Answering Editors and Reviewers

We appreciate the editor and reviewers for their time and effort to review our paper and providing valuable comments to improve our manuscript. Please find below our point-by-point responses which are marked in blue color.

Reviewer #1:

Scientific Quality: Grade C (Good)

Language Quality: Grade B (Very good)

Conclusion: Major revision

Specific Comments to Authors:

The manuscript presents an innovative approach to predicting survival in esophageal cancer patients using radiomics and body composition analysis (BOA) at the T12 vertebral level from both pretreatment and follow-up CT scans. The study addresses an important clinical problem—limited access to L3-level imaging in follow-up scans—by proposing T12-level analysis as a practical alternative. The integration of radiomic, clinical, and BOA features into machine learning (ML) models, with logistic regression achieving strong performance (AUC = 0.91), is a notable contribution. The paper is well-structured, supported by adequate literature, and offers clinically meaningful insights.

Response: We sincerely thank the reviewer for the positive evaluation of our work. We are particularly grateful for the constructive comments and helpful suggestions, which have strengthened both the scientific merit and the clarity of our manuscript. Our detailed, point-by-point responses to the reviewer's comments are provided below.

Major Concerns

1. Lack of Validation: The study does not perform internal cross-validation or external multicenter validation. This limits generalizability and may overestimate model performance. Suggestion: Apply k-fold cross-validation and, if possible, validate with an independent cohort.

Response:

We appreciate the reviewer's emphasis on validation and fully acknowledge the absence of internal cross-validation and external multicenter testing as a limitation of our study. We apologize the original purpose statement led to misinterpretation, and for not applying k-fold cross-validation directly to our dataset and endpoint. Our rationale is as follows:

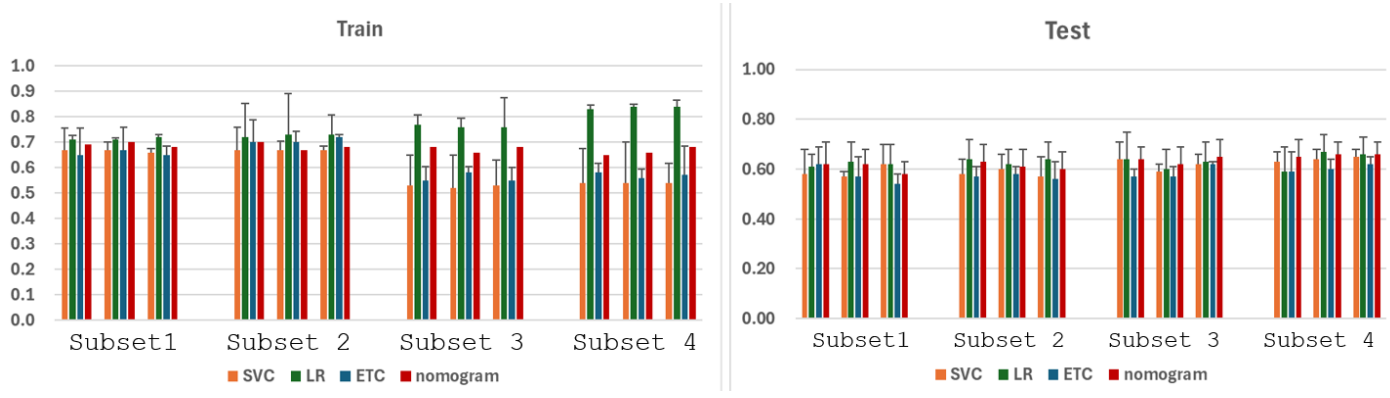
Study purpose. Our primary objective was to investigate the influential features—whether from individual subsets or combined subsets—among demographic, sarcopenia, and radiomics features derived from both initial and follow-up CT scans at the T12 level. This differs from previous studies that focused on the L3 level, which is frequently absent in follow-up CT. To our knowledge, this is the first study to specifically address the prognostic role of T12-level features in esophageal cancer.

Feature selection and leakage concern. We already applied LASSO with fivefold cross-validation for feature selection before Cox modeling and downstream classifiers. Applying an additional naïve k-fold CV on top of this process risks inflating accuracy if feature selection is not re-performed within each fold.

Time-to-event outcomes. Our endpoints rely on time-to-event measures (KM/log-rank and time-dependent AUC). Random k-fold partitioning mixes patients with heterogeneous follow-up intervals and censoring patterns across folds, which can bias both time-dependent AUC and risk calibration. Given our single-center cohort and the wide follow-up CT interval (60–180 days), preserving the natural time structure was preferable to random folds.
Role of models in the study. In our design, machine learning models and the nomogram served as mathematical tools to validate and illustrate the impact of the selected influential features within our patient cohort. The intention was not to produce a generalizable prognostic model for OS across all esophageal cancer populations.

We also note that several recent studies on machine learning–based prognosis in esophageal cancer have been published at the model development stage without internal or external validation. For example, Vogele et al. (Diagnostics 2024) applied CT radiomics of skeletal muscle for sarcopenia/progression assessment without validation (Ref. 26), and Zhang et al. (J Cancer Res Clin Oncol 2023) proposed a deep learning–based staging system for OS in ESCC without validation (Ref. 39). Thus, our approach is consistent with the current stage of methodological development in the field.

As additional context, we conducted an exploratory fivefold internal validation on our dataset. The results showed that AUCs in the training group were slightly lower than in the original analysis, while the logistic regression model in subset 4 (combined features) retained the same high AUC as reported originally. Lower AUCs in the test group were expected, but the relative superiority of subset 4 remained consistent. However, these analyses were performed using a preliminary, non-optimized model and were therefore not sufficiently rigorous for inclusion in the revised manuscript. We instead emphasize this limitation and explicitly note that robust internal validation and external multicenter validation will be undertaken in our planned future studies.



To avoid misinterpretation, we have revised the Purpose statement of the manuscript, modified the subsection title of Comparison of Model Performance to Features in subsets and OS years with corresponding contextual adjustments, and added clarifications in the Limitations section to explicitly note that both internal and external validation will be undertaken in future multicenter collaborations.

Revised Purpose:

"Therefore, this study aims to evaluate the influential features for esophageal cancer prognosis by integrating radiomic and body composition–based indices of skeletal muscle and adipose tissue at the T12 level from both pretreatment and follow-up CT images, combined with clinical and demographic data, in order to assess their value in predicting overall survival (OS)."

Limitations:

Fourth, this study is the absence of both internal cross-validation and external multicenter validation, limiting the generalizability and clinical applicability of the proposed models. While our primary aim was to identify influential features at the T12 level rather than to build a broadly generalizable predictive model, we acknowledge that the lack of validation may introduce optimism in performance estimates. To strengthen and validate our findings, internal and external validation using multicenter data should be conducted in future studies.

2. Sample Imbalance: Gender and histological subtypes are highly skewed (94% male, 97% squamous cell carcinoma). Findings may not apply to females or adenocarcinoma patients. Suggestion: Discuss in more depth how this imbalance may affect model performance and explore subgroup analyses.

Response:

We thank the reviewer for highlighting this important point. We fully agree that the sample imbalance may limit the generalizability of our findings to female patients or those with adenocarcinoma. The predominance of male patients with squamous cell carcinoma in our cohort reflects the epidemiology of esophageal cancer in East Asia. Nevertheless, we acknowledge that this imbalance restricts the broader applicability of our results. We have revised the manuscript to emphasize this limitation and have noted that future multicenter studies with larger and more diverse cohorts are required to validate model performance in female and adenocarcinoma subgroups.

The first limitation in the Limitations section has been modified as follows:

"First, this study was conducted retrospectively at a single center, involving a relatively small and heterogeneous patient cohort. For instance, gender distribution was markedly skewed, with only a small proportion of female patients (n = 12, 5.7%) and adenocarcinoma cases (n = 7, 3.3%), potentially introducing selection bias. Although our cohort consisted predominantly of male patients (94.3%) and squamous cell carcinoma cases (96.7%), this distribution reflects the epidemiology of esophageal cancer in East Asia [47]. Nevertheless, such demographic and histological imbalance may limit the generalizability of our findings to female patients or to those with adenocarcinoma. Future multicenter studies with larger and more balanced cohorts are warranted to validate model performance across diverse patient subgroups."

3. Reproducibility of Radiomics: CT images were acquired from different scanners and protocols, but harmonization techniques (e.g., ComBat) were not applied. Radiomic features are sensitive to such variations. Suggestion: Add harmonization or at least discuss potential bias introduced by scanner variability.

Response:

We thank the reviewer for raising this important point. We acknowledge that radiomic features can be sensitive to variations in scanner type and acquisition protocols. In our study, CT data were acquired from multiple scanners; however, standardized acquisition settings (120 kVp, 500 mm FOV, automatic tube current modulation) were applied to reduce variability. While harmonization methods such as ComBat could further mitigate inter-scanner effects, we did not apply these techniques in the current analysis. We have now revised the Limitations section to explicitly note this issue and to discuss potential bias introduced by scanner variability. We also emphasize the need for future multicenter studies with harmonization procedures to enhance feature reproducibility and generalizability.

The added content in the Limitations section is as follows:

"Second, due to the retrospective nature of the study spanning several years, variability in imaging acquisition parameters and reconstruction protocols may have affected the stability and reproducibility of radiomic features[48].CT images in this study were acquired from multiple scanners and reconstruction protocols. Although standardized acquisition settings were applied, radiomic features are known to be sensitive to scanner- and protocol-related variations[49]. We did not apply image harmonization techniques such as ComBat[50], which may have further improved feature reproducibility. This variability may introduce bias and should be addressed in future multicenter studies using harmonization methods to ensure robust model performance across institutions."

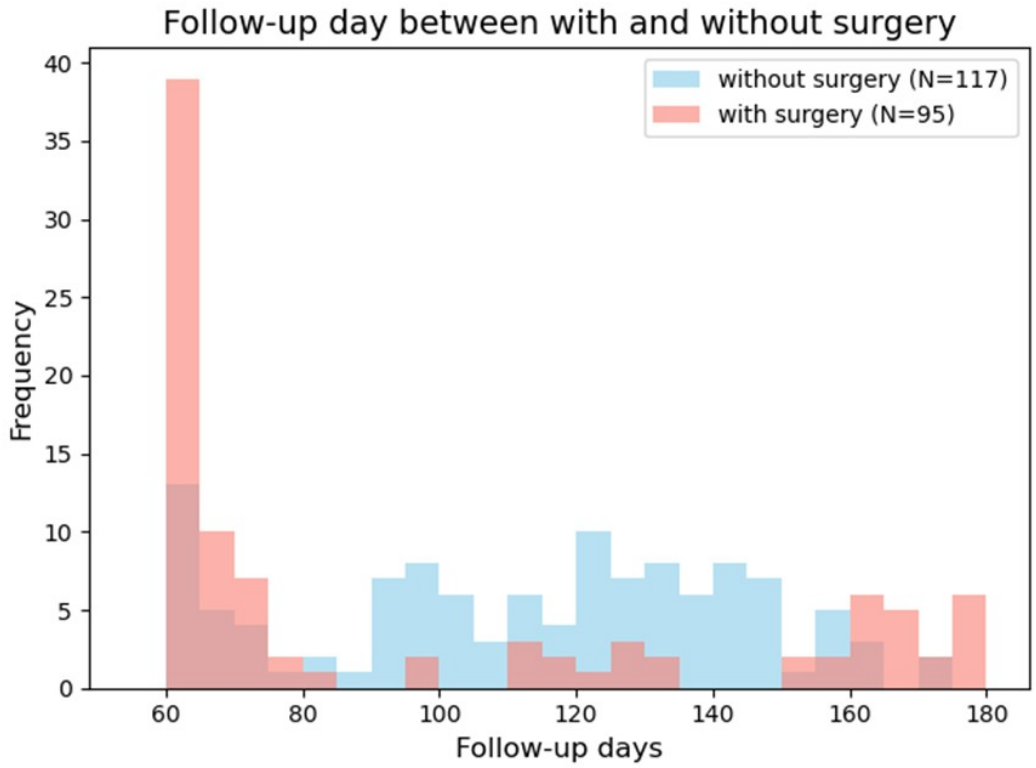
4. Follow-up CT Timing: The follow-up CT window (60–180 days) is wide, introducing variability. The optimal timing for prognostic CT should be investigated. Suggestion: Perform stratified analysis by narrower intervals.

Response:

We thank the reviewer for this thoughtful suggestion and agree that the timing of prognostic CT is an important factor in OS prediction. In our cohort, surgical patients generally had shorter intervals between CCRT and follow-up CT, while non-surgical patients had longer intervals (as shown in the following figure). This real-world variation made stratified analysis by narrower windows difficult, as it would create unbalanced subgroups by surgical status and reduce statistical power due to the small sample size. To maximize cohort size, we therefore included all eligible patients. We acknowledge that the wide interval may introduce variability and have revised the Limitations section to note that future studies should adopt standardized, narrower follow-up windows to better determine the optimal timing for prognostic imaging.

Revised text in the Limitations section:

"Third, the first follow-up CT scans were performed within a relatively wide window (60–180 days) after CCRT, partly reflecting clinical scheduling. This heterogeneity may introduce variability in prognostic assessment. Future studies should stratify patients by narrower intervals to determine the optimal timing for prognostic imaging."



5. Clinical Interpretation of Features: Many selected radiomic features (e.g., VAT_glrlm_RunEntropy1) lack clinical interpretability. While statistically valid, their biological meaning should be discussed to strengthen translational value.

Response:

We thank the reviewer for this valuable suggestion. We agree that many radiomic features, while statistically robust, may appear abstract and lack immediate clinical interpretability. To address this, we have expanded the Discussion to provide biological context for the most predictive features identified in our models. We have added the following text to the Discussion:

"Although certain radiomic features may appear abstract, they capture quantitative aspects of tissue heterogeneity not visible to the naked eye. For example, VAT_glrlm_RunEntropy quantifies heterogeneity in voxel intensity runs, which may reflect irregular adipose distribution caused by cachexia, therapy response, or treatment-induced metabolic and inflammatory changes—processes that are clinically relevant in esophageal cancer prognosis. Similarly, first-order features (e.g., mean intensity) and shape features (e.g., sphericity, SurfaceVolumeRatio) are interpretable as measures of tissue composition and morphology, reflecting atrophy, redistribution, or structural remodeling after treatment. Texture features derived from GLCM, GLDM, and GLRLM are crucial for capturing tissue heterogeneity[34], and histogram entropy from pre-treatment PET/CT has been associated with poorer recurrence-free and overall survival in gastroesophageal junction cancer[34]. Shape features have also shown high reproducibility and strong links to tissue morphology[46]. Furthermore, integrating CT radiomics of skeletal muscle with machine learning models has correlated with sarcopenia and predicted disease progression in gastric and esophageal tumors[26], and skeletal muscle radiomics is increasingly recognized as a potential biomarker for sarcopenia[21]. Thus, while not directly intuitive, these radiomic signatures likely encode meaningful biological processes underlying prognosis."

Minor Concerns

1. Figures (nomograms, KM curves, AUC plots) could be better labeled with higher resolution for clarity.

Response:

We thank the reviewer for this helpful suggestion. All figures (nomograms, Kaplan–Meier survival curves, and AUC plots) have been regenerated with higher resolution and improved labeling to enhance clarity and readability.

2. Some inconsistencies in terminology (e.g., "clinicl_M" typo in tables). Proofreading is needed.

Response:

We appreciate the reviewer for pointing this out. We have carefully proofread the manuscript and corrected all typographical errors and inconsistencies. Specifically, the variable name "clinicl_M" and "clinicl_T" in Tables 4 and 5 has been corrected to "clinical_M" and "clinical_T". We also reviewed all tables, figures, and text to ensure consistent terminology throughout the manuscript.

3. The abstract mentions "sensitivity 0.81, specificity 0.88," but does not specify which time point this corresponds to—clarify.

Response:

We thank the reviewer for this observation. The reported sensitivity (0.81) and specificity (0.88) correspond to the 2-year overall survival prediction using the logistic regression model with the combined input (Feature Subset 4), which demonstrated the best performance across all models (AUC = 0.91). To clarity, we have revised the Abstract as follows:

"The model integrating both BOA and radiomics from pretreatment and follow-up CT, combined with clinical data, achieved the best performance for 2-year overall survival prediction, with an AUC of 0.91, sensitivity of 0.81, and specificity of 0.88 using the logistic regression model."

4. Reference list is strong, but a few more comparative studies on deep learning in esophageal cancer prognosis (2022–2025) could be cited.

Response:

We thank the reviewer for this helpful observation. We have added four studies (2023-2025) on deep learning approaches in esophageal cancer prognosis to strengthen the reference list. Now, there are 13 studies (2022–2025) in References (Ref 24-27, 34–42). We added the context to include these citations in Discussion.

"Recently, a raised number of studies are investigating integrated models that used sarcopenia or radiomics with machine learning model to predict survival outcomes in esophageal cancer patients[24-27; 34-42]. Difference from those previous studies, in this study, we compared the performance of four different feature datasets—comprising demographic and clinical data, BOA indices, and radiomic features extracted at the T12 vertebral level from both pretreatment and follow-up CT images—across four predictive models: nomogram, SVC, LR, and ETC."

5. Consider shortening the discussion slightly—some sections repeat earlier points.

Response:

We thank the reviewer for this helpful suggestion. In the revised manuscript, we have shortened the Discussion and removed overlapping or repetitive content to improve clarity and conciseness.

Recommendation Major Revision: The study is scientifically sound and clinically relevant but requires additional validation, improved discussion of bias/limitations, and stronger focus on clinical interpretability before acceptance in a high-impact journal.
Response:
We sincerely thank the reviewer for the opportunity to revise our manuscript. We have carefully addressed all comments and made the corresponding revisions. We hope that the revised version now meets the acceptance criteria for WJGO.

Editorial Office's comments. Authors must revise the manuscript according to the Editorial Office's comments and suggestions, which are provided below:
(1) Science Editor:
1 Scientific quality: The authors submitted a retrospective study of machine learning survival prediction in esophageal cancer using radiomics and body composition from pretreatment and follow-up T12-level CT.
(1) Classification: Grade C;
(2) Summary of the Peer-Review Report: The paper is well-structured, supported by adequate literature, and offers clinically meaningful insights. However, the following problems still need to be solved. (1) The study does not perform internal cross-validation or external multicenter validation; (2) Sample imbalance; (3) Add harmonization or at least discuss potential bias introduced by scanner variability; and (4) Many selected radiomic features lack clinical interpretability.
Response: We sincerely appreciate the editor's positive assessment of our work, as well as the constructive comments and valuable suggestions provided by the reviewers. These insights have been instrumental in improving both the clarity and scientific merit of our manuscript. We have carefully addressed each of the reviewers' concerns in a point-by-point manner, as detailed in our responses above.

(3) Manuscript Type: After verification, the manuscript type is "Retrospective Study".
Response: Yes, it is.

2 Specific comments
(1) Country/Territory of origin: Taiwan.
Response: Yes, it is.

(2) The language classification is Grade B. Please provide the latest language certificate after Return the Manuscript to Author for Revision. Please visit the following website for the professional English language editing companies that we recommend: https://www.wjgnet.com/bpg/gerinfo/240.

Response: We have carefully addressed this concern and have had our revised manuscript professionally edited. A language editing certificate from a professional English language editing company is provided with our revised submission.

(3) Manuscript Title:
Except for capitalization of the first word, all other words are represented in lowercase.
Response: The title format has been revised so that only the first word is capitalized, with all subsequent words in lowercase, as follows:
Machine learning survival prediction in esophageal cancer using radiomics and body composition from pretreatment and follow-up T12-level CT

(4) Audio Core Tip. In order to attract readers to read the full-text article, we request that the first author make an audio file describing the final core tip. This audio file will be published online, along with the article. The author can invite English language editing company to assist in resolving the language issues of Audio Core Tip.
Response: The first author has recorded an audio file of the final core tip, which has been carefully edited by a professional English language editing company to ensure clarity and correctness.

(5) Reference numbers in the main text.
If there are only 2, use ",". For example, mitigating the risk of disease progression[10,11].
The name of the author(s) of a reference is listed in the sentence, the reference number should be placed immediately after the author(s) of the reference. Example: Mandal et al[8] proposed that retractor aponeurosis disinsertion is the most likely cause of congenital low lid entropion.
Response: Thank you for the detailed instruction. We have carefully revised the reference formatting throughout the manuscript to ensure consistency with the journal's guidelines. All reference numbers have been modified as suggested.

(6) There are issues with the references:
Please provide the PMID numbers (https://pubmed.ncbi.nlm.nih.gov/) and DOI citation numbers (https://doi.crossref.org/simpleTextQuery) to the reference list and list all authors of the references. If a reference has no PMID and DOI, please provide the source website address of this reference.
To ensure the accuracy of the references, please use "Edit References by Auto-Analyser" (https://www.f6publishing.com/Forms/main/ArticleReferenceTool.aspx) to edit the references of the manuscript.
Response: Thank you for the detailed instruction. We have modified all references as per your suggestion, and have provided the PMID and DOI citation numbers for all applicable references. Additionally, we used "Edit References by Auto-Analyser" tool to ensure accuracy and consistency across all references.

(7) Figures.

Abbreviations must be defined upon first appearance in the Figure Legends. Do not use non-standard abbreviations, unless they appear at least two times in the text preceding the first usage/definition.
Original figure documents. In the meantime, authors should provide the original figure documents. Please prepare and arrange the figures using PowerPoint to ensure that all graphs or arrows or text portions can be reprocessed by the editor, and upload it to the file destination of "Image File" in the F6Pubishing system.
Response: All figures have been revised to ensure a uniform style, with consistent labeling. Abbreviations have been defined upon their first appearance in the figure legends. The original figure files have been prepared in PowerPoint format to allow further processing of graphical elements.

(8) Tables.
The Table title needs to be bolded.
Response: All table titles have been revised to bold font as requested.

(9) Abstract.
AIM. The purpose of the study should be stated clearly, with no or minimal background information, must start with "To", following the format of: "To investigate/study/determine…".
Response: The Abstract has been revised in accordance with the instructions. The AIM section now states: "To evaluate the influential features for esophageal cancer prognosis by integrating radiomic and body composition–based indices of skeletal muscle and adipose tissue at the T12 level from both pretreatment and follow-up CT images, in order to assess their value in predicting overall survival (OS)."
3 Recommendation: Conditional acceptance.
Language Quality: Grade B (Very good)
Scientific Quality: Grade C (Good)
(2) Executive Editor-in-Chief:
I recommend the manuscript to be published in the World Journal of Gastrointestinal Oncology (2025 released IF: 2.5, Web of Science).
Response: We sincerely thank you for considering our manuscript for publication in the World Journal of Gastrointestinal Oncology. We gratefully accept the recommendation to transfer our manuscript to the journal and hope that the revised version meets the editorial standards and expectations of the recommended publication.