

## Peer Review File

Article information: <https://dx.doi.org/10.21037/tcr-23-1197>

### Reviewer A

In this study by Jiang et al utilized RF and ANN to identify genes expressed in HBV related HCC and constructed a genetic diagnostic model with the overall goal of distinguishing HBV related HCC patients from HBV-infected patients without HCC. The study is overall interesting, however would benefit from further editing of the English language, as well as some improvement in organization and clarity of content. Specific comments below:

Line 32 - Is it possible to include some of the objective data results in the abstract results section? Otherwise, we are just being told there was high sensitivity and specificity without knowing the values, and that the "ROC curves efficiency was excellent". Although abstracts are limited on space, there needs to be some detailing of the findings within the abstract.

Reply: Yes, we added relevant values and results (see Page 1 line 34-35, line 37)

Line 63 - I am not sure what you mean by 'the deep position of the liver within the body'?

Reply: we have modified our text (see Page 2, line 71)

Line 69 - I would separate Machine Learning into a separate paragraph.

Reply: we have modified our text as advised(see Page 2, line 79)

Could you elaborate a bit better regarding RF and ANN, ie what they are and their differences a bit more, particularly for readers who are not as familiar with the methods?

Reply: we have modified our text as advised(see Page 2, line 86-89, line 94-99)

Line 74 - I assume "ML" is machine learning, but I don't think that was labeled previously.

Reply: we have modified our text as advised(see Page 2, line 79)

In the opening paragraph of your discussion, you state that the AUC efficiency was excellent. This is more of a subjective statement. I would recommend either re-stating the AUC, or provide further elaboration to how it is excellent.

Reply: we have modified our text as advised (see Page 8, line 331-332)

Line 337 - Would recommend moving this paragraph earlier in the discussion.

Reply: we have modified our text as advised (see Page 8, line 324-326, Page 8 332-334)

Line 349 - can you elaborate a bit on what specific outcomes were potentially

predicted?

Reply: we have modified our text as advised (see Page 9, line 380-382)

Overall I think your discussion section would benefit from further cohesiveness of the significance of your findings.

### **Reviewer B**

Hepatocellular carcinoma (HCC) constitutes 90% of primary liver cancer cases, rendering it a formidable public health challenge. In Eastern countries, HBV is the primary risk factor for cirrhosis, with a relatively high prevalence and incidence, contributing to a higher incidence of HBV-related HCC in Asia. Specifically, in 2018, the prevalence of HBV infection in the general population of China was estimated to be approximately 5-7.99%, with over 90% of cases occurring in adults aged 20 and above. This implies that despite reductions in infections due to preventive measures and vaccination programs over the past decade, an estimated 84 million people are still infected, surpassing the numbers in any European country. Thus, it is especially crucial to identify reliable diagnostic markers for distinguishing HBV-related HCC from non-cancerous individuals with HBV.

Jiang et al. utilized expression data from three datasets and employed the random forest method to select important variables. Furthermore, the research team established an artificial neural network model for early diagnosis and screening of HBV-related HCC. Additionally, the study conducted a detailed assessment of the ratio of infiltrating immune cells in non-cancerous liver tissues from HBV patients and HBV-related HCC tissues. These findings provide valuable insights that aid in early differentiation of HBV-related HCC from other HBV-infected individuals through reliable diagnostic markers. Additionally, they offer valuable guidance to enhance the clinical efficacy of HBV-related HCC.

The paper is very well written and they have established an effective diagnostic model for HBV-related HCC based on gene expression data from GEO. Although some of these observations are well documented and convincing, there are numerous weaknesses in this study.

There are some other major concerns listed below. To represent a significant contribution, these comments should be substantially addressed.

1. page4, 142: The article “a” may be incorrect. It should be changed to “an RF.”  
Reply: we have modified our text as advised(see Page 4, line 162)

2. page7, 300: The noun phrase “present study” seems to be missing a determiner before it. It should be changed to “In the present study.”

Reply: we have modified our text as advised(see Page 8, line 326)

3. The sample size used for the construction and validation of this model is relatively small. Can the sample size be increased in the future?

Reply: If there are more sample sizes in the future, we will use them again for this model.

4. The HBV-related HCC diagnosis model using an ANN was solely based on gene expression data. Can the model be further enhanced and incorporate other relevant pathogenic factors in the future?

Reply: Yes, we are trying to incorporate data of different dimensions to predict the prognosis of HBV-related HCC.

5. page8, 367: It should be changed to “activated in.”

Reply: we have modified our text as advised(see Page 9, line 403)

6. Some additional types of diagnostic and predictive models for HBV-related HCC have also been established previously, such as PMID: 2845411, which can be cited in the discussion section to provide further supplementation.

Reply: We cited this paper in the discussion section(see Page 9, line 375-378)

7. Can the model be evaluated through confirmatory experiments in the future?

Reply: Yes, we will conduct confirmation experiments.

8. Recent studies, such as doi: 10.21037/hbsn-21-519, have revealed that tumor-infiltrating ILC2s were found to accumulate in human HCC tissues, and higher infiltrations of ILC2s in HCC were correlated to better patient survival rates, which can be cited in the discussion section.

Reply: We cited this paper in the discussion section(see Page 9, line 395-398)