Response to Reviewer 1 Comments

Point 1: The novelty and innovativeness of the paper must be outlined. The paper uses well-known deep learning models (DenseNet, InceptionNet, ResNet, NasNet, and MobileNet), which have been used several times before by different authors, including for the skin disease recognition tasks. The research paper should go beyond the application of known methods, which is a standard engineering task. The proposed methodology is traditional.

Response 1: The novelty and innovations of the paper are:

- **Applying the Soft-Attention to unsupervisedly extract to heat map feature representing the main part of the lesion.**
- **Using metadata including age, gender, and localization as a feature**
- **Using a new weight loss to overcome the problem of data imbalance.**

They are all outlined in the abstract.

Point 2: The related works subsection is poorly organized and presented. The selection of works seems to be ad hoc. I suggest to add some structural organization (e.g., machine learning based, deep learning based methods) and discuss the state-of-the-art papers published in the previous 2-3 years, which better reflect the trends and achievements in this rapidly evolving research field. The authors are encouraged to discuss, for example, Malignant skin melanoma detection using image augmentation by oversampling in nonlinear lower-dimensional embedding manifold. Extraction of abnormal skin lesion from dermoscopy image using VGG-SegNet. Melanoma segmentation: A framework of improved DenseNet77 and UNET convolutional neural network. Finalize by discussing the limitations of existing methods as a motivation of your study.

Response 2: The related work has already been reorganized. We have divided all papers into two sections containing Deep Learning based method and Machine Learning based method. We also discuss furthermore the limitation of related works. We have also cited and discussed the three papers you suggested.

Point 3: Provide specific values of image augmentation parameters for replicability.

Response 3: We have already added the parameters used in image data augmentation section 3. Results → 3.1 Experimental Setup → Training from lines 292 to 301.

Point 4: Explain how you set the hyperparameter values for training such as training epochs and batch size. Did you use any hyperparameter optimisation/finetuning?

Response 4: For the training epochs, batch size, and learning rate, we have taken reference from some SOTA model training. Then we set the initial epochs to 250. The other SOTA usually use a batch size of 16 and 32, we have tried and found out that a batch size of 32 is much better. The

initial learning rate is set to 10^(-4) decreasing with a factor of 0.2 if the validation accuracy does not increases after 25 epochs. We choose the factor of 0.2 because after trying 0.5 or 0.1, the accuracy does not increase after reducing the learning rate. We have also discussed it from lines 302 to 306.

Point 5: More experimental results should be added such as confusion matrices and ROC plots.

Response 5: We have already added the confusion matrix and ROC plots in section 3.Results → 3.2 Discussion → Figure 10, 11, and 16.

Point 6: Evaluate the computational complexity of the proposed methodology. Report on the total number of trainable parameters in the proposed model.

Response 6: The trainable parameters of the proposed method have been added at section 3. Results → 3.2 Discussion → Table 6. Training time and infer time, on the other hand, are provided for computational complexity at Table 6.

Point 7: Compare your results with the results of other studies using the same datasets.

Response 7: We have already made a comparison between our approach with other studies using the same data set in section 3.Results → 3.2 Discussion → Table 9

Point 8: Add the discussion section and discuss the limitations of the proposed methodology.

Response 8: We have already discussed the limitation in distinguishing melanoma and the black nevus in the last part of section 3. Results → 3.2 Discussion → Figure 17. The limitation is the melanoma and the black nevus sometimes look the same, therefore the proposed model may get confused.

Point 9: The conclusions section just summarizes all findings of this study. What are the deeper implications of this study and its significance to the biomedical research field? Support your claims by the main numerical findings from this study.

Response 9: We have already provide the needed results in the conclusion section.

Point 10: Extend Table 1 by reporting more specific information about the discussed studies such as deep learning models used and accuracy (performance) achieved.

Response 10: We have already extended the Table 1 for more information including type of method and the result provided.

Point 11: The caption of Table 6 is confusing.

Response 11: We have already changed the caption of Table 6 to more understandable one.

Point 12: Why there are missing values in Table 7?

Response 12: We have already filled the missing results.