Peer Review File

Article information: https://dx.doi.org/10.21037/ccts-22-11

Response to Reviewers

We are grateful for reviewers' feedback and the opportunity to resubmit a revised version of our manuscript titled [Developing an Artificial Intelligence-Based Clinical Decision-Support System for Chest Tube Management: User Evaluations & Patient Perspectives of the Chest Tube Learning Synthesis and Evaluation Assistant (CheLSEA) System] for publication. We thank the reviewers for their consideration, efforts, and detailed assessments. In the revised manuscript, we addressed all comments and suggestions to our best ability, and we highlighted them in the revised text. Please see below for point-by-point response to the reviewers' comments.

Line numbers in the "change in text" Refer to the line numbers in the revised final *CLEAN VERSION* Word document.

Response to Reviewer A

Comment A1: I suggested that the result section could be simplified.

Response: We thank the reviewer for their valuable feedback. We have revised the results section and simplified areas to provide a more succinct and concise summary of our results.

Changes in text: Results section, pages 11-16, lines 265-382

Comment A2: For the chest tube system used in this study, I suggested different chest tube system may be a confounding factor. In 2022, a systemic review and meta-analysis discussed that the digital chest tube has the advantage of shorten hospital stay and chest tube use time, and digital chest tube also has the advantage of better portability. This may be needed to be discussed in the discussion.

Response: We are grateful for this suggestion. We agree that there are various chest tube drainage systems available, each with unique advantages and disadvantages. It would be interesting to compare data collected from various commonly used chest tube drainage systems to assess its effect on the quality of the recommendations made by the clinical decision support system. However, this manuscript's focus is to collect feedback from health care providers and patients affected by the implementation of an AI-based clinical decision support system and its user-interface in a clinical setting. Ultimately, the components of our decision support system (CheLSEA) including its interface can be adapted to any type of digital pleural drainage device

as long as there is data of air leak flow.

Response to Reviewer B

Comment B1: What are the biggest strengths and weaknesses of CheLSEA? What other problems need to be overcome?

Response: We thank the reviewer for their comments. We have further elaborated upon the strengths, limitations, and challenges of CheLSEA. Some digital devices lack network connectivity which poses logistical challenges in getting the intra-pleural monitoring data to the decision-support system. Patient safety remains the most important challenge in implementing decision support system in clinical surgery and medicine. However, the larger problem of acceptance and understanding of AI systems in various clinical care settings must be overcome first. To this end, we have undertaken this research to gain insight in the clinical team members understanding of the information presented by the system, and attempt to better understand patient perceptions regarding the use of AI system to assist in postoperative surgical care. The issue of safeguards against potentially unsafe recommendations from the system has been addressed by the creation of a fail-safe component consisting of a rule-based, finite-state machine (FSM) The FSM can override classifications made by CheLSEA which are deemed non-compliant with the clinical protocol. A strength of CheLSEA is the optimization of the chest tube removal process to improve the quality of the treatment since both premature and delayed removal of chest tubes can result in unnecessary suffering and prolonged hospitalization. Another strength is to be able to demonstrate that end users and patients were engaged in the development process which will hopefully promote trustworthiness and credibility towards the clinical implementation of the system.

Changes in text: Methods section, page 7, lines 159-163. Discussion section, page 17, 18, lines 395-397; 415-417

Comment B2: It is suggested to use random samples, and the results may be more perfect.

Response: Thank you for the comment. We began with users and patients from our institution because we are still in the early stages of interface development and system testing, and we plan to conduct multi-institutional trials in the future. For clarity, the system was developed at a single institution and while surveys were implemented in the Phase 1 study, the aim was to start with users from our institution given the early phase of the interface development and system testing to gain a sense of impact that the system had. For phases 2 and 3, the focus was on gaining detailed feedback from likely users and from patients during the development process of CheLSEA system and its interface. Given that the focus was understanding the types of issues, rather than how common those issues are, the purposive sampling in which individuals were targeted based on specific pre-determined characteristics is an entirely appropriate sampling

approach. However, we acknowledge that reference to a 'convenience' sample in the limitations section may have created a misunderstanding about the study populations. As such we have revised the limitations section.

Changes in the text: Methods section, Page 6, lines 134-136 & Discussion section, page 18, lines 422-425

Comment B3: The introduction section of this paper is not comprehensive enough, and some similar papers have not been cited, for example, "Modern day guidelines for post lobectomy chest tube management, PMID: 32274076". It is recommended to cite this article.

Response: Thank you for raising this point. We have revised the introduction and have added information to guide the reader for comprehensiveness and clarity. We have included this citation in the introduction section.

Changes in text: Changes throughout the introduction pages 3-5; Cited in Introduction section, page 3, lines 71-73

Comment B4: What indicators must chest tube management be based on?

Response: We thank the reviewer for their comment. CheLSEA uses a summary of patient follow-up measurements described further in the introduction (paragraph 5). These are accumulated for every 12-hour postoperative time frame and are used as data supplied to the machine learning model to obtain recommendations for chest tube removal in future 12-hour time frames. The data in each 12-hr frame depict the volumes of air and fluid drainage, compliance with clinical criteria for air leak resolution, recurrence of leak of fluid, chest x-ray grading with regards to pneumothorax, subcutaneous emphysema, and pleural effusion. Compliance with air leak resolution criteria, which is one of the criteria used for chest tube removal, is checked by the system. Other criteria, such as the patient's clinical status, which the system cannot determine on its own, are used to determine whether a chest tube can be removed. CheLSEA makes its recommendations for the current and future 12-hour time windows as we have described this in the Introduction section.

Changes in text: Methods, page 6, lines 146-151

Comment B5: This study is a qualitative study with a limited sample size. How to improve its generalizability?

Response: Thank you for the comment. As indicated above, the goal of this work was not to generate generalizable data but to identify key issues that should be considered in the development of AI-CDSS with interface. Participants were information rich and allowed us to reach thematic saturation. Anticipated future work will trial the system in a multi-centre study and we suggest that future multi-centre research would provide beneficial insights into the

commonality of the issues we identified as well as variation between settings. We have revised our limitations section to reflect this:

Changes in text: Discussion, page 18-19, lines 422-423; 431-432

Response to Reviewer C

We appreciate that the reviewer acknowledged the contributions of our study to this area of research, and their suggestions to improve our manuscript, including the clarity in reporting and readability. Our updated manuscript includes the suggested corrections to address the points raised in the comments.

Comment C1: This paper deals with a bread-and-butter issue in thoracic surgery and the much welcome use of AI in aiding management decisions. The reviewers had comments that we encourage the authors to consider in their revisions. It would be helpful if the reader had some idea of the standard care of chest tubes at the authors' institution, e.g., usual care pathways, how decisions are made, and who makes them. Although these may have some similarities between hospitals, there may also be local differences that could have bearing on the paper here.

Response: We thank the reviewer for their comment. The purpose of this study is to assess the interface of a clinical decision aid in postoperative care following lung resection. The standard of chest tube care adheres to a well-defined protocol that has been implemented at our institution and is further described in the discussion section. The protocol was presented in previous papers published by the University of Ottawa's thoracic surgery group, which have been appropriately cited. (Optimal management of postoperative parenchymal air leaks - French - Journal of Thoracic Disease (amegroups.com) & Optimizing postoperative care protocols in thoracic surgery: best evidence and new technology - PubMed (nih.gov). Patients are assessed twice a day and their air leak flow monitoring data evaluated for air leak resolution. The decision to remove chest drains is made based on clinical status of the patient, physical examination for subcutaneous emphysema, volume of liquid drained over time, and CXR data if available as CXRs are not performed routinely in stable patients. Here are some of the defined criteria:

1. Resolution of air leak: target of 30 mL/min over 8 hours regardless of the intrapleural pressure.
2. Pleural liquid output: ≤20% of whole-body lymphatic flow (or approximately less than five times body weight in kilograms in a 24-hour period)
3. Status of subcutaneous emphysema: absent, mild, or stable over 24-hour period.
4. (IV) Results of chest X-ray (CXR): absent or small pneumothorax, absent or small pleural effusion.

Changes in text: Methods section, page 6-7, lines 155-159

Comment C2: The reader would require more background information on the AI system itself. How was this system developed, how effective is it (at least in preliminary evaluations/trials), what are the anticipated advantages, especially in the context of usual protocols at the authors'

institution?

Response: We thank the reviewer for their comments and excellent feedback. The computer engineering principles and methodology underpinning the development of an operational system, and its preliminary performance evaluation have been described in detail elsewhere (Klement et al. ,2022, https://link.springer.com/article/10.1007/s41060-021-00296-8) Briefly, CheLSEA utilizes a machine learning classifier that recommends chest tube removals in several, future 12-hour time frames after surgery. The validation of predictions prior to invoking the safety protocol shows that our model can offer robust performance (low variation), accurate classifications (80 sensitivity, 90+ specificity, 95+ AUC), and safe recommendations (< 10% false chest tube removals). The decision of selecting the random forest method was based on selecting the best performing model, these results were published in (Klement et al. 2019, https://ieeexplore.ieee.org/abstract/document/8964153)

We have demonstrated in a randomized controlled trial study that the clinical decision-making surrounding chest tube duration post-operatively continues to be a challenge for clinical care teams, even when using digital pleural drainage monitoring (Randomized trial of digital versus analog pleural drainage in patients with or without a pulmonary air leak after lung resection - ScienceDirect). Our research group have teamed up with colleagues from the faculty of engineering with specializations in artificial intelligence to potentially develop a system that could reliably predict the course of chest tube management after lung resection. Our main intent was always to develop a decision support system to assist the clinical team rather than trying to fully automate the management of chest tubes.

Change in text: Methods section: page 6-7, lines 143-154

Comment C3: Could the authors describe the different interface components within the main text? Appendix A should be included as a Figure accompanying the main document, as it is extremely relevant and would nicely illustrate the paper.

Response: Thank you for this suggestion. To address this, we have added a new paragraph in the methods section describing the main components of the interface. Appendix A in the initial manuscript has been moved to the main text under the methods section (Figure 2).

Changes in text: Methods section, page 7, lines 164-174. Please see revised Figure 2. (mentioned in introduction and in Methods)

Comment C4: For those less familiar with AI, it may be useful to elaborate further on some of the generic challenges in the clinical adoption of such technologies: the authors do briefly allude to these, for example the obscurity of AI based decisions (lines 416-7), and the importance of evaluating AI using multiple perspectives, including patient perspectives and the issue of perceived effectiveness. Addressing these questions early on may help more clearly frame the way in which study was constructed and how the different phases of the study (lines 103-106) «

fit together », i.e., address these different questions in a hierarchical way. This will also help the reader to avoid feeling overwhelmed by the amount of information in the text and tables (an admittedly inherent characteristic of qualitative research).

Response: We agree with the reviewer and have further described the clinical adoption challenges of AI systems earlier in the introduction section and we have removed the lines 416-417 from the conclusion for clarity. In addition, we have rephrased the aims of our study to better inform the reader of the information presented in the results.

Changes in text: Introduction section, page 3, lines 86-89; 109-113

Comment C5: We agree with the reviewers that the results section should be simplified, if possible.

Response: We thank the reviewer for this appreciated feedback. In line with previous comments made, the results section has been revised for added clarity.

Changes in text: Results section, pages 11-16, lines 266-383.

Comment C6: There are some minor syntax issues throughout.

Response: We thank the reviewer for their comment. The manuscript has been revised to remove any syntax errors.

Comment C7: Lines 644-5?

Response: This line has been removed from the figure caption.

Changes in text: Removed lines 644-5, revised figure 1 caption.

Response to Reviewer D

Comment D1: In the introduction section, I suggest the author highlight the drawback of AI based systems (opacity of its inner working mechanism) and the importance of understanding and confidence from doctors and patients for an AI system.

Response: Thank you for this suggestion. We have included a new paragraph in the introduction to highlight various limitations and challenges to the clinical implementation of AI systems**.**

Changes in text: Introduction section, page 3-4, lines 85-91

Comment D2: In the limitation section, the authors stated that participants were a convenience sample of volunteers. I suggest this statement should also be clarified in the method section.

Response: Thank you for the comment. For clarity, for phases 2 and 3, the focus was on gaining detailed feedback from likely end users and recipients of the CheLSEA system. Given that the focus was understanding the types of issues, rather than how common those issues are, the sequential method is an entirely appropriate sampling approach (indeed similar to the sequential approach taken in a clinical trial should the individual be eligible). However, we acknowledge that reference to a 'convenience' sample in the limitations section may have created a misunderstanding about the study populations. As such we have revised the limitations section. In addition, we clarified the sampling method in the Methods section.

Changes in the text: Methods section, page 6, lines 135-137 Discussion section, page 18, lines 425-430

Comment D3: I think that the conclusion section should be simplified. Generally, this section should concisely summary the main results and clinical significances from the conducted study.

Response: Thank you for this feedback. We apologize for the lack of clarity in this section, we have revised the conclusion section accordingly.

Change in text: Conclusion section, page 19, lines 435-439

Comment D4: This sample size of the current study is very limited based on a single center, dose the authors conduct a following large study to provide more evidence in future?

Response: We thank the reviewer for their comment. As indicated above, the goal of this work was not to generate generalizable data but to identify key issues that should be considered in the development of AI CDSS. Participants were information rich and allowed us to reach thematic saturation. Anticipated future work will trial the system in a multi-center study and we suggest that future multi-center research would provide beneficial insights into the commonality of the issues we identified as well as variation between settings. We have revised or limitations section to reflect this.

Changes in text: Discussion, page 18-19, lines 428-430; 432-433

Comment D5: The survey questions were modified based on the System Usability Scale. How was the scale in the current study developed? Did the item in the scale was discussed in a specialist team? Please detail them.

Response: Thank you for pointing this out. The survey questions and format were developed locally within our team. The research team consisted of a thoracic surgeon, academic computer scientists, research program manager, research coordinator and medical students. Some members

of this team were the developers of CheLSEA and the user- interface. Our research team operates at an academic tertiary center within the Thoracic Surgery Department at the Ottawa Hospital-General Campus. Our team has also sought continuous feedback from the "Ottawa Methods Centre" while developing the survey. In the process, the team came across the widely used System Usability Scale (SUS) developed by Brooke (1996) as a "quick and dirty" measure of usability. We were interested in the concept, simplicity, and the design of the SUS; however, we needed to have a modified version specific to our system and that would capture the perceptions of potential users of our system in a clinical environment. The survey was pretested within the department for content validity and ease of completion. Members who developed the survey were not eligible to participate. We apologize that the survey was unintentionally excluded from the appendix of our initial submission. We have added the survey to appendix A.

Changes in text: Methods section, page 7-8, lines 179-190. See revised appendix A.