target volume delineation and field setup: a practical guide for conformal and intensity-modulated radiation therapy*. Springer Science & Business Media.

[23] Renuka Malik, Julia L Oh, John C Roeske, and Arno J Mundt. 2005. Survey of resident education in intensity-modulated radiation therapy. *Technology in cancer research & treatment* 4, 3 (2005), 303–309.

[24] Lawrence Martin, Karen Saperson, and Barbara Maddigan. 2003. Residency training: challenges and opportunities in preparing trainees for the 21st century. *The Canadian Journal of Psychiatry* 48, 4 (2003), 225–231.

[25] Kevin L Moore, Rachel Schmidt, Vitali Moiseenko, Lindsey A Olsen, Jun Tan, Ying Xiao, James Galvin, Stephanie Pugh, Michael J Seider, Adam P Dicker, et al. 2015. Quantifying unnecessary normal tissue complication risks due to suboptimal planning: A secondary study of RTOG 0126. *International Journal of Radiation Oncology* Biology* Physics* 92, 2 (2015), 228–235.

[26] Anna Newcomb, Amber W Trickey, Elena Lita, and Jonathan Dort. 2018. Evaluating surgical residents' patient-centered communication skills: practical alternatives to the "apprenticeship model". *Journal of surgical education* 75, 3

(2018), 613–621.

[27] CF Njeh. 2008. Tumor delineation: The weakest link in the search for accuracy in radiotherapy. *Journal of medical physics/Association of Medical Physicists of India* 33, 4 (2008), 136.

[28] Kate Rassie. 2017. The apprenticeship model of clinical medical education: time for structural change. *The New Zealand Medical Journal (Online)* 130, 1461 (2017), 66.

[29] Bethany Rittle-Johnson and Jon R Star. 2011. The power of comparison in learning and instruction: Learning outcomes supported by different types of comparisons. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 199–225.

[30] Mike Schaekermann, Carrie J Cai, Abigail E Huang, and Rory Sayres. 2020. Expert discussions improve comprehension of difficult cases in medical image assessment. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.

[31] Joseph R Schneider, John J Coyle, Elizabeth R Ryan, Richard H Bell Jr, and Debra A DaRosa. 2007. Implementation and evaluation of a new surgical residency model. *Journal of the American College of Surgeons* 205, 3 (2007), 393–404.

[32] Richard H Thaler and Cass R Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

[33] Maaike Van Den Haak, Menno De Jong, and Peter Jan Schellens. 2003. Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & information technology* 22, 5 (2003), 339–351.

[34] Chih-Hung Wu, Yueh-Min Huang, and Jan-Pan Hwang. 2016. Review of affective computing in education/learning: Trends and challenges. *British Journal of Educational Technology* 47, 6 (2016), 1304–1323.

[35] Hideya Yamazaki, Hiroya Shiomi, Takuji Tsubokura, Naohiro Kodani, Takuya Nishimura, Norihiro Aibe, Hiroki Udono, Manabu Nishikata, Yoshimi Baba, Mikio Ogita, et al. 2011. Quantitative assessment of inter-observer variability in target volume delineation on stereotactic radiotherapy treatment for pituitary adenoma and meningioma near optic tract. *Radiation Oncology* 6, 1 (2011), 1–6.

[36] Matin Yarmand, Chen Chen, Danilo Gasques, James D. Murphy, and Nadir Weibel. 2021. Facilitating Remote Design Thinking Workshops in Healthcare: The Case of Contouring in Radiation Oncology. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 40, 5 pages. https://doi.org/10.1145/3411763.3443445

[37] Matin Yarmand, Srishti Palani, and Scott Klemmer. 2021. Adjacent Display of Relevant Discussion Helps Resolve Confusion. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–11.

[38] M Yarmand, M Sherer, C Chen, L Hernandez, N Weibel, and JD Murphy. 2022. Evaluating Accuracy, Completion Time and Usability of Everyday Touch Devices for Contouring. *International Journal of Radiation Oncology, Biology, Physics* 114, 3 (2022), S96.