UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

FELIPE CÉSAR LOPES MACHADO
Advisor: Prof. Rodrigo César Pedrosa Silva

APPLICATION OF DEEP LEARNING METHODS TO DETECT GLOBAL PATTERNS IN DERMOSCOPIC IMAGES AND AID THE SKIN CANCER DIAGNOSIS

Ouro Preto, MG
2021

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

FELIPE CÉSAR LOPES MACHADO

APPLICATION OF DEEP LEARNING METHODS TO DETECT GLOBAL PATTERNS IN DERMOSCOPIC IMAGES AND AID THE SKIN CANCER DIAGNOSIS

Undergraduate thesis presented to the Computer Science course at the Federal University of Ouro Preto as partt of the mandatory requirements for obtaining the Bachelor's degree in Computer Science.

Advisor: Prof. Rodrigo César Pedrosa Silva

Ouro Preto, MG
2021

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

M149a Machado, Felipe Cesar Lopes .
Application of Deep Learning methods to detect global patterns in dermoscopic images and aid the skin cancer diagnosis. [manuscrito] / Felipe Cesar Lopes Machado. - 2021.
30 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Rodrigo César Pedrosa Silva.
Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Biológicas. Graduação em Ciência da Computação .

1. Inteligência artificial- Aplicações médicas. 2. Aprendizado profundo. 3. Imagens digitais- Dermatologia. I. Silva, Rodrigo César Pedrosa. II. Universidade Federal de Ouro Preto. III. Título.

CDU 004.8

Bibliotecário(a) Responsável: Soraya Fernanda Ferreira e Souza - SIAPE: 1.763.787

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIENCIAS EXATAS E BIOLOGICAS
DEPARTAMENTO DE COMPUTACAO

FOLHA DE APROVAÇÃO

Felipe César Lopes Machado

Application of Deep Learning methods to detect global patterns in dermoscopic images and aid the skin cancer diagnosis

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal
de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 24 de Agosto de 2021.

Membros da banca

Rodrigo César Pedrosa Silva (Orientador) - Doutor - Universidade Federal de Ouro Preto
Pedro Henrique Lopes Silva (Examinador) - Mestre - Universidade Federal de Ouro Preto
Marcos Antônio Alves (Examinador) - Mestre - Universidade Federal de Minas Gerais

Rodrigo César Pedrosa Silva, Orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 24/08/2021.

Documento assinado eletronicamente por **Rodrigo Cesar Pedrosa Silva**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 24/08/2021, às 15:33, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 , informando o código verificador **0210282** e o código CRC **0E348AD5**.

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35400-000
Telefone: 3135591692 - www.ufop.br

Abstract

Skin cancer is one of the main causes of low life expectancy in the world. To diagnose skin cancer, a specialist looks at different patterns which may appear in a skin lesion. Thus, identifying these patterns is an important step for early diagnosis and prognosis. Previous work in this topic revolves around the use of classical image processing techniques to extract features from the images and building a feature vector for classification. This type of work is laborious and involves a careful selection of the feature extraction techniques to be successful. Convolutional Neural Networks (CNNs), on the other hand, which are a part of Deep Learning, have the ability to learn such feature extractors from the data. Under this perspective, the main goal of this work is to study the use these techniques for the identification of global patterns in dermoscopic images of skin lesions. Although CNNs have already been used in other problems related to skin lesion classification, its suitability for the global pattern classification problem has not been thoroughly assessed yet. Our experiments indicate that the lack of large datasets and their unbalanced nature cause problems for the use of CNNs. Nevertheless, by applying techniques of data augmentation and transfer learning, these problems can be mitigated, and CNNs can become a useful tool. Among the compared CNN architectures, the best performing one was the SqueezeNet. Since it is a compact architecture, it may suffer less from the lack of a large data set. Its weighted average accuracy, sensitivity and specificity was 84.6%, 61.1% and 91.2%, respectively. Top-down hierarchical models were later used but they did not improve the overall sensitivity.

Keywords: CNN, Skin Lesion, Global Patterns, Image Classification, Ensemble of Classifiers.

Resumo

O câncer de pele é uma das principais causas da baixa expectativa de vida no mundo. Para seu diagnóstico, um médico especialista examina os diferentes padrões que podem aparecer em uma lesão de pele. Assim, identificar esses padrões é uma etapa importante para o diagnóstico e prognóstico precoce. Os trabalhos anteriores neste tópico giram em torno do uso de técnicas clássicas de processamento de imagem para extrair recursos das imagens e construir um vetor de características para classificação. Esse tipo de abordagem é trabalhosa e envolve uma seleção cuidadosa das técnicas de extração de características para ter sucesso. As redes neurais convolucionais (CNNs), por outro lado, que fazem parte do *Deep Learning*, têm a capacidade de aprender esses extratores de características a partir dos dados. Sob essa perspectiva, o objetivo principal deste trabalho é estudar a utilização dessas técnicas para a identificação de padrões globais em imagens dermatoscópicas de lesões de pele. Embora os CNNs já tenham sido usados em outros problemas relacionados à classificação de lesões de pele, sua adequação para o problema de classificação de padrão global ainda não foi completamente avaliada. Nossos experimentos indicam que a falta de grandes conjuntos de dados e sendo eles bem desbalanceados dificultam o uso de CNNs. No entanto, aplicando técnicas de aumento de dados e aprendizagem por transferência, esses problemas podem ser mitigados e as CNNs podem se tornar uma ferramenta útil. Entre as arquiteturas CNN comparadas, a de melhor desempenho foi a *SqueezeNet*. Por ser uma arquitetura compacta, ela pode sofrer menos com a falta de um grande conjunto de dados. Sua média ponderada de acurácia, sensibilidade e especificidade foi de $84,6\%$, $61,1\%$ e $91,2\%$, respectivamente. Modelos hierárquicos foram testados posteriormente, mas não obtiveram melhora na sensibilidade geral.

Keywords: CNN, Lesão de pele, Padrões globais, Classificação de imagem, Ensemble de classificadores.

List of Figures

List of Tables

Contents

1 Introduction

Skin diseases can affect 30% up to 70% of the population and it gets higher in rural areas of developing countries Hay e Fuller (2011). They are the 4^{th} biggest cause of low life expectancy, according to the DALY metric (Disability-Adjusted Life Year) Hay et al. (2014).

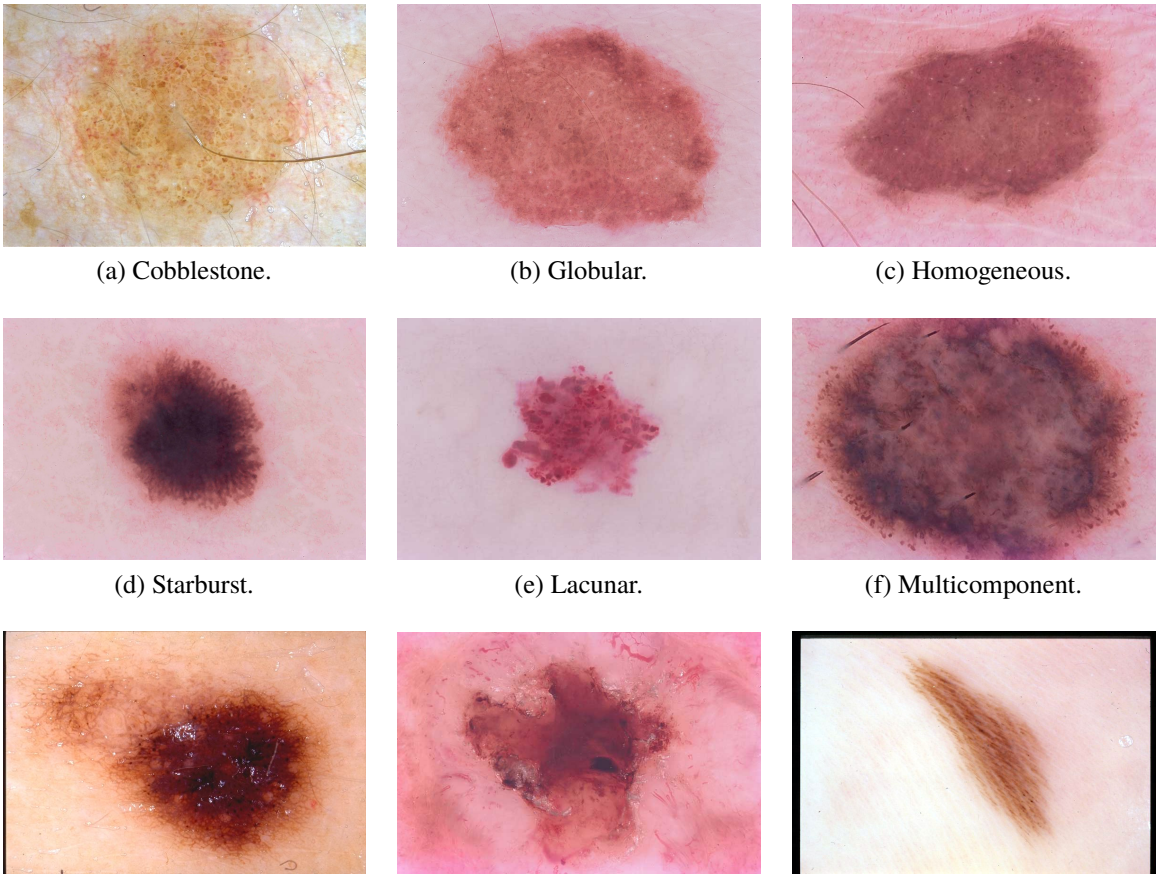
Skin cancer is one of the most diagnosed cancers worldwide Skin... () and melanoma is its most aggressive form, as well as the one with the highest mortality rate. The American Cancer Society American... (2020) estimates that, in 2020, there will be more than $100,000$ new cases of melanoma skin cancer in the United States. These cancers can appear anywhere on the skin surface, mostly due to excessive exposure to ultraviolet rays. The chances of cure increase the earlier the disease is detected in the patient.

Dermoscopy is a non-invasive examination of the skin that provides a detailed view of its layers. Analysis of lesion patterns in dermoscopic images has proven to be a very efficient method for early diagnosis (CARLI et al., 2003; CELEBI et al., 2007; CELEBI; MENDONCA; MARQUES, 2015).

Several clinical diagnostic methods have been developed over the years, among them, the ABCD rule is very commonly used Ali, Li e Yang (2020). This method is based on observing the asymmetry (A), border (B), color (C) and differential structures (D) to differentiate malignant from benign lesions. In addition, recognizing the textured structures observed in dermatoscopic images (see Figure 1.1), known as global skin lesion patterns, is an important step to differential diagnosis and can also be useful to define the courses of treatment (SÁEZ; ACHA; SERRANO, 2014; SERRANO; ACHA, 2009).

The automatic analysis of global patterns on a skin lesion allows for quick screening and helps the assessment of a specialized healthcare professional (CELEBI et al., 2007; CELEBI; MENDONCA; MARQUES, 2015). These patterns can be characterized in the following nine categories (SÁEZ; ACHA; SERRANO, 2014; SERRANO; ACHA, 2009): reticular (patchy network), globular (aggregated globules), cobblestone (similar to aggregated cobblestones), homogeneous (diffuse pigmentation), starbursts (characterized by radial structures), parallel (along furrows), multicomponent (combination of three or more patterns), lacunar (larger globules separated with bigger gaps) and unspecific (used to refer to those who do not belong to any of the previous patterns). Examples of some global patterns can be seen in Figure 1.1.

Argenziano et al. (2003) found that the diagnosis of skin lesions, aided by the information of global patterns, allowed improved diagnosis performance. Under this perspective, Deep Neural Networks, in particular, Convolutional Deep Neural Networks, have been achieving superhuman level in many tasks of computer vision LeCun, Bengio e Hinton (2015). Thus, they become a natural candidate for the task of global pattern recognition.



(a) Cobblestone. (b) Globular. (c) Homogeneous.

(d) Starburst. (e) Lacunar. (f) Multicomponent.

(g) Reticular. (h) Unspecific. (i) Parallel.

Figure 1.1 – Sample of dermoscopic images for each global pattern - Argenziano G, Soyer HP, De Giorgi V, Piccolo D, Carli P, Delfino M, et al.Dermoscopy: A Tutorial. EDRA Medical Publishing NewMedia, 2002.

At this point, we do not know of any work that addresses the classification of all the nine skin patterns depicted in Figure 1.1 combined, using CNNs. In this context, this work aims to study the most popular CNN architectures and their suitability in the classification of global patterns in dermoscopic images.

1.1 Justification

Several studies have been made targeting the classification of skin lesions in the past years. To our knowledge, none of them has considered all the nine global skin patterns (this term is defined in the theoretical background section). The accurate classification of skin patterns impacts the accuracy of the patient's diagnostic by a medical doctor. By improving skin lesion classifiers we can aid several patients and dermatologists to identify early malignant lesions and potentially save their lives.

1.2 Objectives

The main objective of this work is to study and develop methods of deep learning for the identification of skin lesion global patterns in dermoscopic images.

The main objective can be divided into the following specific objectives:

- Review of the state-of-the-art methods proposed for the target problem;
- Evaluate deep learning methods for the target problem;
- Find an artificial neural network architecture which is able to efficiently learn and classify the most important skin lesion patterns;
- Experiment with different types of ensemble learning in order to improve the baseline models performance.

2 Literature Review

As mentioned in the introduction, the identification of global patterns in skin lesions is an important yet challenging task. Classical methods of image processing and pattern recognition have been applied to the identification of global patterns problem in Tanaka et al. (2008), Iyatomi et al. (2008), Serrano e Acha (2009), Mendoza, Serrano e Acha (2009), Situ, Yuan e Zouridakis (2011), Isasi, Zapirain e Zorrilla (2011), Abbas, Celebi e Fondón (2012), Abbas et al. (2013). In these references, techniques such as principal component analysis Iyatomi et al. (2008), image segmentation Tanaka et al. (2008), Mendoza, Serrano e Acha (2009), Isasi, Zapirain e Zorrilla (2011), Situ, Yuan e Zouridakis (2011), statistical analysis of the image color space Serrano e Acha (2009), Abbas, Celebi e Fondón (2012), Abbas et al. (2013), and color intensity Tanaka et al. (2008) are used to extract a feature vector for each image. Then, a classifier is built to identify each pattern in the defined feature space.

Despite the fact that these methods perform reasonably well (reported accuracy of over 90% for most of the cases), there is no consensus about the right feature space. Besides, the construction of the feature spaces is laborious and requires the application of several different techniques which are, in some cases, handcrafted, Situ, Yuan e Zouridakis (2011). Under this perspective, the use of Deep Learning (DL) techniques seems to be a reasonable choice since one of its main strengths is the ability of automating the feature extraction process. Another useful attribute of DL is its ability of reusing feature extractors learned from other applications through transfer learning. Below, we review applications of DL to other classification problems related to dermoscopic lesions.

Demyanov et al. (2016) created a CNN with three convolutional layers among pooling layers, followed by two dense layers (128 neurons ending in 2) and ReLU functions. The aim of the paper was to classify dermoscopic images, provided by ISIC, containing 211 images, manually labeled by professionals, of local skin lesion patterns set as "typical network" or "regular globes". They extracted random parts of the images, labeled as sub-images, and applied scale and rotation as data augmentation in each one of them, yielding almost 30,000 images. These images made four sets: positive and negative sets for each class. A positive set of a class means the images belong to the class, and the negative set are the group of images that does not belong to it. For feature extraction, three algorithms were used: K means clustering, sparse coding and Fischer kernel-feature encoding. In addition, 20 experiments for each algorithm was done, varying the seeds to generate the parameters. The best results had an average accuracy of 88% and 83% for typical networks and regular globes, respectively.

The following papers used pre-trained CNNs to aid the feature vector generated at the end of the convolution phase of the network. Using CNNs trained on unrelated image classes

proved to be efficient to skin lesion classification.

Lopez et al. (2017) compared 3 methods for detecting melanoma in dermoscopic images, using images from ISIC 2016 and VGG16 CNN architecture. In the first method, they used a VGG16 convolutional network, training it from scratch. In the second method, they used the same VGG16 CNN architecture, but pre-trained with the images from ImageNet, a database of tens of millions of images separated into $1,000$ classes. In the third method, they applied fine-tuning to method two. In total, $1,279$ images of benign and malignant skin lesions were used, of which 900 for training and 379 for testing. The third method proved to be more efficient than the other two, yielding an overall average 78.66% sensitivity and 79.74% precision and 81.33% accuracy. Although the authors achieved a good result, they only considered 2 types of lesion patterns and used a set of few images. There are currently more types of skin lesion patterns that need to be included in automatic classifications. Their work made clear the importance of using pre-trained CNNs and opened the possibility to use this method on more classes, other than just use it on a melanoma classifier. In our work, a more recent set of lesion skin patterns were used, containing 9 classes.

Using a pre-trained AlexNet with images from ImageNet, Hosny, Kassem e Foaud (2019) also aimed at the problem of classification of skin lesions. The image set was taken from MEDNODE, DermIs and ISIC 2018 databases. MEDNODE has 70 melanoma and 100 nevus images. DermIs has 119 melanoma images and 87 for nevus, and ISIC provided 374, 254 and $1,372$ for melanoma, seborrheic keratosis and nevus, respectively. The exception of the latter with three classes, the images were divided into two classes (melanoma and nevus). In the pre-processing phase, the image set was enlarged using a data augmentation technique: for each image, 71 copies were made, each one rotated 5° of each other (0°, 5°, 10°, ... 355°), ending with 72 images. In addition, for the same image, 72 random rotations were applied, resulting in 143 images rotated from the same image, plus the original image. Using 10-fold cross-validation, the reported results were 96.86%, 97.70%, and 95.91% of average accuracy for MED-NODE, DermIs–DermQuest and ISIC, respectively. The authors made use of a lot of data augmentations over a significant amount of images, in comparison to the previous work. This gave them excellent results in conjunction to the pre-trained network.

Mahbod et al. (2019) had AlexNet, VGG16 and ResNet-18 architectures for skin lesion feature extraction on images provided by ISIC 2016 and ISIC 2017 for the training phase. These features were used to train non-linear support vector machines for later skin leasion classification. The dataset have 2037 dermoscopic images in high resolution (1022×767 to 6748×4499). Their classifier output one of the three considered classes: melanoma (411), seborrheic keratosis (254) and nevi (1372). After classifying with the different network architectures, a fusion of the results is made averaging the class scores. The final result reported an average of 90.69% of AUC. The authors introduced a hybrid approach, also combining the CNNs architectures for feature extraction. Like the previous works, it still lacks the inclusion of many other skin lesion patterns.

On the HAM10000 dataset, Chaturvedi et al. (2019) used a set of $10,015$ dermoscopy images, including melanocytic nevi, melanoma, benign keratosis, basal cell carcinoma, actinic keratosis, vascular lesion and dermatofibroma addressing the classification of skin lesions. They applied data augmentation to the unbalanced classes using the Keras ImageDataGenerator and resulted in a total of $38,569$ images. Then, after data augmentation, a MobileNet, pre-trained on the 2014 ImageNet Challenge dataset, was used for training the resulted images. They reported an weighted average, among all classes, of precision, recall and f1-score of 89%, 83%, and 83%, respectively. The authors focused their classification on the final diagnose of the skin lesion, which differs from our work that focuses on identifying patterns on the lesions to facilitate future diagnoses.

Regarding relevant CNN architectures for skin cancer classification between melanoma and nevi, Munir et al. (2019) did a bibliographic review and concluded, based on values of accuracy, sensitivity and specificity, that papers that used AlexNet, VGG16, VGG19, Res-Net, SVM and KNN, SkinNet, U-net CNN and DenseNet-201 yielded the most significant results compared to other CNN architectures on different datasets.

In the next sections, we present a theoretical background of everything used in this work, and then proceed to our methodology and results.

3 Theoretical Background

3.1 Dermatoscopy

In dermatology, dermatoscopy is a skin analysis aided by a dermatoscope. It is used to diagnose several skin conditions, such as skin tumors, fungal infections, scalp diseases, skin cancer, and many other issues. The dermatoscope allows the experts to have higher accuracy on the skin condition diagnosis. Further definition is presented on the subsections below.

3.1.1 Dermoscopic images

Dermoscopic images are images of pigmented sections of the skin taken with the aid of a dermatoscope. A dermatoscope is a microscope with a light source, which can enlarge the image by up to 20 times so that it is possible to observe deeper layers of the skin and the presence of multiple lesions or spots (nevi). Digital dermatoscopy, on the other hand, manages to enlarge the image by up to 70 times and allows its storage for the patient's follow-up.

Dermoscopy is used in the diagnosis of skin lesions and is important for the early identification of skin cancers.

3.1.2 Dermoscopic images classification

The diagnosis of skin lesions, as mentioned before, is aided by its dermoscopic images. The classification of these lesions occurs according to the goal of the analysis to be made. In the case of cancer detection, for example, dermoscopic images can be split into benign and malignant lesions, among them basal cell carcinoma (low lethality) Basal… (2021), squamous cell carcinoma (most common skin cancer) and melanoma (highest index of mortality). The classification considered by this work, however, follows the pre-diagnosis objective, that is, in order to assist health professionals in the conclusion of the aforementioned types.

3.2 Deep Learning and CNNs

Deep Neural Networks are computer simulated networks loosely based on the functioning of the human brain, with multiple layers of interconnected neurons, and are part of machine learning methods. These networks can be very effective when you have a large amount of data for training. In the field of computer vision, CNNs are the most commonly used. These networks usually consist of multiple convolution blocks followed by pooling layers, which are responsible for image feature extraction and the decrease of image size. These layers are followed by a dense neural network, or fully connected layers, capable of classifying the images into the expected

groups. A simple schematic diagram is shown in Figure 3.1: an image of a skin lesion is given as an input for a further binary classification between benign and malignant lesion. The aim of the convolution layers is to detect common features in the images that are used to identify their classes LeCun, Bengio e Hinton (2015). Given the input image and a filter, the network produces one convoluted image. These filters or kernels are small two dimensional data that sweep the inputed image, multiplying their pixel values, producing an altered picture which can have certain important features, for later classification, highlighted. Pooling layers are able to decrease the quantity of parameters of the network, reducing the size of the images. The most common pooling layers are called max-pooling, where each previously determined squared area of pixel values is reduced to the highest value within that area. Fully connected networks are typically disjoint sets of layers of neurons connected to the previous and next set of neurons. The last set of neurons in the dense network is responsible to classify the input image based on the most activated output neuron.

It is known that the use of deep neural networks, in particular the CNNs, can help in the classification of different groups of images. Before the use of CNNs became popular, the researchers relied on manual feature extraction of images to input in their various types of classification algorithms. The CNNs changed this scenario by presenting an automatic way of extracting features from images, using convolution.

In the context of this work, it can be used on the classification of the aforementioned global patterns, as long as we have a large amount of dermoscopic images with their global patterns previously identified.

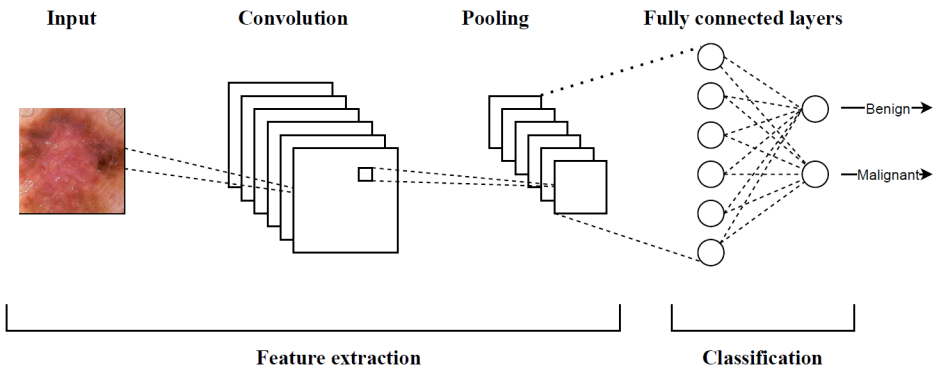


Figure 3.1 – Schematic diagram of a basic convolutional neural network performing a binary classification between benign and malignant skin lesions.

3.2.1 The class imbalance problem

Unfortunately, especially for dermoscopic images with the ground truth shown in Figure 1.1, large balanced sets are scarce. In our case, we have $1,039$ images for 9 classes, which are, comparing to the literature on image classification, very few for a good and generalized outcome

on a neural network to be trained from scratch. Moreover, an important issue is the unbalanced amount of images for each class, defined by the global skin patterns. Often, the number of images on each class differs a lot, which can affect the generalization performance negatively since it tends to be more biased towards the classes with more images Buda, Maki e Mazurowski (2018). Therefore, techniques to train our networks and balance the data set classes are needed.

3.2.2 Transfer Learning

One way to deal with a small number of images, as it is our case, is to use transfer learning. Transfer learning is a technique that uses a pre-trained neural network to benefit from the features it has already learned. The use of previously trained CNN models may prove to be more efficient on classifying images than training a CNN from scratch Shin et al. (2016). Therefore, in the case of neural networks focused on the skin lesion classification, especially those that lack the necessary large amount of images for training, it is essential to use this technique to obtain satisfactory results Menegola et al. (2017). Transfer learning consists of fine-tuning a pre-trained network based on the input data. Fine-tuning consists in unfreezing the weights update of some convolutional layers, during the training phase, in order to the CNN have a better generalization of the classes when extracting their features. In this work, dermoscopic images of skin lesions will be used to feed CNNs pre-trained with the ImageNet dataset.

3.2.3 Data augmentation

As said before, the dataset available to this work is very unbalanced. To help with this issue, data augmentation techniques are used in order to balance the amount of data among the classes and increase their number. These are techniques used to increase the number of images to improve the performance and generalization capacity of the neural network. These techniques are required and effective for small data sets, as well as to improve the results of already large datasets, as in the case of our dermoscopic images Perez et al. (2018). In the literature, there are several attempts to create new methods of data augmentation, such as transferring style Mikołajczyk e Grochowski (2018) and neural augmentation Perez e Wang (2017). The more traditional techniques, as image rotation, zoom and horizontal flip are commonly used along with other methods.

3.2.4 CNN Architectures

The main goal of this work is to assess the suitability of DNNs for the skin lesion global patterns classification problem. To achieve this goal we compare several different, commonly used, CNN architectures, using transfer learning - all of them pre-trained on the ImageNet dataset Deng et al. (2009). The architectures used in this work are listed below.

3.2.4.1 AlexNet

The AlexNet architecture Krizhevsky, Sutskever e Hinton (2012) is a relatively shallow DNN, shown in Figure 3.2. It contains 4 convolutional and 3 max-pooling layers followed by 3 fully connected layers. This architecture is composed of 60 million parameters and was used in the 2010 and 2012 ImageNet's competition, achieving the best result among their competitors. The top-5 test error rate was 17.0% and 15.3%, respectively, against 28.2% and 26.2% as the second place entries. The authors listed ReLU funtion, overlapping max-pooling layers and dropout as its important and distinguishable features. The inclusion of ReLU activation function, which led to its later popularization, played a major role on the training time of large datasets. The new dropout concept allowed their network to be trained with more iterations and less overfitting.

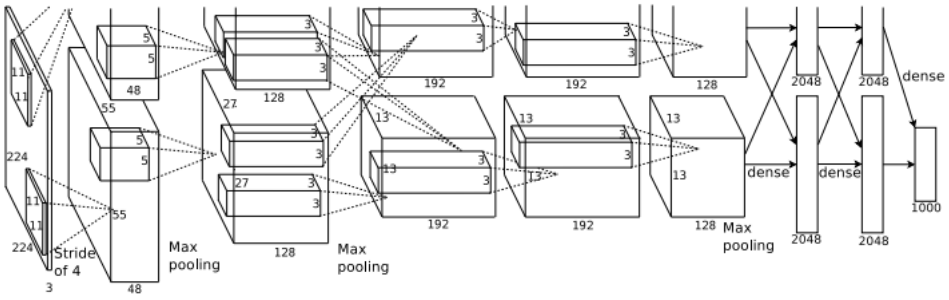


Figure 3.2 – Alexnet architecture scheme. The communication among the convolution layers and their kernal maps occur on two different paths until they merge only at certain layers. Source: Krizhevsky, Sutskever e Hinton (2012)

3.2.4.2 DenseNet

The DensetNet was proposed in 2017 Huang et al. (2017) and was used on CIFAR-10, CIFAR-100, SVHN and ImageNet datasets, achieving significant improvements over the state-of-the-art of that time. In the latter, DenseNet-264 showed performance similar to the ResNet-152 with the half of the number of parameters and execution time. This architecture was built on the idea of the loss of image or gradient information through the layers of very deep neural networks. To address this issue, the authors connected all convolutional layers to each other, in a feed-forward way - every layer obtains the feature maps of all preceding layers. This allows the reuse of previous features and also the use of fewer parameters (20 million). The convolutional blocks containing this implementation are called dense blocks which are depicted in Figure 3.3.

3.2.4.3 VGGs

The VGG16 Simonyan e Zisserman (2014) is a model presented in the 2014 ImageNet competition, which showed a significant improvement of the AlexNet architecture in the previous competitions. The VGG16 achieved a 92.7% top-5 accuracy when testing on the competition dataset. What differentiated the VGG models from other architectures was the use of multiple

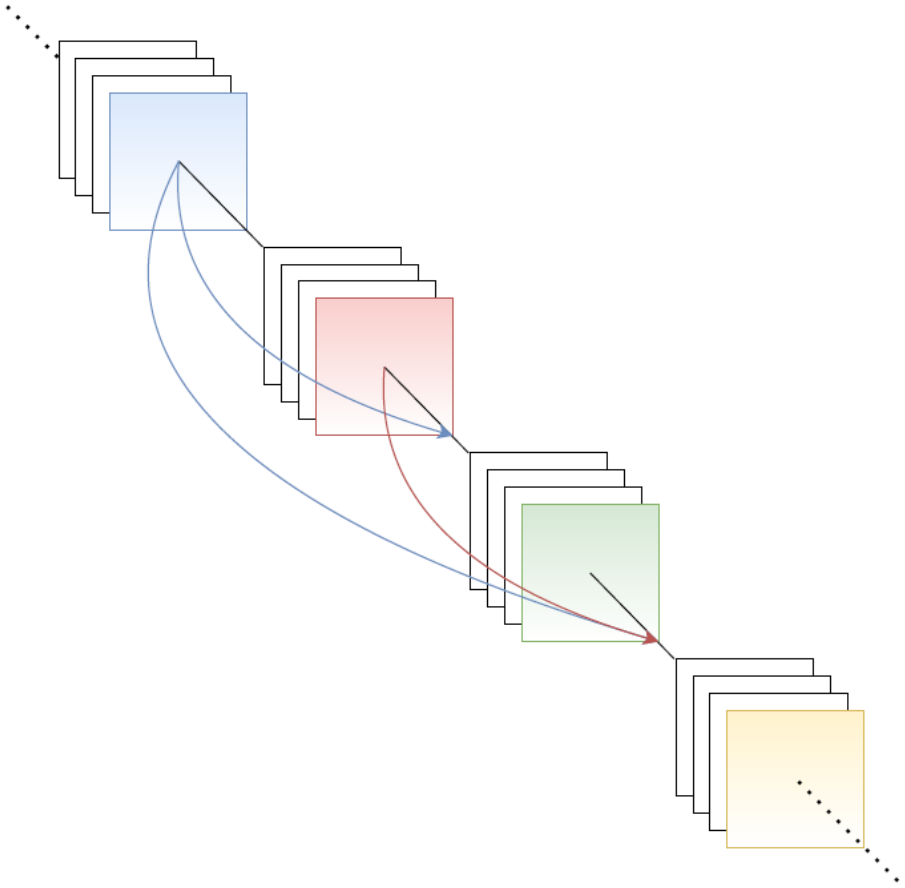


Figure 3.3 – DenseNet convolutional layers building blocks (dense blocks) are based on the idea of connecting the convolutional layers to every preceding others, preserving information of the feature maps generated on the first layers. Adapted from Huang et al. (2017).

3×3 kernels along a very deep convolutional network. Filters this size are the smallest ones which can take into account the adjacent pixel values to alter the pixels in the image, therefore they can make more subtle changes on each layer and have a lower computational cost. The AlexNet architecture, for example, have a filter size of 11×11. The number of the VGG model tells us the sum of convolutional and fully connected layers, all the other parameters are the same (VGG19 is 19 layers deep). By increasing the depth of the network, Simonyan and Zisserman showed that the depth of a CNN was beneficial to the classification test accuracy.

3.2.4.4 GoogLeNet and Inception-v4

GoogLeNet Szegedy et al. (2015), also known as Inception-v1, was presented in the 2014 ImageNet competition, achieving the first place with slightly better results than the VGG networks, and significantly better than AlexNet. The Inception-v4 architecture is a Inception variant, evolved from GoogLeNet. When used in conjunction with InceptionResNets Szegedy et al. (2016), achieved a state-of-the art performance in the 2015 ImageNet competition dataset. Both architectures have in common the Inception Module. This module emphasizes the idea that

a CNN does not need to be a simple linear progression of inputs of convolutional layers. Thus, each Inception block computes the image with different kernel sizes for later concatenation. It is composed of three parallel 1×1 filter size kernels and one 3×3 max-pooling, followed by 3×3 and 5×5 kernels operating on two of them, and 1×1 kernel on the max-pooled output. At the end, all the outputs are concatenated. This was based on the idea that smaller kernels can provide a more precise classification of two similar but different type of objects, for example two dogs of different breed and the type of a boat, and their use with conjunction of larger kernels contributes to a better generalization. The v4 CNN is deeper, using more Inception modules along with batch normalization layers and factorization, which were introduced in the previous Inception versions (v2 and v3). The overall inception module scheme is shown in Figure 3.4.

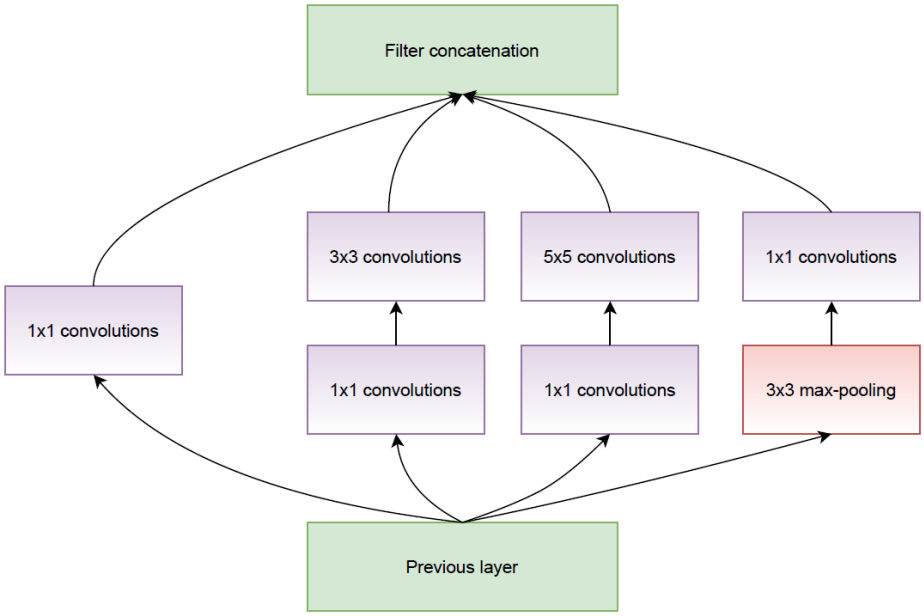


Figure 3.4 – Inception module scheme. Adapted from Szegedy et al. (2016).

3.2.4.5 ResNets

The ResNet He et al. (2016), also called Residual Network, was presented in the 2015 ImageNet competition and achieved a 3.57% top-5 error on the competition's dataset. An important difference of this CNN is the big leap on the number of layers. The previous model with the greatest number of layers was the VGG with 19. The models used in this work are 50 and 101 layers deep. The idea behind the residual network is to make short connections between every other layer. Since the training error tends to increase when the network is too deep, these short connections help mitigating this problem. The ResNet architectures proved that the classification accuracy can still be significantly improved when using residual networks. A residual network scheme is shown in Figure 3.5.

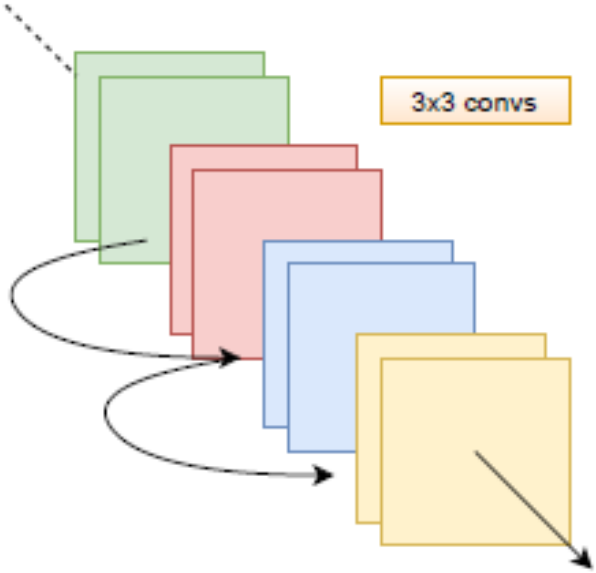


Figure 3.5 – Schematic of a standard residual network. Skipping every other convolutional layer, a shortcut connection is inserted. Adapted from He et al. (2016).

3.2.4.6 SqueezeNet

The SqueezeNet architecture Iandola et al. (2016), released in 2016, was presented as a better alternative to AlexNet. It was based on the idea of constructing a network with fewer parameters while presenting a good classification accuracy. To make this happen, one strategy was to replace the 3×3 filters with 1×1. This strategy is contained in what the authors called the "fire module". The idea of these modules is to squeeze the data in 1×1 convolution filters and feed them into expanded layer with 1×1 and 3×3 filters, decreasing the number of input channels. An example of this filters is shown in Figure 3.6. To compensate for a possible loss in the classification accuracy, iandola2016squeezenet downsample the data late in the network. This way, the image information is kept for longer, extracting better features. One important detail to be noted is the lack of a fully connected layer on top of the fire modules. The authors showed that, as the paper's title suggest, SqueezeNet reaches the AlexNet's levels of accuracy while having $50\times$ less parameters in the same ImageNet database.

3.2.4.7 ShuffleNet

The ShuffleNet architecture Zhang et al. (2018) aims at running under limited computing power, while presenting a competitive accuracy on image classification. The idea behind this network is to shuffle the input between two convolution groups, instead of linear channels. Convolution groups were also used on the AlexNet architecture, where two groups of linear convolution layers, computed in different GPUs, only meet at certain point of the network. In the case of ShuffleNet, the input channels between some layers are randomly shuffled, and the

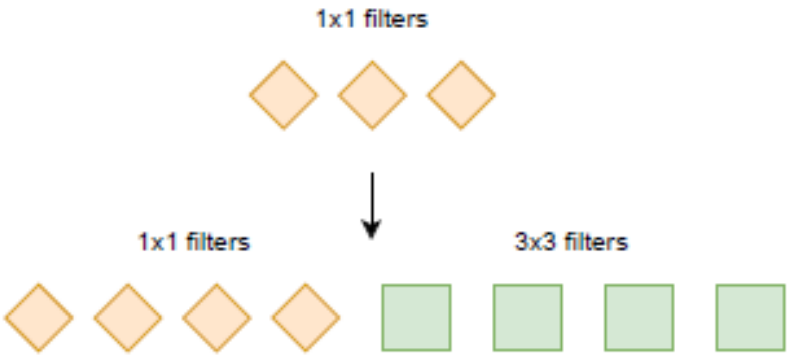


Figure 3.6 – Example of the convolutional filters on the SqueezeNet, which decreases the number of input channels after the expansion on the bottom half set of filters. The top half represents the squeezing phase. Adapted from Iandola et al. (2016).

use of group convolutions allows for a reduction of computational cost of the pointwise (1×1) operations and better performance. This method also allows the exchange of information across convolutional groups, which led to good results. The model was evaluated on the ImageNet 2012 competition dataset achieving a 7.8% top-1 error.

3.2.4.8 MobileNet

The MobileNet Howard et al. (2017) was also developed on the premise of reducing the number of parameters to allow its use on mobile and embedded applications. The key method proposed by the authors was to use depthwise convolutions. Similarly to pointwise convolutions, which performs convolutions on single pixels, depthwise convolutions apply a single filter on the input channel. Later, the pointwise convolution is performed in a linear combination. To decrease the computational cost, a hyperparameter was introduced to make the network thinner. This parameter allows the CNN to reduce the number of input and output channels in each layer. The authors also added a resolution multiplier, which is applied to every internal representation of the inputted image in order to reduce the parameters when running on limited computers. The MobileNet has proved to be especially accurate in object detection and face attributes. The CNN showed better results than GoogleNet and VGG16 on the ImageNet dataset.

3.2.4.9 NasNet

The NASNet architecture Zoph et al. (2018) was developed on the idea of a mutable network. This means that some architectural configurations of the network elements are automatically determined by a reinforcement learning algorithm (RLA). The network consists of normal cells and reduction cells. Normal cells generate a feature map with the same size, while in the reduction cells the feature maps are divided by two since it is applied a stride of 2. This model

was tested on the CIFAR-10 dataset, using part of it as the input to the RLA. This approach was based on the fact that creating a new CNN architecture is time and resource demanding, since this architecture is not made by humans. On the ImageNet dataset, NASNet achieved state-of-the-art results with significant lower computational costs, exceeding the human designed models.

4 Methodology

The methodology used in this work is presented below. In section 4.1, hardware used in the experiments is presented. The section 4.2 details the dataset used in this work. In section 4.3, the methods used to pre-process our data and tackle the class imbalance problem are presented. Finally, in sections 4.4 and 4.5 the CNNs parameters and the training procedures are presented, respectively.

4.1 Hardware

The experiments done in this work used an NVIDIA GeForce GTX 1070 graphics card and the PyTorch machine learning library.

4.2 Dataset

The dataset used in this work is the EDRA-CDROM (2002) G Soyer HP (2002) and it is comprised of $1,039$ dermoscopy images of the 9 previously labeled global skin pattern classes, having 768×512 pixels each, in RGB color space. Their labels and amount of images are the following: 30 Cobblestone, 82 Globular, 78 Homogeneous, 15 Lacunar, 332 Multicomponent, 47 Parallel, 344 Reticular, 46 Starburst and 64 Unspecific. One of the images in our dataset does not have a defined global pattern and, therefore, was ignored. The images in the lacunar pattern were also not used in this work due to their disproportionate low number, which is half of the second-lowest (cobblestone pattern). Using them would result in too much data augmentation over the same image when balancing the classes, which could contribute to underfitting. Thus, it was used $1,023$ images of the original dataset, covering 8 out of the 9 classes. Samples of each class are shown in Figure 1.1.

4.3 Pre-processing

As mentioned in the previous section, our dataset is very imbalanced. To address this issue, a pre-processing step is needed to balance the classes before the training phase. In addition, we used data augmentation techniques to further increase the number of samples available for the models.

Our dataset needs to be divided into three sets: training, validation and test. The test dataset was gathered by taking 30% of each global pattern class, randomly, resulting in 307 images (716 for training and validation); no data augmentation technique was done. To balance the dataset, before the training and validation phases, the following data augmentation techniques were used:

zoom, rotation, and horizontal flip. We will divide the pre-processing phase into two steps. The first step is the creation of the set of new images, achieved by data augmentation techniques. The second step is the addition of the original images followed by more data augmentation in order to balance the modified dataset. The first step is shown in Figure 4.1 and explained below:

- 450×450 cuts

  On the first step, all the 716 images were cut on the 450×450 square at their center, in order to remove the non-relevant information, mostly skin that is not part of the lesion. These cuts were also done in Abbas et. al. (2013) Abbas et al. (2013) on dermoscopic images of skin lesions. The authors named them regions of interest (ROI).

- Zoomed cuts

  On the previous altered dataset, it was applied 4 zoomed cuts of 224×224 to every image on each of their quadrants, shifted by 57 pixels from the closest extremities to the center. This was made as long as the amount of data in its class does not surpass the one with the highest image count (reticular pattern, 240 images). As an example, the cobblestone pattern had 105 images after this process (21 initial$+21 \times 4$) and the reticular pattern images was not altered. The image size choice was done considering the fact that most of our pre-trained CNNs were trained using this image resolution. The slight shift of the cuts to the center was due to the fact that our region of interest is located near the center of every image. At the end of this process, some classes have the maximum quantity of images (240) except for cobblestone, parallel, starburst and unspecific patterns.

- $5°$ rotations

  To correct the class imbalances, $5°$ rotation is applied randomly on the class dataset, and no more than 2 rotations per image, until it reaches 240. At the end of this, we had $1,920$ images in total.

On the second step, all the original images, except for the test dataset, were added to the altered dataset from the first step. To balance the dataset again, the horizontal flips (mirror the image on the vertical axis) and vertical flips ($180°$ rotation) were used simultaneously and randomly on this new set, until each class reaches 480 images. After the described process, our dataset for training and validation had $3,840$ images. From this, 20% of each class was taken randomly for the validation dataset, finally leaving $3,072$ images for training, 768 for validation and 307 for testing.

4.4 CNN parameters

On every CNN architecture, the classifier block, the top dense layers ending in $1,000$ neurons, was removed. The added classifier has two dense layers: 500 neurons connected to 8

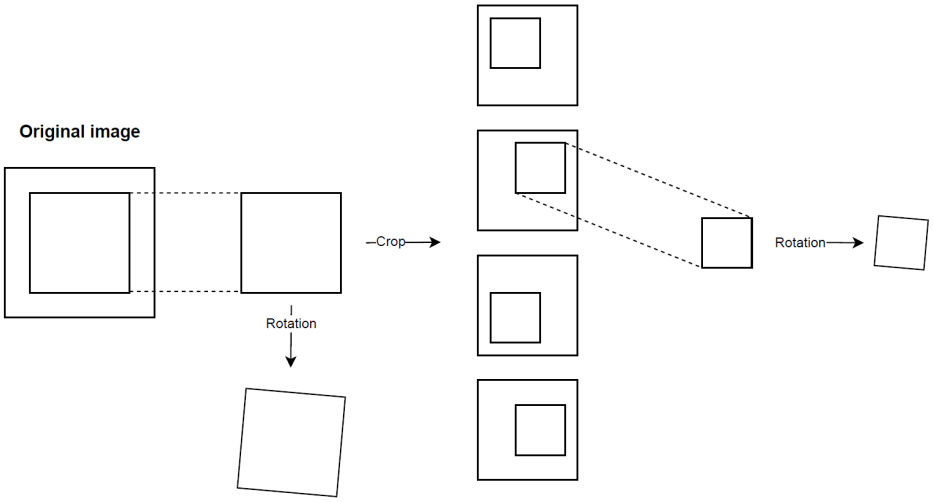


Figure 4.1 – The first step of the pre-processing phase consists on cropping the original images in their inner quadrants and applying 5 degree rotations.

(equal to the number of classes). The only exception was the SqueezeNet that is built differently, since it has a convolutional layer on its classifier block. Thus, a fresh new convolutional layer with the same parameters was added, except for the 8 output classes. The image input size is 224×224 for all architectures except for the Inception-V4 (299×299). Resizing was done using bilinear image scaling.

4.5 Training

The training phase is the same for every architecture. It begins with the base (the convolutional layers) frozen, i.e. the weights are not updated during the training. This is done because, since the training loss is high at the beginning, some features learned from the ImageNet on the base layers can be altered dramatically or lost. For 20 epochs, batch size of 32 and using Adam optimizer Kingma e Ba (2014), the network only updates the weights of the added classifier block (fully-connected layers) and batch normalization layers. Then, for fine-tuning, the last two layers of the base are unfrozen with $1/3$ and $1/9$ of the fully-connected layers learning rate, the last one with the highest value.

The initial learning rate is calculated using a similar approach to the Cyclical Learning Rates Smith (2017). Before the training phase, over an epoch, the network starts with a learning rate of 10^{-8} and grows exponentially at each batch of 32 images. The algorithm halts when the loss value calculated for the current batch surpasses the best one in 4 times. In other words, it breaks out if the loss value explodes. In addition, a data vector is built with information on the learning rate used for the batch and the loss value calculated. A graph of learning rate (in logarithmic scale) $\times$ loss value is built. Then, ignoring the first 10 and the last 5 points of the vector, the learning rate is chosen where the graph shows the steepest decline. The learning rate graph of the WideResNet50 is shown in Figure 4.2.

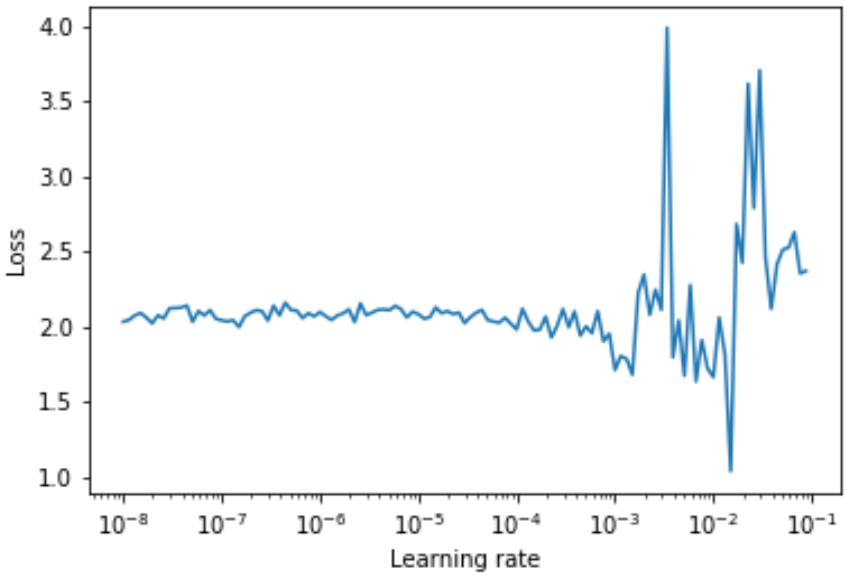


Figure 4.2 – Learning rate $\times$ Loss value over an epoch of the WideResNet-50. The greatest decline was calculated at learning rate of approximately 0.0034.

For evaluating the results, we compared the accuracy, sensitivity and specificity values among the architectures, and the average of these metrics for each class among the CNNs. We call true positives (TP) those images that were correctly identified in their class; false positives (FP) are the images that, among all images classified as a given class, were incorrectly identified as belonging to that class; false negatives (FN) are the images that, among all images that belong to a given class, were incorrectly identified; and true negatives (TN) are the images that, considering one class, were correctly classified as not belonging to that class. That said, the metrics used are as follows:

- $accuracy = (TN + TP)/(TP + FP + TN + FN)$;
- $sensitivity = TP/(TP + FN)$;
- $specificity = TN/(TN + FP)$.

Finally, an ensemble of all the architectures was made, with increased weight on the architecture with the best result, and the weighted average results for each class is shown in Table 5.5. Our ensemble is a mix of all architectures trained, using a majority vote on their classification of each image in the test dataset.

The next goal, we will call it approach number 2, was to use the CNN which yielded the best results to build a hierarchical model for image classification. Thus, for each class on our pre-processed dataset, we built a binary classifier splitting our target class and all the others not yet analyzed. The target class and the other class, containing a set of the remaining classes, were balanced between them and among the classes of the latter. Therefore, as shown in 4.3, in

the first step, as we separate the parallel lesion from the others, both the classes had the same quantity of images in the training phase (480). The class named "other" is balanced considering the remaining classes. The order in which the classes were separated was determined subjectively after looking at the confusion matrix of the SqueezeNet (Table 5.2): we chose to start by the easier to classify.

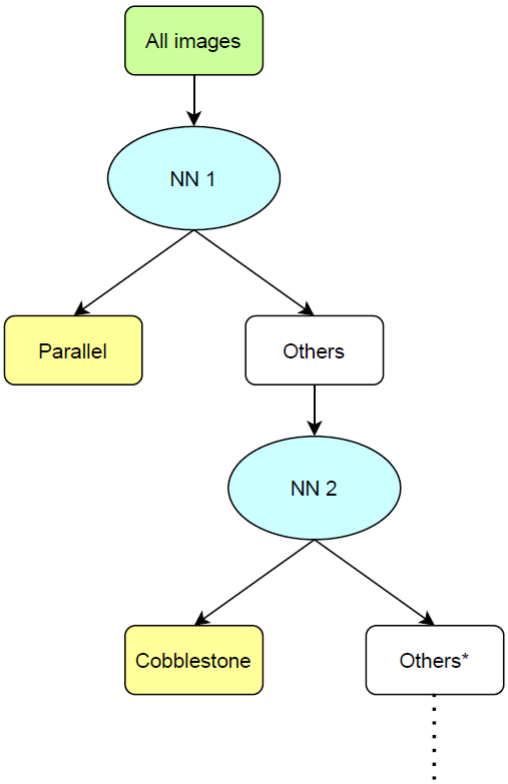


Figure 4.3 – Hierarchical model containing seven neural networks to classify eight global skin patterns. (*) Contains the other classes, excluding the already identified.

Finally, in approach number 3, although the classes don't seem to have a natural hierarchical division, we divided the classes based on the visual similarities of their images, thus not considering the confusion matrices, as shown in Figure 4.4. On each of its neural networks, all the classes were balanced accordingly.

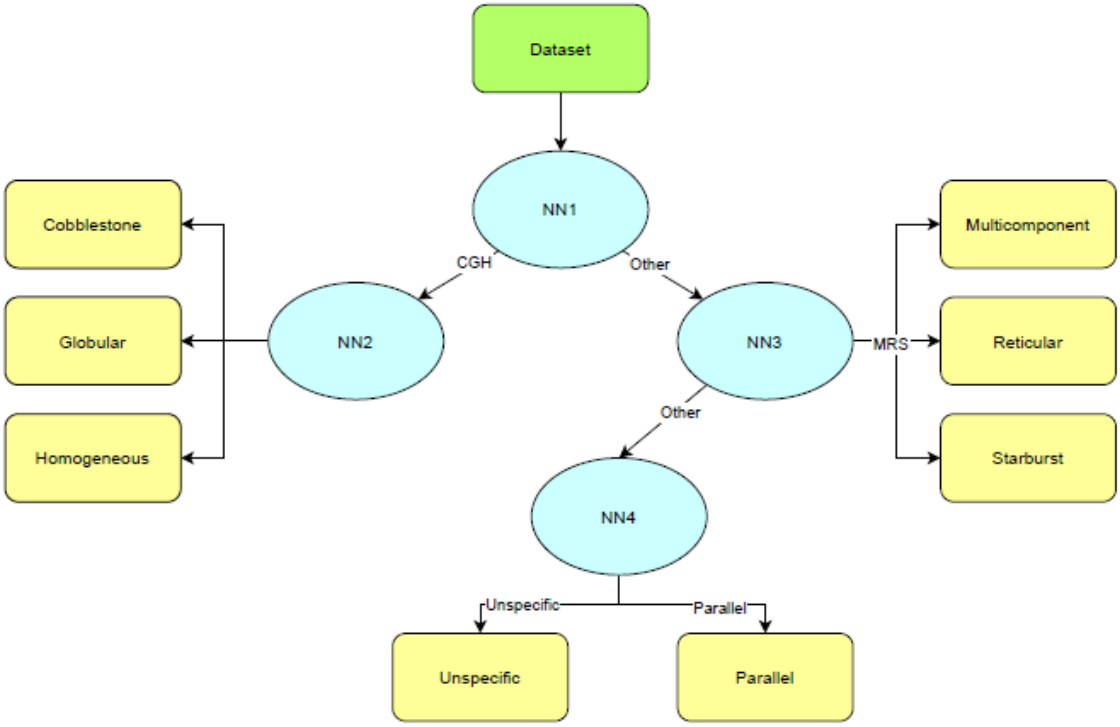


Figure 4.4 – Hierarchical model of approach number 3. It contains four neural networks dividing hand selected groups of classes.

5 Results and Discussion

Table 5.1 shows the weighted average accuracy, sensitivity and specificity, since the dataset is not balanced, using the test dataset of 307 images, for every trained architecture. Each row represents the weighted average of all the eight global patterns. Table 5.3 highlights the result of the most accurate architecture, SqueezeNet. Table 5.4 is the condensed individual global pattern results, averaging the metrics of all architectures. In Table 5.5, we show the result of the ensemble architecture; SqueezeNet with weight 3.

Table 5.1 – Weighted average results for all the global patterns considered, using the test dataset.

Architecture	Accuracy	Sensitivity	Specificity
AlexNet	0.819	0.563	0.870
DenseNet-201	0.816	0.482	0.942
DenseNet-169	0.819	0.511	0.931
GoogLeNet	0.789	0.396	0.934
Inception-V4	0.767	0.396	0.902
MobileNet-V2	0.782	0.399	0.928
NASNet Mobile	0.798	0.450	0.927
ResNet-101	0.786	0.421	0.930
ResNet-50	0.810	0.479	0.941
ShuffleNet	0.801	0.463	0.924
SqueezeNet	0.846	0.611	0.912
VGG16	0.814	0.486	0.938
VGG19	0.799	0.470	0.915
Wide ResNet-50	0.776	0.386	0.926

Table 5.2 – Confusion matrix for the test dataset on the SqueezeNet CNN.

Global Pattern	Cob.	Glob.	Hom.	Mul.	Par.	Ret.	Star.	Uns.
Cobblestone	6	1	1	1	0	0	0	0
Globular	1	11	1	3	3	2	2	2
Homogeneous	0	2	13	7	0	0	2	0
Multicomponent	5	9	5	61	0	8	8	4
Parallel	0	0	0	0	15	0	0	0
Reticular	0	11	2	17	1	67	3	3
Starburst	0	0	3	1	0	1	9	0
Unspecific	0	2	3	2	1	4	0	8

The results in Table 5.1 show that shallower architectures with fewer parameters were a better fit for our dataset: SqueezeNet, AlexNet and DenseNets yielded the best overall results. The difference between shallow and deep NN's results over smaller datasets is evident both in the average results for all global patterns and the global patterns individually (Tables 5.1

Table 5.3 – Classes results for the SqueezeNet using the test dataset.

Global Pattern	Accuracy	Sensitivity	Specificity
Cobblestone	0.971	0.667	0.980
Globular	0.875	0.440	0.913
Homogeneous	0.916	0.542	0.948
Multicomponent	0.775	0.610	0.853
Parallel	0.984	1.000	0.983
Reticular	0.833	0.644	0.928
Starburst	0.936	0.643	0.950
Unspecific	0.933	0.400	0.969

Table 5.4 – Average results for each global pattern using the test dataset.

Global Pattern	Accuracy	Sensitivity	Specificity
Cobblestone	0.936	0.603	0.946
Globular	0.841	0.431	0.877
Homogeneous	0.899	0.554	0.927
Multicomponent	0.730	0.351	0.910
Parallel	0.946	0.862	0.950
Reticular	0.773	0.440	0.941
Starburst	0.908	0.760	0.915
Unspecific	0.897	0.536	0.922

Table 5.5 – Results for each global pattern on ensemble architecture.

Global Pattern	Accuracy	Sensitivity	Specificity
Cobblestone	0.971	0.667	0.980
Globular	0.868	0.600	0.892
Homogeneous	0.913	0.625	0.937
Multicomponent	0.765	0.480	0.900
Parallel	0.958	0.933	0.959
Reticular	0.823	0.567	0.952
Starburst	0.932	0.714	0.943
Unspecific	0.926	0.650	0.945
Weighted average	0.837	0.579	0.930

and 5.4, respectively). The most variance of the results is more evident when comparing the sensitivities. Our dataset did not seem to benefit from wider implementations of networks nor parallel operations (Wide ResNet, GoogLeNet, Inception). CNNs that addresses the vanishing-gradient problem, which occurs when information of the input is lost after several layers of computation, yielded unsatisfactory results (ResNets), even the shallower ones (DenseNet). Even though MobileNet being a small CNN, we believe that the presence of too few parameters, also provided by the use of depthwise separable convolutions, contributed to a not proper learning, considering a small dataset. SqueezeNet is highlighted as it presents the best predictions.

Given the difference in the number of samples among the classes, data augmentation was

Table 5.6 – Results for each global pattern on the hierarchical model architecture.

Global Pattern	Accuracy	Sensitivity	Specificity
Cobblestone	0.942	0.556	0.954
Globular	0.849	0.360	0.892
Homogeneous	0.936	0.583	0.965
Multicomponent	0.785	0.420	0.957
Parallel	0.958	0.867	0.963
Reticular	0.775	0.625	0.850
Starburst	0.913	0.786	0.919
Unspecific	0.910	0.350	0.948
Weighted average	0.825	0.534	0.915

heavily applied on the minority classes (i.e. Cobblestone, Starburst, Parallel and Unspecific). The results for each global pattern in Tables 5.3, 5.4 and 5.5, show outstanding values compared to the majority classes. Keep in mind that the test dataset was untouched and was not used at any moment during the training phase. This test on a completely independent dataset indicates that the augmentation was successful, providing a satisfactory generalization of the network. In the future, more data augmentation can be done on the majority classes to improve the results. One idea could be to apply augmentation to them, but extract multiple subsets with its initial class size, randomly. Then, make multiple training with a same architecture with different subsets of the majority classes and ensemble the results.

The SqueezeNet results are very similar to the ensemble architecture. However, sensitivity among the minority classes show a significant improvement on the ensemble architecture compared to SqueezeNet, whilst better in all metrics compared to the overall results. The improvements are highlighted in Table 5.5. This means that the network is more accurate within a global skin pattern. Table 5.2 shows the confusion matrix for the test dataset on the SqueezeNet CNN. We can see that homogeneous and reticular patterns were often mistaken for multicomponent (highlighted in red), and the CNN struggled to identify globular and unspecific patterns in general.

Regarding the approach number 2, although the networks yielded interesting results, the overall sensitivity actually decreased significantly, as shown in 5.6. The approach number 3 did not perform well from the beginning, which resulted in worse accuracy and sensitivity compared to the SqueezeNet and similar sensitivity: 0.829, 0.569 and 0.903, respectively.

6 Final Considerations

6.1 Conclusion

This work presents a literature review and compares some well known convolutional neural networks architectures for the problem of identifying global patterns in dermoscopic images of skin lesions. Since the classes are imbalanced, the lacunar global pattern was dismissed due to lack of images and data augmentation was applied to the other minority classes. By using the pre-trained models of the CNNs, the classifier layer was replaced in every one of them by a standard dense layer. All their initial parameters were the same except for the learning rate, which was calculated using the same method. Later, an ensemble architecture was developed using the CNN with the best individual results with increased weight, using majority vote predictions. The SqueezeNet architecture yielded the best results with our dataset, achieving and average of 84.6% accuracy, 61.1% sensitivity and 91.2% specificity. The ensemble architecture improved the sensitivity of the minority classes compared to the average results of the CNNs and the individual SqueezeNet results.

The two subsequent approaches were based on the hierarchical top-down model. The first one separates each class from the pool of all classes individually, reducing the number of classes at each step. Each of those steps have their own CNN which were trained considering the remaining classes. The second approach separates groups of classes, instead of one by one. These groups were chosen by the visual similarities of their images on the training dataset. Both of these approaches yielded a worse sensitivity, 0.534 and 0.569, respectively, and similar accuracy and specificity.

To the best of our knowledge, the experiments that were closer to our work regarding the number of global skin patterns were done by Abbas et al. (2013), which considered 7 classes (Reticular, Globular, Cobblestone, Homogeneous, Parallel, Starburst and Multicomponent). Our methods differ, since the authors used AdaBoost.MC algorithm, instead of CNN. They reached 89.28% sensitivity, 93.75% specificity and 0.986 area under the curve (AUC). Our overall sensitivity was significantly lower.

6.2 Future work

For future work, it is imperative that we seek ways to improve the sensitivity values of our results, which is the quantity of correct classifications on a given class over the total images that belongs to that class. One idea, as said in the previous chapter, is to increase data augmentation and train different CNNs on disjoint sets of the augmented dataset. Then, an ensemble of these

CNNs would be built. Another idea for future works is to insert a one-shot or few-shot learning networks as alternatives for image classification.

Bibliography

ABBAS, Q.; CELEBI, M. E.; FONDÓN, I. Computer-aided pattern classification system for dermoscopy images. *Skin Research and Technology*, Wiley Online Library, v. 18, n. 3, p. 278–289, 2012.

ABBAS, Q.; CELEBI, M. E.; SERRANO, C.; GARCÍA, I. F.; MA, G. Pattern classification of dermoscopy images: A perceptually uniform model. *Pattern Recognition*, Elsevier, v. 46, n. 1, p. 86–97, 2013.

Ali, A. H.; Li, J.; Yang, G. Automating the abcd rule for melanoma detection: A survey. *IEEE Access*, v. 8, p. 83333–83346, 2020.

AMERICAN Cancer Society - Key Statistics for Melanoma Skin Cancer. 2020. <https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html>. Accessed on: 05-05-2020.

ARGENZIANO, G.; SOYER, H. P.; CHIMENTI, S.; TALAMINI, R.; CORONA, R.; SERA, F.; BINDER, M.; CERRONI, L.; ROSA, G. D.; FERRARA, G. et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the internet. *Journal of the American Academy of Dermatology*, Elsevier, v. 48, n. 5, p. 679–693, 2003.

BASAL Cell Carcinoma - The Skin Cancer Foundation. 2021. <https://www.skincancer.org/skin-cancer-information/basal-cell-carcinoma/>. Accessed on: 02-02-2021.

BUDA, M.; MAKI, A.; MAZUROWSKI, M. A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, Elsevier, v. 106, p. 249–259, 2018.

CARLI, P.; QUERCIOLI, E.; SESTINI, S.; STANTE, M.; RICCI, L.; BRUNASSO, G.; GIORGI, V. D. Pattern analysis, not simplified algorithms, is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology. *British Journal of Dermatology*, Wiley Online Library, v. 148, n. 5, p. 981–984, 2003.

CELEBI, M. E.; KINGRAVI, H. A.; UDDIN, B.; IYATOMI, H.; ASLANDOGAN, Y. A.; STOECKER, W. V.; MOSS, R. H. A methodological approach to the classification of dermoscopy images. *Computerized Medical imaging and graphics*, Elsevier, v. 31, n. 6, p. 362–373, 2007.

CELEBI, M. E.; MENDONCA, T.; MARQUES, J. S. *Dermoscopy image analysis*. [S.l.]: CRC Press, 2015. v. 10.

CHATURVEDI, S. S.; GUPTA, K.; PRASAD, P. et al. Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using mobilenet. *arXiv preprint arXiv:1907.03220*, 2019.

DEMYANOV, S.; CHAKRAVORTY, R.; ABEDINI, M.; HALPERN, A.; GARNAVI, R. Classification of dermoscopy patterns using deep convolutional neural networks. In: IEEE. *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. [S.l.], 2016. p. 364–368.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. ImageNet: A Large-Scale Hierarchical Image Database. In: *CVPR09*. [S.l.: s.n.], 2009.

G SOYER HP, D. G. V. A. Dermoscopy: A tutorial. In: *EDRA Medical Publishing New Media*. [S.l.: s.n.], 2002.

HAY, R. J.; FULLER, L. C. The assessment of dermatological needs in resource-poor regions. *International journal of dermatology*, Wiley Online Library, v. 50, n. 5, p. 552–557, 2011.

HAY, R. J.; JOHNS, N. E.; WILLIAMS, H. C.; BOLLIGER, I. W.; DELLAVALLE, R. P.; MARGOLIS, D. J.; MARKS, R.; NALDI, L.; WEINSTOCK, M. A.; WULF, S. K. et al. The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. *Journal of Investigative Dermatology*, Elsevier, v. 134, n. 6, p. 1527–1534, 2014.

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*. [S.l.: s.n.], 2016. p. 770–778.

HOSNY, K. M.; KASSEM, M. A.; FOAUD, M. M. Classification of skin lesions using transfer learning and augmentation with alex-net. *PloS one*, Public Library of Science, v. 14, n. 5, p. e0217293, 2019.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

HUANG, G.; LIU, Z.; MAATEN, L. V. D.; WEINBERGER, K. Q. Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 4700–4708.

IANDOLA, F. N.; HAN, S.; MOSKEWICZ, M. W.; ASHRAF, K.; DALLY, W. J.; KEUTZER, K. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.

ISASI, A. G.; ZAPIRAIN, B. G.; ZORRILLA, A. M. Melanomas non-invasive diagnosis application based on the abcd rule and pattern recognition image processing algorithms. *Computers in Biology and Medicine*, Elsevier, v. 41, n. 9, p. 742–755, 2011.

IYATOMI, H.; OKA, H.; CELEBI, M. E.; OGAWA, K.; ARGENZIANO, G.; SOYER, H. P.; KOGA, H.; SAIDA, T.; OHARA, K.; TANAKA, M. Computer-based classification of dermoscopy images of melanocytic lesions on acral volar skin. *Journal of Investigative Dermatology*, Elsevier, v. 128, n. 8, p. 2049–2054, 2008.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015.

LOPEZ, A. R.; NIETO, X. Giro-i; BURDICK, J.; MARQUES, O. Skin lesion classification from dermoscopic images using deep learning techniques. In: IEEE. *2017 13th IASTED International Conference on Biomedical Engineering (BioMed)*. [S.l.], 2017. p. 49–54.

MAHBOD, A.; SCHAEFER, G.; WANG, C.; ECKER, R.; ELLINGE, I. Skin lesion classification using hybrid deep neural networks. In: IEEE. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.], 2019. p. 1229–1233.

MENDOZA, C. S.; SERRANO, C.; ACHA, B. Scale invariant descriptors in pattern analysis of melanocytic lesions. In: IEEE. *2009 16th IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2009. p. 4193–4196.

MENEGOLA, A.; FORNACIALI, M.; PIRES, R.; BITTENCOURT, F. V.; AVILA, S.; VALLE, E. Knowledge transfer for melanoma screening with deep learning. In: IEEE. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. [S.l.], 2017. p. 297–300.

MIKOŁAJCZYK, A.; GROCHOWSKI, M. Data augmentation for improving deep learning in image classification problem. In: IEEE. *2018 international interdisciplinary PhD workshop (IIPhDW)*. [S.l.], 2018. p. 117–122.

MUNIR, K.; ELAHI, H.; AYUB, A.; FREZZA, F.; RIZZI, A. Cancer diagnosis using deep learning: A bibliographic review. *Cancers*, Multidisciplinary Digital Publishing Institute, v. 11, n. 9, p. 1235, 2019.

PEREZ, F.; VASCONCELOS, C.; AVILA, S.; VALLE, E. Data augmentation for skin lesion analysis. In: *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. [S.l.]: Springer, 2018. p. 303–311.

PEREZ, L.; WANG, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.

SÁEZ, A.; ACHA, B.; SERRANO, C. Pattern analysis in dermoscopic images. In: *Computer vision techniques for the diagnosis of skin Cancer*. [S.l.]: Springer, 2014. p. 23–48.

SERRANO, C.; ACHA, B. Pattern analysis of dermoscopic images based on markov random fields. *Pattern Recognition*, Elsevier, v. 42, n. 6, p. 1052–1057, 2009.

SHIN, H.-C.; ROTH, H. R.; GAO, M.; LU, L.; XU, Z.; NOGUES, I.; YAO, J.; MOLLURA, D.; SUMMERS, R. M. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, IEEE, v. 35, n. 5, p. 1285–1298, 2016.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

SITU, N.; YUAN, X.; ZOURIDAKIS, G. Assisting main task learning by heterogeneous auxiliary tasks with applications to skin cancer screening. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. [S.l.: s.n.], 2011. p. 688–697.

SKIN Cancer Statistics. <https://www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics>. Accessed on: 07-05-2020.

SMITH, L. N. Cyclical learning rates for training neural networks. In: IEEE. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. [S.l.], 2017. p. 464–472.

SZEGEDY, C.; IOFFE, S.; VANHOUCKE, V.; ALEMI, A. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261*, 2016.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2015. p. 1–9.

TANAKA, T.; TORII, S.; KABUTA, I.; SHIMIZU, K.; TANAKA, M. Pattern classification of nevus with texture analysis. *IEEJ Transactions on Electrical and Electronic Engineering*, Wiley Online Library, v. 3, n. 1, p. 143–150, 2008.

ZHANG, X.; ZHOU, X.; LIN, M.; SUN, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 6848–6856.

ZOPH, B.; VASUDEVAN, V.; SHLENS, J.; LE, Q. V. Learning transferable architectures for scalable image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 8697–8710.