Our point-by-point responses to the reviewers’ comments are given at the bottom of this letter.

All the changes are highlighted using the Yellow color in the revised manuscript.

Reviewer 1: Thank you for submitting your manuscript to Intelligence-Based Medicine. Please find my general and specific comments below.

General comments:

1. The introduction does a good job communicating the role that ML can play for solving this clinical problem and the importance of CAD approaches. There is strong tie-in between the use of CAD techniques and solving this clinical problem.

**Our Reply:** Thank you very much for your complement.

2. It is clear the authors spent considerable time researching existing methods and approaches. One way this section could be strengthened is by including some discussion about how their approach differs from these other approaches/the paper's unique contributions. A summary table of all the approaches, with columns such as approach, dataset size, performance, etc. would also be interesting to see in this section.

**Our Reply:** Thank you very much for your nice suggestions. We have added a summary table of all the approaches. Please see the section 2 in the revised manuscript.

3. Please use as consistent nomenclature as possible when describing the neural networks - e.g. in Figure 1 "DenseNet121" is referenced, but throughout the rest of the paper it is "DenseNet201". Additionally, please use a consistent format with/without dashes for the neural network architecture names - i.e. "VGG-19 and DenseNet-201" (with a hyphen) OR "VGG19 and DenseNet201" (without a hyphen). In 3.3.3 you also mention using VGG-16 in DFU\_MultiNet while in other sections of the manuscript you mention VGG-19.

**Our Reply:** Thank you very much for your nice suggestions. We have changed these mistake. Please see the revised manuscript.

4. A brief primer on how deep neural networks, specifically MTL and transfer learning, would be valuable to the audience. Please include 2-3 paragraphs of this in the introduction. Discussion can be kept at a relatively high level, but these are concepts that readers would benefit from seeing a brief exposition on.

**Our Reply:** Thank you very much for your nice suggestions. We have discussed the importance of deep neural networks, specifically MTL and transfer learning. Please see the revised manuscript.

5. In the discussion section, please add 1-2 paragraphs of what steps would need to be taken as well as how this approach could be adapted for clinical use. Please describe a workflow for how clinicians could use this framework as a future Computer-Aided Diagnostic (CAD) tool.

**Our Reply:** We have discussed the adaptation steps of this approach for clinical use and also described how clinicians could use this framework as a future Computer-Aided Diagnostic (CAD) tool. Please see the discussion section in revised manuscript.

6. To the greatest extent possible, please provide more details about the dataset used for this study. I recommend including it as a separate sub-section in Section 4. Please include demographic details of the patients, what hospital(s) these images were taken from, the original resolution of the images, etc. This information can provide readers with a sense of the heterogeneity of the dataset utilized for this study.

**Our Reply:** Thank you very much for your nice suggestions. We have provided more details about the dataset used for this study in sub-section 4.1. Please see the revised manuscript.

Specific comments:

0. Abstract:

a. Abstract is strong - my only recommendation is to include additional metrics for classification such as precision, recall, F1 score, and AUROC. These metrics will help to better communicate the efficacy of this approach.

**Our Reply:** Thank you very much for your nice suggestions. Accordingly, additional metrics for classification have been included into the abstract.

1. Introduction:

a. Please provide citations for the reports you are referencing in the first paragraph.

**Our Reply:** We have provided citations for the reports in the first paragraph.Please see the revised manuscript.

b. When you introduce ML and DL ("The performance of ML and DL approaches..."), please spell out "Machine Learning" and "Deep Learning" before you use these acronyms.

**Our Reply:** We have spelled out the full terms of ML and DL before using the acronyms.Please see the revised manuscript.

c. For the 3rd crucial point of this research work, I recommend adding 1-2 sentences on why a segmentation-free approach is advantageous for this use case.

**Our Reply**: We have added two advantages of the segmentation-free approach in the 3rd crucial point of this research work. Please see in the revised manuscript.

2. Related Literature:

a. Please see General Comment 2 above.

**Our Reply:** The summary table of all the approaches has been added after the related literature.Please see the new updated Table 1 in the revised manuscript.

3. Methodology:

3.1. DFU Dataset:

a. Good description of the approach as well as the training methodology. Could the authors please briefly describe how the datasets were split (e.g. at random)? Please incorporate into the paper for completeness as well.

**Our Reply:** Thank you very much for your nice suggestions. We have briefly described how the datasets were split. Please see the revised manuscript.

3.2. Data preprocessing:

a. Good description of dataset pre-processing. Figures 3a and 3b are also good for visualizing the dataset distribution samples.

**Our Reply:** Thank you very much for your complement.

b. Table 3 is helpful for understanding some of the data augmentation techniques used. Please briefly describe how samples are selected for Data techniques strategies in which the parameter value is not True/False - is a sample parameter value taken from a uniform distribution from a range of values? For instance, for image rotation range, is the rotation for an augmented sample simply selected as \theta ~ U([0, 90])?

**Our Reply:** Thank you very much for your nice suggestions. We have briefly described how samples are selected for Data techniques strategies. Please see the revised manuscript.

3.3. Building the MTL Model:

a. Figure 5 is very helpful for visualizing the individual pre-trained CNNs and how subsequent model layers are structured and connected.

**Our Reply:** Thank you very much for your nice suggestions. We have described the structure and connection process of subsequent model layers in section 3.3. Please see the revised manuscript.

b. Please precisely describe the dataset(s) that each of the 3 CNNs used for this model were pre-trained on. Please also explain why you chose to use pre-trained models rather than re-training them from scratch (i.e. the use of transfer learning).

**Our Reply:** Thank you very much for your nice suggestions. We have precisely described the dataset(s) that each of the 3 CNNs used for this model were pre-trained on and also explained why we choose pre-trained models rather than re-training them from scratch. Please see the revised manuscript.

3.3.3. VGGNet:

a. Please use "ILSVRC competition" rather than "ILSVRC competitiveness".

**Our Reply**: We have used "ILSVRC competition" rather than "ILSVRC competitiveness". Please see in the revised manuscript.

b. In the other sections you discuss using VGG-19, however here you state VGG-16. Please keep this consistent throughout the manuscript. Please see General comment 3 as well.

**Our Reply**: In section 3.3.3, it would be VGG19 instead of VGG16 and we have corrected this mistake. Please see in the revised manuscript.

3.4. Fine-tuning process:

a. This sub-section is a good description of the fine-tuning process. Please include Table 4 in an Appendix - the reviewers appreciate the inclusion of this table (it is a good example of details to include in research papers such as these) but it should be in the supplemental section of the paper.

**Our Reply:** Thank you very much for your nice suggestions. We have included Table 4 in an Appendix A. Please see the revised manuscript.

4. Experimental setup, performance metrics, and results analysis:

4.1. Experimental Setup:

a. The reviewers appreciate the authors including their experimental setup and computing specifications. Could the authors please also incorporate this into an Appendix/supplemental table and provide citations for Keras and Python?

**Our Reply:** Thank you very much for your nice suggestions. We have incorporated a supplemental table for the experimental setup with specifications and also provided citations for Keras and Python. Please see the experimental setup section in the revised manuscript.

4.2. Performance Metrics:

a. Please note that a formula/description for recall (REC) was not provided. Please include for consistency.

**Our Reply:** Thank you very much for your nice suggestions. We have provided the recall (REC) formula with description. Please see the revised manuscript.

b. In Formula 7 (Matthews's Correlation Coefficient / MCC), in the first term of the denominator, it should be "(TP + FP)", not "(TP + TP)".

**Our Reply:** Thank you very much for your nice suggestions. We changed the first term of the denominator in the MCC formula. Please see the revised manuscript.

c. Lastly, for MCC this should be "Matthews Correlation Coefficient", not "Matthews's Correlation Coefficient".

**Our Reply:** Thank you very much for your nice suggestions. We have changed "Matthews's Correlation Coefficient" to "Matthews Correlation Coefficient". Please see the revised manuscript.

4.3. Training and parameter optimization:

a. In the first paragraph, please change "a small amount of learning rate" to "a small learning rate". In this same sentence, please also clarify that this makes the CNN classifier training slower, not the CNN classifier itself.

**Our Reply:** Thank you very much for your nice suggestions. We have changed "a small amount of learning rate" to "a small learning rate". Please see the revised manuscript.

b. In this section, please also provide 1-2 sentences on the importance of utilizing a separate validation set for hyperparameter tuning and why hyperparameters shouldn't be tuned using evaluation on the test set.

**Our Reply:** Thank you very much for your nice suggestions. We have discussed the importance of utilizing a separate validation set for hyperparameter tuning and why hyperparameters shouldn't be tuned using evaluation on the test set. Please see the revised manuscript.

c. Table 5 is helpful. Please briefly explain what the "Factor" parameter is. For hyperparameter tuning, was a structured approach such as Grid Search or Random Search used? If so, please provide the tables of values used during selection in addition to what you already have (i.e. the final hyperparameter values). If a search was not used, how were optimal hyperparameters selected?

**Our Reply:** Thank you very much for your nice suggestions. We have briefly explained the "Factor" parameter and the optimal hyperparameters selected approach. Please see the revised manuscript.

d. The heading for this sub-section is larger than the other sub-section headings (guessing just a slight formatting detail).

**Our Reply:** Thank you very much for your nice suggestions. We have changed size of the heading for this sub-section. Please see the revised manuscript.

4.4. Results analysis:

a. First sentence, please change "receiver operating characteristics" to "receiver operating characteristic".

**Our Reply:** Thank you very much for your nice suggestions. Accordingly, we have changed "receiver operating characteristics" to "receiver operating characteristics". Please see the revised manuscript.

b. Table 6 does a good job demonstrating the advantages that can occur through ensembling. Please bold the best-performing metric for each category in the respective row (please bold multiple if there is a tie).

**Our Reply:** Thank you very much for your nice suggestions. We have bold the best-performing metric. Please see the new updated Table 7 in the revised manuscript.

5. Discussion:

a. In the first sentence, please change "less obeying" to "not obeying".

**Our Reply:** Thank you very much for your nice suggestions. Accordingly, we have changed "less obeying" to "not obeying". Please see the revised manuscript.

b. Figure 8 does an excellent job highlighting the improvement in performance that can be seen from transfer and ensemble learning.

**Our Reply:** Thank you very much for your complement.

c. For Table 7, please emphasize that while the proposed framework performs the best, this comparison cannot be made completely rigorously if the same dataset(s) and data splits are not used between all of the models. If it is the case that all these other models were evaluated on the same dataset and dataset splits, please make this very clear.

d. In Table 7, please also include the number of total training and testing samples if there is column space to do so.

**Our Reply:** Thank you very much for your nice suggestions. Accordingly, we have added two columns (training and testing samples) to the new updated Table 8. Please see the revised manuscript.

e. This dataset is well-balanced between positive and negative samples. Please discuss in either a couple sentences or paragraphs how the authors anticipate ensembling would aid with classification performance in the presence of a highly imbalanced dataset (assuming data augmentation is still used).

**Our Reply:** Thank you very much for your nice suggestions. We have discussed about the classification performance of the ensemble technique used in this study in the presence of a highly unbalanced dataset. Please see the discussion section in the revised manuscript.

f. Have the authors considered any other ensemble formulations/architectures? Additionally, the idea of creating ensemble neural networks that have learned features after the output of neural networks is a powerful technique that can be leveraged across other clinical machine learning applications - please take 1-2 sentences to discuss how one of the contributions of this study is as an example use case of combining different MTL networks into a learned ensemble to improve overall classification performance.

**Our Reply:** We have discussed the contributions of this study in other clinical machine learning applications. Please see the discussion section in the revised manuscript.

5.1. Limitations of the study:

a. This is a great limitation to point out. It may be valuable to include computational metrics such as needed inference time. It may also be worth discussing how future work could be in applying this framework and optimizing it for use in clinical areas where there is not significant computing power available.

**Our Reply:** We have discussed the application of this framework in future work and the optimization process for use in clinical areas where there is no significant computing power available. Please see the section 5.1 in the revised manuscript.

b. Please also discuss if the individual pre-trained models you compared the DFU\_MultiNet approach to were trained/fine-tuned at all - if they were not, I recommend discussing this as another limitation of the study, because in this case the DFU\_MultiNet has had additional fine-tuning/training opportunities for the particular image classification task while the individual pre-trained networks have not. If this is not the case, please describe how the individual pre-trained networks are trained in this comparative evaluation (e.g. the last few layers are unfrozen).

**Our Reply:** We have discussed the training process of the individual pre-trained networks. Please see the section 5.1 in the revised manuscript.

c. While the dataset size still allows the results to be treated as rigorous, having access to larger datasets would be great. Please consider briefly discussing this point (it doesn't have to be phrased as only through the lens of a limitation, but also through the lens of future work).

**Our Reply:** We have discussed the future contributions of the proposed model in larger datasets. Please see the section 5.1 in the revised manuscript.

6. Conclusion:

a. Please change "fabricated" to "based on" and "parallely" to "in parallel".

**Our Reply:** Thank you very much for your nice suggestions. Accordingly, we have changed "fabricated" to "based on" and "parallely" to "in parallel". Please see the revised manuscript.



All the changes are highlighted using the Yellow color in the revised manuscript.