

PEER-REVIEW REPORT

Name of journal: *Artificial Intelligence in Gastroenterology*

Manuscript NO: 82283

Title: Risk factor profiles for gastric cancer prediction with respect to *Helicobacter pylori*:
A study of a tertiary care hospital in Pakistan

Provenance and peer review: Invited Manuscript; Externally peer reviewed

Peer-review model: Single blind

Reviewer's code: 03769068

Position: Editorial Board

Academic degree: PhD

Professional title: Adjunct Professor, Professor

Reviewer's Country/Territory: Brazil

Author's Country/Territory: Pakistan

Manuscript submission date: 2022-12-14

Reviewer chosen by: AI Technique

Reviewer accepted review: 2022-12-19 22:41

Reviewer performed review: 2022-12-27 16:57

Review time: 7 Days and 18 Hours

Scientific quality	<input type="checkbox"/> Grade A: Excellent <input type="checkbox"/> Grade B: Very good <input checked="" type="checkbox"/> Grade C: Good <input type="checkbox"/> Grade D: Fair <input type="checkbox"/> Grade E: Do not publish
Language quality	<input type="checkbox"/> Grade A: Priority publishing <input checked="" type="checkbox"/> Grade B: Minor language polishing <input type="checkbox"/> Grade C: A great deal of language polishing <input type="checkbox"/> Grade D: Rejection
Conclusion	<input type="checkbox"/> Accept (High priority) <input type="checkbox"/> Accept (General priority) <input type="checkbox"/> Minor revision <input checked="" type="checkbox"/> Major revision <input type="checkbox"/> Rejection
Re-review	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Peer-reviewer statements	Peer-Review: [<input checked="" type="checkbox"/>] Anonymous [<input type="checkbox"/>] Onymous
	Conflicts-of-Interest: [<input type="checkbox"/>] Yes [<input checked="" type="checkbox"/>] No

SPECIFIC COMMENTS TO AUTHORS

I have now reviewed your paper and recognize the importance of your research question. Manuscript NO. 82283 aimed to identify and incorporate Gastric Cancer (GC) risk factors into a computer model to predict the likelihood of GC development. The main and short titles accurately reflect the major topic and content of the study. However, there is no clear delineation of the study's main objective in the Abstract. This section highlights only the incorporation of GC risk factors into a dynamic computer tool for the prediction of GC in the "BACKGROUND" subsection. The BACKGROUND and AIM subsections should be separate and the study's AIM should be clearly stated. It is also necessary to better contextualize the research motivation in the "BACKGROUND" subsection. The "RESULTS" subsection should provide detailed important data from the research findings. Finally, the "CONCLUSION" subsection of the Abstract should further explore the limitations of the study and future prospects in the research field. The INTRODUCTION should be improved. It is recommended to structure the risk factors for CG in a more cohesive way. The justification for building prediction models - i.e., avoiding unnecessary exposure of patients to invasive procedures - is well stated. Nevertheless, the "state-of-the-art" application of machine learning models in gastroenterology and gastrointestinal oncology should be further explored. It is very interesting that one of the motivations for the construction of the tool was the perception of an overload in referrals to endoscopic procedures in your institution. The MATERIALS AND METHODS are not sufficiently described. The ethics-related aspects of the research are no problem. However, there is no sufficiently detailed description of the inclusion criteria for the study. There is also no justification for the exclusion criteria.



**Baishideng
Publishing
Group**

7041 Koll Center Parkway, Suite
160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
<https://www.wjgnet.com>

Methods for assessing *H. pylori* infection status are detailed. But what are the "symptoms suggestive of upper gastroduodenal endoscopy"? This should be explained in detail. The elaboration of a scheme for the diagnostic approach would also be interesting. The Machine-learning algorithm should be further explored in the METHODS section. But it is well-detailed in the RESULTS alongside Principal components analysis (PCA) and decision trees. The reduction of some categories to 'yes' and 'no' in the Machine-learning algorithm is justifiable but constitutes an important limitation of the study that should be further discussed. The RESULTS provide sufficient experimental data on GC risk factors. This is a major strength of the manuscript. However, the findings on the accuracy of the computer model are not well presented. In this sense, the clinical-epidemiological characteristics of the enrolled patients are well approached in the DISCUSSION. The risk factors found are well-compared to the findings of other studies. The model correctly classified 80% of the cases; 10% of the cases from GC, 98% from GDD, and 30% of the NGM participants. Therefore, this results in >80% test accuracy. Although the model shows high predictive power for GDD, it has a questionable performance in predicting GC itself. These limitations should be clearly addressed. In conclusion, this study presents interesting data about the risk factors for the development of gastric cancer and the clinical-epidemiological characteristics of patients affected by GDD. These data constitute the major part of the findings of this manuscript. Although the RF GC-predictive model is an interesting tool, its construction, results, and limitations are not well exposed or discussed. The manuscript has language issues and the text construction is somewhat repetitive. Manuscript formatting should be revised according to BPG guidelines. The tool has potential and should be further optimized.

PEER-REVIEW REPORT

Name of journal: *Artificial Intelligence in Gastroenterology*

Manuscript NO: 82283

Title: Risk factor profiles for gastric cancer prediction with respect to *Helicobacter pylori*:
A study of a tertiary care hospital in Pakistan

Provenance and peer review: Invited Manuscript; Externally peer reviewed

Peer-review model: Single blind

Reviewer's code: 05194798

Position: Editorial Board

Academic degree: MD

Professional title: Director

Reviewer's Country/Territory: Japan

Author's Country/Territory: Pakistan

Manuscript submission date: 2022-12-14

Reviewer chosen by: AI Technique

Reviewer accepted review: 2023-01-08 23:21

Reviewer performed review: 2023-01-11 11:04

Review time: 2 Days and 11 Hours

Scientific quality	<input type="checkbox"/> Grade A: Excellent <input type="checkbox"/> Grade B: Very good <input type="checkbox"/> Grade C: Good <input checked="" type="checkbox"/> Grade D: Fair <input type="checkbox"/> Grade E: Do not publish
Novelty of this manuscript	<input type="checkbox"/> Grade A: Excellent <input checked="" type="checkbox"/> Grade B: Good <input type="checkbox"/> Grade C: Fair <input type="checkbox"/> Grade D: No novelty
Creativity or innovation of this manuscript	<input type="checkbox"/> Grade A: Excellent <input checked="" type="checkbox"/> Grade B: Good <input type="checkbox"/> Grade C: Fair <input type="checkbox"/> Grade D: No creativity or innovation

Scientific significance of the conclusion in this manuscript	<input type="checkbox"/> Grade A: Excellent <input checked="" type="checkbox"/> Grade B: Good <input type="checkbox"/> Grade C: Fair <input type="checkbox"/> Grade D: No scientific significance
Language quality	<input type="checkbox"/> Grade A: Priority publishing <input checked="" type="checkbox"/> Grade B: Minor language polishing <input type="checkbox"/> Grade C: A great deal of language polishing <input type="checkbox"/> Grade D: Rejection
Conclusion	<input type="checkbox"/> Accept (High priority) <input type="checkbox"/> Accept (General priority) <input type="checkbox"/> Minor revision <input checked="" type="checkbox"/> Major revision <input type="checkbox"/> Rejection
Re-review	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
Peer-reviewer statements	Peer-Review: <input checked="" type="checkbox"/> Anonymous <input type="checkbox"/> Onymous
	Conflicts-of-Interest: <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

SPECIFIC COMMENTS TO AUTHORS

This manuscript is an original article that retrospectively reported the prevalence of H. pylori infection in Pakistan along with its association with various risk factors related to gastroduodenal diseases, and incorporated the risk factors into a dynamic computer tool for the prediction of gastric cancer (GC). The authors identified that age, income level, vomiting, bloating and medication had significant association with gastroduodenal disorders and GC. In addition, the authors developed a dynamic RF GC-predictive model with >80% test accuracy. As the authors analyzed various risk factors, endoscopic and histopathological findings in detail, the results contain informative knowledge, which will be of interest to researchers and clinicians in the field. However, the following major and minor issues require clarification: Major 1. The definitions in the severity of gastritis (mild, moderate, marked) and ulcer (moderate, marked) are unclear and less objectivity. Minor 1. Study period should be provided. 2. Was H. pylori status proven using all provided modalities (RUT, UBT, biopsy, HPE) or one of them? Please describe in detail. 3. Were the patients after H. pylori eradication included in the study? 4. Unabbreviated words with abbreviation should be provided at the first

appearance (NGM, GDD, RUT, HPE, et al). 5. As the authors use many abbreviations, a list of abbreviations should be provided. 6. (Figure 1) Why did the authors select cross-correlation bar charts with respect to gender, age and H. pylori infection status based on RUT. 7. Conclusion should be more summarized. Especially, the content in the second paragraph should be described in Discussion section.

PEER-REVIEW REPORT

Name of journal: *Artificial Intelligence in Gastroenterology*

Manuscript NO: 82283

Title: Risk factor profiles for gastric cancer prediction with respect to *Helicobacter pylori*:
A study of a tertiary care hospital in Pakistan

Provenance and peer review: Invited Manuscript; Externally peer reviewed

Peer-review model: Single blind

Reviewer's code: 06485091

Position: Peer Reviewer

Academic degree: MD

Professional title: Doctor

Reviewer's Country/Territory: China

Author's Country/Territory: Pakistan

Manuscript submission date: 2022-12-14

Reviewer chosen by: AI Technique

Reviewer accepted review: 2023-01-08 01:32

Reviewer performed review: 2023-01-13 08:32

Review time: 5 Days and 7 Hours

Scientific quality	<input type="checkbox"/> Grade A: Excellent <input type="checkbox"/> Grade B: Very good <input type="checkbox"/> Grade C: Good <input checked="" type="checkbox"/> Grade D: Fair <input type="checkbox"/> Grade E: Do not publish
Novelty of this manuscript	<input type="checkbox"/> Grade A: Excellent <input type="checkbox"/> Grade B: Good <input checked="" type="checkbox"/> Grade C: Fair <input type="checkbox"/> Grade D: No novelty
Creativity or innovation of this manuscript	<input type="checkbox"/> Grade A: Excellent <input type="checkbox"/> Grade B: Good <input checked="" type="checkbox"/> Grade C: Fair <input type="checkbox"/> Grade D: No creativity or innovation

Scientific significance of the conclusion in this manuscript	<input type="checkbox"/> Grade A: Excellent <input type="checkbox"/> Grade B: Good <input checked="" type="checkbox"/> Grade C: Fair <input type="checkbox"/> Grade D: No scientific significance
Language quality	<input type="checkbox"/> Grade A: Priority publishing <input checked="" type="checkbox"/> Grade B: Minor language polishing <input type="checkbox"/> Grade C: A great deal of language polishing <input type="checkbox"/> Grade D: Rejection
Conclusion	<input type="checkbox"/> Accept (High priority) <input type="checkbox"/> Accept (General priority) <input type="checkbox"/> Minor revision <input checked="" type="checkbox"/> Major revision <input type="checkbox"/> Rejection
Re-review	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
Peer-reviewer statements	Peer-Review: <input checked="" type="checkbox"/> Anonymous <input type="checkbox"/> Onymous
	Conflicts-of-Interest: <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

SPECIFIC COMMENTS TO AUTHORS

The authors studied risk factor profiles for gastric cancer prediction with respect to *Helicobacter pylori* and constructed a computer model to predict the likelihood of developing GC. There is some novelty in the present study; however, there are some serious problems concerning the data and the results. 1. The original data provided is inconsistent with the results. 2. The title does not cohere with the content. The research focused on risk factors associated with gastric cancer (GC), but there were only 28 GC samples. Most of the samples (262, 77%) were about gastroduodenal diseases (GDD), which was not closely related to the purpose of the study. Are the results of machine learning obtained from such a small sample size reliable? 3. Some of the statistical methods are not appropriate. For example, in Table 1, the author used Chi-Square (X^2) tests to calculate the p value for the Medication factor, but 4 (26.7%) of the cell counts in the 5×3 table is less than 5, in which circumstance the Fisher's test should be used. 4. Sodium intake in Table 6 is divided into four categories with accuracy to a milligram (mg) which are difficult to be accurately defined during the questionnaire. 5. In the regression analysis, it is necessary to ensure that there is no correlation between all

factors, but the education level and the income level in Table 2 are obviously correlated. 6. The necessity of this study is not stated in the Background section of the abstract. 7. The authors stated that they constructed a computer model to predict the likelihood of developing GC with high sensitivity and specificity, but the validation of the sensitivity and specificity of the model was not shown. 8. Several minor points. 1) Table 6 appeared before Table 1. 2) Many abbreviations, such as PAN, RUT, and HPE, was not defined at first mention. 3) Language inaccuracy, e.g. 'little sleeps'.