Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-2025-1012

Reviewer A

This manuscript built a multi-modal model to predict patient outcome in EGFRmut NSCLC who received osimertinib therapy. The model takes features from genomic alteration, clinical and demographic information and also pathology image features which were processed by pre-trained model to predict patient outcome. The overall framework and methodology of the model is sound, but there are a few major limitations for this work and the real implication of the model and results is not clear.

Thank you for the helpful feedback, and we appreciate the opportunity to revise and resubmit our manuscript. We have addressed all the issues raised. Please find our response below in blue text.

1. The dataset used in training and cross-fold validation only contained 42 patients, with such as limited sample size, it is very hard to conclude the significance and robustness of the C-index models between uni-modal and multi-modal models.

Reply 1:

Thank you for your thoughtful comment. We completely agree that the sample size in this study is relatively small, and this does limit the statistical power to draw strong conclusions about the significance. We have acknowledged this as a key limitation in the discussion section (lines 285–291), and future work will focus on expanding the cohort, ideally through collaborations with additional clinical centers, to validate and build on these findings.

That said, in our analyses presented in this manuscript, we did observe some encouraging results suggesting the potential benefit of the multimodal approach:

- In our model evaluation, when comparing each model with random chance (lines 182–188):
The multimodal model was significantly better than chance ($p = 0.01$), while the single-modality models were not. This suggests that combining modalities can lead to stronger predictive performance, even within a limited dataset.

- In the evaluation of models' risk stratification capacity (lines 195–202):
The log-rank test showed a statistically significant separation in survival between predicted risk groups for the multimodal model ($p = 0.04$), and it correctly identified the lowest-risk group. In contrast, both single-modality models had p-values > 0.05. This adds further support for the added value of integrating multiple data types.

-To ensure the robustness of our results, we included standard deviation for all c-indexes and conducted stratified site-level analysis for all models (line 189-193), which showed that the trend of improvement after integrating different modalities was not only true for the general cohort, but also validated on the individual sites.

Finally, we fully recognize that the small cohort size prevents us from making broad claims. However, the trends we observed are promising and offer evidence that multimodal models may improve resistance prediction tasks. We see this study as a first step that lays the groundwork for larger-scale validation in future work.

Changes in the text 1:
Added detailed discussion on this limitation from line 291 to 297.

2. The authors found the ethnicity and RB1 mutations are associated with resistance, would it be easier to identify such biomarkers using univariate subgroup analysis? What is the advantage of using a multi-modal machine learning approach ?

Reply 2:
Thank you for this insightful question. Indeed, univariate subgroup analyses are commonly used to identify associations between individual variables and outcomes. While such analyses can identify important biomarkers, they examine each variable in isolation and do not account for the complex interactions and dependencies among multiple features. Our multimodal machine learning approach offers several key advantages beyond univariate analyses:

- Capturing Complex Relationships: Resistance to osimertinib is influenced by a combination of clinical, genomic, and histologic factors. multimodal machine learning models can learn nonlinear and high-order interactions between these features, which may not be evident when variables are assessed independently.

- Predictive Capacity: Rather than focusing solely on associations, our model was designed to predict treatment resistance. It is capable of providing data-driven estimation of patient resistance development risk based on a comprehensive analysis of patient data.

- The primary purpose of the feature importance analysis is to provide insights into how the model predicts resistance development. Rather than functioning as a black box, an interpretable model offers greater transparency and builds trust, making its predictions more meaningful and actionable for both clinicians and patients. Thus, the nature of this analysis is to support the interpretability of the model and not to identify biomarkers similar to univariate analysis.

To summarize, while univariate analyses are useful for identifying individual risk factors, our multimodal machine learning framework offers a more powerful,

integrative, and predictive approach. It not only identifies potentially important biomarkers but also supports clinical decision-making through patient-level risk prediction through an interpretable approach.

Changes in the text 2:
Added line 204-205; Updated line 271-272

3. Adeno-squamous histology switch has been reported as a resistance mechansim to EGFR TKI, does the image model detect any patient that has pre-existing squamous-like tumor based on image data? (even though the patients are adeno by diagnosis).

Reply 3:
Thanks for pointing out this resistance mechanism. Unfortunately, we do not have access to this subtype information in this dataset. This direction will definitely be included in our future studies.

Changes in the text 3:
Added line 300-302

4. The 42-patient cohort is from two clinical sites in both first-line and later-line setting, are clinical site and line-of-therapy associated with resistance/treatment outcomes?

Reply 4:
Thanks for this question. Clinical site was not included as an input feature in our model and therefore was not assessed in the feature importance analyses. This was a deliberate design choice, as our primary goal was to develop a resistance prediction model that is generalizable across sites and can be deployed in various clinical settings (including resource-limited settings) without introducing potential confounders. Including site as a predictor would limit the model's applicability to new institutions without prior labeled data, which contradicts our intention of creating a broadly accessible clinical tool.
That said, site-specific differences in outcomes are often mediated by more fundamental variables such as genetic mutations and demographic factors, many of which were included in our input features and analyses.
Regarding line-of-therapy, this information was unfortunately not available in our dataset and was therefore not included in our analyses. Additionally, given the modest cohort size, conducting subgroup analyses based on treatment line would have substantially reduced statistical power. We recognize the potential value of investigating the impact of treatment setting and plan to incorporate this variable in future studies with larger cohorts.

Changes in the text 4:
Added line 300-302

Reviewer B

1. Figures
    1) Figure 2: Please add unit to the x-axis.
       Added in the figure
    2) Figure 3: Please indicate the staining method and magnification in the legend.
       Added line 552-553

2. Table 1: Please add unit for “Age” and “Follow-up Time”.
   Added in the table