Machine Learning-Based Classification of Skin Lesions

Joan Carles Montero Jiménez

Abstract— This study focuses on the classification of seven different types of skin lesions to create a preliminary assessment tool for identifying skin cancers, particularly melanomas. The data, sourced from the HAM10000 dataset on Kaggle, contains 10015 jpg images of various skin lesions. The study first involves reducing image dimensionality and ensuring balanced representation across all lesion classes. Two primary methodologies are employed: classical machine learning algorithms, including Random Forest and logistic regression-based one-vs-rest and one-vs-one approaches, and Convolutional Neural Networks (CNNs) using Google Colab's GPU computing.

The classical approach uses a K-Means algorithm to simplify images into principal colours and their standard deviations, while CNNs employ layers like BatchNormalisation, Conv2D, MaxPooling2D, Flatten, and Dense with Dropout for multi-class classification. Furthermore, the study adapts the CNN model for binary classification to distinguish between melanoma and non-melanoma cases. Evaluation metrics prioritize recall to ensure accurate cancer detection, with Grad-CAM used to visualize the decision-making process of the CNNs.

This work culminates in a Streamlit app enabling users to upload lesion images for preliminary melanoma diagnosis, demonstrating the practical application of these AI techniques in healthcare. This app will use the trained CNN multi class model reaching a 50% recall with melanomas.

Keywords—Image classification, cancer detection, CNNs, Streamlit

1 Introduction

Skin cancer, a prevalent and potentially life-threatening condition, commonly manifests in skin areas frequently exposed to sunlight, such as the face, neck, hands, and arms. Among its various forms, melanomas stand out for being the most dangerous skin cancer. Although it is not the most common type, their aggressive nature makes early detection crucial for improving the prognosis and extending the lifespan of those affected. This disease predominantly affects individuals between 65-74 years old, with a higher incidence in white males. Melanomas originate from malignant transformations in melanocytes, the skin cells responsible for producing melanin, the pigment that gives skin its natural colour[1].

Artificial intelligence has demonstrated to be effective in plenty of fields, including image classification. For those reasons, it would be of high interest to perform a 7-class classification of lesions, with the final objective of creating a preliminary assessment tool for identifying those lesions.

The data used comes from a Kaggle dataset[2], containing 10015 jpg images of seven different skin lesions, as well as their respective masks, to just focus on where the lesion is. Those lesions are abbreviated and may contain several diseases per group. Although it depends on each specific case, those lesions can be sorted by survival rate in the following table, so that then we can focus on the most critical ones.

Lesion	Brief explanation
Melanoma (MEL)	Most serious form of skin cancer. They can spread to other parts of the body and can be deadly if not caught early
Basal Cell Carcinoma (BCC)	Common form of skin cancer. It grows slowly and rarely spreads to other parts of the body. However, if left untreated, it can reach the bone causing serious damage.
Bowen's Disease (AKIEC)	These are early forms of skin cancer that can develop into more serious types if not treated. However, when caught early, they are highly treatable.
Benign Keratosis-like Lesions (BKL)	Common benign skin growth. Include solar lentigines, seborrheic keratoses, etc. They are generally not dangerous but can be monitored for changes.
Dermatofibroma (DF)	A benign skin lesion that usually does not pose a threat to life. It is more of a cosmetic concern.
Melanocytic nevi (NV)	Common and usually harmless. However, they can sometimes develop into melanoma, which is why they are monitored for changes in size, colour, or shape.
Vascular Lesions (VASC)	Including angiomas, angiokeratomas, pyogenic granulomas, etc. Most of them are benign and have no impact on life expectancy.

Table 1. Lesions explanation by survival rate.

Originally, the data comes from the HAM10000 dataset[3], and contained other information such as the patient age, gender, location of the lesion and other relevant data that it was not considered in this study, to just focus on the images analysis.

An example of how these looks like, with its corresponding mask (included on the Kaggle dataset) is the following.

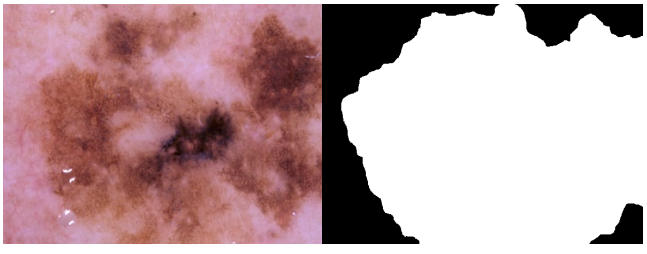

Fig. 1 and 2. Example of a melanoma and its respective mask. Image ID: ISIC_0024449

Also, it is key to consider the number of samples of each class, as we have much more NV images than DF for instance. These initial proportions were the following.

Lesion	MEL	BCC	AKIEC	BKL	DF	NV	VASC
Prop.	0.11	0.05	0.03	0.11	0.01	0.67	0.02

Table 2. Proportions (expressed on a per unit basis) of the lesions in our original dataset.

2 METHODOLOGY

To address the initial problem, it was necessary to reduce the dimensionality of the images. Originally sized at 600x450 pixels, each was downscaled to 300x225 pixels, with each pixel containing three values corresponding to the RGB channel. To further simplify, a subset of 1001 images from the original dataset was selected, ensuring an equal representation of 143 images per class, approximately 14% for each. For classes with fewer images, some were duplicated to maintain balance, resulting in a smaller yet more uniformly distributed dataset.

In solving the image multi-class classification challenge, two strategies were employed. The first utilised algorithms such as Random Forests in combination with one-vs-rest and one-vs-one approaches, implemented through Logistic Regressions. Despite size reduction, each image still consisted of 300x225x3 data points, necessitating data extraction. The K-Means algorithm was then implemented to extract the first three principal colours, and the standard deviation of each colour. This approach effectively simplified each image to 12 values: 9 for the principal colours and 3 for the standard deviations of the red, green, and blue values. The data was split into training and testing sets, and a standard scaler was used for normalisation. This process was applied to images with their respective masks to exclude non-lesion principal colours.

Concurrently, convolutional neural networks (CNNs), a common technique for image classification, were employed. For this, the Google Colab platform was utilised to leverage GPU computing, significantly reducing execution time. The GPU used was the Tesla T4, based on the Turing architecture, featuring 40 Streaming Multiprocessors (SMs), 6MB of L2 cache memory, and 16GB of GDDR6 VRAM. The CNN architecture comprised five parts:

1. A *BatchNormalisation* layer to normalise the input and accelerate learning.
2. Two sets of *Conv2D* layers for complex feature learning from the images, each followed by a *MaxPooling2D* layer to reduce spatial dimensions.
3. A *Flatten* layer to convert 2D feature maps into a 1D feature vector, connected to *Dense* layers.
4. A *Dense* layer, regulated with L2 regularisation and followed by a *Dropout* layer to prevent overfitting, processing the features from the *Conv2D* layers.
5. A final *Dense* layer with seven units and a *SoftMax* activation function for the seven-class classification task.

The model was trained using the Adaptive Moment Estimation (Adam) optimiser and the categorical cross-entropy loss function. Data augmentation, including adding noise and slight rotations, was crucial to manage image repetitions from less represented classes and prevent overfitting.

For efficient image uploading to Google Colab, the images were transformed into NumPy objects and uploaded to Google Drive using the Google Cloud API service.

To translate the work into a real-world application, the same model was adapted to distinguish between melanoma and non-melanoma cases. This required resampling the classes, following the previously discussed technique, but with 1008 images: 504 of melanomas and 84 of each other class, maintaining a 50% representation for each group. Some minor changed were needed to the model, such as changing the last layer to have only one dense unit with the sigmoid activation.

After developing these models, a Streamlit app was created that allows users to upload a lesion image for a preliminary skin lesion identification.

3 RESULTS AND ANALYSIS

Dealing with the sensitive topic of skin cancers, it's crucial to focus on recall metrics, which indicate how many actual cases are correctly identified by the model. High recall is essential, as missing a true positive – not detecting a skin cancer case – can have serious consequences. In this context, while precision is important to avoid false alarms, the priority is to maximize recall to ensure effective detection of skin cancer. Early detection is key, and this approach prioritizes patient safety and accurate diagnosis.

3.1 Classical machine learning approach with reduced dimensionality

In the analysis of the three models - Random Forest, One-vs-Rest, and One-vs-One classifiers - they show similar performance but with distinct characteristics. All of them classify accurately the vascular lesions class, as it those skin lesions are characteristic for their pink/purple's shades. In terms of global metrics, the Random Forest provides the highest precision and recall values, almost reaching 50%. On the other hand, the other two models obtain better melanoma recalls.

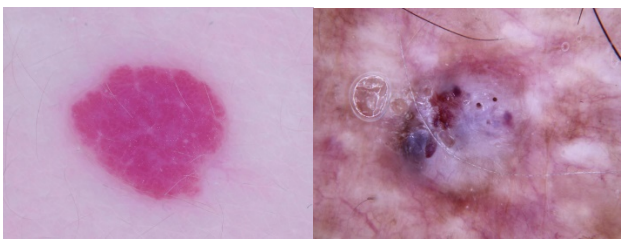

Fig. 3 and 4. Example of a vascular lesion (left) against a basal cell carcinoma (right). Vascular lesions follow a pattern with homogeneous red/pink stains.

Comparing the other two, the One-vs-Rest classifier exhibits a stronger performance, especially with a 50% recall for melanomas and an overall recall of 47%. These figures position it as the most effective model among the three. This is the confusion matrix obtained from the model.

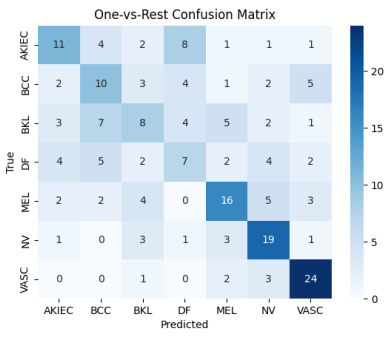

Fig. 5. Confusion matrix for the One-vs-Rest model.

Despite the efficiency and accuracy of these models, there is a keen interest in exploring more advanced techniques, such as Convolutional Neural Networks (CNNs), for further enhancement of the analysis. This suggests a move towards more sophisticated methods that could potentially offer better prediction accuracy and recalls.

3.2 CNNs for multi class classification

The CNN resulted in a more robust model with a global precision and recall of about 55%, with a recall value of 59% in melanomas and 64% in basal cell carcinomas, the two most dangerous skin lesions. The network also recognises with ease the vascular lesions. The respective confusion matrix is the following.

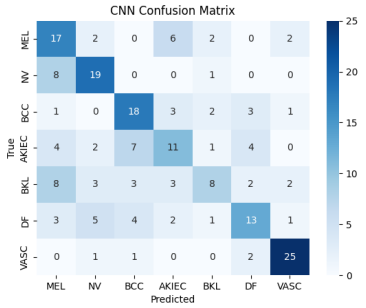

Fig. 6. Confusion matrix for the initial CNN model.

Achieving a high melanoma recall would be something interesting. As this model does not classify two classes, the process of predicting a class it is being done by getting the highest value of each array of normalized probabilities. Because of that, it is not possible to just lower the threshold so that we get the desired recall value, so other alternative methods should be done.

The one chosen consisted of setting up a threshold for melanomas in which when the predicted value exceeds it is automatically assigned to be one of it, if not, the standard process of assigning the class with the maximum value is being done. So, to first compute this value each PR class was being represented as following.

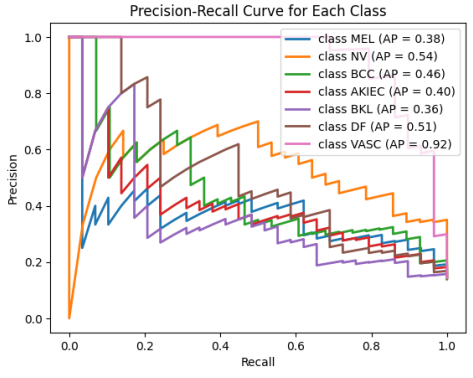

Fig. 7. PR curves for each class of the CNN model.

It is important to remark the high performance of the vascular lesions class, as it was previously commented, that for being different, it can be easily classified, followed by the melanocytic nevi lesion. The worst performances are attributed to melanomas and basal cell carcinomas, that indeed are the most dangerous skin cancers.

Then, after setting the recall for melanoma to 90%, it was only needed to get the nearest threshold for that value obtained from the individual PR curve. Changing the class assignment to the explained methodology, the new confusion matrix shows as following.

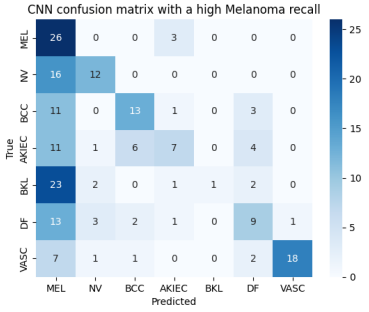

Fig. 8. Confusion matrix for the CNN model with a high Melanoma recall.

Despite the model is getting confused identifying basal cell carcinoma as melanomas, this could be a possible skin cancer identification what might work in some cases, almost capturing every melanoma and sometimes predicting other skin lesions.

3.3 CNNs Grad-CAM

Convolutional Neural Networks (CNNs) can often seem like a "black box" in terms of understanding how they make predictions. To demystify this process and gain insight into what the network is "seeing," techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) are used.

When Grad-CAM is applied to such a model, it focuses on the last convolutional layer of the CNN. This is crucial because this layer captures the most abstract and detailed features of the skin lesions in the input images, which are essential for accurate classification.

For each of the seven skin cancer classes, Grad-CAM generates a unique heatmap that highlights the most influential areas in the prediction of the specific class. For instance, in the case of melanoma, the heatmap might highlight heterogeneous colouring, or in case of vascular lesions its sharply defined area.

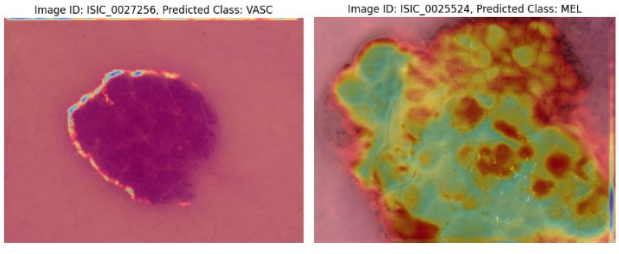

Fig. 9 and 10. GRAD_CAMs maps for two images, predicted as vascular lesion and melanoma respectively. For each class, the model considers different aspects of the image.

3.4 CNNs for binary classification

As it was seen, it is of great importance to be able to discern between having or not a Melanoma, then it is reasoning creating a binary model, just to discriminate both cases. The results obtained when splitting the data into both classes are satisfactory, with a global precision of 81% and a great MEL recall up to 95%, without initially modifying the threshold.

By plotting up the PR curve and lowering the threshold, it will be achieved a specific MEL recall, in this example set to 99%. This will allow the model to minimize the problematic false negatives while still be able to predict not having a skin cancer in a surprisingly great way.

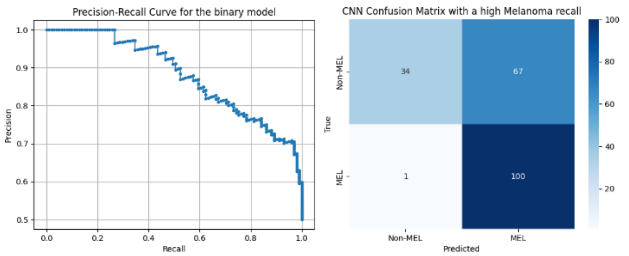

Fig. 11 and 12. PR curve of the binary CNN model (left) and the confusion matrix for this model with a high Melanoma recall value (right).

3.5 Streamlit implementation

A streamlined Streamlit app has been created to effectively utilize the insights from the analysis. This user-friendly application prompts users to upload an image, which is then evaluated by a binary CNN model to determine whether it is indicative of melanoma or another type of skin lesion. The app not only reveals the model's predictions but also displays the associated confidence percentages. Additionally, it includes a Grad-CAM heatmap to visually highlight the areas the model focused on while making its assessment. For those seeking more information about melanomas, the app features an informative section. It has been successfully containerized using Docker, allowing easy access for anyone interested in testing the app. To use it, simply pull the image from Docker Hub by entering: ***docker pull mjoancarles/skincancer:latest*** in your terminal. Below are some screenshots demonstrating how the app works.

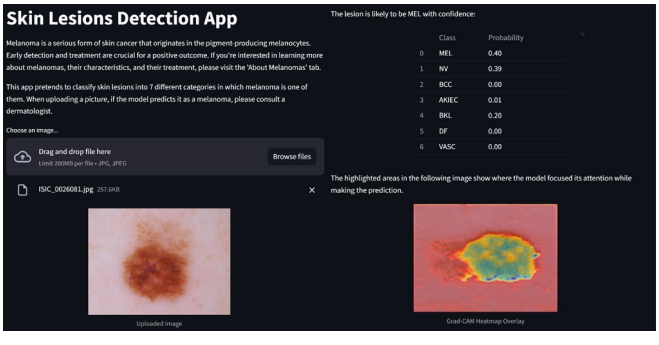

Fig. 13 and 14. App testing with a MEL image, showing the predictions and GRAD-CAM associated with the prediction.

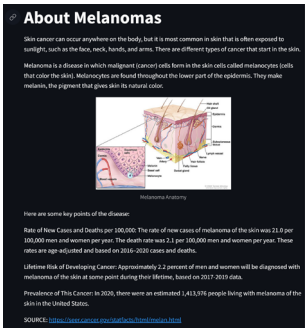

Fig. 15. App information screen about melanomas, to divulgate the importance of early prevention and the origins of it.

4 Conclusions

In this research, the use of Convolutional Neural Networks (CNNs) proved effective in classifying melanomas from skin lesion images, a task notably challenging due to the lack of distinct features in melanomas. The CNN approach, which demands more computational resources, particularly GPUs, showed superior performance compared to simpler techniques like the K-Means algorithm, which was initially employed for image dimensionality reduction.

The study was conducted with a constrained subset of images, indicating that extending the analysis to the full 10,000-image dataset could potentially yield more accurate and reliable results. Furthermore, the incorporation of additional data extracted from the images, such as specific lesion locations and patient demographic information, could significantly enhance the effectiveness of the models, especially for methods that focus on reduced image complexity.

Overall, the findings from this study highlight the potential of applying machine learning techniques in the field of dermatology, specifically for the early detection of melanomas. Future research, leveraging a more extensive dataset and a broader range of image features, is expected to refine these diagnostic tools further, offering valuable support in medical decision-making and potentially contributing to improved patient outcomes in skin cancer care.

Bibliography

[1] SEER Cancer Stat Facts: Melanoma of the Skin. National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/statfacts/html/melan.html

[2] Volodymyr Pivoshenko. 2023. Skin Cancer: Lesions Segmentation. Retrieved from https://www.kaggle.com/datasets/volodymyrpivoshenko/skin-cancer-lesions-segmentation/data

[3] 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, V4, UNF:6:KCZFcBLiFE5ObWcTc2ZBOA== [fileUNF]