

Peer Review File

Article information: <https://dx.doi.org/10.21037/tcr-23-1367>

Reviewer A

Comment 1: I think that out of an infinite number of patient and tumor factors, the 11 factors used in the analysis are quite small for using machine learning. Are there any other important features hidden?

Reply 1: Thanks for your suggestion. In patients with malignant tumors, there is a certain heterogeneity in the factors affecting prognosis, due to the significant differences in the biological behavior and molecular characteristics of different solid malignant tumors. We first included baseline features, and in blood tests, we included CEA and CA19-9, which are specific for gastric cancer. At the same time, the pathological type is the most important influencing factor for the prognosis of gastric cancer, and we included all of them in the analysis. In addition, we have included the Borrmann type of macroscopic. These factors are currently recognized and are the most prognostic factors for gastric cancer. Therefore, the current 11 factors are the most important characteristics of gastric cancer patients, and the combination of these characteristics can help to accurately predict the prognosis. Your suggestion also inspires us that the field is dynamic and new factors may emerge as research progresses. This emphasizes that feature selection in machine learning is an ongoing process, with researchers constantly refining the model to incorporate more insights and improve prediction accuracy. We supplemented the study limitations.

Changes in the text: See Page 9-10, line 41-2.

Comment 2: INTORODUTION, the 5-year survival rate after gastric cancer surgery is 40% and the recurrence rate is 60%, is this common? I think the results are too poor, but what do you think?

Reply 2: Thanks for your suggestion. Gastric cancer is a malignant tumor with a high mortality rate and a high recurrence rate, partly due to the abundant blood supply around the stomach and the complex lymphatic drainage. We base our reporting on *Lancet* and *Gastric Cancer* and cite the current literature (PMID: 20409751; 35933683; 33495964).

Changes in the text: None.

Comment 3: Why did you exclude multiple recurrences in your patient selection? Many postoperative recurrence cases of gastric cancer cases have multiple recurrences.

Reply 3: Thank you for your suggestion, which is essential for us to improve the quality of our manuscripts. We excluded multisite recurrence because of the extremely poor prognosis for patients with multisite recurrence and the extremely complex nature of tumor heterogeneity. At the same time, due to the nature of retrospective studies, we were unable to judge the chronological order of multisite recurrence. Therefore, we excluded multisite recurrence to ensure that the results were as accurate as possible. For multi-site recurrence, we will make further predictions based on machine learning

models in future prospective studies. We supplemented the study limitations.
Changes in the text: See Page 9, line 37-41.

Comment 4: Is the Krukenberg tumor a direct invasion of the ovary? Is it an ovarian metastasis?

Reply 4: Thank you for your suggestion. In the current study, we included patients with R0 resection, which means that no direct tumor invasion or metastasis was found during surgery. Krukenberg tumors often metastasize to the ovaries from the primary site (most commonly the stomach). Tumour cells usually spread from the primary site to the ovaries through the bloodstream or lymphatic routes. Features of Krukenberg tumors include the presence of signet ring cells, and histologically, they are often associated with mucus-producing adenocarcinomas.

Changes in the text: None.

Comment 5: Are the 336 patients in Figure 1 who are recurrence-free or unknown wrong?

Also, is the recurrence rate of pN3b cases this high?

Reply 5: Thank you for your suggestion. We recheck and confirm that there are no errors. Due to the nature of retrospective studies, we need to ensure the continuity of the included samples. More importantly, loss to follow-up is an inevitable problem in retrospective studies, and we have had adequate follow-up, but there are still cases of loss to follow-up. Therefore, in order to ensure data continuity, we grouped the non-recurrence with the unknown, which is also in line with the principle of statistical assignment. We reviewed the literature on the recurrence rate of pN3b. First, the incidence of pN3b is rare and few studies are conducted. Secondly, pN3b is the highest grade of lymph node grade in gastric cancer, which means that patients have a heavy burden of lymph nodes, and tumors are more likely to metastasize through lymphatic drainage. In some small studies, the recurrence rate of pN3b was as high as 100% (PMID: 28456896). Therefore, we determined that the recurrence rate of pN3b was high.

Changes in the text: None.

Comment 6: Is the machine learning only analyzing cases that have a recurrence? If so, it is just a ranking of risk factors, which is not very meaningful. If this AUC is derived with only 11 features after examining all cases, it would be an important finding.

Reply 6: Thank you for your suggestion. We analysed all patients and derived AUC values. We analysed patients as a whole and analysed these AUC values for patients overall.

Changes in the text: None.

Reviewer B

1. Please recheck the full name of “ROC” in the abstract.

prediction model. Model performance was evaluated using the receiver operating characteristic curve (ROC) and the area under the curve (AUC). Python was used for the analysis of machine

Response: The ROC has been re-corrected

2. A **highlight box** is needed to highly summarize the key findings/recommendations, innovation, and potential implications of the study. Please provide a highlight box for your manuscript.

- No reference should be cited in the highlight box. The box should be concise with no more than 250 words.

- Here is the template:

<p>Key findings</p> <ul style="list-style-type: none">• Report here about key findings of the study. <p>What is known and what is new?</p> <ul style="list-style-type: none">• Report here about what is known.• Report here about what does this manuscript adds. <p>What is the implication, and what should change now?</p> <ul style="list-style-type: none">• Report here about implications and actions needed.

Response: Highlight Box has been added

3. Please provide the full name of “CT” “pTNM” “CEA” “CA19-9” “PET” “XGBoost” “KMO” in the main text. Please also check through your article to make sure **all** the abbreviated terms have been defined when they **FIRST** appear in the Abstract and the main text.

Response: Abbreviations have been added.

4. Please recheck the full name of “VEGF” in the main text.

nicotine components in tobacco can significantly induce the expression of angiogrowth factor (VEGF) and cyclooxygenase 2 (COX-2), thereby promoting angiogenesis[33]. In addition,

Response: All errors have been corrected.

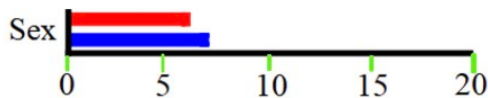
5. Figures and Tables

- **All abbreviations** in figures/tables and legends should be explained. “HMU” “PCA” in Figure 2, and “CEA” “CA19-9” in Table 1 for example. Please check all abbreviations and provide the full names in the corresponding legends.
- Please provide an **editable** version of Figure 1, 2 as a stand-alone **WORD/PPT** file, so that the editor can slightly and properly adjust the lines and structures, and text during the editing.
- Please resubmit Figures 3-5 in **jpg/tiff** format with higher resolution to us.
- Please recheck the data in Figure 1.

Recurrence (n=205)

No recurrence or unknown (n=336)

- Please add scale bars as below in the x-axes of Figure 3.



- Please add description for the x-axes in Figure 3.
- Please add description for the y-axes in Figure 5.
- Please indicate how data are presented in Table 1.
- Please check if units needed to add for “CEA” and “CA 19-9”.

CEA	
≤5	229(68.1)
>5	107(31.8)
CA19-9	

- Please recheck the data in Table 1, whether they should be reversed.

CEA			
≤5	229(68.1)	100(76.3)	129(62.9)
>5	107(31.8)	31(23.6)	76(37.0)
CA19-9			
≤37	232(69.0)	96(73.2)	136(66.3)
>37	104(30.9)	35(26.7)	69(33.6)

- Please recheck the data in Table 1, whether they should be “21” “102”.

Total stomach	31(9.2)	17(8.3)	4(3.1)
Tumor size (mm)			
≤50mm	84(25.0)	55(26.8)	29(22.1)
>50mm	252(75.0)	150(73.2)	103(77.9)

- Please recheck the data in Table 1, whether they should be “39.5” “68.2” after rounding. Please also check all other data in Table 1.

Age (years)		
≤60	190(56.5)	124(60.5)
>60	146(43.5)	81(39.8)
CEA		
≤5	229(68.1)	

- Please recheck the data in parentheses.

Categories of recurrences ↵	205(61.0)↵	205(61.0)↵
Locoregional ↵	69(33.7)↵	69(33.7)↵
Distant ↵	83(40.5)↵	83(40.5)↵
Peritoneal ↵	53(25.9)↵	53(25.9)↵

Response: All errors have been corrected.