Classification of skin lesion images

Carolina Bellani[1], Breno Morais[2] & Sofia Jerónimo[3]

[1]M200170098 m20170098@novaims.unl.pt
[2]M20170035 m20170035@novaims.unl.pt
[3]M20170070 m20170070@novaims.unl.pt

Abstract:
Skin cancer is the most prevalent human malignancy and is firstly inspected visually in the initial clinical screening, being followed by dermascopic analysis, histopathological examination, follow-up examination, expert consensus and confirmation by in-vivo microscopy. Being able to identify the lesion from the first stage using dermascopic images can fasten the diagnosis process and consequently follow the appropriate treatment. In addition to that, it does not require a sample of the skin tissue and it is a cheaper diagnosis option. Although it is a challenging task due to the different appearances of the same skin lesion 1 and therefore hard to predict. In this project, we tried different multi-classification models and built several neural network models to classify seven different types of lesions. We think that more techniques should be applied to solve this complex problem; improvements in image preprocessing phase and in the models.

Keywords: Dermatoscopy; Image Processing; Machine Learning

Statement of Contribution: The tasks were equally distributed between the three members of the group and the interest in the topic was strong for all. Particular dedication in the CNN models was of Breno.

I. Introduction

A skin lesion is an abnormal lump, bump, ulcer, sore or colored area on the skin. Most skin lesions are benign though some, such as actinic keratoses and certain moles, can be an indicator of skin cancer. A correct diagnosis of the skin lesions are extremely important to start the correct treatment and prevent skin cancer.

In 1994, Binder and colleagues used dermatoscopic images successfully to train an artificial neural network to differentiate melanomas, the deadliest type of skin cancer, from melanocytic nevi. Although the results were promising, the study, like most earlier studies, suffered from a small sample size and the lack of dermatoscopic images other than melanoma or nevi.

Because of the limitations of available datasets, past research focused on melanocytic lesions (i.e the differentiation between melanoma and nevus) and disregarded non-melanocytic pigmented lesions although they are common in practice. The mismatch between the small diversity of available training data and the variety of real life data resulted in a moderate performance of automated diagnostic systems in the clinical setting despite excellent performance in experimental settings. Building a classifier for multiple diseases is more challenging than binary classification. Currently, reliable multi-class predictions are only available for clinical images of skin diseases but not for dermatoscopic images.

The lesions included in the study are of following types:

- Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec): This type of lesion is invasive, variants of squamous cell carcinoma that can be treated locally without surgery. Actinic keratoses are more common on the face and Bowen's disease is more common on other body sites. The diagnosis of these tumors is mainly based on the assessment of vascular patterns. The architectural arrangement and distribution of the vessels within the lesion and the correlation with the clinical assessment (e.g. texture, firmness) may provide improved specificity. Other associated, but nonspecific features are erythema, scale, erosion or keratin. 1
- Basal cell carcinoma (bcc): Basal cell carcinoma is a common variant of epithelial skin cancer that rarely metastasizes but grows destructively if untreated . The skin's basal cells are the deepest layers of the epidermis (which is the outermost layer of the skin). BCCs often look like open sores, red patches, pink growths, shiny bumps, or scars. 2
- Benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl): Solar lentigines are sharply circumscribed, uniformly pigmented macules that are located predominantly on the sun-exposed areas of the skin. Seborrheic keratoses are benign epithelial lesions that can appear on any part of the body except for the mucous membranes, palms, and soles. Clinically, early seborrheic keratoses are light- to dark brown oval macules with sharply demarcated borders (solar lentigo). As the lesions progress, they transform into plaques with a waxy or stuck-on appearance. The surfaces of these lesions have a warty and keratotic appearance. Often, the lesions have follicular plugs scattered over their surfaces. The size of the lesions varies from a few millimeters to a few centimeters. Lichen planus-like keratosis (LPLK) is a relatively common skin lesion found on actinically damaged skin. Clinically, LPLK usually occurs as a solitary lesion with sharply demarcated borders. It may be macular or slightly raised, and has brown, tan-brown, violaceous, or red-brown color. There is also a pink variant of LPLK that resembles a basal cell carcinoma. 3
- Dermatofibroma (df): The appearance is a round bump that is mostly under the skin, the normal range's size is about the size of the tip of a ballpoint pen to a pea, and it usually remains stable, the color may be pink, red, gray, light brown or purple in varying degrees, and may change over time and the most commonly found on the legs, but sometimes on the arms, trunk, and less common elsewhere on the body. 4
- Melanoma (mel): These tumors originate in the pigment-producing melanocytes in the basal layer of the epidermis. Melanomas often resemble moles; some develop from moles. The majority of melanomas are black or brown, but they can also be skin-colored, pink, red, purple, blue or white. Melanoma is caused mainly by intense, occasional UV exposure (frequently leading to sunburn), especially in those who are genetically predisposed to the disease. 5
- Melanocytic nevi (nv): Melanocytic nevi is a type of melanocytic tumor that contains nevus cells. A mole can be either subdermal (under the skin) or a pigmented growth on the skin, formed mostly of a type of

cell known as a melanocyte. The high concentration of the body's pigmenting agent, melanin, is responsible for their dark color. 6

- Vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc): The most common vascular lesions in childhood are the hemangiomas of infancy and, in adulthood, the cherry angiomas. Hemangiomas and angiomas are benign proliferations of blood vessels. Lacunae (also known as lagoons) are well-demarcated, round to oval red, reddish-brown or reddish-blue areas that commonly vary in size and color within a given lesion. Lacunae may be either tightly clustered or loosely scattered throughout the lesion and they are often located on a background of red, red-blue, or red-white homogeneous color. No vascular structures are seen inside the lacunae. Occasionally, individual dilated blood vessels or a red network of vessels may be visible. Hemangiomas may develop a partial thrombosis, acquiring a focal blue-black color, or a total thrombosis manifesting a jet-black color. Angiokeratomas are dark lacunae: sharply demarcated, ovoid structures with a dark blue, dark violaceous or black color. They have Blue-White veil: confluent blue pigmentation with an overlying whitish haze. Ulceration and surface scale may be present. Rarely, rainbow pattern may be present. Pyogenic granulomas have reddish or red-whitish homogeneous area: a structureless zone whose color varied from completely red to red with whitish zones. They also have white collarette, a ring-shaped or arcuate squamous structure that is usually located at the periphery of the lesions and white rail lines, white streaks that intersect the lesion. They have vascular structures (dotted vessels, hairpin vessels, linear-irregular vessels, telangiectasias, polymorphous vessels). Ulceration may be present. 7

Our project used a open source data set consisting of 10,015 dermastocopic images. Our goal is to train a model and classify every image into seven pigmented lesions. More than 50% of the target values have been confirmed by pathology, while the ground truth for the rest of the cases was either follow-up, expert consensus, or confirmation by in-vivo confocal microscopy.

II. Data

The 10,015 images were extracted from the dataverse project, a Open source research data repository software 8. The images were collected over a period of 20 years from two different sites, the Department of Dermatology at the Medical University of Vienna, Austria, and the skin cancer practice of Cliff Rosendahl in Queensland, Australia. If necessary, an expert dermatologist performed manual histogram correction to adjust images for visual inspection by human readers 9. The seven skin lesion classes covered more than 95% of all pigmented lesions examined in daily clinical practice of the two study sites

a. Description

The Austrian image set consists of lesions of patients referred to a tertiary European referral center specialized for early detection of melanoma in high risk groups. This group of patients often have a high number of nevi and a personal or family history of melanoma. The Australian image set includes lesions from patients of a primary care facility in a high skin cancer incidence area. Australian patients are typified by severe chronic sun damage 9.

Chronic sun damaged skin is characterized by multiple solar lentigines and ecstatic vessels, which are often present in the periphery of the target lesion. Very rarely also small angiomas and seborrheic keratoses may collide with the target lesion. We did not remove this "noise" and we also did not remove terminal hairs because it reflects the situation in clinical practice. In most cases, albeit not always, the target lesion is in the center of the image 9.

Studying the age of the patient, 0.69% is missing and it is possible to impute it with the median by lesion's type; 9 bkl types imputed with the median age of 65yo, 1 mel type with 65yo and 42 nv with 45 years old. The following plot shows that the age could be use an input in the model.

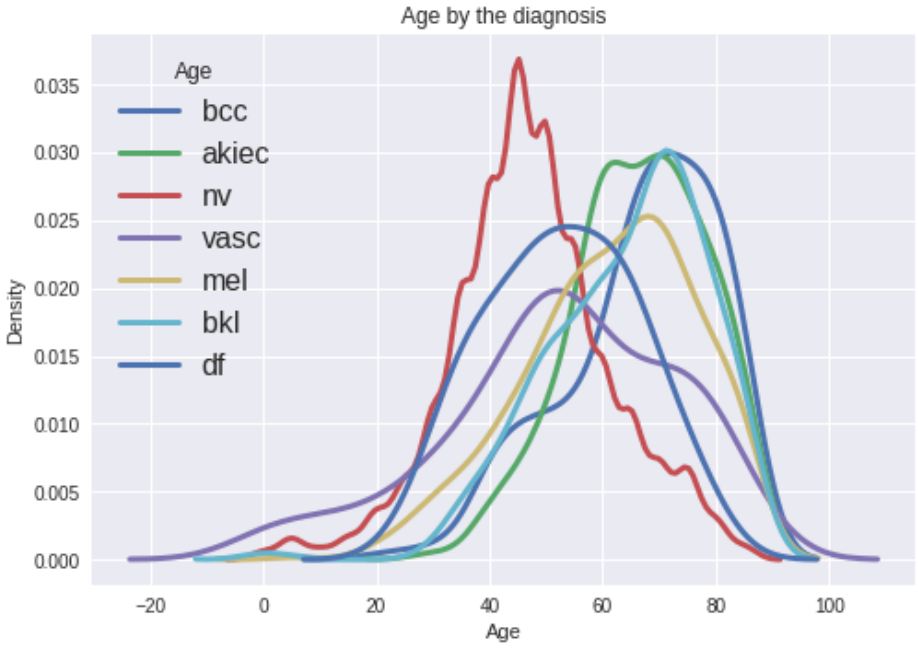

Image 1. Age by lesion's type

Considering the patient's sex, 0.67% is unknown and it will be imputed considering the mode by class; 9 of the class bkl and 41 of the class nv be imputed as male following the mode by lesion's type.

dx	akiec	bcc	bkl	df	mel	nv	vasc	All
sex								
female	80	122	304	38	243	2584	48	3419
male	148	205	414	35	371	2778	50	4001
unknown	0	0	9	0	0	41	0	50
All	228	327	727	73	614	5403	98	7470

Table 1. Cross table between age and lesion's type.

The incidence of the lesions occur equally in both sexes and the pattern seems very similar.

The 15 possible locations are: abdomen, acral, back, chest, ear, face, foot, genital, hand, lower extremity, neck, scalp, trunk and upper extremity. Most of the lesions occur in the back, trunk and lower extremity.

In conclusion, age, localization and sex can be used as features in the model, but we are expecting that they will just help with a minimal contribution.

b. Extraction

There we could extract the two zip folders containing the images and the metadata file with the following features: lesion_id, image_id, dx (the type of skin lesion), dx_type (the method used to classify the type), age, sex and localization of the lesion. The name of each image corresponds to the image_id present in the metadata file and it contains unique values. Each image had a size of 600X450 pixels in 3 different colour channels: red, blue and green.

To automatize the extraction of the images from the zip file, we unzipped them with for loop and store each image name.

c. Transformation

-Image vectorization

There are many techniques that allow to transform the image information to a vector-based information. We needed to consider the information of the individual pixel and the position of the image. In order to reduce the size of the data set and improve the speed of the model training, we decreased the number of pixels of each image to 1/3 of the original size. The final size of each image was 200X150.

To decrease the dimensionality of the image, we transform the rgb image space (10015, 200, 150, 3) into greyscale space (10015, 150, 200, 1) maintaining the pixel position in the image using the following formula:

$$gray=0.2989\times r+0.5870\times g+0.1140\times bgray=0.2989\times r+0.5870\times g+0.1140\times b$$

where r is red, g is green and b is blue channel. After, we reshaped the image matrix (200X150) into a vector of 30,000 elements and we concatenate all images into a single matrix of dimensions 10,015X30,000.

-*Merge the image matrix with the metadata*

We merged the image matrix with the metadata by the image_id feature, this would had personal data that can improve the performance of the model. After this step, we checked the type of the personal data and converted them in the category type. Because of the collinearity problem, when the dummy variables are used, we don't select the first dummy.

-*Removal of the duplicate*

s In the provided dataset, there are 7,470 lesions over the 10,015 images, then 25.41% of the lesions are from the same patient. The same lesion can be present in the dataset up to six times. In order to guarantee the ability of generalization, we only maintained the unique lesions and dropped randomly the duplicates. The image matrix was then reduced to 7,470 rows.

-*Data Splitting*

Considering the distribution of the target and in order to maintain the proportions of the target values in the training and test sets, we randomly selected the rows in a stratified way (by target's classes) with the following proportions:

-Training Set (70%): the models will be trained using this set.

-Test Set (30%): the models will be tested their performance in this set and the results will be compared with the true target values considering various measures. Even if accuracy apparently can be good, through the cross table between the predicted and the actual classes, we calculated some more important measures and we compared them for the different models and different sets.

-*Data Cross Validation*

For some models, we tried three different data partitions and then different test sets.

III. Results and Discussion

Random Forests

-Reduction of dimensionality

The presence of many features in the model can deteriorate the results of our models due to the presence of redundant and irrelevant features.

Data reduction techniques can be applied to obtain a reduced representation of the data set that is much smaller in volume, yet closely maintains the integrity of the original data.

The goal of PCA is to summarize the information contained in the quantitative data with a new set of variables called principal components. These new variables are uncorrelated (no redundancy) and are ordered so that the first of them account for most the variation in all the original variables. Before it is good practice to standardize the features. After applying the optimized orthogonal transformation, we considered to keep the components with which it is possible to explain 95% of the total variance.

-First study case of Random Forests

Considering the first 162 variables and the patient's data, the random forests model with 1,000 decision trees has an accuracy of 74.7%. Accuracy metric is not the correct one to consider in this case. In fact, looking to the cross table between predicted classes and actual classes, it is possible to notice that only 3 classes, on 5 possible ones, are predicted from the algorithm. Moreover, the precision of, for example, mel is only 27.27%.

Predicted	bkl	mel	nv
Actual			
nv	2	2	536
mel	3	3	55
bcc	8	0	25
bkl	19	5	49
vasc	0	0	10
akiec	11	1	11
df	0	0	7

Table 2. Cross table of the predicted and actual classes in the unbalanced dataset.

We thought that the main problem was the unbalanced classes distribution.

-Unbalanced classes

It is possible to see the unbalancing of the classes.

Lesion's Type	Percentage	Count
akiec	3.05%	228
bcc	4.38%	327
bkl	9.73%	727
df	0.98%	73
mel	8.22%	614
nv	72.33%	5403
vasc	1.31%	98

Table 3. The seven lesion types and respective percentage and absolute frequency.

There are many techniques that offer a solution to this problem: undersampling and oversampling.
A priori, it is not possible to know which one can perform better for the problem. For simplicity, we decided to perform the following undersampling technique: considering one third of the majority class, nv, and not considering df and vasc classes.

Lesion's Type	Percentage	Count
akiec	6.17%	228
bcc	8.85%	327
bkl	19.66%	727
mel	16.61%	614
nv	48.72%	1801

Table 3. The seven lesion types and respective percentage and absolute frequency after undersampling.

The models was training and tested in this, more or less, balanced sample. We think this doesn't reflect the reality; the test should be at least with the similar priori distribution. We also think that losing so much information is not good; it could be possible to use different training sets including always the minority classes and changing the majority class. The final decision could be based on the ensemble of different training sets models.

Oversampling technique probably is more appropriate for this kind of problem and without loss of information. It could be possible to create new sample of the minority classes considering a noise of the image or also, reconsidering some delete duplications of the original dataset.

-Second study case of Random Forests

We performed a new one principal components analysis in the balanced dataset and the components needed to represent the 95% of the total variance in this case, were 175.

It is possible to check how the image changes considering how the image changes considering only the first 175 components instead of the 30 000 dimensions.

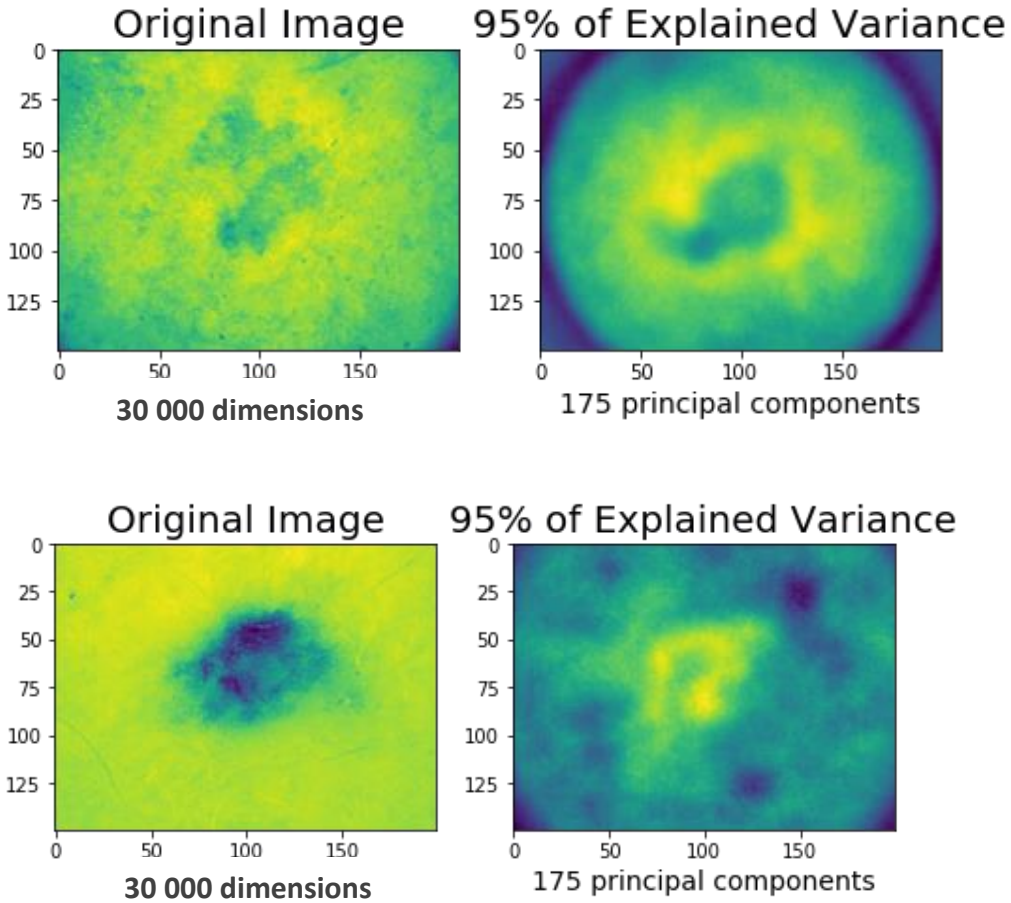

Image 2. How the image changes considering the reduced features.

Applying also in this case a random forest model with 1000 decision trees, the accuracy is of 94.32% and the results in the test set are the following:

Predicted	akiec	bcc	bkl	mel	nv
Actual					
bcc	0	20	0	13	0
bkl	0	0	73	0	0
nv	1	0	0	0	179
mel	0	0	0	61	0
akiec	16	0	0	0	7

Table 4. Cross table of the predicted values and the actual values in the balanced dataset.

It seems balancing the classes is better, as we were hoping. But we need to have in mind that, unfortunately, also the test was balanced and then the results could be different. Moreover, it is not possible to classify two classes.

Neural Networks

Neural Network is another possible used algorithm for multi-classification problems.

We used it to have a benchmark for the Convolutional Neural Network and we run it just in the unbalanced dataset.

Convolutional Neural Networks

Convolutional neural networks are often used in images' classification because it is possible to give as input a n-dimensional space. Inspired by biological processes as neural network, the model simulate the visual cortex of an animal. Metaphorically explaining, it is possible to see the algorithm as the eye plus the neurons in a person. The eye receives the information and it passes through different stimuli to various neurons. For this reason, the image could be passed without detailed pre-processing.

We tried to consider 11 layers with different activation functions, different output's shape and internal parameters. Using Adam optimizer and categorical cross entropy loss function. We set the annealer learning rate. In additional, we performed data augmentation to prevent overfitting using "ImageDataGenerator" function.

-First study case CNN

With unbalanced classes the accuracy was exactly the percentage of the majority class, nv. Then we tried to give weights to the classes but the results were worst than expected (6% of accuracy).

-Second study case CNN

Using the balanced dataset, the results are not better as in the random forest. The accuracy in the test set is of 49.1% and the cross table of the predicted and actual classes shows only one class predicted, the most frequent.

IV. Conclusions

Because of the complexity of the problems, many other techniques could be explored. We think that a better image pre-processing could make the prediction improves: considering, for example, the spatial correlation, the reduction of the space by pixel positions, some filters of the image's noises and many other treatments. Moreover, it would be possible to explore and test better different undersampling/oversampling techniques and maybe, also ensemble different models with different datasets to have a more stable classification in the validation. Additionally, many parameters could be tested and tuned to better personalize the model but always considering a good trade-off between the goodness of the model and the ability of generalization.

V. References

1. Dimitriou F., Deinlein T., Zalaudek I., 2018, "Actinic keratosis / Bowen's disease / keratoacanthoma / squamous cell carcinoma", https://cs.stanford.edu/people/esteva/nature/
2. Skin Cancer Foundation, 2018, "Basal Cell Carcinoma (BCC)", https://www.skincancer.org/skin-cancer-information/basal-cell-carcinoma
3. Ralph P. Braun, 2018, "Solar lentigines / seborrheic keratoses / lichen planus-like keratosis", https://dermoscopedia.org/Solar_lentigines_/_seborrheic_keratoses_/_lichen_planus-like_keratosis_(full_text)
4. Johnson J., Marcin J., 2017, "Dermatofibroma: Causes, images, and treatment", https://www.medicalnewstoday.com/articles/318870.php
5. Skin Cancer Foundation, 2018, "What Is Melanoma?", https://www.skincancer.org/skin-cancer-information/melanoma
6. Wikipedia, 2018, "Melanocytic nevus", https://en.wikipedia.org/wiki/Melanocytic_nevus
7. Zaballos P., Martín I. G., 2018, "Vascular lesions", https://dermoscopedia.org/Vascular_lesions
8. Tschandl, Philipp, 2018, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions", https://doi.org/10.7910/DVN/DBW86T, Harvard Dataverse, V1,

9. Jiang Y., Kim J. B., Sill C. J., Kerns B. K., Kline J. D., Cunninghan P. G., 2018, "Inter-comparison of multiple statistically downscaled climate datasets for the Pacific Northwest, USA", https://www.nature.com/articles/sdata201816