Response to Reviewer 2 Comments

Point 1: The sentence "In this paper, a combination of SOTA model such as DenseNet, 8 InceptionNet, ResNet, NasNet, and MobileNet and Soft-Attention is proposed" in the abstract is not correct. You didn't use combination of all them in a specific unique structure. Different combinations of these networks are used in your proposed approach. Discuss about it or correct this sentence.

Response 1: Actually, as you commented our proposed method is not the combination of all models in a specific unique structure. Our proposal is a combination of one Deep Learning model (DenseNet, ResNet, etc) with Soft-Attention. We have changed the sentence in the abstract section from line 8 to line 9.

Point 2: All of the equations should be numbered.

Response 2: All of the equations are numbered.

Point 3: How do you propose the weights formula? (Section 2.2.5). Is there any related reference? How much is the size of output W in this equation?

Response 3: We have already added the citation to the paper that inspires us to propose the W formula. The size of output W is equal to the number of classes in the data set. In this research, the W size is 7.

Point 4: As I know, the input size of the used CNNs such as mobileNetV2, mobileNetv3, ResNet, etc are not same. Do you resize all of the images to the same size to start process? Or do you run each CNN with different input size?

Response 4: The ways of image processing are given in the 2.Material and Methods → 2.2 Methodology → 2.2.2 Input Schema from lines 226 to 234. Images are processed correspondingly to the backbone model input requirement.

Point 5: I think your proposed approach can be used widely in medical applications. For example, it can be used in DNA classification, etc. For example, I find a paper titled "DNA Repair Genes (APE1 and XRCC1) Polymorphisms–Cadmium interaction in Fuel Station Workers", which has enough relation. Cite this paper and discuss about it as one the advantages of your proposed approach.

Response 5: We agree that our proposed method can be used in many different approaches. However, the paper you suggested is about DNA Repair Genes so their approach is to analyze the effect of Cadmium on Fuel Station Workers, therefore visualizing the genes is needed. DNA classification, on the other hand, uses another approach. Since the above image is not the real

image of genes (hard to extract the main pattern), DNA sequence ("AGXTTTATTX" for example) is applied.

Point 6: It is suggested to discuss about the runtime of your proposed method briefly (Compare performance with other methods is not needed)

Response 6: The runtime of the proposed method is given in section 3.Results → 3.2 Discussion → Table 6.

Point 7: In scientific papers, usually, the title of the tables is written above them.

Response 7: Owing to the Latex template provided by MDPI, the table captions are auto-generated below them.

Point 8: The pre-process of skin lesion recognition is skin detection process. For example, I find a paper titled "An innovative skin detection approach using color based image retrieval technique", which has relation. Cit this paper and discuss about the necessary pre-process in this scope briefly.

Response 8: The paper you suggest is really fantastic. We have already cited it and discussed briefly it and the need for image preprocessing before feeding it into the model. It is added in section 2. Materials and Method → 2.2 Methodology → 2.2.2 Input Schema from lines 217 to 225.