Peer Review File

Article information: https://dx.doi.org/10.21037/jgo-23-18

**Review Comments**

**Comment 1:** First, the title needs to indicate the development and validation of the machine learning based postoperative VTE prediction model.

**Reply 1:** The title has been changed to: Development and validation of machine learning models for postoperative venous thromboembolism prediction in colorectal cancer inpatients: a retrospective study (see Page 1, line 3-5).

**Changes in the text:**

Development and validation of machine learning models for postoperative venous thromboembolism prediction in colorectal cancer inpatients: a retrospective study (Page 1, line 3-5).

**Comment 2:** Second, the abstract is not adequate and needs further revisions. The background did not explain why the ML algorithm can improve the predictive accuracy and the clinical needs for this research focus. The methods need to describe the inclusion of subjects, the assessment of baseline clinical factors, the generation of training and validation samples, follow up procedures, and diagnosis of postoperative VTE. The results need to describe the clinical characteristics of the study sample, the variables used in the prediction model, and the AUC, specificity, and sensitivity in both the training and validation samples. The conclusion needs to have comments for the clinical implications of the findings, not to repeat the significance of this study again.

**Reply 2:** We have modified our abstract as advised. The advantage of ML algorithm has been added to the background of abstract (see Page 2, Line 33-38). The inclusion of subjects, the assessment of baseline clinical factors, the generation of training and validation samples, follow up procedures, and diagnosis of postoperative VTE have been added to the methods of abstract (see Page 2, Line 39-45). The clinical characteristics of the study sample, the variables used in the prediction model, and the AUC, specificity, and sensitivity of the ML model with best performance have been

described in the results of abstract (see Page 3, Line 54-60). We highlighted the clinical value of the XGBoost model in VTE prophylaxis decision-making and the proposed risk factor used in this model in VTE-risk stratification (see Page 3, Line 61-64).

**Changes in the text:**

**Background:** Colorectal cancer (CRC) is a heterogeneous group of malignancies distinguished by distinct clinical features. The association of these features with venous thromboembolism (VTE) is yet to be clarified. Machine learning (ML) models are well suited to improve VTE prediction in CRC due to their ability to receive the characteristics of a large number of features and understand the dataset to obtain implicit correlations.

**Methods:** Data were extracted from 4,914 patients with colorectal cancer between August 2019 and August 2022, and 1191 patients who underwent surgery on the primary tumor site with curative intent were included. The variables analyzed included patient-level factors, cancer-level factors, and laboratory test results. Model training was conducted on 30% of the dataset using a ten-fold cross-validation method and model validation was performed using the total dataset. The primary outcome was VTE occurrence in postoperative 30 days. Six ML algorithms, including logistic regression (LR), random forest (RF), extreme gradient boosting (XGBoost), weighted support vector machine (SVM), a multilayer perception network (MLP), and a long short-term memory network (LSTM), were applied for model fitting. The model evaluation was based on six indicators, including receiver operating characteristic curve-area under the curve (ROC-AUC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), and Brier score. Two previous VTE models (Caprini and Khorana) were used as the benchmarks.

**Results:** The incidence of postoperative VTE was 10.8%. The top ten significant predictors included lymph node metastasis, C-reactive protein, tumor grade, anemia, primary tumor location, sex, age, D-dimer level, thrombin time, and tumor stage. In our results, the XGBoost model showed the best performance, with a ROC-AUC of 0.990, a SEN of 96.9%, a SPE of 96.1% in training dataset and a ROC-AUC of 0.908, a SEN of 77.5%, a SPE of 93.7% in validation dataset. All ML models outperformed the

previously developed models (Caprini and Khorana).

**Conclusions:** This study developed postoperative VTE predictive models using six ML algorithms. The XGBoost VTE model might supply a complementary tool for clinical VTE prophylaxis decision-making and the proposed risk factors could shed some light on VTE risk stratification in CRC patients.

**Comment 3:** Third, the introduction of the main text needs to list the AUC, sensitivity and specificity of the available prediction models, review the predictors used and methods for developing the models, and have comments on the limitations of these models. Please describe the rationale for machine learning and explain why machine learning could improve the predictive accuracy of the model.

**Reply 3:** We have modified our introduction partly as advised. The AUC of Khorana model in previous studies has been added (see Page 4, line 85-88). The AUC of Caprini model in previous studies has been added (see Page 4, line 93-95). The predictors used in Khorana and Caprini model have been described in Page 4, line 82-84 and Page 4, line 96-98. Further, all predictors used in Khorana and Caprini model were listed in Table S3. The model development methods have been added for either Khorana score (see Page 3, line 79-81) or Caprini score (see Page 4, line 98-99). The rationale that ML models could improve the performance of VTE model has been described in Page 4, line 105-113. Briefly, due to the ability to receive large number of features and obtain implicit correlations, ML models is well suitable to improve VTE prediction in CRC.

**Changes in the text:**

The Khorana score was initially developed using multivariate logistic regression method in ambulatory cancer patients and was further validated in hospitalized cancer patients (7); (Page 3, line 79-81)

Due to the limitations that some potential laboratory biomarkers (such as D-dimer) were not involved as predictors, the receiver operating characteristic curve-area under the curve (ROC-AUC) values of the Khorana model were previously in a range of 0.5-0.7, and a value of over 0.8 is expected (9); (Page 4, line 85-88)

Several studies have been performed to validate the predictive ability of the Caprini

model for surgical patients with CRC, the ROC-AUC values were in a range of 0.6-0.7 (11,12); (Page 4, line 98-99)

It is worth noting that Caprini score is developed by a summary of risk factors from 538 patients not statistical method (13). (Page 4, line 98-99)

Related References

7.Mulder FI, Candeloro M, Kamphuisen PW, et al. The Khorana score for prediction of venous thromboembolism in cancer patients: a systematic review and meta-analysis. Haematologica 2019;104:1277-87.

9.van Es N, Ventresca M, Di Nisio M, et al. The Khorana score for prediction of venous thromboembolism in cancer patients: An individual patient data meta-analysis. J Thromb Haemost 2020;18:1940-51.

11.Lu X, Zeng W, Zhu L, et al. Application of the Caprini risk assessment model for deep vein thrombosis among patients undergoing laparoscopic surgery for colorectal cancer. Medicine 2021;100:e24479.

12.Yao J, Lang Y, Su H, et al. Construction of risk assessment model for venous thromboembolism after colorectal cancer surgery: a Chinese single-center study. Clin Appl Thromb Hemost 2022;28:10760296211073748.

13.Caprini JA, Arcelus JI, Hasty JH, et al. Clinical assessment of venous thromboembolic risk in surgical patients. Semin Thromb Hemost 1991;17:304-12.

**Comment 4:** Fourth, in the methodology of the main text, please correctly describe the clinical research design of this study, the generation of training and validation samples, and the threshold predictive accuracy parameters such as sensitivity for a good predictive model. In statistics, the authors must explain why chained equations method is suitable to impute the missing data and please ensure $P<0.05$ is two-sided.

**Reply 4:** The clinical research design of this study has been described (see Page 5, line 122). The generation of training and validation samples were described in Page 6, line 182-183 and Page 7, line 184. In statistics, the predictive performance of the ML models was assessed using ROC-AUC and the threshold values have been specifically described (see Page 7, line 208-210). Multiple imputation by chained equation method

is the most popular method for handling missing data. The method replace each missing value with samples from its posterior predictive distribution. The missing values are imputed multiple times in order to account for the uncertainty of imputation and each imputed dataset is used to fit the analysis model. In the presence of general missing data patterns, the chained equations, also known as MICE method, has been shown to achieve superior performance in practice (1). Additionally, we highlighted that all tests were two-sided and P-values were two-sided in this study (see Page 7, line 207-208).

Related references

1. Chang C, Deng Y, Jiang X, et al. Multiple imputation for analysis of incomplete data in distributed health data networks. Nat Commun 2020;11:5467.

**Changes in the text:**

This is a single-center, retrospective observational study (Page 5, line 122).

All tests were two-sided; P values less than 0.05 were considered statistically significant (Page 7, line 207-208).

The model performance was considered excellent for ROC-AUC values 0.9-1, good for ROC-AUC values 0.8-0.9, fair for ROC-AUC values 0.6-0.8, and poor for AUC values 0.5-0.6 (Page 7, line 208-210).

**Comment 5:** Finally, please consider to cite the below related paper: Wei Q, Wang Y, An YB, Yang ZY, Liu YS, Zhang X, Gu ZC, Yao HW; the CRC-VTE investigators. Rationale and design of a prospective, multicenter, cohort study on the evaluation of postoperative Venous ThromboEmbolism incidence in patients with ColoRectal Cancer (CRC-VTE trial). Transl Cancer Res 2022;11(5):1406-1412. doi: 10.21037/tcr-21-1860.

**Reply 5:** We have added the above mentioned paper to reference list (see Page 14, line 428-431).

**Changes in the text:**

16. Wei Q, Wang Y, An YB, et al. Rationale and design of a prospective, multicenter, cohort study on the evaluation of postoperative Venous ThromboEmbolism incidence in patients with ColoRectal Cancer (CRC-VTE trial). Transl Cancer Res 2022;11:1406-12 (Page 14, line 428-431)