

Response to Reviewers

We would like to express our sincere gratitude to all three reviewers for their insightful, constructive, and detailed comments on our manuscript. Each suggestion has been invaluable in improving the clarity, scientific rigor, and overall quality of the work. We have carefully studied and fully incorporated the reviewers' feedback, and accordingly made meticulous revisions throughout the manuscript to address all concerns. We especially appreciate the reviewers' thorough reading and thoughtful critiques, which not only highlighted areas requiring clarification but also provided valuable guidance that significantly enhanced the robustness and clinical relevance of our study. The reviewers' professional expertise and dedication to advancing research in this field have played a vital role in shaping this revision, and we deeply value the time and effort they devoted to providing such comprehensive feedback.

08478492

1) The authors should clarify the novelty of the study in the abstract.

We have revised the Background and Conclusions sections of the abstract to emphasize that this study is the first to propose an ML-clinical combined nomogram model for predicting myelosuppression following first-line chemotherapy in colorectal cancer. Furthermore, we highlight our methodological innovation in integrating multiple machine learning algorithms with FMA, achieving both accuracy and interpretability. Unlike previous studies that relied solely on regression analysis or a single ML approach, our work is distinguished by a cross-method integration strategy.

2) Provide literature study just on the novelty of your work in the background.

We have revised the background to explicitly highlight the novelty of our study. Specifically, we emphasized that our feature mapping algorithm (FMA) integrates multiple ML algorithms with conventional nomogram construction, thereby combining interpretability with robust predictive performance and addressing the limitations of both traditional nomograms and single-model ML approaches.

3) Please clarify the aim of the study at the end of background.

We have refined the statement of the study aim and incorporated it into the Background section to better highlight the objective of this research.

4) Provide more numerical data in abstract and conclusion.

We have supplemented the abstract and conclusion with more detailed numerical data, including AUC, sensitivity, specificity, and other relevant metrics.

5) It is better to combine figures 4 and 5 and discuss more and find a connection.

We have combined Figures 4 and 5 and added further discussion. The revised results show that both RF and DT consistently identified BSA, BMI, and ALB as the most important predictors (importance >0.10), while tumor markers such as CEA, CA19-9, and CA125 ranked moderately high (0.05 - 0.10). Position and Chemotherapy Regimens also carried weight in both algorithms. This integration highlights the consistency across methods and strengthens the robustness and clinical relevance of our findings.

6) For Fig.7, you should discuss much more focusing that better AUC means

what? For example more AUC means better ranking not necessary better prediction.

In the revised manuscript, we have further clarified the interpretation of AUC in Fig. B. Specifically, while a higher AUC value indeed reflects superior discriminative ability, meaning the model is more effective at ranking positive cases ahead of negative ones across different thresholds, it does not necessarily guarantee optimal calibration or absolute predictive accuracy in probability estimation.

7) Considering discussion section: The use of the FMA to construct the clinic-ML nomogram demonstrates methodological innovation and clinical relevance, though further external validation would strengthen the generalizability of the findings.

We have revised the Discussion to highlight that the clinic - ML nomogram, constructed using the FMA algorithm, integrates multiple ML models with clinical variables, thereby improving both predictive accuracy and interpretability compared with single-model approaches. This demonstrates methodological innovation and clinical relevance. At the same time, we acknowledge that external validation in independent cohorts will be necessary to further confirm its generalizability.

08653643

1) The manuscript lacks citations for several key conclusions.

We have carefully revised the manuscript and added the missing references to support key conclusions. These updates strengthen the validity and scientific rigor of our work.

2) In the formula explanation, the sentence “is the prediction value produced by the th ML” appears to be incorrect. It likely contains a typographical error in “th,” which should either specify an ordinal number (e.g., “5th”) or be replaced with “the.” Clarification is needed to avoid confusion.

We have corrected the formatting errors by replacing potentially garbled formula content with the appropriate text and have applied consistent revisions to similar formatting issues. These adjustments make the formula explanation clearer and help avoid confusion for readers.

3) It is recommended to report one additional performance metric, specifically the AUPRC (Area Under the Precision-Recall Curve), as this is particularly informative in datasets with class imbalance and can complement AUC-ROC.

We have supplemented the performance metrics by reporting both AUC-ROC and AUPRC. While AUC-ROC reflects overall discrimination, AUPRC is particularly meaningful under class imbalance as it emphasizes precision - recall trade-offs. The inclusion of both metrics strengthens methodological rigor and enhances the clinical reliability of our predictive tool.

4) It is necessary to explicitly provide the parameters and hyperparameter settings of each machine learning method, including random seeds, feature

scaling approaches, and cross-validation settings, to ensure reproducibility.

To ensure reproducibility, we have detailed the parameter settings as follows: 10-fold cross-validation (KFold(n_splits=10, shuffle=True, random_state=42)); all random seeds set to 42; CEA, CA199, CA125 scaled by division ($\times 0.1$), BSA by multiplication ($\times 10$), others by Min-Max normalization. LR used liblinear; DT and RF with 400 trees; ET with 600 trees; KNN default; ANN with two hidden layers (50, 50) and 1000 iterations; SVM with probability estimation; XGBoost with use_label_encoder=False and logloss. The clinic - ML nomogram combined RF (400 trees) and ET (600 trees).

5) Several algorithms (e.g., DT, ET, RF, XGBoost) achieve perfect AUC=1.0 in the training set, which is a strong indicator of potential overfitting despite the use of cross-validation. This issue should be discussed in detail, and hyperparameter tuning procedures should be fully described to demonstrate that performance inflation was minimized.

During model construction, we employed rigorous 10-fold cross-validation to effectively reduce random error and enhance model stability. In addition, systematic hyperparameter tuning was performed for each algorithm, with optimization of key parameters such as maximum tree depth. For tree-based models (RF, ET, XGBoost), we restricted maximum tree depth, set the minimum leaf node size to greater than one, and introduced early stopping where applicable. On the other hand, models that achieved perfect performance in the training set did not maintain such results in the validation set. Specifically, RF and ET achieved an AUC of 1.0 in the training set, but their AUC values decreased to 0.94 and 0.96, respectively, in the validation set. This pattern is consistent with normal model generalization and indicates that the models did not simply memorize the training data. Moreover, calibration curves and decision curve analyses demonstrated stable risk estimation and net clinical benefit across cohorts, further supporting the reliability of the models.

6) Some tables and figures contain apparent formatting or logical inconsistencies.

We apologize for the formatting and logical inconsistencies identified in some tables and figures. These issues arose due to errors during the initial data entry process, which led to minor deviations in numerical presentation. We have now thoroughly re-checked and corrected all affected tables and figures to ensure consistency. Importantly, we confirm that the original clinical data remain authentic, and the revised version has undergone rigorous cross-verification to guarantee both the reliability and accuracy of the reported results.

7) The layout of Table 1 is not well-organized, making it difficult to read. It is suggested to refine its structure—possibly grouping variables by category and summarizing values more concisely—to enhance clarity and readability.

We have carefully revised the layout of Table 1. Variables are now grouped by category, and the presentation of values has been summarized more concisely. These refinements were made to enhance clarity, logical flow, and overall readability, while

maintaining the integrity and accuracy of the original data. We believe the revised table is now better structured and more user-friendly for readers.

8) The feature selection process—combining LASSO, RF, and DT—needs to be explained more clearly. Specifically, it should be clarified how discrepancies between methods were resolved, why certain predictors were included or excluded, and how clinical expertise influenced final variable selection.

In our study, we employed three complementary algorithms—LASSO, RF, and DT—for feature selection. To address discrepancies among methods, we first focused on the intersection of predictors identified by all three algorithms, which yielded eight consistently selected variables: BSA, BMI, Position, ALB, CEA, CA19-9, CA125, and Chemotherapy Regimens. These were considered robust predictors given their reproducibility across different selection techniques. In addition, we incorporated clinical expertise into the decision-making process. Through evaluation by three senior oncologists and based on accumulated clinical experience, it was concluded that Age and Chemotherapy Cycles are highly relevant to the risk of chemotherapy-induced myelosuppression in colorectal cancer patients, despite not being simultaneously selected by all three methods. Therefore, these two predictors were also included. As a result, the final set of ten predictors comprised: Age, BSA, BMI, Position, ALB, CEA, CA19-9, CA125, Chemotherapy Regimens, and Chemotherapy Cycles. This combined approach ensured methodological rigor while integrating clinical plausibility, thereby enhancing both the robustness and interpretability of the predictive model.

08184318

1) Similar studies that enhance the discussion

we have carefully revised the manuscript by incorporating a discussion of similar studies in both the Background and Discussion sections. These additions provide a broader context for our work, highlight methodological similarities and differences with previous research, and further substantiate the novelty and clinical significance of our study. We believe this revision has strengthened the depth and comprehensiveness of the manuscript.

2) Better highlighting in the discussion the importance of detailed evaluation of adverse events associated with chemotherapy

We have revised the Discussion to include evidence from similar studies showing that XELOX and FOLFOX, while comparable in efficacy, differ in hematologic toxicities, with thrombocytopenia more common in XELOX and neutropenia more frequent in FOLFOX. We also highlighted that such adverse events critically affect treatment adherence, quality of life, and continuation of therapy. Therefore, we emphasized the importance of detailed evaluation of chemotherapy-associated adverse events, particularly hematologic toxicities, for guiding individualized treatment and improving long-term outcomes.

3) Adding a conclusion at the end of the discussion

We have added a conclusion, summarizing the main findings: the clinic–ML nomogram developed in 855 colorectal cancer patients showed high discrimination, good calibration, and consistent net clinical benefit. We also highlighted its methodological innovation, alignment with the role of nomograms as individualized decision aids, and its potential clinical application in guiding prophylaxis and dose modification strategies.