Review of: "Application of Ensemble Learning in CXR Classification for Improving COVID-19 Diagnosis"

Lenin G. Falconi[1]

1 Escuela Politécnica Nacional

Potential competing interests: No potential competing interests to declare.

First, let me congratulate the research team for their effort and dedication to researching a matter of interest such as the classification of Covid-19 images. My observations on the written work I have received to check are the following:

1. **About the general presentation of the document:** I suggest writing in LaTeX if possible and if there are templates available for that purpose. If this is not possible, please use the same font and font size according to the sections you are writing and use adequate space between objects like equations. Also, variable names, when referred to in a paragraph, should be written with a different font. In LaTeX, that is easy because you write the variable between $ signs. But if you are writing from Word, you should come up with a way to help the reader notice that those symbols are not part of the literature but from an equation.

2. **About document references:** I think references are missing in the document. There are several statements made by the authors that should be referenced, especially if they are not from their original work or they are known to the authors by means of experience or study. For instance, statements like: a)"Artificial intelligence is revolutionizing the detection of medical conditions like breast cancer…" Such statements should include reference to evidence of such applications in breast cancer, brain tumors, or covid-19 detection? Prediction? b) SVM excels in text data analysis… (according to whom? Did the authors study this before?) c) Naive Bayes aiding in fruit classification(Is NB the only model for that? Who says it is good for fruit classification? What kind of classification? Without a reference, it is hard to understand. Moreover, a reference would make clear why I could read the title of the research and have an idea of why NB was applied or used for fruit classification)

3. **About the literature review:** It would be good for the document to provide evidence on the methodology followed in order to conduct the literature review. For instance, you could consider the methodology by Barbara Kitchenham on Systematic Literature Reviews. How many works did the authors read for this research? How were they retrieved? How were they searched? In the sample of works studied, how many works do not use KNN, SVM, and etc.? Otherwise, there is not enough evidence to sustain some of the claims in the work. Also, the writing of the literature review could be improved. All sentences are written in the same structure: Authors et. al. did ….. Remember, you have to keep your reader interested in your work. You could try to establish a discussion, and if you have lots of the same to say, it could go on a Table.

4. **Equations:** Equation 1 talks about an *r* which never appears in Equation 1.

5. **Datasets:** It is not clear what Kaggle dataset was used. A link to the dataset card or dataset web page should be included.

6. **Figures:** Figure 3 may be the first figure, so the reader first understands what images you are working with, and then present Figure 1 to make more clear for the reader the effects of HOG.

7. **Methodology:** This whole section should be re-written because of the following reasons:

   1. The structure of the text is not clear and creates confusion for the reader. For instance, since the authors are working with images, how was SMOTE applied? SMOTE is not a technique applied to image data augmentation but to tabular data. This confusion is triggered because the two-stage strategy used is not well described. I think a figure could help the reader understand that the authors took images of X-ray lungs, then applied image augmentation techniques, then the dataset was split, then HOG features were obtained? Unfortunately, the writing of this section is rather confusing, and the reader will not understand what the authors did. As a matter of fact, this confusion makes one wonder about the results obtained by the authors.

   2. There is no clear indication of the size of the dataset after data augmentation. It is not clear what techniques were used and with what values.

   3. There is no indication of the setup of the classifiers.

   4. It is claimed that cross-validation was used, but this does not appear in a proper section about training setup. This is claimed in the results.

   5. It is said that SMOTE is used. Was it used in the images of COVID, or was it used on the HOG feature vector?

   6. It is said that removing nonnumeric columns was applied, but to what? To the image dataset? To the HOG features? And why were some columns of the feature vector removed?

   7. What happened to the different sizes of images: Were the images rescaled to the same size? Or the size of the image does not matter for the methodology proposed, but the cell size matters. If so, what was the size of the cell?

8. **Results:** About the results, I understand that the authors made different experiments using different splits of train, validation, and test. However, the tables presented are confusing because, to the best of my knowledge, the idea to prove a machine learning model or artifact is to prove that it can generalize. This means that evaluation should only be made on the test dataset. Tables and Figure 8 make me think that the results presented are those of the train dataset and not the test dataset, and make me think that probably the test results were bad and because of that, they are not presented.

Please keep in mind that the Methodology section of the paper should be so clear that any other scientist could replicate your work.