

Peer Review File

Article information: <https://dx.doi.org/10.21037/jmai-23-67>

Reviewer A

The paper introduces a transfer-learning based system for early detection of melanoma lesions, utilizing a convolutional neural network trained on ImageNet weights and fine-tuned for melanoma detection. While the paper is technically sound, there are, however, numerous typos, awkward sentences and grammatical errors (please use a spell checker!!!). The complete list of such errors is too long to present in its entirety in this review (see the annotated attached file for some comments). Also, Section II provides a deficient analysis of existing melanoma detection and noise removal prediction approaches. Among the missing references:

[1] MoNuSAC2020: A multi-organ nuclei segmentation and classification challenge; IEEE Transactions on Medical Imaging, 2021.

[2] Melanoma detection using adversarial training and deep transfer learning; Physics in Medicine & Biology, 2020.

[3] A Variational approach to maximum a posteriori estimation for image denoising; Proc. International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, 2001.

Reply: The references were added

* The structure of the manuscript can be improved. For instance, a separate related work section should be included.

Reply: More parts was added to the related work

* Please carefully review the paper for language errors, grammatical mistakes, and typos.

Reply: Done

* Figures and tables should be placed as near as possible to where they first referred to in the paper, preferably in the same page.

Reply: The format was given by the journal that figures and tables should be added at the end

* The confusion matrix needs to be further explained.

Reply: More description was added to the confusion matrix

* A "Discussion" section should be added to provide a more compelling and argumentative analysis of the obtained results.

Reply: Done

* What are the limitations of the proposed retrieval system?

Reply: Described in the paper

Changes in the text: The only drawback that our method has is it takes more time if the program is used in sequential way. Using parallel programming will omit the drawback and it will be as fast as other methods.

Reviewer B

The authors presented in this paper an interesting, innovative methodology and outstanding results that are superior to the AI-based Skin Cancer classification current state of the art.

However, the work needs careful review. The contents are not very well presented and organized, making comprehension more difficult. Terms that are either too general or too technical are often used, and references are lacking. Text formatting and syntax are poor

I strongly advise the authors to work on this aspect, because the proposed results are of interest to the scientific community.

Below I list in detail the points to be reviewed.

Lines 32 “highest rate” and 34 “deadliest type of skin cancer”: do you have updated prevalence/incidence rate and mortality/survival rate estimates to add? These would help to quantify “the problem” and your possible impact on it.

Reply: These were the most recently papers I could find

Line 37 “medical mistake”: do you have any reference quantifying how often melanoma are misdiagnosed or about skin pathology classification complexity?

Reply: reference was added

Changes in the text: Due to medical mistakes [6]

Lines 39 and 40 “Using computer science, the process of detecting the cancer can be much faster. In recent 40 decades Deep Learning has been considered as one of the best methods in detection and recognition”: please add some relevant examples/references.

Reply: reference was added

Changes in the text: In recent decade, Deep Learning is considered to be one of the best methods in detection and recognition [5].

Lines 47 and 48 “Many methods to increase the images 48 have been discussed in the paper”: Do you mean your own work or are you referring to other previous publications? If so, please add citations.

Reply: My own work

Lines 40, 43 and 49 “Deep Learning [...] machine learning [...] deep learning”: please be consistent in the use of capital letters. As field names, not proper names, they can be written in lower case.

Reply: Done

Changes in the text: All instances changed to lower case.

Lines 44, 49, 121, 124, 125, 126, 128, 138 and 144“overfitting [...] feature extraction [...] weights of a network [...] layer [...] softmax [...] back-propagation [...] learning rate [...] segmentation [...] pooling”. These are very technical terms, please add brief explaining sentences or references to help non expert readers.

Reply: Done

Changes in the text: When the model cannot generalize well and it fits completely or closely to the training data, overfitting happens.

Feature extraction is the second important part of a classification system. By feature extraction, one tries to extract the most important information from the data.

In transfer learning a network will use the weights of a neural network that was trained on another image dataset similar to what we want to use, and the dataset have had enough samples.

In this network, the last layer (The last layer of the fully connected layer) that is designed for classification 1000 classes

In softmax layer the number of nodes is equal to the number of outputs, and the value shows the probability of each output.

In the next step back-propagation (A method that tries to adjust the weights based on the error in the output) is applied to network to fine-tune the weights

Learning rate, is a hyperparameter used in neural network to control the change in model based on the error.

Segmentation divides image into two parts, the cancerous part and other part of the image, but it has its own drawbacks.

pooling layer reduces the dimensions and finds the highest response in local parts of the feature map

Line 52 “In skin cancer datasets most of the skin parts that needs attention 53 are located at

the center of the image”: Why, how can this be said? Explain this concept further or use an image to exemplify it (Figure 1?).

Reply: Images was refereed.

Changes in the text: In skin cancer datasets most of the skin parts that needs attention are located at the center of the image (refer to figure 1).

Line 67 “SVM”: please define the abbreviations before the use -> correct with Support Vector Machine (SVM)

Reply: Done

Changes in the text: Dorj et al. used Support Vector Machine (SVM) with deep SVM to ...

Line 68 “basal cell carcinoma (BCC), squamous cell carcinoma(SCC)”: please add spaces before parenthesis.

Reply: Done

Changes in the text: basal cell carcinoma (BCC), squamous cell carcinoma (SCC)

Line 71 “SK”: do you mean Seborrheic keratosis (SK)? Please define abbreviation.

Reply: Done

Changes in the text: Seborrheic keratosis (SK)

Lines 86, 89 and Table 1: please add “%” to accuracy scores digits.

Reply: Done

Changes in the text: They reached to 86.1% accuracy [16]. Ali et. al. used self attention based PGAN to detect vascular, pigmented benign keratosis, pigmented Bowen’s, nevus, dermatofibroma, etc. ISIC 2018 dataset was used. A generative model was enhanced with a stabilization technique. The accuracy was 70.1% [17].

Line 92 “III. METHODOLOGY” and line 147 “IV. RESULTS”: is there and extra space? Please use everywhere same numbering formatting -> 2.2 Methodology ... 2.3 Results

Reply: Done

Line 63 “II. LITERATURE REVIEW”: as before -> correct with 2.1 Literature review.

Concluding comment on section “Literature review”: You have well-presented many noteworthy works in the literature and the AI methodologies used in them. However, I would enrich this paragraph by indicating what kind of data these works used (always images or also numerically coded clinical features?) and what kind of tasks they defined (classification? segmentation? etc.). Also, it would be interesting to know for all the papers what metrics they adopted and what performances they obtained with their models.

Reply: Done and more information was added to the literature review

Line 97 “proprocessing”: typo -> processing (or to avoid repetition, rather use image preparation)

Reply: Done

Changes in the text: One is an image preparation part that fix ...

Lines 101, 102 “For building a strong CNN many images are needed for training and testing which is a huge challenge for skin cancer detection problem where there are not many images”: please avoid repeating the word “image”. Please make clear why is difficult to have large dataset (low prevalence, low availability of image sets, scarce labelled sets, etc.?)

Reply: Done

Changes in the text: For building a strong CNN many images are needed for training and testing which is a huge challenge for skin cancer detection problem. Creating large datasets with variety in different cancer types is time consuming.

Line 103: add citation for the dataset ISIC2020

Reply: Reference was added

Line 103, 104 “1400 of these images”: usually is preferred to express test and training size also with percentages. Please add.

Reply: Done

Changes in the text: 1400 (Around 4%) of these images were chosen for testing and set aside.

Line 104, 105 “The data augmentation should not be applied on the testing images”: an inexperienced reader cannot understand the motivation. I would rephrase it to saying that only the training set is augmented to provide more examples to the model and increase its robustness (while the test set does not need to be augmented since real images are needed to test performance).

Reply: Rephrased as mentioned

Line 106 “ 35×31726 ” and line 109: please use the \times (multiplication sign symbol)

Reply: Done

Changes in the text: In order to increase the number of training images, each image was rotated 10 degrees which resulted in 35×31726 new images.

Line 107 “Another useful technique was used from [19]. 5 percent and 10 percent [...]”: please rephrase first presenting the description on the augmenting strategy (cropping and zooming), then the citation. Please use % symbol instead of term “percent”.

Reply: Done

Changes in the text: Another technique on the augmenting strategy was used from [19]. In their work, 5% and 10% of the image was cropped from the top, bottom, left and right part of the image.

Line 112 “class with higher samples”: specify the label of the most prevalent class.

Reply: Done

Changes in the text: The model has bias towards the class with higher samples, in skin cancer detection, most images belongs to benign samples.

Line 112 “To overcome this problems other techniques were used”: So in the end, how large did the training set fed by the model become? Please indicate the number of total images.

Reply: Done

Changes in the text: After applying these techniques, the total number of training images become $16 \times 8 \times 35 \times 31726$.

Lines 117-120: this part is a repetition of what has already been written above. It could be shortened to one sentence to then introduce the techniques of transfer learning.

Reply: Done

Changes in the text: Training CNN is not an easy task, the reason is CNN needs many images for training and testing. Transfer learning is used to overcome the problem of not having many images in addition to image enhancement.

Line 122, 123 “The way that I used transfer learning is as follows: First the CNN weights for ImageNet are used as the initial weights for the network”: please do not repeat the verb “use” (replace with exploit, adopt, implemented...). Add citation for ImageNet work. Lastly, in scientific publication is better to avoid use of “I”/“We”, passive form is preferable (e.g. In this work transfer learning was exploited as follows).

Reply: Done

Changes in the text: In transfer learning a network will use the weights of a neural network that was trained on another image dataset similar to what we want to exploit, and the dataset have had enough samples. In this work transfer learning was exploited as follows: First the CNN weights for ImageNet [24] are used as the initial weights for the network.

Line 126: 2 classes softmax.. hence you are doing a binary classification right? Specify this is your task for non-technical readers.

Reply: Done

Changes in the text: In this network, the last layer (The last layer of the fully connected layer) that is designed for classification 1000 classes (Number of classes in ImageNet) is changed with a softmax layer with 2 classes (Melanoma versus non-Melanoma), which is binary classification (just to mention instead of softmax sigmoid function can also be used).

Line 127-131 “The other important factor is that we do not need to apply the back-propagation to the whole network”: is it really necessary to reiterate it in this paper? it is a feature of the transfer technique.

Reply: Wanted to mention that this can be a hyper-parameter

Changes in the text: Removed

Line 133 “Three different networks was”: correct verb -> were

Reply: Done

Changes in the text: Three different networks were trained based on what mentioned above.

Line 134 “on what mentioned above”: please be clearer

Reply: Sentence was changed

Changes in the text: Three different networks were trained based on different images and transfer learning.

Line 137 “There are twos ways”: correct typo -> two

Reply: Done

Changes in the text: There are two ways to overcome this problem.

Line and Figure 4 “The Three networks were trained and fine-tuned”: this methodological aspect is not very clear: you defined 3 different networks and then you fed them with 3 different transformed subsets from the original images. But is not clear if you defined 3 different new architectures or if the nets are 3 re-trained/finetuned versions of ImageNet (exploiting transfer learning). In case you have defined 3 completely new networks, please detail the architectures (input layer, conv, pooling,.activation func.. softmax, output). In any case include details of loss function, epochs, batch size, learning rate, optimizer algorithm, and other hyperparameters essential to understand how the training was done. Lastly, were techniques used to reduce overfitting, other than data augmentation (regularization, early stopping, etc)???

Reply: They are not new structures, 3 fine tuned versions of imagenets

Lines 149-152 and 154-159: these sentences should be moved to the methods section (not properly about results).

Reply: Moved to method parts

Lines 155, 156 and 157: please use formula formatting for the equations of accuracy, recall and precision (no inline text)

Reply: Done

Tables 1 and 2: please fix label numbering: both tables are labeled with “Table 1”

Reply: Done

Concluding comment on section “Results”: from the figures and tables it is clear that the

results of this work are superior to the state of the art. However in the text section this aspect is not well discussed. The performance metrics obtained from the model must be clearly presented to the reader here, reporting the numbers of accuracy, recall, precision and AUC as percentage. Any intermediate results obtained in the experiments carried out (eg for hyperparameter tuning) must also be presented. Moreover, a discussion section is completely missing. The authors must add this section in which to compare their results and model design choices in detail with those already known in the literature. The drawbacks of such AI models (generability issues, data bias, eg sex o ratial biases etc.) and the feasibility of deploying such models to effectively use them in clinical practice should also be mentioned. Finally, for greater transparency and reproducibility of the results, it would be advisable to insert a link with the availability of the code (e.g. GitHub repository)

Reply: Discussion was added

General comment on text formatting: please use consistent formatting for headings and numbering of sections, subsections, tables and figures (Roman numerals, Arabic numerals, and letters are mixed, titles are written both with all-capital text or only with first initial capitals etc.). I also suggest a general revision of the syntax of the text, so as to make some somewhat generic sentences more formal, precise, or technical (you often made use of words “some”, “usually”, “ways”, “use” ect., belonging to an informal linguistic register that make you look inaccurate). The presence of many technical terms without explanations makes it difficult for clinicians and non-AI expert readers to understand. In my opinion, the authors should decide well what target audience this work of theirs is aimed at and arrange the text and content accordingly.

Reply: Changes were applied to the requested parts