Requires Changes

2 specifications require changes

Dear student

Great start on this proposal! This is the mentor (Graham) that's been working with you in the study group on this proposal. By chance, I was assigned your proposal for review, and I'm really happy with how things are coming together.

I've noted a few areas where you should add a bit more detail, but I think that you've picked a great project and you're definitely on the right track. I think you'll see that most of these things shouldn't take long to update. Almost there...keep going!

Cheers!

Project Proposal

Student briefly details background information of the domain from which the project is proposed. Historical information relevant to the project should be included. It should be clear how or why a problem in the domain can or should be solved. Related academic research should be appropriately cited. A discussion of the student's personal motivation for investigating a particular problem in the domain is encouraged but not required.

Great job giving the reader an introduction to the problem domain!

Suggested:

- Since you've included a link to your data source in this section, you can directly lift it into the 'Project Overview' section of your capstone report (to save a bit of time when you write the report).

Student clearly describes the problem that is to be solved. The problem is well defined and has at least one relevant potential solution. Additionally, the problem is quantifiable, measurable, and replicable.

I think that this is a fairly clear overview of how the problem is structured (it might also be a good idea to specifically note that you're treating this as a classification problem).

The dataset(s) and/or input(s) to be used in the project are thoroughly described. Information such as how the dataset or input is (was) obtained, and the characteristics of the dataset or input, should be included. It should be clear how the dataset(s) or input(s) will be used in the project and whether their use is appropriate given the context of the problem.

Nice start! Here are a few things to be sure to include in this section next time you submit your project:

- Please be sure to include the link to the dataset itself.

- Please explain a bit more about how the images are formatted. Are there layers? Are the dimensions consistent?
- How many images are there in the wrist-study data that you'll be using?
- Please describe how the classes in this subset are balanced (how many images are normal? how many are abnormal?). If each class has roughly the same number of instances, then we can say that the dataset classes are balanced. This is important for pre-processing and choosing a good metric.
- How will you split the data into training/validation/testing sets? Will you do anything to maintain class balances across each subset (e.g. upsampling or stratifying the dataset using the labels)?

Student clearly describes a solution to the problem. The solution is applicable to the project domain and appropriate for the dataset(s) or input(s) given. Additionally, the solution is quantifiable, measurable, and replicable.

I think that this is a pretty good high-level summary of how your solution will work. Great job!

A benchmark model is provided that relates to the domain, problem statement, and intended solution. Ideally, the student's benchmark model provides context for existing methods or known information in the domain and problem given, which can then be objectively compared to the student's solution. The benchmark model is clearly defined and measurable.

As I mentioned in the study group, this is a great baseline for your project. One thing that you should be sure to note in your report is how this model performed with the specific type of data that you're using. It's important to provide a very clear threshold that you're trying to beat.

Student proposes at least one evaluation metric that can be used to quantify the performance of both the benchmark model and the solution model presented. The evaluation metric(s) proposed are appropriate given the context of the data, the problem statement, and the intended solution.

Accuracy (number of patients accurately classified/total number of patients for wrist study) could be a primary metric.

Recall, precision, and the F1 score should all be fine for this project. However, please note that accuracy is only appropriate for classification problems where the dataset classes are balanced:

https://towardsdatascience.com/accuracy-paradox-897a69e2dd9b

Please be sure to note how your classes in the dataset are balanced. If there are similar numbers of each class, then you don't need to change anything here. Otherwise, if the classes are imbalanced, then you should avoid using accuracy as a metric for the project.

Also, one other thing you'll need to add:

- According to the paper where the benchmark model is published, the model was quantified using the Cohen's kappa metric. You'll need to use this metric to compare your model's results with the published results from the paper:

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.cohen_kappa_score.html

Student summarizes a theoretical workflow for approaching a solution given the problem. Discussion is made as to what strategies may be employed, what analysis of the data might be required, or which algorithms will be considered. The workflow and discussion provided align with the qualities of the project. Small visualizations, pseudocode, or diagrams are encouraged but not required.

Great job here!

Suggested:

- I think that you've got the right idea here. Keep in mind that there are some other pre-trained models that you might want to consider (ResNet50 for example).
- Data augmentation might be useful (especially if you need to upsample one of the classes).

Proposal follows a well-organized structure and would be readily understood by its intended audience. Each section is written in a clear, concise and specific manner. Few grammatical and spelling mistakes are present. All resources used and referenced are properly cited.

The template format is followed and the proposal is well written.