# Classification of melanoma in images of skin lesions

Karlo Kada
FER
1191244222
Zabok, Croatia
karlo.kada@fer.hr

Mihael Kožul
FER
0016141878
Zagreb, Croatia
mihael.kozul@fer.hr

Leon Zrnić
FER
0036527005
Zagreb, Croatia
leon.zrnic@fer.hr

Karlo Šutalo
FER
0036522932
Zagreb, Croatia
karlo.sutalo@fer.hr

*Abstract*—**Today, the most frequent type of cancer is skin cancer, of which 75% is melanoma. As is the case with other cancers and diseases, early detection can lead to much higher chances of recovery, alongside lower repercussions. Leveraging deep learning techniques, researchers have made significant strides in developing robust models for accurate and timely melanoma diagnosis. Convolutional Neural Networks (CNNs) have emerged as a cornerstone in melanoma classification, weighing in their ability to automatically learn intricate features from skin images. The same techniques were used in this project using a truncated ISIC melanoma Kaggle competition dataset. Three models were built, each being an improved version of the previous model. The results obtained from the final model were deemed acceptable with the validation accuracy of 80%.**

**Keywords—image classification, melanoma, CNN, cancer, neural network**

## I. INTRODUCTION

The most prevalent type of cancer today is skin cancer. Melanoma, specifically, is responsible for 75% of skin cancer deaths, despite being the least common skin cancer. In 2022, it was estimated that almost 16.5 thousand people died from melanoma skin cancer in Europe [1]. As with other cancers, early and accurate detection—potentially aided by data science—can make treatment more effective.

Deep learning models, particularly convolutional neural networks (CNNs) show great potential in image classification and pattern analysis. In the context of melanoma identification, these models leverage vast datasets of skin images to learn complex features associated with malignant lesions. By training on diverse datasets that include various skin types and melanoma presentations, such models enhance their ability to discern subtle nuances indicative of malignancy.

Doctors may use these models to validate their findings or seek second opinions, contributing to a more robust diagnostic process. By incorporating deep learning models into their diagnostic workflows, doctors benefit from enhanced accuracy, efficiency, and valuable decision support, ultimately leading to more effective cancer diagnosis and treatment planning.

The integration of deep learning into melanoma identification processes has shown promising results, exhibiting comparable or even superior performance to traditional diagnostic methods. These advancements pave the way for automated and efficient screening, potentially reducing the time and expertise required for accurate melanoma diagnosis [2].

## II. OVERVIEW OF THE EXISTING WORK

The classification of melanoma in images of skin lesions has been one of the most essential problems in the intersection of data science and healthcare. Leveraging deep learning techniques, researchers have made significant strides in developing robust models for accurate and timely melanoma diagnosis. In this chapter we will go through some of the existing work in this domain, highlighting key methodologies, datasets, and advancements that contribute to the ongoing progress in melanoma classification.

Convolutional Neural Networks (CNNs) have emerged as a cornerstone in melanoma classification, leveraging their ability to automatically learn intricate features from skin images. Studies such as Esteva et al. (2017) demonstrated dermatologist-level classification accuracy using CNNs, underscoring the transformative potential of deep learning in dermatology [3]. A scheme of the CNN architecture is shown in the following image.
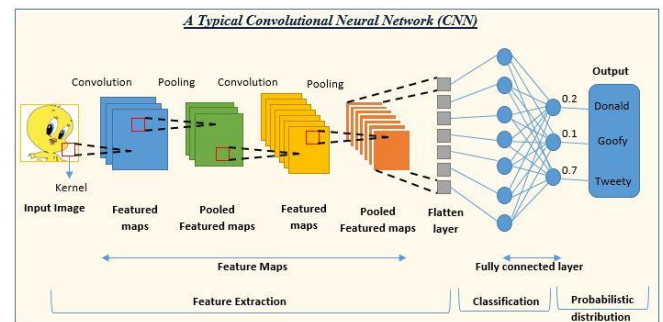


Scheme of the CNN architecture with two convolutional layers ("Conv2D") [4]

Architectures like DenseNet [5], ResNet [6], and EfficientNet [7], which are commonly used as pretrained models, have been tailored for melanoma detection tasks through techniques such as transfer learning and fine-tuning on melanoma-specific datasets. The availability of aforementioned large-scale annotated datasets, notably the International Skin Imaging Collaboration (ISIC) dataset, has been pivotal in training robust melanoma classification models. Transfer learning techniques, where pre-trained models on general image datasets like ImageNet are fine-tuned for melanoma classification, have become prevalent.

The combination of Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) in medical image analysis, including cancer detection, is often called a CNN-RNN architecture. This approach is particularly useful when dealing with sequences of medical images or volumetric data. RNNs are designed to capture sequential dependencies, making them suitable for understanding the temporal context in sequences of images. Such combination approach, which combines the geographical data collected by the CNN and the temporal dependencies collected by the RNN, has been proven to increase the precision of lesion categorization [8].

Another important approach to this problem are ensemble methods, which combine predictions from multiple models. Researchers have found that Convolutional Neural Network (CNN) based ensemble methods can improve the existing performance of lesion segmentation [9]. Furthermore, there is a growing emphasis on the interpretability of deep learning models in medical applications. Researchers are developing techniques to explain the decision-making process of CNNs, providing insights into the features influencing melanoma classification.

Recent studies explore the integration of clinical information, such as patient history and demographic data, with deep learning models for melanoma classification. Melanoma can manifest differently in various demographic groups, and integrating this knowledge into the model may aid in refining predictions. Patient history, including previous skin cancer diagnoses or family history of melanoma, can also be crucial in assessing the risk level of a given skin lesion. This holistic approach aims to improve the overall accuracy and clinical relevance of the models, aligning with the broader trend toward personalized and context-aware healthcare solutions [10].

In summary, the existing body of work showcases the transformative potential of data science and deep learning in melanoma classification. The integration of advanced CNN architectures, large-scale datasets, transfer learning, ensemble methods, and a focus on interpretability collectively contribute to the evolution of robust and clinically relevant models for melanoma diagnosis.

### III. SOLUTION DESCRIPTION

In this project, we used a publicly available dataset from Kaggle [11]. This set consists of 2357 images of malignant and benign oncological diseases, which were formed from The International Skin Imaging Collaboration (ISIC). The images are split into two separate subsets – training data (2239 images) and test data (118 images). All images were sorted according to the classification taken with ISIC, and all subsets were divided into the same number of images, with the exception of melanomas and moles, whose images are slightly dominant. The data set contains the following diseases: *actinic keratosis, basal cell carcinoma, dermatofibroma, melanoma, nevus, pigmented benign keratosis, seborrheic keratosis, squamous cell carcinoma and vascular lesion*. The original image size is 600×450 pixels.

The goal was to build a CNN based model which can accurately detect melanoma. For this task, we used TensorFlow which is an open-sourced end-to-end platform, a library for multiple machine learning tasks and Keras, a high-level neural network library that runs on top of TensorFlow.

We started with splitting the training data into two datasets with ratio of 80:20 – training dataset which contains 1792 images and validation dataset which contains 447 images, both with the batch size of 32. All the images were resized to 180×180 pixels.

In the next step, we visualized one instance of all the nine classes present in the dataset and normalized all of the images, so that all of the pixel's values were in the range [0, 1].

The model was built using the keras.Sequential class, which allows grouping a linear stack of layers into a model. The first layer was a 2D convolution layer with the number of filters set to 16, kernel size of 3, even padding and rectified linear activation function (ReLU). Next, we added a max pooling operation for 2D spatial data, which downsamples the input along its spatial dimensions (height and width) by taking the maximum value over an input window. This process reduces the computational load. We repeated these two layers two more times, but we changed the number of filters in the second convolution layer to 32, and to 64 in the third convolution layer.

We then added a Flatten layer, which is used to flatten the output from the last convolutional layer into a one-dimensional vector. This vector represents the extracted features from the convolutional layers and is suitable for input to the subsequent dense layers. The Flatten layer is followed by two Dense layers with ReLU activation, providing a fully connected structure. These layers can capture more complex relationships between the high-level features learned by the convolutional layers. The final Dense layer with 9 units (one for each class) and no activation function is used for classification. This layer produces the raw output scores for each class.

We chose the Adam optimizer, a stochastic gradient descent method that is based on adaptive estimation of first order and second-order moments. This method is computationally efficient, has little memory requirement, invariant to diagonal rescaling of gradients, and is well suited for problems that are large in terms of data/parameters [12]. The loss was defined through the SparseCategoricalCrossentropy class, which is used when there are two or more label classes, and it computes the cross-entropy loss between the labels and predictions.

We trained the model through 20 epochs and got the following results. On the x-axis we can see the number of the epoch.
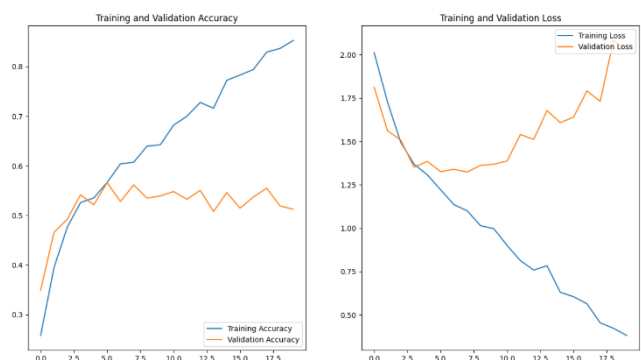


In the first graph, the training accuracy is increasing linearly over time, whereas validation accuracy stalls around 50% in the training process. Also, the difference in accuracy

between training and validation accuracy is noticeable, which is a sign of overfitting. When there are a small number of training examples, the model sometimes learns from noises or unwanted details from training examples to an extent that it negatively impacts the performance of the model on new examples. To solve this problem, we used data augmentation to generate additional training data from the existing images, by performing random transformations on the images, such as flipping, rotating, and zooming in the image. We also added a Dropout layer after the convolutional layers, as a form of regularization. The Dropout layer randomly sets input units to 0 with a frequency of rate at each step during training time, which helps prevent overfitting. Inputs not set to 0 are scaled up by $\frac{1}{(1-rate)}$ such that the sum over all inputs is unchanged. We chose the rate of 0.2, and after training the model through 20 epochs the results were better, but still not satisfying.
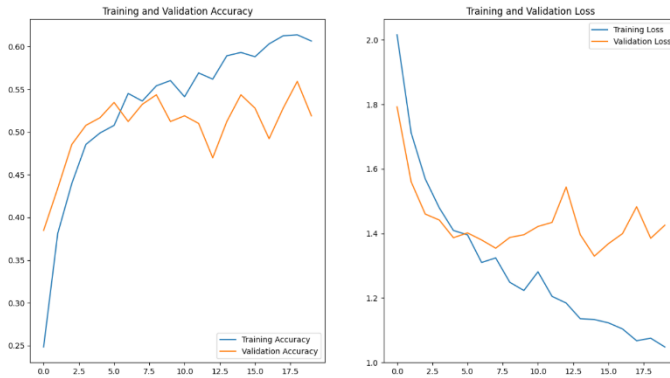


We can clearly see that the training and validation accuracy do not differ greatly, but the final model accuracy is around 50 %, which means that the model is underfitting – it is not able to capture the underlying patterns in the data.

To find the source of this problem, we visualize the training data. On the x-axis, we can see the names of each of the classes, and on y-axis, we can see number of images for each class.
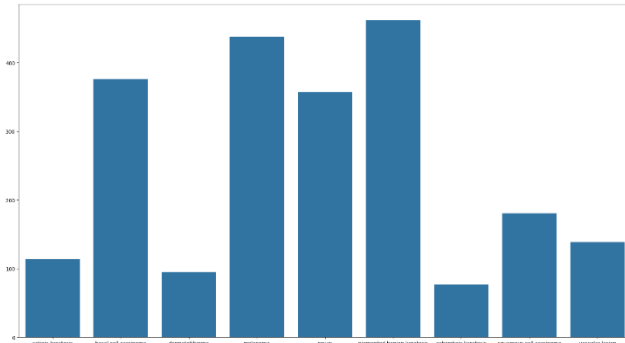


It is clear that the training dataset isn't balanced – the class with the lowest number of images is *seborrheic keratosis* (77 images) and the class with highest number of images is *pigmented benign keratosis* (462 Images).

To balance the training dataset, we used the Augmentor python library to add 500 additional image samples per class to all of the nine classes. Now we can see that the dataset is more balanced than before. The complete training dataset now consists of 6739 images, of which we now use 5392 in for training of the model and 1347 for validation of the model.
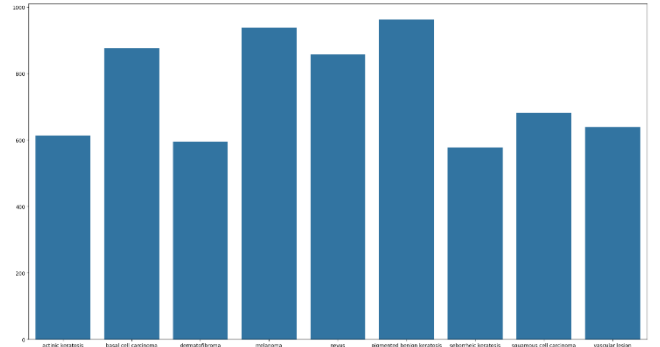


## IV. RESULTS

For the final step, we trained the model through 50 epochs, and got the following results:
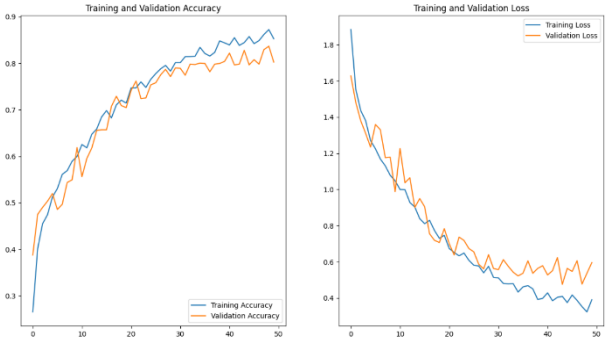


As we can see, the final training accuracy is 85%, and the validation accuracy 80%. The training loss is equal to 0.3907, and the validation loss to 0.5957. We find these results acceptable and satisfactory. The finished model is now both generalizable and accurate thanks to the synthetically enlarged dataset and the use of dropout layers.

## V. CONCLUSION

In conclusion, this project delves into the ever-evolving problem of skin cancer – specifically melanoma, and the need for both accurate and timely diagnosis in order to save lives or prevent unnecessary hardships. The integration of deep learning models such as the one implemented here presents a promising start for enhancing melanoma detection.

Three models were developed, each of which was an improved version of the previous one. The results obtained from the final model were deemed acceptable with the training accuracy being 85%, but more importantly validation accuracy being 80%. Of course, this model could be improved with further hyperparameter optimization and learning on a larger and higher quality dataset. Given being resource restricted the dataset that was used was a truncated one from the 2020 ISIC challenge consisting of approx. 5% of original dataset, which is around 2400 pictures. Additionally, experimenting with various other model architectures and technologies can also yield same, if not better and more satisfactory results.

### REFERENCES

[1] Estimated number of melanoma skin cancer deaths in European Union countries in 2022. (January 2024). https://www.statista.com/statistics/1388292/number-of-melanoma-skin-cancer-death-in-eu-by-country/#statisticContainer

[2]    Haenssle, Holger A., et al. "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists." Annals of oncology 29.8 (2018): 1836-1842.

[3]    Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." nature 542.7639 (2017): 115-118.

[4]    Shah, Saily. "A Typical Convolutional Neural Network (CNN)." Convolutional Neural Network: An Overview, 2012, https://www.analyticsvidhya.com/blog/2022/01/convolutional-neural-network-an-overview/. Accessed 8 Jan. 2024.

[5]    Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[6]    He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[7]    Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.

[8]    Kumar, RP Ram, et al. "Feasible Skin Lesion Detection using CNN and RNN." E3S Web of Conferences. Vol. 430. EDP Sciences, 2023.

[9]    Ali, Redha, et al. "Deep learning ensemble methods for skin lesion analysis towards melanoma detection." 2019 IEEE National Aerospace and electronics conference (NAECON). IEEE, 2019.

[10]    Han, Seung Seog, et al. "Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm." Journal of Investigative Dermatology 138.7 (2018): 1529-1538.

[11]    Andrey Katanskiy. 2020. Skin Cancer ISIC, [Version of the dataset]. Retrieved 3 Jan. 2020. from https://www.kaggle.com/datasets/nodoubttome/skin-cancer9-classesisic/data

[12]    Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).