



**UNIVERSIDAD
DE ANTIOQUIA**

Melanoma Classification

Oscar Nicolas Gomez Giraldo

Nestor Ivan Arbelaez Lopez

Universidad de Antioquia

Facultad de ingeniería, Departamento ingeniería en
sistemas

Medellín, Colombia

2021



Melanoma Classification

Oscar Nicolas Gomez Giraldo

Nestor Ivan Arbelaez Lopez

Monografía presentada como requisito parcial para optar al título de: **Especialización en
Analítica y Ciencia de Datos**

Asesor(a):

Lina María Sepúlveda Cano PhD. Ingeniera Electrónica (MSc)

Universidad de Antioquia

Facultad de Ingeniería, Departamento de ingeniería en sistemas

Medellín, Colombia

2021

Melanoma Classification

Néstor Iván Arbelaez López
Especialización Analítica
Universidad de Antioquia
Medellín, Colombia
nestor.arbelaez@udea.edu.co

Oscar Nicolás Gómez Giraldo
Especialización Analítica
Universidad de Antioquia
Medellín, Colombia
onicolas.gomez@udea.edu.co

Abstract—We presented our solution for the SIIM-ISIC melanoma classification challenge. This is a multi-class multi-modal classification model using images and metadata and, we tested both binary and multi-class image-only models and a binary multi-modal model. The keys to success for our solution were the selection of the target variable, using the available metadata, and the data augmentation strategy. Achieving AUC values of 0.95 and F1 of 0.71 for the validation data.

Index Terms—Melanoma, multi-modal, CNN, data augmentation

I. INTRODUCTION

Skin cancer is one of the most common types of cancer. Melanoma, although it is one of the less common types of cancer, is responsible of 75% of deaths related with skin cancer. Similar to other cancer types, an early detection may allow the treatment to be more effective. Although melanoma can be diagnosed in its first stages by visual inspection, many people are still diagnosed lately. As a consequence, it is estimated that 57000 victims worldwide died by melanoma in 2020 [1].

The use of computational tools related to machine learning, such as computer vision deep learning problems, has reached human-level performance in many cases. The usefulness of these models allows the scalability of the proposed solution. Besides, it can be used in regions in which access to a specialist is difficult. Furthermore, it can be used as a diagnosis assistant [2].

The availability of new datasets, with large image quantities that are reviewed by specialists, might boost the development of new models, which can be an excellent help for dermatologists, allowing identify lesions suspicious of been melanoma, making it possible to have a better performance in their medical labor [1], [2].

In this work, we show our approach to solve the SIIM-ISIC Melanoma Classification challenge¹. The goal of this competition is to identify melanoma from a set of skin lesion images with contextual data. The paper structure is outlined as follows. Section II describes the dataset that we work with and provides an overview of the methodology used for this work. In Section III, we present the results of this study. Lastly, in Section IV, we present our conclusions.

¹<https://www.kaggle.com/c/siim-isic-melanoma-classification>

II. MATERIALS AND METHODS

A. Data

For this problem, we have mixed data which consists on images of skin lesions and their corresponding contextual tabular data with information of the approximate age, location of the lesion, sex, diagnosis, and a Benign/Malignant labeling according to the lack or presence of Melanoma in the diagnosis. In Figure 1, we show a sample of lesion images from the dataset.

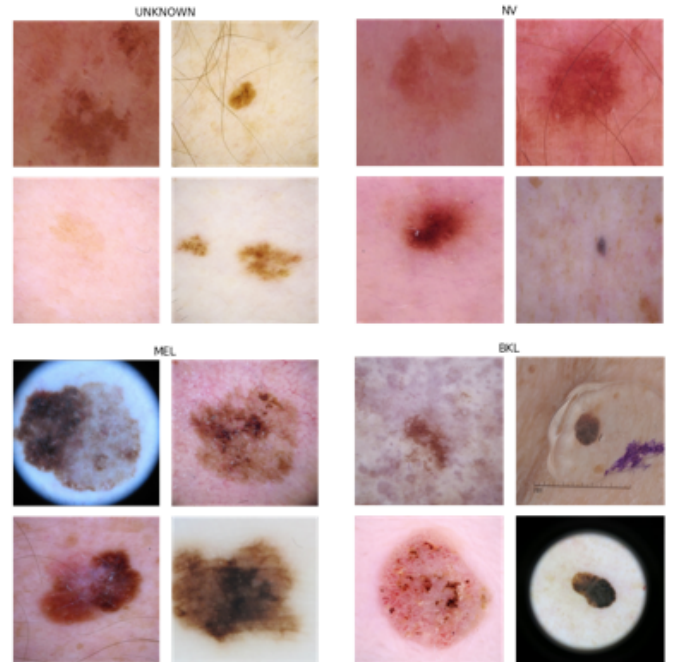


Fig. 1. Sample of skin lesion images by diagnosis (UNKNOWN, NV, MEL, BKL).

The 2020 competition data is composed of 33k images, from which 1.76% are marked as melanoma. Since the original competition data has a very high-class imbalance, we also added the 2019 data (including 2018) [3], increasing the sample to 55k with a class imbalance of 8.66%. Besides, by general rule, neural network models fit better when you have a larger number of samples, becoming this into a significant improvement from the original dataset.

B. Methodology

To test different approaches and thus validate which one better fits our problem, the project was developed in different iterations. In these iterations, we tested several neural network architectures, input data types, data pre-processes, and target variables. We will detail them below.

1) *Iteration 1*: In this iteration, we sought to define the architecture of the convolutional neural network to be used as a baseline for following iterations. For this, we defined our task as a binary classification problem, as 2020 SIIM-ISIC challenge suggested. In addition, we also performed the following processes, which will be used in subsequent iterations.

Preprocessing the dataset images have different sizes; in order to be used in neural networks, they must have the same dimensions. For this, we resized dataset images to the input dimension of the defined neural network architecture. In addition, we normalize pixel values to the range between 0 and 1.

Data augmentation for unbalanced problems, it is common to use oversampling techniques for the minor class. We used the Keras library² with the following transformations: shear, zoom, rotation, width shift, height shift, horizontal flip, vertical flip, brightness, and Gaussian blur. In Figure 2, we show the effect of the transformation on the dataset images.

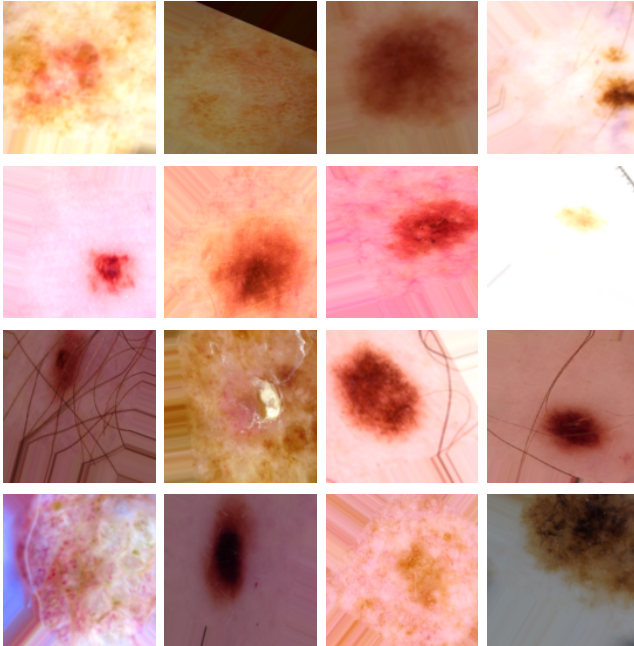


Fig. 2. Train augmentation for the dataset. The Figure shows the effect of the image transformations on the original images.

Models in this iteration, we tested different pre-trained convolutional neural network architectures with the ImageNet dataset [7], the tested architectures were MobileNetV2 [4],

²https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator

ResNet50, ResNet101 [5], and Xception [6]. From these architectures, we selected Xception since we found it has the best performance. Figure 3 shows the Xception architecture. After choosing the architecture, we went on to test some of its hyperparameters, such as different specific cost functions for unbalanced problems, however, the Keras cost function performs better for classification problems. We also tested other optimizers, obtaining a faster convergence with the Adam optimizer.

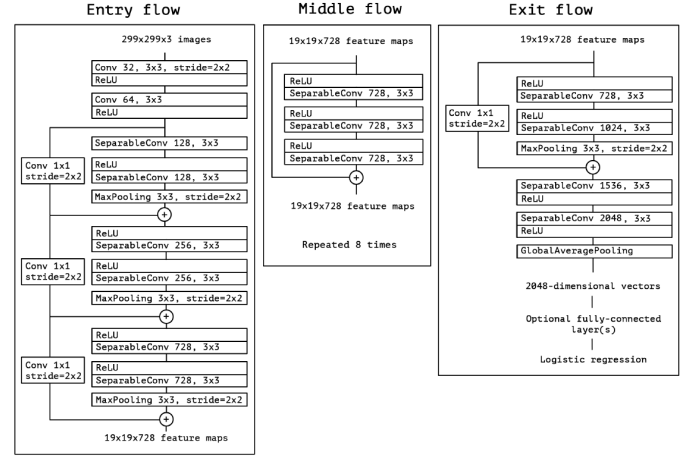


Fig. 3. The Xception architecture [6].

2) *Iteration 2*: We moved from a binary classification model to a multi-class approach.

Preprocessing we apply a mapping to the diagnosis variable for the 2020, and 2019 data [3] since the output variable in both datasets has some differences. In Table I, we show the mapping used in this work. Moreover, Figure 4 shows the diagnose distribution once the mapping is applied.

TABLE I
TARGET MAPPING FOR 2019 AND 2020 DATA.

2019	2020	Output
NV	nevus	NV
MEL	melanoma	MEL
BKL	seborrheic keratosis solar lentigolichenoid keratosis	BKL
BCC AK SCC VASC DF	atypical melanocytic proliferation cafe-au-lait macule unknown	UNKNOWN

3) *Iteration 3*: In this iteration, we moved from an image only to a multi-modal model, where, in addition to images as input variables, we also have tabular data as input.

Preprocessing we performed a mapping of the skin lesion location as described in Table II. Besides, we normalized the age variable to the range between 0 and 1.

Models the neural network is adjusted to be multi-modal, i.e., in addition to the convolutional neural network, a dense neural network is included, then the output of both branches

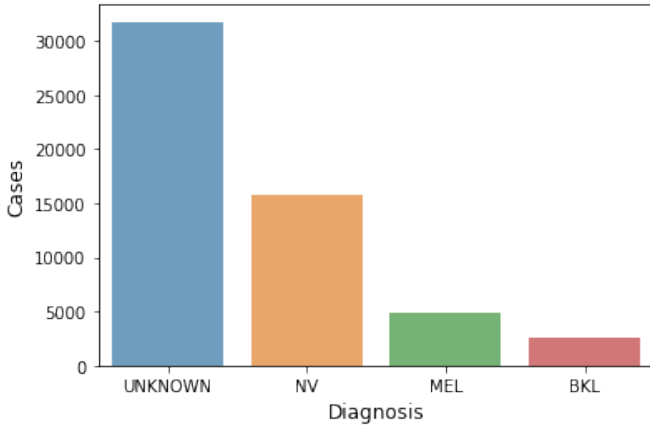


Fig. 4. Final diagnose distribution once the mapping summarized in Table I is applied.

TABLE II
LOCATION MAPPING.

2019	2020	Output
head/neck	head/neck	head/neck
lower extremity	lower extremity	lower extremity
oral/genital	oral/genital	oral/genital
upper extremity	upper extremity	upper extremity
anterior torso	torso	torso
lateral torso		
posterior torso		
torso		

are concatenated to generate later the outcome of the multi-class model, Figure 5 illustrates model architecture.

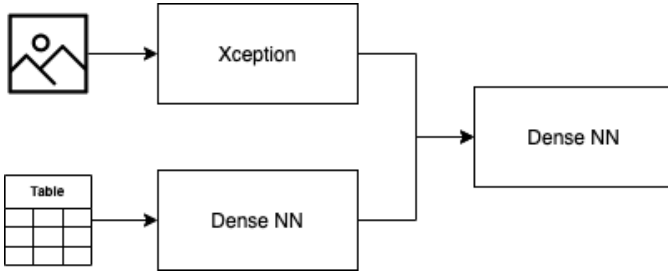


Fig. 5. Multi-modal architecture

4) *Iteration 4:* In this iteration, we test the multi-modal model with a binary output.

Preprocessing the target variable was changed, turning the problem into binary classification problem.

III. RESULTS

In Table III, we summarize the model iterations used in this work. For all approaches, we used 2019+2020 dataset. The architecture can be either the single convolutional neural network (CNN) or the multi-modal model. For data augmentation, we used the approach described in Subsection II-B. Finally, we also checked the effect of the binary or multi-class output. For all experiments, we trained up to 18 epochs.

TABLE III
SUMMARY OF MODEL ITERATIONS ANALYZED IN THIS WORK.

	Architecture	Output
iteration 1	CNN	Binary
iteration 2	CNN	Multiclass
iteration 3	Multimodal	Multiclass
iteration 4	Multimodal	Binary

For the SIIM-ISIC Kaggle competition, we have the area under the curve (AUC) of the ROC curve between the predicted probability and the observed target as the evaluation metric. In Figure 6, we plot the AUC metric versus the epoch number for all iterations summarized in Table III, both for train and validation data. As can be seen from the Figure, for iterations 1 and 4, although they have higher AUC values for train data, they also exhibit the presence of high overfitting. On the other hand, we found for iterations 2 and 3 that overfitting is negligible, meaning that, for this problem, a proper way to tackle overfitting is to change the output of the model architecture to a multi-class approach. This is because, for the diagnosis column, value distribution is more granular than the binary case, turning this into a way to add information to the model [3].

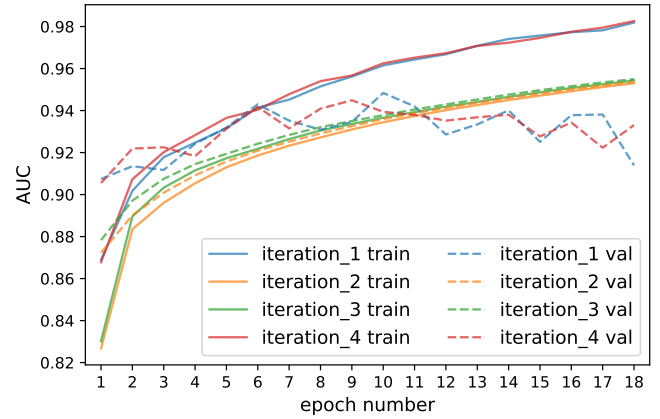


Fig. 6. AUC score versus epoch number for the 4 iterations described in Table III. Train and validation data results are shown with solid and dashed line respectively.

In Figure 7, we show the F1 score metric obtained for the validation data versus the test data for each one of the iterations described in Table III. As we can see from the Figure, the best result either for validation or test data is obtained for iteration 3, i.e., the use of metadata slightly increases the predictability power of the model. Consequently, the use of contextual data helps the model to perform better.

IV. CONCLUSIONS

This work summarizes the process performed for the challenge of skin lesion analysis for melanoma detection, where with the use of different methods based on deep neural networks a competitive classification of skin lesions for melanoma detection can be achieved.

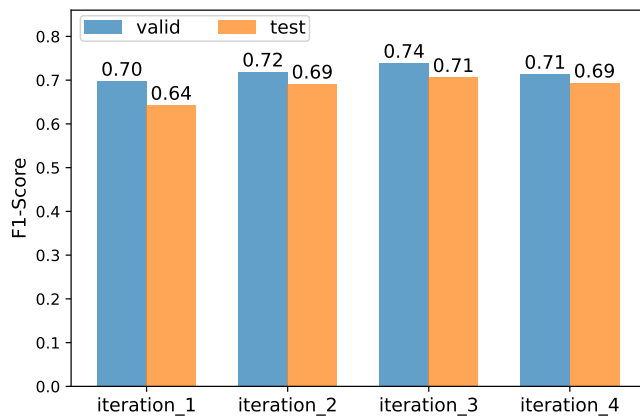


Fig. 7. F1 score metric of validation (blue bar) and test (orange bar) data for the iterations summarized in this work.

- We found in the multi-class classification models lower overfitting with respect to the binary classification models, that although they presented a better performance in train with respect to the multi-class in test the performance was lower.
- We found that the models in which metadata was used showed an improvement in F1, obtaining 0.74 for the validation data.
- The use of data augmentation helps the model to better generalize since it has a wide variety of input images.

The code used to develop this work is available on https://github.com/nicolasgiraldo/melanoma_classification_project.

REFERENCES

- [1] International Skin Imaging Collaboration (ISIC), “About ISIC: Background.” [Online]. Available: <https://www.isic-archive.com/#!/topWithHeader/tightContentTop/about/aboutIsicBackground> [Accessed: 20-Nov-2021].
- [2] I. Castiglioni et al., “AI applications to medical images: From machine learning to deep learning,” *Physica Medica*, vol. 83, pp. 9–24, March 2021.
- [3] Q. Ha, B. Liu, F. Liu, “Identifying melanoma images using efficientnet ensemble: Winning solution to the siim-isic melanoma classification challenge,” *arXiv preprint arXiv:2010.05351*, 2020.
- [4] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018 pp. 4510–4520.
- [5] K. He, X. Zhang, S. Ren, J. Sun, “Identity Mappings in Deep Residual Networks,” *European conference on computer vision 2016*, 2016 pp. 630–645. Springer, Cham.
- [6] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017 pp. 1800–1807.
- [7] O. Russakovsky et al., “ImageNet Large Scale Visual Recognition Challenge”, *International Journal of Computer Vision (IJCV)*, 2015 pp. 211–252.