ANSWERING REVIEWERS

May 1, 2021

Dear Editor,

Please find enclosed the edited manuscript in Word format (file name: **WJG 65160** revision.docx).

Title: Artificial intelligence for hepatitis evaluation

Author: Wei Liu, Xue Liu, Mei Peng, Gong-Quan Chen, Peng-Hua Liu, Xin-Wu Cui, Fan Jiang, Christoph F. Dietrich

Name of Journal: *World Journal of Gastroenterology*

ESPS Manuscript NO: 65160

We are very grateful about the reviewers' valuable comments. We believe that these comments have made our manuscript more comprehensible. We have made all the changes as required. We have made the required changes in the main text, abstract, references, legends and figures. We tracked the changes in the text.

Reviewer 05738094:
Generally speaking, the present review is well written, it covers a very interesting topic and it adequately reviews the existing literature.
Reply: Thank you very much.
Only minor adjusments need to be done:
- strongly suggest to include in the introduction section a very small description of the main differences between classic machine learning (shallow models) and deep learning, e.g. which is the input in the two cases? which are the main differences in the workflow? Which are the benefits?
Reply: Thank you very much for your valuable suggestion, we included it as you suggested, and it was shown as follows: Compared with traditional AI, deep learning can directly apply the image to the learning process without manual feature extraction, while traditional machine learning needs manual recognition and extraction of different features. Therefore, deep learning has the advantage of no manual extraction of various features, which makes its learning process faster, intelligent, and accurate, in addition, deep learning can iterate and improve from past mistakes, but it needs more big data and more results analysis to fully show its robust and precise efficiency. (Page 3, line 5)
- Please include a critical discussion of the limitation of the reviewed studies and propose some future perspectives
Reply: We improved it as you suggested. Its subheading was named as "LIMITATIONS AND FUTURE PERSPECTIVES", and its contents were presented as follows:

At present, all reviewed studies using AI are mostly single-center research, and the amount of data in the training sets may be insufficient. In the future, people should build an internet database or hospital, and disease-related information from various countries or regions could be uploaded to the internet database or hospital so that AI researchers can obtain more disease-related information with different demographics, geographic areas, etc., to carry out multicenter research, and establish a more robust and inclusive AI model for clinical use. (Page 7, line 32)

- Please include in Table 1 and Table 2 the information about the sample size and the type (and size, if necessary) of the input to the AI models.

Reply: Thank you very much for giving us valuable suggestions, which would make my manuscript better. We included the information about the size and the type of the input to the AI models, and it can be seen in the appropriate location of Table 1 and Table 2. (Page 10, Page 11, respectively)

- For each study, please include in the text the population size and a brief description of the employed features, which is missing in some points, e.g. section 1 – predicting the incidence of hepatitis

Reply: We corrected it as you suggested. The sentences "They extracted data concerning the incidence of hepatitis A from 1987 to 2001 in the above-mentioned Disease Control and Prevention Center. The incidence of hepatitis A from 1981 to 1997 was taken as the training group and from 1998 to 2001 as the validation group." (Page 3, line 27) was ued to desceribe Guan P et al's employed features and population size. The paragraph "From January 2012 to August 2019, the local health bureau reported 486,983 cases of hepatitis B. Hepatitis B incidence from January 2012 to December 2018 was used to build and train ARIMA (0,1,1) model and ElmanNN model with 8 neurons, and hepatitis B incidence from January 2019 to August 2019 was used to validate ARIMA (0,1,1) model and ElmanNN with 8 neurons, with RMSE (root-mean-square error) and MAE (mean absolute error) applied to evaluate the prediction effect of the model." (Page 3, line 32) was used to desceribe Zheng Y et al's employed features and population size. The paragraph "From 2003 to 2012, 10,486,959 cases of hepatitis B were reported by National Health Commission. Compared with the above data, the prediction effect of a hybrid model, Grey Model (1,1), Grey Model (2,1) was calculated." (Page 3, line 40) was used to describe Gan R et al's employed features and population size. The paragraph "All above-mentioned data were collected from January 2005 to December 2017 in the local CDC (Centers for Disease Control). The monthly incidence and cases from January 2005 to June 2015 were used as the input variable to establish an AI model of the ARIMA, SVM (supporter vector machine), and LSTM (long-short time memory neural network), and that from July 2015 to December 2017 as the validating group." (Page 3, line 44) was uded to describe Guo Y et al's employed features and population size. The phrase "with 52 patients included," (Page 3, line 54) was used to describe Ahmad G et al's employed features and population size. The phrase "Obtaining 119 confirmed HBV (hepatitis B virus) infected samples from histopathology department," (Page 4, line 3) was used to describe Khan S et al's employed features and population size. The phrase "With 120,023 HCV (hepatitis C virus) patients and 9,601,900 non-HCV patients used as modeling data," (Page 4, line 12) was used to describe Doyle OM et al's employed features and

population size. The phrase "With 466 patients (401 with chronic hepatitis B, 65 without fibrosis) undergoing partial hepatectomy used as the data of training group and test group of transfer learning radiomics," (Page 5, line 12) was used to describe Xue LY et al's employed features and population size. The phrase "with 513 chronic liver disease patients and 45 healthy liver subjects taken as modeled data," (Page 5, line 35) was used to describe Son JH et al's study.

- Minors: a. Please carefully review the acronyms and make sure to define the first time they are used. The following definitions are missing: BP (back-propagation), support vector machine (SVM), long short-term memory (LSTM) network, multilayer mamdani fuzzy inference system (ADHB-ML-MFIS), radial basis function (RBF), recursive feature elimination (RFE), area under the curve (AUC), area under the receiver operating curve (AUROC), region of interest (ROI), least absolute shrinkage and selection operator (LASSO).

Reply: We carefully reviewed the acronyms and make sure to define the first time they were used (in green color) in the manuscript. All the acronyms and full names were shown as follows:

artificial intelligence (AI), hepatocellular carcinoma (HCC), hepatic encephalopathy (HE), artificial neural network (ANN), autoregressive integrated moving average (ARIMA), ElmanNN (Elman neural network), RMSE (root-mean-square error), MAE (mean absolute error), BP-ANN (back propagation artificial neural networks), GM (Gey Model), CDC (Centers for Disease Control), SVM (supporter vector machine), LSTM (long-short time memory neural network), ADHB-ML-MFIS (automated diagnosis of hepatitis B using multilayer mamdani fuzzy inference system expert system), HBV (hepatitis B virus), RBF (radial basis function), HCV (hepatitis C virus), aspartate aminotransferase-to-platelet ratio indexes (APRIs), CS (Cross-sectional), CHC (chronic hepatitis C virus), National Veterans Health Administration (VHA), Interferon alfa (IFN), Ribavirin (RIB), CT (computed tomography), MRI (magnetic resonance imaging), AAH (alcohol-associated hepatitis), AST (aspartate aminotransferase), RFE-RF (recursive feature elimination using random forest), original radiofrequency (ORF), contrast-enhanced micro-flow (CEMF), AUCs (area under the curves), transfer learning (TL), AUROC (area under the receiver operating curve), CT texture analysis (CTTA), random forest classifier (RFC), convolution neural network (CNN), (VolS (spleen volume), VolL (liver volume), shear wave elastography (DLRE), SWE (shear wave elastography), FIB-4 (fibrosis index based on four factors), focal nodular hyperplasia (FNH), arterial phase (AP), portal venous phase (PVP), ROIs (region of interests), ICCs (inter- and intra- class correlation coefficients), LASSO (least absolute shrinkage and selection operator), EASL (European Association for the Study of the Liver), LI-RADS (Liver Imaging Reporting and Data System), microvascular invasion (MVI), AFP (alpha-fetoprotein), delayed phase (DP), concordance correlation coefficient (CCC), CF (clinical factors), HBP (hepatobiliary phase)

- Minors: b. Please revise the following sentence: "Computer science and technology, such as cloud computing, big data, make radiomics and deep learning are hot topics in the field of medical imaging diagnosis" (Hepatitis evaluation based on radiology)

Reply: It has been revised like this: With the rapid development of computer technology and AI, radiomics and deep learning have become hot topics in the field of medical imaging. (page 4, line 39)

We would like to thank the reviewer again for taking the time to review our manuscript. Thank you again for publishing our manuscript in the *World Journal of Gastroenterology*.

Sincerely yours,

Prof. Dr. med. Xin-Wu Cui
Email : cuixinwu@live.cn