scientific reports

Check for updates

OPEN

# Assessing clinician performance using a multi-modality clinical decision-support system for lung cancer prognostication

Jaryd R. Christie[1,2], Karen Eddy[2], Richard A. Malthaner[3], Mehdi Qiabi[3], Saurav Verma[4], Daniel Breadner[4], Pencilla Lang[4], Viswam S. Nair[5,6] & Sarah A. Mattonen[1,2,4,7]✉

**Surgery is the primary treatment for early-stage lung cancer. Adjuvant therapy is offered to patients who are at a high risk of recurrence, however, determining the patients that would benefit from additional therapy is often difficult. In this study, we aimed to develop a clinical decision support system (CDSS) for post-surgery lung cancer prognostication integrating a multi-modality deep learning model (DLM). Pre-operative medical images and clinical, surgical, and pathological information were fed into an externally validated DLM. A CDSS was then developed to display the patient information and DLM results for potential clinical use. Four oncologists evaluated each patient's recurrence probability, their confidence level, and their post-surgery recommendations both with and without the DLM information. The CDSS DLM information demonstrated the potential to improve user prediction performance and confidence. This exploratory study is the first to integrate a multi-modality DLM for prognostication with a CDSS, as well as the first study of clinician attitudes towards CDSSs for lung cancer.**

**Keywords** Deep learning, Clinical decision-support system, Multi-modality, Lung cancer, Computed tomography, Positron emission tomography

Lung cancer is the leading cause of cancer death in North America and worldwide[1,2]. Treatment can include surgery, chemotherapy, radiation therapy, and immunotherapy. Prognosis plays an important role in selecting the treatment regimen[3,4]. The tumor-node-metastasis (TNM) staging system is the current gold standard for predicting prognosis in lung cancer, where patients in different stage cohorts experience different survival and recurrence rates[5]. However, these rates can vary within the same stage cohort and this has led to the American Joint Committee on Cancer (AJCC) TNM staging system being described as "too broad" when attempting to make individualized predictions for treatment guidance[6]. Studies have demonstrated that clinical features, such as age, gender and smoking status, can improve outcome stratification after treatment and allow for more personalized treatment decisions[7]. This laid the foundation for studies to explore alternative sources of information in an attempt to create more accurate and precise prognostic models.

Deep learning (DL) has been investigated for its ability to develop intricate associations between clinical variables and outcomes[8,9]. These methods have been shown to provide risk predictions for various clinical endpoints to better inform treatment decisions and allow clinicians to understand the factors contributing to the model. Additionally, medical images such as computed tomography (CT) and positron emission tomography (PET) used during the clinical workflow, have been shown to hold prognostic information and augment clinical variables[10,11]. DL has also been shown to be able to identify important information in these medical images to provide predictions that can augment clinical information[12,13]. Therefore, there is a need to develop clinical tools

[1]Department of Medical Biophysics, Western University, 1151 Richmond Street, London, ON N6A 3K7, Canada. [2]Baines Imaging Research Laboratory, London Health Sciences Centre Research Institute, London, ON, Canada. [3]Division of Thoracic Surgery, Department of Surgery, Western University, London, ON, Canada. [4]Department of Oncology, Western University, London, ON, Canada. [5]Clinical Research Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. [6]Division of Pulmonary, Critical Care & Sleep Medicine, University of Washington School of Medicine, Seattle, WA, USA. [7]Verspeeten Family Cancer Centre, Room A4-821, 800 Commissioners Road East, London, ON N6A 5W9, Canada. ✉email: sarah.mattonen@uwo.ca

incorporating imaging and clinical variables to aid clinicians in their decision-making and further improve treatment outcomes.

The clinical implementation of prognostic models in oncology is facilitated by clinical decision-support systems (CDSSs). CDSSs aim to strengthen clinical decision making by allowing the clinician to combine their high-level knowledge with the information provided by the prognostic model[14]. These systems allow for patient-specific assessments enabling clinicians to migrate from population-based evidence to more personalized approaches[15]. Additionally, studies have shown that the incorporation of prognostic models into clinical practice can enhance predictive confidence and improve communication with patients[16]. Previous studies have attempted to develop CDSSs for lung cancer care[8,9,17–20]. However, many of these studies do not incorporate imaging and none have attempted a multi-modality imaging approach with clinical data alongside multiple imaging modalities. Additionally, surgery and pathologic information have been shown to be independently prognostic of outcomes in resected lung cancer, and information such as the margin status and the presence of lymphovascular invasion have yet to be incorporated into these CDSSs[21]. CDSSs incorporating all of this information may be able to better identify patients who require more aggressive adjuvant treatment options.

In this study, we describe a novel CDSS designed for outcome prognostication of resected non-small cell lung cancer (NSCLC) patients. This system incorporates an externally validated deep learning model (DLM) using patients' pre-treatment[18]F-fluoro-2-deoxy-D-glucose (FDG)-PET/CT image, clinical, surgical, and pathological information to determine prognosis after surgery and subsequently stratify them into risk groups. Additionally, we present an exploratory feasibility user-study evaluating clinician performance with and without the CDSS DLM information, the usability of the system by clinicians, and clinician attitudes toward CDSSs. Through this investigation, we aim to obtain an initial assessment of the CDSS's predictive accuracy, user-friendliness, and alignment with clinical decision-making. To our knowledge, this is the first CDSS incorporating a multi-modality model for NSCLC prognostication as well as the first study of clinician attitudes towards CDSSs within the Canadian clinical context. This study will inform future clinical trials and best practices of multi-modality CDSSs for NSCLC prognostication.

## Methods

### Patient population

Institutional review ethics board approval was obtained, and patient informed consent was waived for this retrospective study. All research conducted was performed in accordance with the Declaration of Helsinki. The study included 10 NSCLC patients referred for upfront surgical resection between 2012 and 2018 at the Verspeeten Family Cancer Centre. This cohort of patients was randomly selected from the testing dataset of a previous study conducted for the development of the DLM[22]. In summary, the patients included pathologically confirmed stage I to III NSCLC according to the AJCC 7th and 8th edition and each patient received adjuvant therapy based on institutional clinical guidelines.

### Imaging and clinical information

The information collected for each patient included the preoperative FDG PET/CT scan, clinical (age, sex, and smoking status), surgical (surgery type and margin status), and tumor characteristics (primary site, histology, pathological AJCC TNM stage, lymphovascular invasion, and visceral pleural invasion). The outcome of interest predicted by the DLM was recurrence-free survival (RFS) and was defined as the interval from surgery to evidence of disease recurrence or death. Recurrence was diagnosed radiologically and/or pathologically, and was defined as any local, regional, or distant disease after surgery.

### Clinical decision-support system design

The CDSS was developed to integrate a multi-modality DLM, incorporating clinical, surgical, pathological, and imaging information[22]. The CDSS could be navigated by the clinicians and provide them with the patient's prognosis using the validated DLM (Fig. 1). The CDSS was created using MATLAB R2022B (The MathWorks, Natick, MA) for the development of the interface and Python 3.8.18 for integration of the validated DLM model. The interface was modified to facilitate the user-study where it could collect the user's risk predictions, decision confidence, and treatment recommendations (Fig. 2).

The interface consists of four panels: (1) the imaging panels—used to navigate the patient's registered FDG PET/CT scan in the axial, sagittal, and coronal planes, (2) the patient information panel—used to display the patient's clinical, surgical, and pathological information, (3) the CDSS DLM information panel—used to display information from the DLM on the recurrence, RFS, and overall survival (OS) of the patient, and (4) the patient assessment panel—used to record the user's prediction, confidence, recommendation, and the treatment failure probability of the patient.

### Clinical decision-support system characteristics

The CDSS first loads the PET and CT images, as well as the clinical information of the patient. The user can view all slices of the images in the axial, sagittal and coronal planes, adjust the image fusion weight, and adjust the window and level values for the CT image and the standardized uptake value (SUV) limits for the PET image. The user can also hover the cursor over regions in the image to obtain the SUV value of the area. The patient's pathology report and PET/CT report are also available to view on the top panel of the interface. For this study, the CDSS was designed with a button labelled "Show Model Risk" which loads the patient data into the DLM and produces a prognosis score and risk category for each patient. The three available risk categories were high, medium, and low. The developed CDSS takes approximately ten seconds per patient to initialize and run the DLM. The color of the CDSS information panel then changes based on the result: red for high-risk patients, orange for medium-risk patients and green for low-risk patients. Lastly, the user has access to

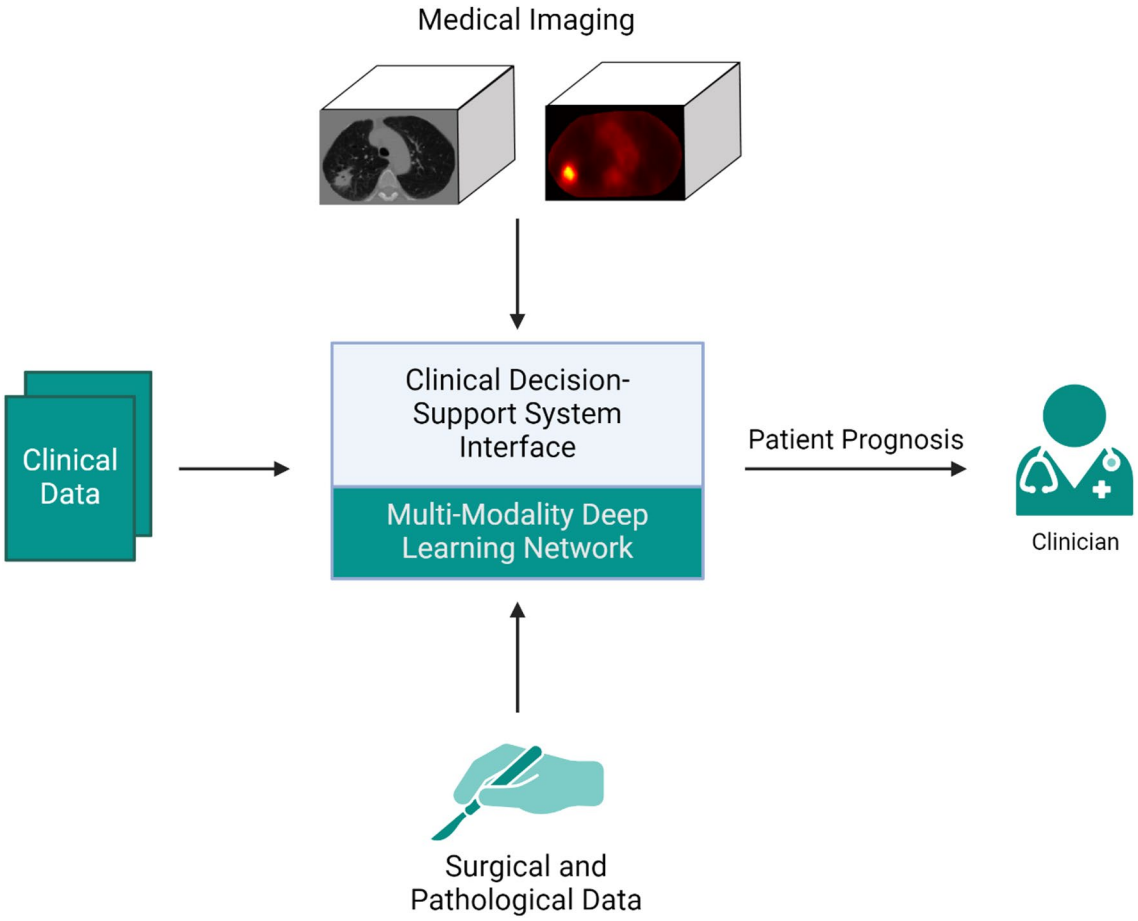


**Fig. 1**. Diagram of the functionality and interactions of the clinical decision-support system (CDSS). The system is composed of a validated multi-modality neural network which uses the patient's clinical, surgical, pathological, and imaging information to generate an output of recurrence-free survival (RFS) which is presented to the end-user (e.g. the clinician).

the recurrence, RFS, and OS Kaplan-Meier curves for the patients in the specified risk group as determined by external validation of the DLM ($n = 195$)[22].

## User study

*Participants*

Four clinicians were recruited to participate in the study. This included two thoracic surgical oncologists (users 1 and 2) and two medical oncologists (users 3 and 4) specializing in the treatment of NSCLC. All users were blinded to the clinical outcomes and each patient was anonymized and randomized for each user. All users were staff clinicians but spanned a wide range of experience levels and backgrounds which was key in obtaining a wide range of perspectives.

*Decision-support system study protocol*

All users were first given a brief orientation and introduction to the study design, the areas and functionalities of the CDSS, the DLM, number of cases to be read, the expected duration, and the outcomes of interest. To mitigate the assumption of a low clinical prevalence based on prior experience, users were informed that the 10-patient cohort contained 20–30% recurrences and 40–50% deaths: an event rate between 50–60%[23].

The users were then asked to utilize the CDSS on ten NSCLC patients. Following the navigation of the patient's information, the user was asked to score the treatment failure (defined as local, regional, or distant metastasis) risk for all patients as high-risk, intermediate-risk, or low-risk. The user was also asked to give a personal risk score for the patient between 0 and 100 indicating their assessment of the patient's probability of treatment failure. Additionally, the user selected their confidence level in their prediction on an ordinal scale which ranged from none, somewhat, moderately, very, and completely. The user was then asked to indicate their post-surgery recommendations from the following options: surveillance, adjuvant therapy (which could indicate adjuvant chemotherapy or radiotherapy), and multi-disciplinary team (MDT) review. This process was completed for all patients without the aid of the CDSS DLM information, then subsequently for all patients with the CDSS DLM information in a new randomized order. After entering the data for each patient, the responses were locked, and the next patient was automatically loaded into the system.

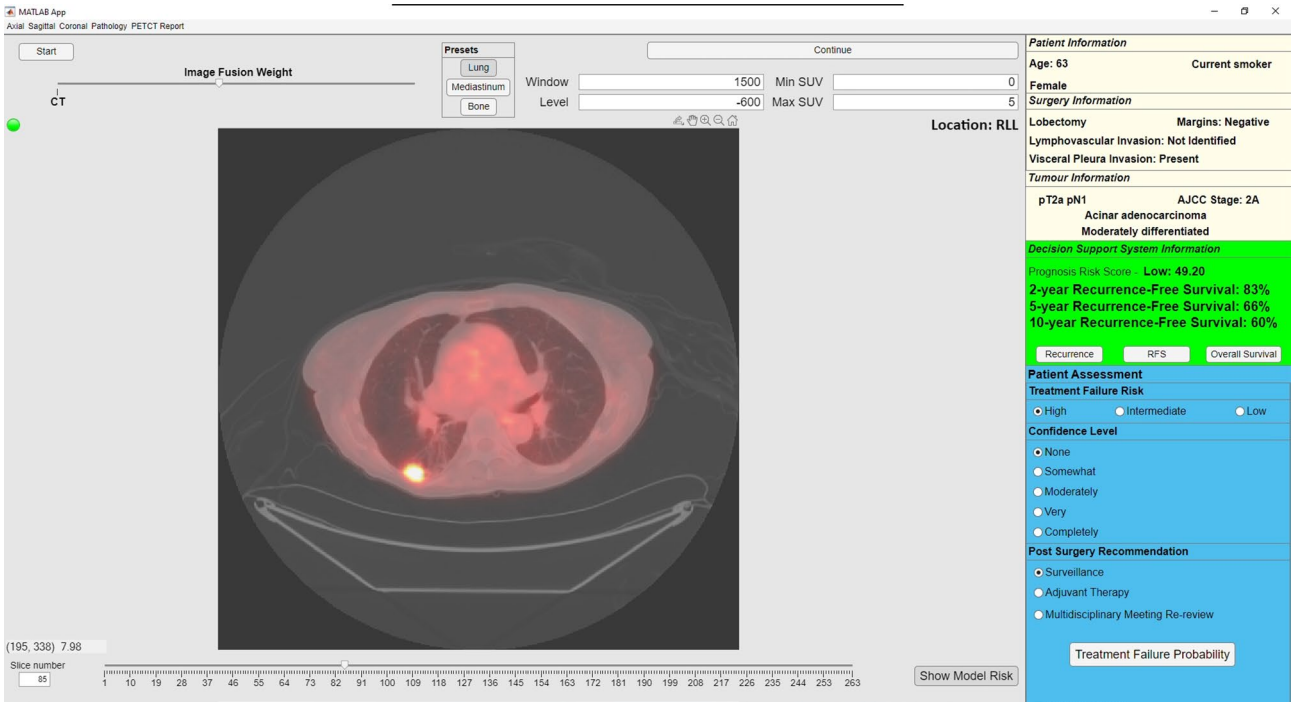


**Fig. 2**. Clinical decision-support system (CDSS) interface utilized for the assessment of treatment failure risk prediction, confidence level and post-surgical recommendation in the user study.

The area under the receiver operating curve (AUC) for predicting treatment failure was determined for all users with and without the CDSS DLM information and was compared to the performance of the CDSS DLM alone. The DeLong test was used to compare each user's prognostic performance to chance as well as the prediction before the use of the DLM to that with the DLM. Paired t-tests were conducted to compare the changes in all users' predictions, confidence levels, and recommendations before and after the use of the CDSS DLM information. The inter-user agreement for the user's prediction, confidence level, and recommendation for each patient was assessed both with and without the CDSS DLM information using the Fleiss-Kappa metric ($\kappa$) where the value was between 0 and 1 as follows: $\kappa < 0$ – no agreement, $0 < \kappa < 0.2$ – slight agreement, $0.2 < \kappa < 0.4$ – fair agreement, $0.4 < \kappa < 0.6$ – moderate agreement, $0.6 < \kappa < 0.8$ – substantial agreement, and $0.8 < \kappa < 1$ – almost perfect agreement[24,25]. Statistical analysis was conducted in MATLAB R2022B (The MathWorks, Natick, MA)[26].

*Qualitative study protocol*

The four users were interviewed in a semi-structured format after completing the CDSS portion of the study. The questions covered the following topics: their experience using the CDSS (usability, efficiency, deficiencies, and added clinical value), attitudes towards CDSSs integrating prognostic models, barriers to CDSS clinical implementation, the role of explainable artificial intelligence (AI) in supporting successful clinical implementation, and the role of CDSS and prognostic models in patient communication[15]. The structured questions can be found in Online Appendix A. The interviews were conducted by J. R. C and were audio-recorded once consent from the user was obtained.

## Results

### Patient population

A summary of the demographics of the patient population ($n = 10$, mean age ± standard deviation = 68 years ± 9; 6 women) used in the study can be found in Table 1. The median RFS was 52 months, and the median follow-up time was 52 months.

### User-study results

*Participants*

The four staff clinician participants recruited for the study had a range of clinical experience, ranging from 4 to 30 years, with an average of 11 years. All clinicians had at least basic familiarity with AI models and the concept of CDSSs for improving lung cancer care.

*Comparison of CDSS to clinician performance*

All users' scores were found to be significant predictors of recurrence in the 10-patient cohort with the CDSS DLM information when either predicting a patient's risk score or risk group (Table 2). Each user's stratification

| Clinical variable | All patients ($n = 10$) |
|---|---|
| Mean age [range] | 66 [46–86] |
| Sex | |
| Men | 4 (40%) |
| Women | 6 (60%) |
| Histology | |
| Squamous cell carcinoma | 1 (10%) |
| Non-squamous cell carcinoma[a] | 9 (90%) |
| Smoking status | |
| Former | 2 (20%) |
| Current | 8 (80%) |
| Pathological stage[b] | |
| I | 5 (50%) |
| II | 3 (30%) |
| III | 2 (20%) |
| Post-op therapy | |
| Yes | 4 (40%) |
| No | 6 (60%) |
| Surgery type | |
| Sublobar resection | 3 (30%) |
| Lobectomy | 7 (70%) |
| Primary site | |
| RUL | 5 (50%) |
| RLL | 1 (10%) |
| LUL | 3 (30%) |
| LLL | 1 (10%) |
| Margin status | |
| Negative (> 5 mm) | 5 (50%) |
| Close (1–5 mm) | 3 (30%) |
| Positive (< 1 mm) | 2 (20%) |
| Lymphovascular invasion | |
| Absent | 9 (90%) |
| Present | 1 (10%) |
| Visceral pleural invasion | |
| Absent | 7 (70%) |
| Present | 3 (30%) |

**Table 1**. Clinical characteristics for the NSCLC patients analyzed in the study. Unless otherwise indicated, data is presented as number of patients and data in parentheses are percentages. [a]Includes adenocarcinoma, large cell carcinoma, or non-small cell cancer not otherwise specified. [b]Based on American Joint Committee on Cancer (AJCC) 7th and 8th Editions.

| | | Surgical oncologists | | | | Medical oncologists | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | User 1 | | User 2 | | User 3 | | User 4 | |
| | CDSS[a] | No DLM | DLM | No DLM | DLM | No DLM | DLM | No DLM | DLM |
| Risk group | | | | | | | | | |
| AUC[b] (95% CI) | 0.83* (0.58, 1) | 0.71 (0.39, 1) | 0.79** (0.59, 0.98) | 0.74 (0.45, 1) | 0.76 (0.50, 1) | 0.59 (0.18, 1) | 0.69 (0.32, 1) | 0.57 (0.24, 0.90) | 1.00** (1, 1) |
| Delta AUC | – | + 0.08 | | + 0.02 | | + 0.10 | | + 0.43 | |
| [c]Delta *p*-value | – | 0.48 | | 0.84 | | 0.77 | | 0.01 | |
| Risk score | | | | | | | | | |
| AUC[b] (95% CI) | 0.83* (0.53, 1) | 0.81* (0.51, 1) | 0.62 (0.25, 0.99) | 0.81* (0.52, 1) | 0.95** (0.84, 1) | 0.74 (0.36, 1) | 0.90** (0.68, 1) | 0.62 (0.24, 1) | 0.86** (0.60, 1) |
| Delta AUC | – | − 0.19 | | + 0.14 | | + 0.16 | | + 0.24 | |
| [c]Delta *p*-value | – | 0.27 | | 0.35 | | 0.44 | | 0.10 | |

**Table 2**. Clinical decision-support system deep learning model (DLM) compared to user assessment of post-surgical treatment failure with and without DLM information. [a]Result is that of the DLM alone with no input from user. [b]DeLong Test: $*p < 0.05$, $**p < 0.01$. [c]Paired DeLong Test.

performance improved after using the CDSS DLM information, however, only one of these improvements was shown to be statistically significant. This overall improvement was demonstrated numerically by the positive delta-AUCs observed when comparing the risk group (high, intermediate, low) AUCs achieved by each user with and without the DLM information as seen in Table 2. Additionally, most users saw a numerical, non-significant improvement in their ability to attribute a treatment failure probability, or risk score, corresponding to the true event of the patient. Medical oncologists saw a greater improvement in stratification performance after using the CDSS DLM information, however, this trend did not demonstrate significance ($p = 0.07$). The DLM integrated into the CDSS, without user input, was found to be a significant predictor of recurrence in the 10-patient cohort achieving an AUC of 0.83 [95% CI 0.53–1, $p = 0.03$]. Compared to the users without the CDSS, the DLM was able to better stratify patients into treatment failure risk groups. However, three out of the four users were able to achieve a higher AUC than the DLM alone when using the DLM information for the risk score prediction.

*Impact of clinical decision-support system on user predictions, confidence and recommendations*

The predictions of treatment failure risk for the 10 patients without the CDSS DLM information were relatively aligned between the clinicians (Fig. 3). There was a moderate level of agreement among the four users, with a κ value of 0.42 [95% CI 0.37–0.47, $p < 0.001$]. Although performance improved for most users after using the CDSS information, the agreement between the users decreased. With the CDSS DLM information provided to the user, there was only a slight agreement among the four users, with a κ of 0.18 [95% CI 0.12–0.23, $p = 0.08$]. Intermediate and high-risk groups predictions were higher on average after using the DLM information, however, this trend did not demonstrate significance ($p = 0.23$).

An example of the change in the clinician predictions due to the use of the CDSS DLM information is seen with the stage IB patient. Without the CDSS DLM information, one user identified the patient as low risk while three users identified them as intermediate risk. With the CDSS information, three users now identified the patient as high-risk while one identified them as intermediate risk. This patient did not present with any conventional high-risk features after surgery and did in fact recur.

The confidence level of each clinician's prediction both with and without the CDSS DLM information varied widely, as shown in Fig. 4. The confidence level of the users increased on average numerically with the use of the CDSS DLM information, demonstrated by the increase in proportion of the "very" selection and the decrease in the proportion of the "moderate" selection. However, this trend did not demonstrate statistical significance ($p = 0.22$). Additionally, without the CDSS DLM information, there was a slight agreement among the four users, with a κ value of 0.15 [95% CI 0.10–0.20, $p = 0.14$]. This agreement decreased with the use of the CDSS DLM information, with a κ value of 0.12 [95% CI 0.06–0.17, $p = 0.24$].

The post-surgical recommendations for the 10 patients without the CDSS DLM information were relatively aligned between the clinicians as shown in Fig. 5 and there was a moderate level of agreement among the four users, with a κ value of 0.57 [95% CI 0.52–0.62, $p < 0.001$]. However, after using the CDSS DLM information, the agreement between the users decreased but still demonstrated a significant agreement. With the CDSS DLM information provided, there was a fair level of agreement among the four users, with a κ of 0.33 [95% CI 0.27–

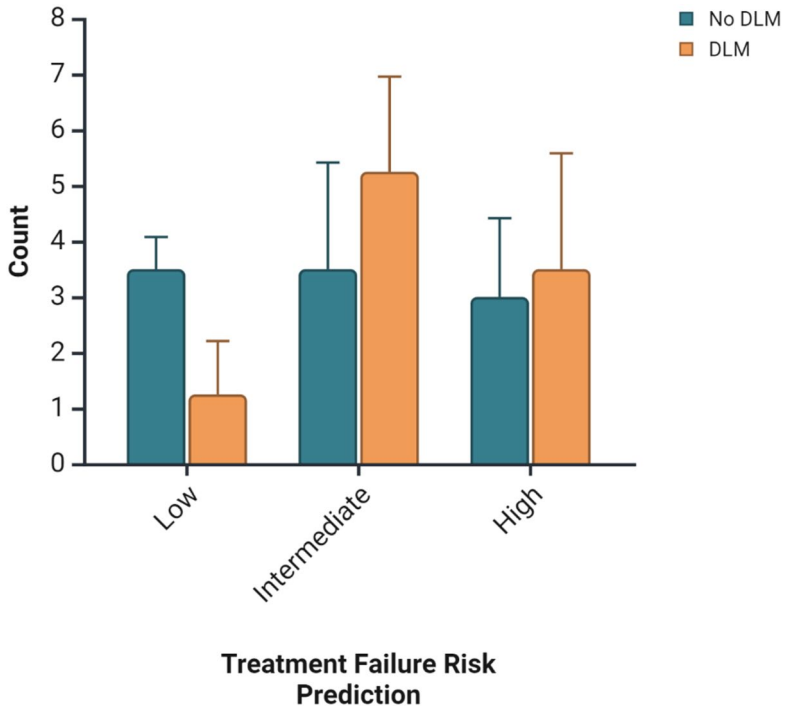


**Fig. 3**. Average and standard deviation of clinician assessments for post-surgical treatment failure risk with and without the clinical decision-support system deep learning model (DLM) information.

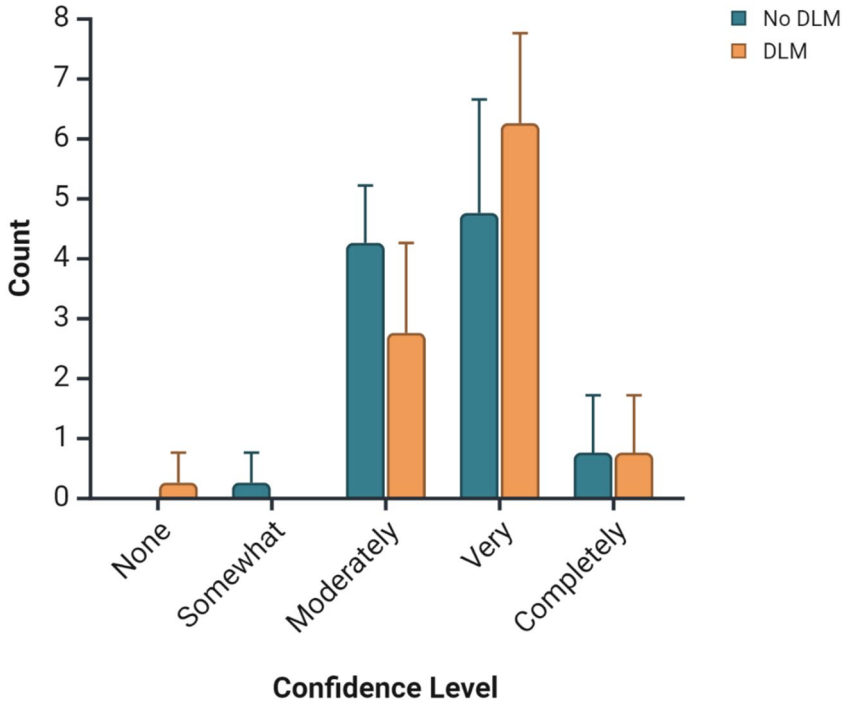


**Fig. 4**. Average and standard deviation of clinician confidence levels for treatment failure risk prediction with and without the clinical decision-support system deep learning model (DLM) information.

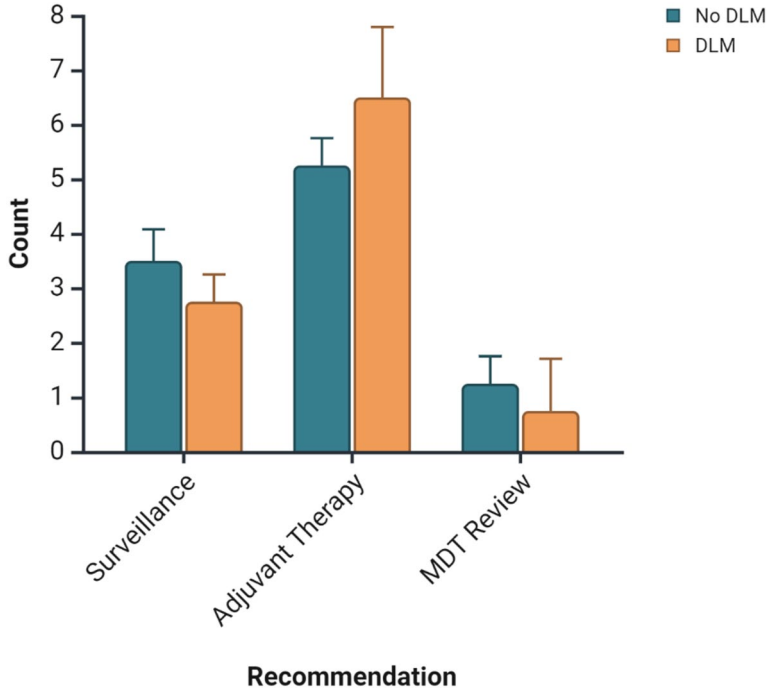


**Fig. 5**. Average and standard deviation of clinician post-surgical recommendations with and without the clinical decision-support system deep learning model (DLM) information.

0.38, $p < 0.01$]. The impact of the CDSS DLM information on the recommendations agreement within each specialty was different. The surgical oncologists saw a decrease in their agreement of recommendations when using the CDSS DLM information (without DLM: $\kappa = 0.48$ [95% CI 0.35–0.61, $p = 0.05$, moderate agreement]; with DLM: $\kappa$ = -0.03 [95% CI -0.15–0.10, $p = 0.91$, poor agreement]). Conversely, the medical oncologist saw a significant increase in their agreement of recommendations when using the CDSS DLM information (without

DLM: κ = 0.49 [95% CI 0.37–0.61, $p = 0.04$, moderate agreement]; with DLM: κ = 0.73 [95% CI 0.57–0.90, $p = 0.02$, substantial agreement]). It is important to note that these results were found with a $n = 2$ for each sub-specialty.

On average, numerically, the users were more likely to recommend adjuvant therapy to patients after using the CDSS DLM information. However, this trend did not demonstrate statistical significance ($p = 0.25$). The medical oncologists on average increased their proportion of adjuvant treatment recommendations when using the DLM information from 55% of the patients to 75% of the patients, respectively. For the surgical oncologists, the average proportion of each recommendation stayed relatively consistent, however, they were more likely to swap patients from one group to another. This is represented by their decrease in agreement. Additionally, on average, there was no significant difference in the patients referred for MDT review, however, some patients were referred for individual users.

## Qualitative study results

*User attitudes towards developed CDSS interface*

All users had an extremely positive experience using the CDSS interface, with an average rating of eight out of ten when describing the user-friendliness of the system. Common themes that contributed to the positive experience were the presence of the pathology and imaging reports, the ability to investigate the SUV of the PET imaging on a voxel-by-voxel basis, the color presentation of the interface, and the time efficiency. The average rating among clinicians of the CDSS with regards to time efficiency was nine out of ten. However, users 2 and 4 highlighted a concern of the added time that would be incurred when implementing a system like this clinically due to the interaction with a clinic's picture archiving and communication system (PACS). The acceptable time a CDSS should take varied among clinicians from two to five minutes.

In terms of added clinical value, all users indicated that the CDSS interface and information provided a clinical benefit with making their decision. User 1 stated that when the CDSS DLM information indicated a patient was at a high risk of experiencing an event, contrary to their initial belief, it made them more likely to recommend the patient to be discussed by the multidisciplinary team. Those patients may not have been previously seen or recommended to receive adjuvant therapy. However, that clinician was less likely to believe the CDSS DLM information if it predicted a patient as having a low risk, contrary to their prediction. Additionally, user 4 identified that the CDSS information was a useful tool when other sources of information cannot be relied on as heavily, for example a lack of genetic markers or poor imaging quality.

The importance of external validation of prognostic models and large prospective clinical trials was emphasized by all clinicians. Multiple users (users 2 and 4) stated that it may even be beneficial to provide the validation results alongside the model for clinician inspection. Additionally, most users highlighted the importance of the patient's age and performance status in recommending treatment. All users were interested in how age was incorporated into the DLM and stated that performance status metrics would be important to include in future systems. Lastly, the majority of users highlighted the difficulty in interpreting the prognostic score provided by the CDSS. Users stated that if implemented clinically, an improvement to the system would be the elimination of the prognosis score and only showing the predicted risk group or only showing the recurrence rates. However, two out of four clinicians (users 2 and 4) used the recurrence, RFS, and OS information when making their decisions, highlighting the importance of clinical endpoints other than recurrence.

*User attitudes towards prognostic models*

All users stated that they believe prognostic models and CDSSs can add value in improving the treatment of lung cancer care but highlighted the lack of rigorous validation and prospective clinical trials as a barrier to the clinical implementation and adoption of these systems. User 4 indicated that they believe these models have a particularly important role in the adjuvant setting. However, their concern was not only limited to lack of external validation but also lack of validation on local data. They stated that differences in acquisition protocols, image quality, and data quality may affect these models, and local validation is crucial before implementation in local treatment management.

In terms of trust in prognostic models, all users trusted or wanted to trust prognostic models. Although the experience with DLMs varied between the users, all users were willing to implement and use these models clinically if adequately validated. Additionally, all users showed hesitation when asked if explainable AI would improve their trust and understanding of prognostic models. Most users highlighted that explainable AI images would need further validation and could confuse the user rather than augment the clinical value of the model.

Users 2, 3, and 4 were strong advocates of using prognostic models for patient communication. They highlighted that patients may feel more comfortable seeing more personalized prognostication and it may summarize that vast amount of information they are exposed to into one digestible number. User 1 emphasized that patients may become overwhelmed when presented with these DLMs which may lead to more questions than answers.

## Discussion

The development of prognostic models for improving lung cancer treatment outcomes has been widely studied[10,12]. More specifically, they have been shown to improve resected NSCLC patient outcome stratification, potentially identifying patients that would benefit from adjuvant therapies. While these models promise to augment clinical decision-making, few studies have explored how best these models should be integrated into the clinical workflow or assessed clinical utility to maximize their clinical value. In this study, we developed a user-friendly and time efficient CDSS interface integrating a validated DLM for resected NSCLC prognostication. We observed a numerical improvement in clinician performance when using the CDSS for post-resection recurrence risk stratification, although this change did not reach statistical significance. Additionally, the CDSS interface received overwhelmingly positive reviews for its user-friendliness, time efficiency, and presentation. To

our knowledge, this is the first study investigating the impact of a DLM for resected NSCLC prognostication on clinician performance, as well as the first study to investigate how a multi-modality CDSS can affect clinician decision-making and prediction confidence across multiple specialties.

The results from this study suggest that a validated DLM, once packaged with the right information, has the potential to provide clinical utility in a risk-stratification setting. There was a modest, non-significant improvement in stratification performance by all clinicians, where three out of the four users achieved an improved prediction result when using the CDSS DLM information compared to the CDSS on its own. However, while a directional improvement in performance was demonstrated, the agreement between the users decreased with the use of the CDSS DLM information. This was also the case for the confidence level and post-surgical recommendations. This may represent the differing degrees of adoption by each user and may cause more cases to undergo MDT review, which may be the best use of these CDSSs in the short to medium term. Conventionally low-risk patients who would not have been previously referred for adjuvant treatment or MDT review may now have a better opportunity to receive additional discussion and possible treatment, whether proven or experimental. However, given a tool like this may influence clinician behavior, caution must be taken to avoid the risk of overtreatment. While the quantitative results of this study were promising, due to the limited number of users and patients in the study, these results are prone to random variation and overfitting and thus need to be validated in future studies. However, these results may indicate a pathway to developing a system which can impact clinician decisions.

Qualitatively, the positive experience from all the users showed that while model performance is important, the presentation of the information to the clinician is equally so. Poor clinical integration can impact the adoption of these systems by clinicians and derail the best performing models. The key takeaway from the users was the importance of having all the patient information integral in decision-making available in one interface. Users enjoyed the ability to navigate the imaging, pathology, and clinical information while utilizing the CDSS. This demonstrates that future CDSSs should aim to provide not only the information used by the model, but all information relevant to the patient in one system. This will allow for an easier integration of DLMs into the clinical workflow.

There has been previous work in the development of user-interfaces for models in lung cancer prognostication. Lei et al. and She et al. both developed interfaces integrating DeepSurv models for NSCLC survival prediction using only clinical information[8,9]. These systems require the user to input variables into the interface and did not include imaging or pathology information. Other CDSS studies involve the validation of the underlying model, but do not attempt to integrate the model into clinical practice with a frontend interface[17,19]. To our knowledge, no studies have attempted to develop a CDSS integrating imaging, pathological, surgical, and clinical information. Additionally, previous studies have aimed to assess clinicians' attitudes towards CDSSs qualitatively. However, this work builds on those studies by having users interact with a validated CDSS DLM, evaluating the perspectives of medical and surgical oncologists, and obtaining the perspective from a Canadian lens[15,27,28].

Many of the users highlighted the ever-changing landscape of lung cancer care. Due to this and the retrospective nature of the study, a transition to neoadjuvant/perioperative chemoimmunotherapy, or the addition of adjuvant immunotherapy after chemotherapy, and genomic alterations (i.e. EGFR and ALK) which are now integral in decision-making were not included in the study[29]. Future work should incorporate these markers into models and interfaces for clinician access. Another limitation of this study is the limited number of users involved. A larger number of users would allow for a better characterization of the CDSS's impact on decision-making and allow for more substantial conclusions about clinician attitudes towards the CDSS's integration into the clinic to be made. Additionally, users from other specialties such as radiation oncology would provide another perspective of the utility of the CDSS. The ten-patient cohort assessed in this study is another limitation. A larger cohort of patients analyzed would allow the users to experience the CDSS under a wider range of clinical scenarios and give a better indication of the CDSS's added value. Additionally, the patients in the study may not have been representative of the clinical outcomes seen clinically; the death rate of patients without a recurrence was higher than normally seen for NSCLC patients (4/10). This representation is difficult to achieve with a limited sample of patients and can affect the generalizability of the findings in this study. Future studies should look to compile a cohort of patients with similar outcome demographics seen clinically. Lastly, the development of the CDSS in MATLAB meant there was no system integration with the clinical PACS. Future iterations of the CDSS interface must be integrated into the hospital's system to optimize the exchange of patient data and recommendations. This may facilitate the integration of more detailed clinical data into the CDSS interface such positive nodal locations and margin distance information.

## Conclusions

In summary, this exploratory study demonstrated the potential of CDSSs implementing multi-modality DLMs for resected NSCLC treatment decisions when designed with a clinician-friendly and time efficient interface. The developed CDSS showed non-significant, numerical improvements in clinician prognostic performance and prediction confidence, although the study was limited by the sample size. Clinicians demonstrated a willingness to adopt the validated model predictions and provided valuable feedback that will improve CDSS implementation for future multi-modality models.

## Data availability

The dataset analysed during the current study is not publicly available but is available from the corresponding author on reasonable request.

## Code availability

The code to the CDSS interface is available at https://github.com/baines-imaging-mattonen-lab/Multi-Modality-NSCLC-Outcome-Prediction-GUI as a git repository. This code needs to be integrated with a deep learning model using Python, medical images, and clinical information from a patient database.

Received: 24 July 2025; Accepted: 1 December 2025
Published online: 20 December 2025

## References

1. Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *Cancer J. Clin.* **73** (1), 17–48. https://doi.org/10.3322/caac.21763 (2023).
2. Brenner, D. R. et al. Projected estimates of cancer in Canada in 2022. *CMAJ* **194** (17), E601–E607. https://doi.org/10.1503/cmaj.212097 (2022).
3. Howington, J. A., Blum, M. G., Chang, A. C., Balekian, A. A. & Murthy, S. C. Treatment of stage I and II Non-small cell lung cancer: Diagnosis and management of lung Cancer, 3rd ed: American college of chest physicians Evidence-Based clinical practice guidelines. *Chest* **143** (5), e278S–e313. https://doi.org/10.1378/chest.12-2359 (2013).
4. Ramnath, N. et al. Treatment of stage III Non-small cell lung cancer: Diagnosis and management of lung Cancer, 3rd ed: American college of chest physicians Evidence-Based clinical practice guidelines. *Chest* **143** (5), e314S–e340. https://doi.org/10.1378/chest.12-2360 (2013).
5. Goldstraw, P. et al. The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for lung cancer. *J. Thorac. Oncol.* **11** (1), 39–51. https://doi.org/10.1016/j.jtho.2015.09.009 (2016).
6. Wang, Y., Lin, X. & Sun, D. A narrative review of prognosis prediction models for non-small cell lung cancer: What kind of predictors should be selected and how to improve models? *Ann. Transl. Med.* **9** (20), 1597. https://doi.org/10.21037/atm-21-4733 (2021).
7. Garinet, S. et al. Updated prognostic factors in localized NSCLC. *Cancers* **14** (6), 1400. https://doi.org/10.3390/cancers14061400 (2022).
8. Lei, J. et al. The predictive value of modified-DeepSurv in overall survivals of patients with lung cancer. *iScience* **26** (11), 108200. https://doi.org/10.1016/j.isci.2023.108200 (2023).
9. She, Y. et al. Development and validation of a deep learning model for Non–Small cell lung cancer survival. *JAMA Netw. Open.* **3** (6), e205842. https://doi.org/10.1001/jamanetworkopen.2020.5842 (2020).
10. Christie, J. R. et al. Artificial intelligence in lung cancer: Bridging the gap between computational power and clinical decision-making. *Can. Assoc. Radiol. J.* **72** (1), 86–97. https://doi.org/10.1177/0846537120941434 (2021).
11. Lambin, P. et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.*, **14**(12), 12. https://doi.org/10.1038/nrclinonc.2017.141 (2017).
12. Avanzo, M., Stancanello, J., Pirrone, G. & Sartor, G. Radiomics and deep learning in lung cancer. *Strahlenther. Onkol.* **196** (10), 879–887. https://doi.org/10.1007/s00066-020-01625-9 (2020).
13. Chen, B., Zhang, R., Gan, Y., Yang, L. & Li, W. Development and clinical application of radiomics in lung cancer. *Radiat. Oncol. (London England)*. **12** https://doi.org/10.1186/s13014-017-0885-x (2017).
14. Sutton, R. T. et al. An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digit. Med.* **3**, 17. https://doi.org/10.1038/s41746-020-0221-y (2020).
15. Ankolekar, A. et al. Clinician perspectives on clinical decision support systems in lung cancer: Implications for shared decision-making. *Health Exp.: Int. J. Public. Particip. Health Care Health Policy*. **25** (4), 1342–1351. https://doi.org/10.1111/hex.13457 (2022).
16. Hallen, S. A. M., Hootsmans, N. A. M., Blaisdell, L., Gutheil, C. M. & Han, P. K. J. Physicians' perceptions of the value of prognostic models: The benefits and risks of prognostic confidence. *Health Exp.: Int. J. Public. Particip. Health Care Health Policy*. **18** (6), 2266–2277. https://doi.org/10.1111/hex.12196 (2015).
17. Niraula, D. et al. A clinical decision support system for AI-assisted decision-making in response-adaptive radiotherapy (ARCliDS). *Sci. Rep.* **13**, 5279. https://doi.org/10.1038/s41598-023-32032-6 (2023).
18. Zhang, S. et al. Intelligent prognosis evaluation system for stage I-III resected non-small-cell lung cancer patients on CT images: A multi-center study. *eClinicalMedicine* **65** https://doi.org/10.1016/j.eclinm.2023.102270 (2023).
19. Dekker, A. et al. Rapid learning in practice: A lung cancer survival decision support system in routine patient care data. *Radiother. Oncol.* **113** (1), 47–53. https://doi.org/10.1016/j.radonc.2014.08.013 (2014).
20. Clay, R. et al. Computer aided nodule analysis and risk yield (CANARY) characterization of adenocarcinoma: Radiologic biopsy, risk stratification and future directions. *Transl. Lung Cancer Res.* **7** (3), 313–326. https://doi.org/10.21037/tlcr.2018.05.11 (2018).
21. Wang, C., Wu, Y., Shao, J., Liu, D. & Li, W. Clinicopathological variables influencing overall survival, recurrence and post-recurrence survival in resected stage I non-small-cell lung cancer. *BMC Cancer*. **20** (1), 150. https://doi.org/10.1186/s12885-020-6621-1 (2020).
22. Christie, J. R. et al. Thorax-encompassing multi-modality PET/CT deep learning model for resected lung cancer prognostication: A retrospective, multicenter study. *Med. Phys.* **52** (6), 4390–4402. https://doi.org/10.1002/mp.17862 (2025).
23. Mattonen, S. A. et al. Detection of local cancer recurrence after stereotactic ablative radiation therapy for lung cancer: Physician performance versus radiomic assessment. *Int. J. Radiat. Oncol. Biol. Phys.* **94** (5), 1121–1128. https://doi.org/10.1016/j.ijrobp.2015.12.369 (2016).
24. Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* **76** (5), 378–382. https://doi.org/10.1037/h0031619 (1971).
25. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33** (1), 159–174. https://doi.org/10.2307/2529310 (1977).
26. Shah, J. *Fleiss Kappa calculator & visualisation of video annotations* [Computer software].https://www.mathworks.com/matlabcentral/fileexchange/78023-fleiss-kappa-calculator-visualisation-of-video-annotations (2024).
27. Pluyter, J. R. et al. Looking through the eyes of the multidisciplinary team: The design and clinical evaluation of a decision support system for lung cancer care. *Transl. Lung Cancer Res.* **9**(4). https://doi.org/10.21037/tlcr-19-441 (2020).
28. Klarenbeek, S. E., Schuurbiers-Siebers, O. C. J., van den Heuvel, M. M., Prokop, M. & Tummers, M. Barriers and facilitators for implementation of a computerized clinical decision support system in lung cancer multidisciplinary team Meetings—A qualitative assessment. *Biology* **10** (1), 9. https://doi.org/10.3390/biology10010009 (2020).
29. Verma, S. et al. An updated review of management of resectable stage III NSCLC in the era of neoadjuvant immunotherapy. *Cancers* **16** (7), 1302. https://doi.org/10.3390/cancers16071302 (2024).

## Acknowledgements

The authors would like to acknowledge Dr. Aneesh Dhar, Dr. Edward Wang, and Dr. Tricia Chinnery for their

assistance and contributions to the conception and execution of this study.

## Author contributions

J.R.C. developed the clinical decision support system (CDSS), recruited participants to the study, conducted the study, analyzed the results, and was major contributor in writing the manuscript. K.E curated the patient and imaging data for use in the CDSS. R.A.M., M.Q., S.V. and D.B., participated in the study and assessed all patient information using the CDSS. V.S.N. and P.L. provided key clinical and interface development insight during the creation of the CDSS to ensure a frictionless clinician experience. S.A.M was responsible for overseeing the study, including conception of the study and research questions, interpretation of the results, and approval of the final manuscript. All authors read and approved the final manuscript.

## Funding

We acknowledge funding support from the Gerald C. Baines Foundation, donor support through the London Health Sciences Foundation, the Cancer Research Society (Grant 152218), the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants program (RGPIN-2020-06498), the Lawson Health Research Institute's Internal Research Fund (IRF) and the National Cancer Institute (NCI) (T32CA009515).

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-31204-w.

**Correspondence** and requests for materials should be addressed to S.A.M.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.