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

Flávio Codeço Coelho[1]

1 Fundação Getulio Vargas

Potential competing interests: No potential competing interests to declare.

Feature reduction (selection) technique employed is too crude. Authors should use model-based feature selection methods, such as sequential feature selection based on feature importances, or recursive feature elimination. These methods take into account the contribution of each feature to the overall performance of the model.

When the authors list the features, they should do so in a table where they can also indicate the type of each variable: continuous numeric, discrete numeric, ordinal, or categorical. Proper units for each of the values should also be included.

The authors also do not properly describe the dataset. One key aspect that must be informed is how well balanced the dataset is, i.e., how many positive vs. negative malaria cases there are in the data. All the models compared in this manuscript require well-balanced datasets in order to give reliable results.

For the methodology, it is not necessary to describe each model, since they are well-known models. It suffices to add references to their description and to the software implementations used. It is important, however, to include the regression equation proposed in the article by the authors.

Regarding the results of the model comparison, the authors should show an ROC curve and confusion matrices for the results of each model. It would also be good to see measures of sensitivity and specificity for each model, as these metrics are the most used in epidemiology. Comparing their best model, LR, with the sensitivity and specificity of diagnostic tests the model aims to replace will help to convince readers of the advantages of the proposed malaria test.

Regarding the GUI, very little information is given as to how it was built, etc.