Peer Review File

Article information: http://dx.doi.org/10.21037/atm-20-4704

Reviewer Comments

The manuscript entitled 'An application of machine learning based on real-world data: Mining features of fibrinogen in clinical stages of lung cancer between genders' aimed to find new indicators for the early diagnosis of lung cancer and their diagnostic performance from large samples of clinical data in the real world. The results show that an integrated application based on supervised and unsupervised machine learning algorithms could effectively explore the potential links contained in the clinical data and reveal the differences in fibrinogen level in different clinical stages of lung cancer between the genders, which could provide a new reference basis for lung cancer staging. This study has certain clinical application guidance value. However, there are still some minor issues that need to be addressed before the paper is accepted for publication.

Minor issues:

Comment 1: In this study, we only used data from our thoracic medical center. From January 1, 2013 to July 17, 2019, we retrospectively collected EMR data of each lung cancer patient admitted to our ward. How can bias be reduced or avoided by using a single center retrospective study?

Reply 1: We agree with the reviewer's comments. Although it is difficult to avoid bias in a single-center retrospective study, we have tried to reduce bias by following methods: strictly controlling the inclusion or exclusion criteria of the subjects to reduce selection bias; using objective data to reduce information bias in later data analysis as possible as we can; multivariate analysis and artificial elimination of known confounding effects to reduce confounding bias; setting validation cohort to validate the model to avoid bias caused by over-fitting in data mining. Further studies will be carried out to continue exploring the distribution of the characteristics of fibrinogen and gender in control

population and different stages of lung cancer (I, II, III, IV) based on multicenter data research, which would provide new clues and improved research support for the diagnosis of lung cancer.

Changes in the text: We have explained the bias in limitations (see Page 22, Line 10-20).

Comment 2: In men, fibrinogen level of 2.94 g / l can initially be used as the upper limit to determine the early lung cancer group, but 3.91 g / l can be initially used as the reference threshold for the lower limit of middle and advanced lung cancer. The latter level can also be used as the upper limit of the critical value of early lung cancer in women. Can fibrinogen be a biomarker for predicting the prognosis of lung cancer?

Reply 2: In our paper, by applying several machine learning methods including MDL, NB, KM, NMF and DT, we built classification models which could predict lung cancer clinical stages more effectively based on fibrinogen level and genders. Specifically, we provide a more precisive decision tree considering both fibrinogen level and gender for lung cancer staging, but not predicting with a single marker and a single threshold. Previous investigators have reported that patients with higher plasma fibrinogen levels tend to have poor progression-free survival (PFS) and overall survival (OS)(1,2). Therefore, fibrinogen may serve as a candidate biomarker for disease monitoring and prognostic evaluation in patients with lung cancer. In future studies, we will attempt to obtain follow-up data for further analysis to figure out whether fibrinogen can be an appropriate biomarker for predicting the prognosis of lung cancer.

Changes in the text: We have added some statements in discussion of characteristics of fibrinogen in different lung cancer stages (see Page 21, Line 7-11).

Comment 3: There were significant differences in fibrinogen level, pleural effusion, chlorine content, A-G ratio, aspartate aminotransferase, alkaline phosphatase and gender composition between early lung cancer group and middle advanced lung cancer group. In addition to fibrinogen, what are the potential biomarkers for lung cancer staging?

Reply 3: More than 100 lung cancer prognostic or predictive markers have been published including serum markers (CEA, CA-125, CYFR A 21–21, chromogranin A, NSE, RBP, etc.), molecular markers (KRAS, EGFR, MET, P53 and so on), radiomics signature, etc. They are all potential biomarkers for lung cancer staging, but most do not reach clinical implementation. Moreover, predictive or diagnostic effects using a single marker are usually limited and weak for lung cancer staging. We therefore established a 3-marker-naive Bayesian model as well as a 2-marker-decision tree model which could predict lung cancer staging more accurately and reliably. So far hundreds and thousands of biomarkers for lung cancer staging have been identified, obviously it will be more effective that using artificial intelligence technologies to comprehensively establish a predictive model or system rather than a single biomarker.

Changes in the text: We have added some explanations in discussion (see Page 18, Line 9-20 and Page 19, Line 1).

Comment 4: The classification model with gender, fibrinogen and alkaline phosphatase as variables can verify the different expression characteristics of gender, fibrinogen and alkaline phosphatase in early, middle and advanced lung cancer, and has good prediction accuracy and potential clinical application value. Can these markers be biomarkers for early diagnosis of lung cancer?

Reply 4: Our study using several markers aimed to predict lung cancer stages was conducted in a lung cancer cohort. If we want to explore whether these biomarkers can be used for early diagnosis of lung cancer, further studies should be conducted in healthy or high-risk population. Fibrinogen and alkaline phosphatase might have the potential for early diagnosis of lung cancer, but currently we cannot conclude based on our data. That's a good idea and we'll look into that in future studies.

Changes in the text: We have added some explanations in discussion (see Page 22, Line 17-20).

Comment 5: Fibrinogen can be used as an important indicator to monitor the progress of metastasis, curative effect and prognosis of malignant tumors. Is it an indicator to

monitor the progress, efficacy and prognosis of malignant tumor metastasis?

Reply 5: Fibrinogen is an essential constituent of the coagulation system. It is mainly synthesized in the liver and released into the circulation in response to systemic inflammation and malignancy(3). An increasing body of evidence has indicated the association between fibrinogen and tumor clinical stage, angiogenesis, metastatic spread, and response to therapy in patients with cancer(1). And as a noninvasive method, measurement of fibrinogen is simple, economical, and less traumatic in clinical practice. Therefore, fibrinogen is obviously a potential indicator to monitor the progression of malignant tumor metastasis, therapeutic efficacy and prognosis. However, few studies combine multiple markers with fibrinogen and use machine learning model to predict the progress, efficacy and prognosis of malignant tumor, which might be a highlight of our study.

Changes in the text: We have modified the text about the relationship between fibrinogen level and malignant tumors (see Page 20, Line 2-5).

Comment 6: At present, it is believed that the increase of fibrinogen level in patients with malignant tumor is mainly due to the interaction between tumor and coagulation system. Elevated plasma fibrinogen levels may lead to coagulation disorders, which are closely related to the occurrence, development, recurrence and metastasis of malignant tumors. Are there any drugs targeting this pathway for the treatment of malignant tumors?

Reply 6: Currently, few anticoagulants are used directly in the treatment of malignant tumors, but there is a consensus that anticoagulants can be used to treat cancer-related thrombosis. The ASCO guidelines strongly recommend the use of low-molecular-weight heparin (LMWH) for the treatment of cancer-associated thrombosis (CAT)(4). A number of clinical trials of oral anticoagulants in the treatment of cancer-related thrombosis are in full swing. On the other hand, many studies have shown that platelets play an important role in promoting tumor development and metastasis(5-7), and some have shown that inhibition of platelet targets such as GPVI can cause intratumor hemorrhage(8), thus achieving the purpose of anti-tumor. We believe targeting

coagulant pathway might be a promising strategy for cancer therapy.

Changes in the text: We have added some relative statements of anti-coagulation drugs in cancer (see Page 20, Line 6-14).

Reference

1. Zhu LR, Li J, Chen P, et al. Clinical significance of plasma fibrinogen and D-dimer in predicting the chemotherapy efficacy and prognosis for small cell lung cancer patients. Clin Transl Oncol 2016;18:178-88.

2. Liu X, Shi B. Progress in research on the role of fibrinogen in lung cancer. Open Life Sciences 2020;15:326-30.

3. de Moerloose P, Casini A, Neerman-Arbez M. Congenital Fibrinogen Disorders: An Update. Semin Thromb Hemost 2013;39:585-95.

4. Key NS, Khorana AA, Kuderer NM, et al. Venous Thromboembolism Prophylaxis and Treatment in Patients With Cancer: ASCO Clinical Practice Guideline Update. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2020;38:496-520.

5. Rachidi S, Metelli A, Riesenberg B, et al. Platelets subvert T cell immunity against cancer via GARP-TGFβ axis. Science immunology 2017;2.

6. Placke T, Salih HR, Kopp H-G. GITR Ligand Provided by Thrombopoietic Cells Inhibits NK Cell Antitumor Activity. The Journal of Immunology 2012;189:154-60.

7. Mammadova-Bach E, Zigrino P, Brucker C, et al. Platelet integrin α6β1 controls lung metastasis through direct binding to cancer cell-derived ADAM9. JCI Insight 2016;1:e88245.

8. Volz J, Mammadova-Bach E, Gil-Pulido J, et al. Inhibition of platelet GPVI induces intratumor hemorrhage and increases efficacy of chemotherapy in mice. Blood 2019;133:2696-706.