Review of: "A Graphical User Interface Based on Logistic Regression Approach for Malarial Detection"

Albert Akuno[1]

1 University of Notre Dame

The work is a good step towards malaria prediction, surveillance, and control. While I appreciate the efforts of the authors in putting together this potentially informative and interesting research, the following are my comments that, if addressed, can help improve the article.

General comments

1. There is no clear gap that is being filled by this research. The authors should indicate, in clear detail, the gap and significance of the present work both in the abstract and the introduction. Perhaps a systematic literature review in the introduction may help them achieve this.

2. There is little connection between the title and the contents of the manuscript, especially the results section. Indeed, from the title, it sounds like the authors have developed a graphical user interface (GUI) for predicting malaria. However, there is no significant mention of a GUI until in the last part of the manuscript, and even at this point, there is no major description of such, with the authors only giving a snapshot of the same in Figure 2. This presentation is very casual, and it is unimaginable that it should be the objective of the manuscript, as is depicted by the title. A detailed description of the GUI should be provided, including the code that the authors used to develop it and to conduct the general analysis in the manuscript. In addition, the authors could consider giving a procedure and example of how to work with the GUI. Importantly, where is the link to the GUI, and how can readers access it?

3. I find the descriptions and analysis in the manuscript to be very shallow and sketchy. For instance, the results and discussion and conclusion sections are very short, and do not quantitatively and qualitatively amount to a publishable original research article. Indeed, the whole manuscript gives an impression that the authors did not put much effort into the research.

4. The authors should consider citing more literature, especially in the introduction, and indeed the entire manuscript.

5. In the data description section, the authors should indicate from where and how they obtained the data, whether there were any privacy and confidentiality issues with it. If there were any of such important data aspects, the authors should indicate whether and how they complied, and if there was any agreement that they needed to sign to obtain the data. In addition, what was the geographical scope of this data? Is it local, country, regional, or world data? The authors should indicate the geographical resolution of Ranchi, Jharkhand, India. Additionally, in this section, the authors may consider using some pictorial presentations like pie charts, bar graphs, and even frequency tables to present the demographic characteristics of the data.

6. The authors should write a detailed description of the performance metrics. In particular, what does Classification Accuracy (CA), Precision, Recall, and F1 Score mean? Their definitions and examples of where and when they are applicable should be given.

Specific comments

1. The statement "Billions of people in the world die from this disease annually" in the first line of the first paragraph in the introduction is not factual. In fact, the falsity of this statement is clear in line 6 of the same paragraph where the authors indicate that "According to WHO, in 2022, there were globally about 249 million cases and 608,000 deaths worldwide". This statement should therefore be revised.

2. I propose that the authors consider citing some literature to back up the statement "Importantly, the disease commonly occurs in tropical countries" in the first paragraph of the introduction.

3. In the second paragraph of the introduction where the authors write "In recent years, the epidemiology of malaria has been expressed mathematically from unrealistic to high-level models," what do they mean by the term "unrealistic?" It is not clear what unrealistic models look like, and whether researchers can set out to conduct any research using unrealistic models.

4. In the statement "Thus, R_0 furnishes the intensity measure of transmission and reports that an area is disease-endemic areas if its $R_0 > 1$" in the second paragraph of the introduction, I suggest that the authors delete "areas" between "disease-endemic" and "if."

5. In paragraph 3 of the introduction, where the authors write "In order to reduce the affected rate of the disease, effective and timely diagnosis of the disease is essential," did they mean to write "infection rate" instead of "affected rate?"