Diagnosis of Skin Cancer Using Hierarchical Neural Networks and Metadata

Beatriz Alves(✉), Catarina Barata, and Jorge S. Marques

Institute for Systems and Robotics, Instituto Superior Técnico, Lisbon, Portugal
beatriz.c.alves@tecnico.ulisboa.pt

Abstract. Skin cancer cases have been increasing over the years, making it one of the most common cancers. To reduce the high mortality rate, an early and correct diagnosis is necessary. Doctors divide skin lesions into different hierarchical levels: first melanocytic and non-melanocytic and then as malignant or benign. Each lesion is also assessed taking into consideration additional patient information (*e.g.*, age and anatomic location of the lesion). However, few automatic systems explore such complementary medical information. This work aims to explore the hierarchical structure and the patient metadata to determine if the combination of these two types of information improves the diagnostic performance of an automatic system. To approach this problem, we implemented a hierarchical model, which resorts to intermediary decisions and simultaneously processes dermoscopy images and metadata. We also investigated the fusion of a flat and hierarchical model to see if their advantages could be brought together. Our results showed that the inclusion of metadata has a positive impact in the performance of the system. Despite hierarchical models performing slightly worse than flat models, they improved certain lesion classes, and can narrow down the lesion to a sub type, as opposed to the flat model.

Keywords: Skin lesion diagnosis · Hierarchical model · Metadata · Deep neural networks

1 Introduction

Skin cancer impact has been increasing due to the alarming growth in the number of cases over the years. According to the World Health Organization, 2/3 million cases of non-melanoma and 132,000 cases of melanoma occur worldwide every year. These numbers reinforce that skin cancer is one of the most common cancers [1]. Melanoma in particular has a high mortality rate when detected in the latest stages. Therefore, an early and accurate diagnosis must be achieved.

This work was supported by the FCT project and multi-year funding [CEECIND/00326/2017] and LARSyS - FCT Plurianual funding 2020–2023; and by a Google Research Award'21.

A. J. Pinho et al. (Eds.): IbPRIA 2022, LNCS 13256, pp. 69–80, 2022.
https://doi.org/10.1007/978-3-031-04881-4_6

The creation of the ISIC Challenges [2] promoted the development of the skin cancer classification models using deep learning techniques. While a variety of deep learning methods were proposed to diagnose dermoscopy images, they were still found to lack in one or more aspects. Deep learning techniques do not understand the information they are dealing with, they simply try to detect patterns or correlations among the different skin lesions. When dealing with only images, important traits can be neglected, such as the anatomical site, which can be a decisive diagnostic feature in some lesions. This conveys that a significant portion of the knowledge acquired by doctors during their medical training is not being put to use in the existing models. This is also true for lesion taxonomy, *i.e.*, the hierarchical organization of lesion types defined by dermatologists which could be used to provide a better understanding of the diagnosis made by the system. While some works on recent literature try to innovate how hierarchical classifiers are used [3,4], others study which method between flat and hierarchical models is better [5], or which hierarchical model prevails among all the different hierarchical structures [5,6]. However, none of these works has explored the incorporation of patient metadata.

Most skin lesion diagnosis systems are based on the analysis of dermoscopic images using flat classifiers. Due to the lack of two previously mentioned topics in the current literature, this work has two main goals. First, it aims to complement dermoscopic images with clinical information/metadata (age, gender, location of the lesion) and to evaluate differences in performance. The second objective is to use medical information about the taxonomic structure of the lesions shown in Fig. 1, using hierarchical classifiers. It is also intended to combine these two types of information and assess whether they allow to improve the performance of the system. The hierarchical structure that this work used can be seen in Fig. 1. This figure displays the eight skin lesions found in ISIC 2019 training dataset [7–9], which are Nevus (NV), Melanoma (MEL), Dermatofibroma (DF), Basal cell carcinoma (BCC), Squamous cell carcinoma (SCC), Actinic keratosis (AK), Benign keratosis (BKL) and Vascular (VASC).

This document is organized as follows. Section 2 explains all the methodologies used to classify skin lesions with and without metadata. Section 3 describes the experimental setup and presents the results with their associated analysis. Lastly, Sect. 4 discusses the conclusion and future works topics.

2 Methodology

In this work we compared 3 approaches: (i) a flat model, that performs a single decision, (ii) a hierarchical model, which resorts to intermediary decision before predicting the final diagnosis, and (iii) a mixed model which combines the two previous models. Additionally, we compare different ways to combine images and metadata for each of the previous approaches. Note that, in all models, images are processed using a convolutional neural network (CNN), while metadata are processed using fully connected layers (FCLs).

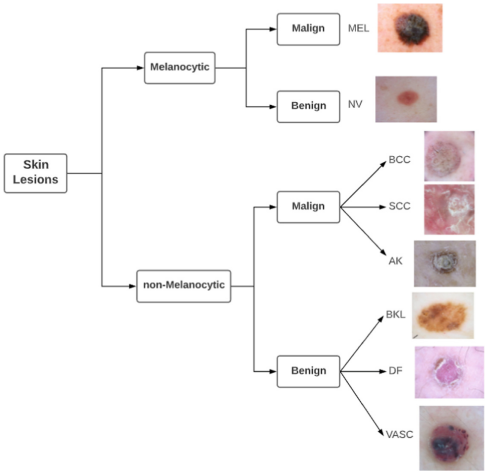

Fig. 1. Hierarchical organization of skin lesions in ISIC 2019 dataset. Dermoscopy images taken from [7–9].

2.1 Flat Classifier

Flat classifiers are simple, straight-forward models. They only need a single classifier to predict all of the categories as it does not take into account the inherent hierarchy among them. Therefore, this model is only required to, given an image and the corresponding metadata, predict one of the classes out of the eight-possible.

In the training phase, the flat classifier is trained with all the images and/or metadata of the eight classes. To train the model we use the training dataset and to choose hyperparameter values we use the validation dataset. The training is done with categorical cross-entropy loss function and Adam optimizer [10].

2.2 Hierarchical Classifier

Unlike the flat model, this one takes into account the taxonomy of several lesions by making intermediary decisions before reaching the final decision. With Fig. 1 in mind, we consider 3 levels of hierarchy: (i) melanocytic vs non-melanocytic, (ii) benign vs malignant, and (iii) the final diagnosis. We can see the visual representation of the proposed hierarchical model in Fig. 2. The model is an aggregation of 5 different classifiers (a, b, c, d and e) with (a) and (c) as intermediary decision classifiers and (b), (d) and (e) as final decision classifiers. The inside architecture of the individual classifiers is the same as its flat counterpart, with the only difference being in the $softmax$ block. Since classifiers (a), (b) and (c) distinguish between 2 types of lesion, their $softmax$ has 2 neurons and classifiers (d) and (e) have three, because they separate three lesions.

To predict a lesion, it is not mandatory that the data passes through all classifiers. There are 3 possible paths that it can go through. Path 1 starts in

classifier (a) and ends in classifier (b), path 2 goes from classifier (a) to (c) and ends in classifier (d) and path 3 is the same as path 2 but ends in classifier (e) instead. For example, if classifier (a) predicts the lesion to be melanocytic, then the lesion goes to classifier (b) and never reaches classifier (c), (d) and (e). Thus, the adopted path depends on the predictions of the intermediate classifiers (a) and (c).

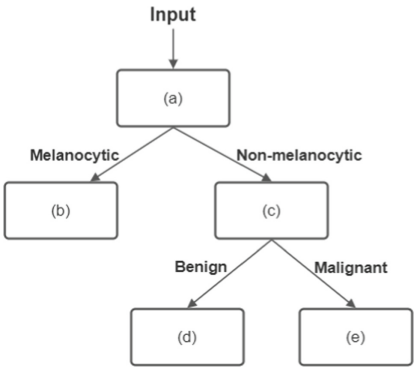

Fig. 2. Generic structure of the classifier block diagram in the hierarchical model. Classifier (a) distinguishes melanocytic lesions from non-melanocytic, (b) NV from MEL, (c) benign non-melanocytic lesions from malignant non-melanocytic lesions, (d) distinguishes between BKL, DF and VASC lesions and classifier (e) sets apart AK, BCC and SCC.

The training phase of the hierarchical models has the same loss function and optimizer as the flat classifier. However, the hierarchical model's classifiers (a)-(e) are only trained with the corresponding subset of the lesions it diagnoses. For example, classifier (d) is trying to diagnose BKL, DF and VASC so it is only trained with data of these three lesions. It is important to note that each classifier is independently trained.

2.3 Methods to Combine Images and Metadata

To combine the metadata with images, we opted to use only early fusion methods, where the combination is done at the feature level. We developed 3 different methods.

The first method is the **concatenation** between both features. First, the features are extracted from the images and from the metadata. Then they are concatenated into a single vector and sent to the classification block, which returns the diagnosis.

The second method is the **multiplication of features** as suggested in [11]. It tries to replicate an attention module by making the model learn which feature maps are less relevant and assigning lower values or even zero to their respective

positions. Here, the metadata goes through a FCL of d neurons, where d represents the dimension of the last convolutional layer of the CNN. After the feature extraction, they are multiplied element-wise and the corresponding vector is sent to the classification block, which returns a diagnosis.

The third method was inspired on [12] and it consists on **reducing the number of image features**. It fixates the number of metadata features, m, and changes the number of image features according to hyperparameter r. Its relation is shown in Eq. 1, where n is the number of the reduced image features, m is the number of metadata features and r is the ratio (0–1) of image features present in the combined feature vector.

$$n = \frac{m}{1-r} - m \tag{1}$$

We also set the number of metadata features, m, as an additional hyperparameter. To reach the number of reduced image features, n, the output of the CNN goes through an extra FCL layer with n neurons and ReLU activation. Simultaneously, the metadata goes through a similar layer but with m neurons instead. Both results are concatenated to form the combined feature vector and sent to the classification block to diagnose the lesion. Figure 3 illustrates this method.

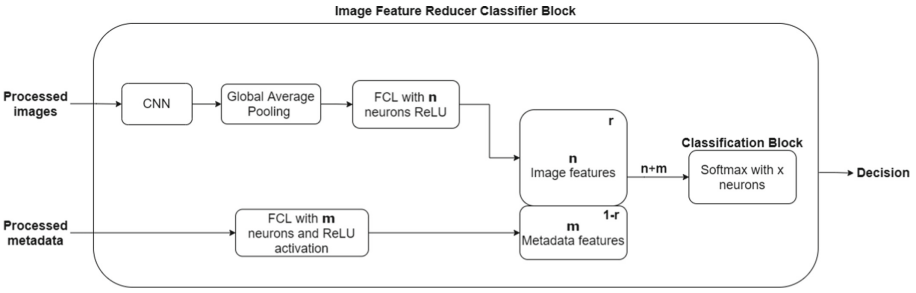

Fig. 3. Classifier block diagram of third method, image feature reducer.

2.4 Selection Between Flat and Hierarchical Models

Hierarchical and flat classifiers have different strengths. By considering the diagnosis from both classifiers, we tested whether both models' advantages can be brought together and if their individual weaknesses can be eliminated. For this, we created three new models which are based on the confidence of the diagnosis of both models.

The first mixed method is a direct competition between the hierarchical and the flat classifiers and it returns the decision of the model with the higher confidence. In this case, each classifier returns the class with the highest probability. Note that the probability is given by the $softmax$ layer, present in both models. The higher the probability, the higher is the confidence of the model in its

decision. For each data, the flat probability is compared to the correspondent hierarchical individual probabilities. If one of these probabilities is lower than the flat, the flat classifier makes the final decision. Otherwise, the decision falls upon the hierarchical classifier. In this and the following mixed models, the flat model has only one probability, while the hierarchical model has either 2 or 3, depending on which path the data goes through. If the data follows path 1, then it will have 2 probabilities, correspondent to classifiers (a) and (b). Otherwise, it will have 3 individual probabilities, which are from classifier (a), (c) and (d) or (e).

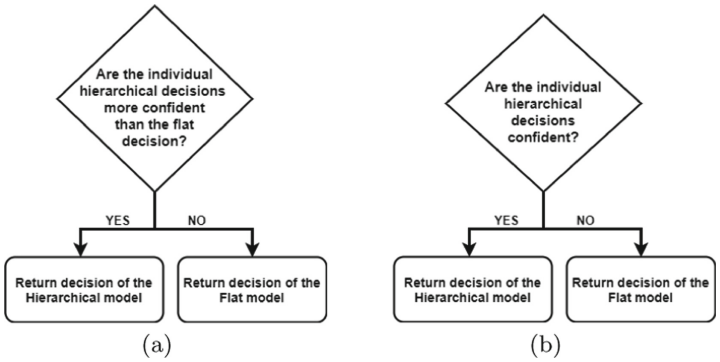

Fig. 4. Diagram of first mixed model, (a), and of the second and third mixed models, (b).

The second mixed method works similarly to the first one. However, rather than using the flat probability, the hierarchical probabilities are compared to a threshold T. This parameter is a percentage that determines how much the hierarchical decision should be preferred above the flat decision, with values ranging from 0% to 100%. If all of the equivalent hierarchical individual probabilities are higher than T, then the hierarchical classifier makes the final decision. Otherwise, the flat model makes it.

In the third mixed model, the confidence of the decision relies in the difference between the two highest $softmax$ probabilities outputted by each classifier. The greater the difference, the higher is the likelihood of the decision being correct. In order to find what is the value of the optimal difference, we created a parameter P. Same as mixed 2, it represents a percentage used to figure out to what extent the hierarchical decision should be prioritized over the flat decision. Similar to mixed 2, the decision is made by the hierarchical classifier unless one of its probabilities is below P, in which case the decision passes to the flat classifier. Figure 4 presents the diagrams of the mixed models.

3 Results

This chapter starts by introducing the dataset, followed by the metrics used to evaluate all the results. Then, it presents the experimental results and its discussion of all the methods proposed. In all tests, 5 CNN architectures were considered ResNet50 [13], ResNet101 [13], DenseNet121 [14], EfficientNet-B0 [15] and EfficientNet-B2 [15] and were evaluated using the validation set unless otherwise specified.[1]

3.1 Dataset

This work used the database of ISIC 2019 challenge [7–9] to train and evaluate the proposed models. The training data has 25,331 dermoscopic images and 8 different classes: NV, MEL, DF, VASC, BCC, SCC, AK, BKL. To perform all of the experiments we divided the original training dataset into 2 subsets, the training set with 80% and the validation set with 20%. The best model obtained after training will be tested using the held out test set that is also provided by the challenge organizers. Table 1 presents the total number of images and how they are divided among all skin lesion classes in the reduced training set, the validation and test sets. No labels are provided for the test set, so there is no information about the number of cases in each class.

Table 1. The total number of samples in training, validation and test sets. The number of samples per class in the training and validation set.

Dataset	MEL	NV	BCC	AK	BKL	VASC	DF	SCC	Total
Train	3,617	10,300	2,658	693	2,099	202	191	502	20,262
Validation	905	2,575	665	174	525	51	48	126	5,069
Test	-	-	-	-	-	-	-	-	8,238

3.2 Performance Metrics

In this work we will use two metrics to evaluate the results. Sensibility (SE), also known as recall or True positive (TP) rate, measures the ratio of all the positive samples that were correctly classified as positive for each class.

To give equal importance to all classes, we used Balanced Accuracy ($BACC$) instead of the weighted accuracy. $BACC$ is the average of the SE obtained for each class and it is given by Eq. 2, where N represents the number of classes. In this work, N is set to eight, as there are eight classes.

$$BACC = \frac{\sum_{i=0}^{N-1} SE_i}{N} \tag{2}$$

[1] Source code to reproduce all experiments is available at https://github.com/Bia55//Skin_cancer_AI.

3.3 Impact of Metadata on Hierarchical and Flat Models

We tested the five previously mentioned CNN as well as different combinations among them in the hierarchical classifiers for all the methods. These combinations proved to be better and their results can be seen in table 2 alongside the results of the models that only use metadata or images. Table 3 reports the CNNs used in each configuration.[2]

Table 2. *BACC* scores of all the models best performance. Cells highlighted in green represent the best score of each classifier. The highlights of the individual classifiers, (a)-(e), also represent the classifiers of the Combined 4 model. Note that the flat score is from the method the Combined model belongs to.

Model\Classifier	**(a)**	**(b)**	**(c)**	**(d)**	**(e)**	**Final score**	**Flat**
Only Metadata	74.55	71.85	59.75	59.73	49.93	27.83	35.30
Only Images	90.69	86.95	86.66	91.75	69.30	66.48	73.05
Concatenation	92.21	88.10	89.40	94.64	77.76	71.98	79.05
Multiplication	91.94	87.75	90.54	95.05	80.57	73.99	79.02
Image feature reducer	91.84	88.00	90.12	96.12	78.96	72.92	79.08
Combined 4	92.21	88.10	90.54	96.12	80.57	73.82	-

Table 3. *CNN* configurations of the best hierarchical and flat models shown in table 2. Only configurations of the models that incorporate images are presented.

Model\Classifier	(a)	(b)	(c)	(d)	(e)	Flat
Only Images	EffNetB2				ResNet101	EffNetB2
Concatenation	DenseNet	EffNetB2	ResNet50	EffNetB0		ResNet101
Multiplication				ResNet101		
Image feature reducer			ResNet101	ResNet50		
Combined 4			ResNet50			-

There were 3 CNN that stood through. When dealing with only images, the EfficientNet-B2 distinguished itself from the other networks with the best results in every classifier except classifier (e) and the flat one. However, when we add metadata into the mixture, the two ResNet networks take over as the best networks. These two together lead to 5 out of 7 best performances. While ResNet-101 has the best results in the individual classifiers (c) and (d), ResNet-50 has the best results in the flat, the overall hierarchical classifier and classifier (e). Classifier (a) best score belongs to DenseNet121 and EfficientNet-B2 holds the best score for classifier (b).

The inclusion of the metadata proved to be beneficial regardless of the combination method or the CNN in use. It performed particularly well in classifier

[2] The results for each CNN model can be found in our supplementary material https://github.com/Bia55/Skin_cancer_AI/blob/main/Supplementary_Results.pdf.

(e) and the flat classifier. Despite this, the flat models consistently outperformed their hierarchical model counterpart. While the hierarchical model has a higher accuracy in diagnosing melanocytic lesions (MEL and NV), the flat classifier performs better in non-melanocytic lesions, particularly malignant ones.

Additionally, we created the hierarchical model Combined 4, which besides combining different CNN, it also combines different image-metadata fusion approaches. Despite having a slightly lower *BACC* score than the multiplication method, it misdiagnosed MEL as a benign lesion less frequently so we considered it the best hierarchical model.

3.4 Comparison of the Mixed Models

Using the mixed models, we tested if it was possible to improve the results by combining the flat and hierarchical models (recall Sect. 2.4). We used the ResNet101 flat classifier from the image feature reducer model with $r = 0.8$ and $m = 200$, and Combined 4 model as the hierarchical classifier. Table 4 presents the results.

Table 4. Comparison of the mixed models with the best hierarchical and flat models. Flat transfers represent the number of cases, in percentage, that the hierarchical model passed to the flat model and the last column represents the *BACC* of the flat model in the transferred lesions. The cells highlighted in green represent the best result for each column.

Model	BACC (%)	Flat transfers (%)	Flat BACC (%)
Mixed 1	80.76	59.12	84.58
Mixed 2	80.62	49.60	70.40
Mixed 3	80.57	49.10	71.40
Flat	79.08	-	-
Hierarchical	73.53	-	-

The mixed models performed very well with mixed 1 being the best one. They had a 7% improvement from the hierarchical model and outperformed the flat model by 1% to 2%. The model with more transfers from the hierarchical to the flat model was mixed 1. However, this model has a 15% higher chance of a lesion being correctly diagnosed when sent to the flat classifier. The results of mixed 2 and 3 were nearly identical. Thus, the third mixed method becomes redundant.

We also observed that the mixed 1 model diagnosed malignant lesions as malignant more often than benign. Hence, even if the diagnosis is incorrect, the model determines that it is a detrimental lesion. On the other hand, the false positives of the malignant lesions are not so concerning, as they could be further analyzed by a pathologist and end up being correctly diagnosed by them.

3.5 Elimination of Classifiers (d) and (e)

Throughout the hierarchical model experiments we observed that classifier (c) confuses malignant with benign lesions very often. We decided to investigate what would happen if all the non-melanocytic lesions were to be diagnosed in classifier (c), thereby eliminating classifiers (d) and (e). Three different architectures were tested for the new classifier (c). They were chosen due to each one being the best model for the previous individual classifiers (c), (d) and (e). The final score uses the same classifiers (a) and (b) as the model Combined 4.

The new classifier (c) with the best performance belongs to ResNet50 using the multiplication method with a ReLU activation, which is the same model as the best previous classifier (c). Its final score is extremely similar to Combined 4, less than 1% of difference. This modified hierarchy only altered the performance of the AK class with a 6% improvement and BKL with a decrease of the same order.

We also tested the modified hierarchy with the previous best model, mixed 1. It reached a performance of 80.45%, just 0.3% less than the original hierarchy. Overall, the number of malignant lesions diagnosed as malignant increased and significant changes can be seen in MEL, where this lesion was 4% worse and greatly augmented the chances of being diagnosed as a benign lesion, especially BKL and NV. These results show that the division of malignant and benign non-melanocytic lesions does not have the impact previously thought and may not be needed. Although more testing would be necessary to know for certain, these findings cause one to reconsider the previously defined hierarchy.

3.6 Evaluation on Held-out Test Set

As mentioned previously, ISIC Challenge provides a test dataset with 8,238 images with no ground truth. The evaluation of the best models established previously was performed in the Challenge online platform [2]. Tables 5 and 6 present the results using the test and validation datasets, respectively.

Overall, the models performed in a similar way in both datasets regarding their order from best to worst, that is, the best model is still mixed 1, followed by the flat and then the hierarchical models. However, in the test set they suffered a decrease of roughly 25% in their final *BACC* scores. The individual performance of the lesions decreased sharply in every lesion except NV, MEL and BCC for all the different models.

NV and BCC were the lesions with the best performance in the test set as opposed to the validation set. MEL closely followed them having the third best perfomance. The other lesions suffered very drastic downgrades with performances below 50%. While SCC remained one of the hardest lesions to diagnose, VASC went from being the best in the validation set to one of the worst in the test set.

In the test set, the original hierarchy was better at diagnosing malignant lesions and the modified hierarchy was better at diagnosing benign lesions. While in the validation set there is no clear best mixed 1 model, in the test set it is

Table 5. *SE* scores of each lesion for the best hierarchical and flat image and metadata models using the test dataset. Cells highlighted in green represent the best result in each column.

	MEL	NV	BCC	AK	BKL	DF	VASC	SCC	BACC
Mixed 1	65.60	79.20	76.70	47.10	45.40	48.90	49.50	38.90	56.41
Mixed 1 with new Hier	66.10	79.30	75.60	47.60	43.80	54.40	50.50	40.10	57.18
Flat	59.80	77.10	74.90	46.80	42.20	51.10	47.50	32.50	53.99
Hier	65.60	75.30	68.10	38.00	39.70	36.70	31.70	28.70	47.98
New Hier	62.60	78.20	63.20	38.20	37.20	42.20	39.60	26.80	48.50

Table 6. *SE* scores of each lesion for the best hierarchical and flat image and metadata models using the validation dataset. Cells highlighted in green represent the best result in each column.

	MEL	NV	BCC	AK	BKL	DF	VASC	SCC	BACC
Mixed	78.67	87.77	85.86	70.11	77.52	83.33	92.16	70.63	80.76
Mixed 1 with new Hier	74.25	89.20	85.86	72.99	75.62	83.33	94.12	68.25	80.45
Flat	73.59	84.82	86.17	71.84	74.10	83.33	92.16	66.67	79.08
Hier	77.90	86.83	81.65	59.20	70.10	66.67	82.35	65.87	73.82
New Hier	77.90	86.83	78.95	65.52	65.33	66.67	82.35	64.23	73.53

a different story. Here, the majority of the lesions do better with the modified hierarchy in the mixed 1 model.

4 Conclusion

This work addressed the shortage of the current literature on hierarchical models and metadata. To achieve this, we used two types of models, hierarchical and flat. Each model was tested using only metadata, only image, and image with metadata. Furthermore, we developed three models that combine hierarchical and flat models and tested their performances using images and metadata.

The experimental setup consisted in the comparison of five different CNN architectures in the hierarchical and flat models of only images and image and metadata. Our results show that the inclusion of metadata has a deep positive impact either in flat or hierarchical models as it always improves their performances. However, it is not sufficient on its own. Despite hierarchical models performing slightly worse than flat models, they provide a rationale for the decision.

The first mixed model ended up being the model with the best performance, which makes the models compete directly with each other. This shows that the combination of flat and hierarchical models can increase both the individual performance of the lesions and the overall score of the model.

Additionally, we studied the prospect of reducing the original three-level hierarchy to a two-level hierarchy, eliminating the intermediary decision of non-melanocytic lesions between malignant and benign. The modified hierarchy

showed that the pruning the original tree can be beneficial to some lesions as it ended up performing better in the test set.

We believe that the findings of this work point towards several research directions, in particular: i) using the hierarchical model to see if it helps with the diagnosis of the unknown category/lesion; and ii) explore a different hierarchical structure, where the first level is benign and malignant lesions and the second melanocytic and non-melanocytic.

References

1. Skin cancer statistics from skin cancer foundation. https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/. Accessed November 2020
2. ISIC Challenge. https://challenge.isic-archive.com/. Accessed November 2020
3. Barata, C., Marques, J. S., Emre Celebi, M.: Deep attention model for the hierarchical diagnosis of skin lesions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2019
4. Esteva, A., et al.: Dermatologist level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017)
5. Barata, C., Marques, J. S.: Deep learning for skin cancer diagnosis with hierarchical architectures. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 841–845 (2019)
6. Kulhalli, R., Savadikar, C., Garware, B.: A hierarchical approach to skin lesion classificationn. ser. CoDS-COMAD 2019, New York, NY, USA: Association for Computing Machinery (2019). https://doi.org/10.1145/3297001.3297033
7. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data **5**, 180161 (2018). https://doi.org/10.1038/sdata.2018.161
8. Codella, N.C.F, et al.: Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC) (2017). arXiv:1710.05006
9. Combalia, Marc, et al.: BCN20000: Dermoscopic Lesions in the Wild (2019). arXiv:1908.02288
10. Kingma, D. P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
11. Li, W., Zhuang, J., Wang, R., Zhang, J., Zheng W. S.: Fusing metadata and dermoscopy images for skin disease diagnosis. In: 17th IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1996–2000. (2020). https://doi.org/10.1109/ISBI45749.2020.9098645
12. Pacheco, A.G., Krohling, R.A.: The impact of patient clinical information on automated skin cancer detection. Comput. Biol. Med. **116**, 103545 (2020)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
14. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K. Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
15. Tan, M., Le, Q.: EfficientNet: rethinking model scaling for convolutional neural networks. In: Chaudhuri, K., Salakhutdinov, R. (eds.) Proceedings of the 36th International Conference on Machine Learning, ser. Proceedings of Machine Learning Research. PMLR, vol. 97, 09–15, pp. 6105–6114 (2019)