

IMAGE CLASSIFICATION MODELS FOR SKIN CANCER DETECTION

Fabio Berrettoni, Marco D'Ercole, Gianluca Frezza,
Ludovica Mazza, Elisa Pierini

December 27, 2022

1 Abstract

In this project, we aimed to classify skin cancer images as either malignant or benign using three different machine learning models: a Convolutional Neural Network, based on the model of the **VGG-16** CNN, a Support Vector Machine, and a Naive Bayes classifier.

We imported a dataset of dermatoscopic images labeled as benign or malignant, balanced it by increasing the number of malignant images through data augmentation, and applied k-fold cross validation to evaluate the performance of the models.

The CNN model achieved the highest accuracy of 86%, followed by the SVM model at 78% and the Naive Bayes classifier at 74%.

2 Introduction

Skin cancer is one of the most common forms of cancer in the world, and early detection is crucial to increase the likelihood of effective treatment. However, detection of skin cancer can be difficult, especially for non-specialists, because of the variety of types and aspects it can take.

At the origin of some of these cancers there are melanocytes, usually of the skin and mucous membranes, and for this reason it takes the name Melanoma: 15% of these turn out to be fatal. Cases have been found to be more frequent in light-skinned in-

dividuals of European stock who have exposed the skin for a long time in very sunny areas.

In the last five years, deaths attributed to cutaneous melanoma in Italy were 4,000 in males and over 3,000 in females.

The characteristics of a mole that may indicate the occurrence of melanoma are summarized in the acronym **ABCDE**:

- A → Asymmetry in shape;
- B → irregular and indistinct Borders;

- C \rightarrow variable Color;
- D \rightarrow increasing Diameter;
- E \rightarrow Evolution of the mole which shows changes in appearance.

Anyway, quick detection of these tumors is not easy, as they do not give rise to symptoms in the early stages. Only when they become large enough and are in a more advanced stage do they begin to give problems such as

bleeding or causing itching and pain. While dermatologists are trained to identify skin cancers through visual examination, there are potential benefits to using automated image classification methods to aid in the diagnosis process.

In this project, we used three popular machine learning models for image classification to distinguish between malignant and benign tumors in skin cancer images.

3 Dataset and proposed method

The dataset is fetched from kaggle, which in turn fetched it from the ISIC (International Skin Image Collaboration) Archive.

It consists of 1800 pictures of benign moles and 1497 pictures of malignant classified moles.

It was already divided into train set and test set: the test set consists of 360 images of benign tumors and 300 images of malignant tumors, while the train set consists of 1440 images of benign tumors and 1197 of malignant tumors.

The pictures have all been resized to low resolution (224x224x3) RGB.

As mentioned in the previous section, three different machine learning models were used:

- CNN model \rightarrow is a type of deep learning model that is particularly well-suited for image classification tasks due to its ability to learn features from the input data automatically;
- SVM model \rightarrow is a supervised learning model that can be used for classification tasks by finding

the hyperplane that maximally separates the different classes in the data;

- The Naive Bayes classifier is a probabilistic model that makes predictions based on the probability of each feature belonging to a particular class.

Once we had chosen the methods we wanted to use, we followed the procedures below in order to arrive at results that were as accurate as possible:

- Import the data;
- Balance the dataset by increasing the number of malignant images through data augmentation which involved generating additional images from the existing images through random transformations such as rotation, scaling, and shifting. After that, we applied a data augmentation once again to all the dataset to increase the number of images available for training and improved the accuracy of the predictions model's;

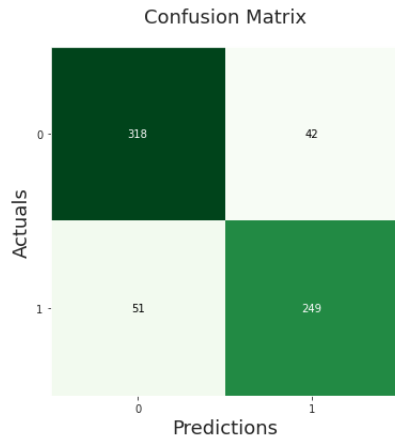
- Use k-fold cross validation to evaluate the performance of the three models. In this technique, the dataset is divided into k equal folds, and the model is trained and tested k times, each time using a different fold as the test set and the remaining folds as the training set. The final

model performance is calculated as the average of the performance across all k folds.

- Train and test the models using the preprocessed and balanced dataset: we calculated the accuracy of the models on the test set to evaluate their performance.

4 Results

The first results we are going to analyze are CNN's.



The accuracy of the model is relatively high at around 86%: this indicates that the model is making correct predictions for the majority of the instances in the dataset.

It's worth noting that the model appears to be performing slightly better on class 0 (benign) than on class 1 (malignant).

This could be due to a variety of factors that can be summarized as difficulty of the task, because it's generally more difficult to classify malignant tumors than benign ones since malignant tumors tend to have more subtle and

varied features that can be harder to discern, and lack of capacity, because it may be that the CNN is not powerful enough (e.g., it has an insufficient number of layers or filters), and as such it may not be able to learn the complex features needed to accurately classify malignant tumors.

Furthermore, observing the classification report it looks like the CNN model is performing well in terms of both precision and recall.

	precision	recall	f1-score	support
0	86.18%	88.33%	87.24%	360.00
1	85.57%	83.00%	84.26%	300.00
accuracy	85.91%	85.91%	85.91%	0.86
macro avg	85.87%	85.67%	85.75%	660.00
weighted avg	85.90%	85.91%	85.89%	660.00

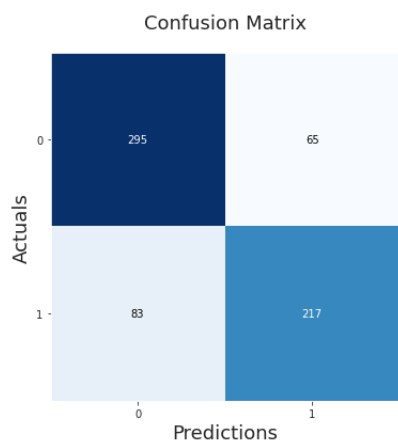
The f1-score, which is a measure of the balance between precision and recall, is also relatively high.

This suggests that, in general, the model is doing a good job of correctly identifying both positive and negative instances.

Let us now turn to the analysis of what happens using the SVM model.

The SVM model has a score of 0.78,

which means that it is making correct prediction for 78% instances of the dataset, which can be considered a good accuracy score. The confusion matrix shows that the SVM model made 295 correct predictions for class 0 and 217 correct predictions for class 1.



Again we can see that the model makes more correct than incorrect predictions for both classes, although, just as in the case of the CNN, it appears to be quite better at predicting class 0 than class 1.

This observation is confirmed by the fact that, for class 0, the indices reported in the classification report are higher, particularly recall, which represents the proportion of actual positive that was identified correctly.

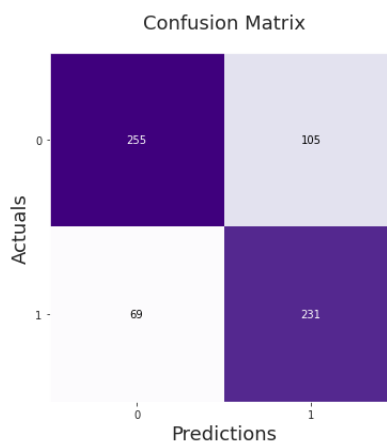
	precision	recall	f1-score	support
0	78.04%	81.94%	79.95%	360.00
1	76.95%	72.33%	74.57%	300.00
accuracy	77.58%	77.58%	77.58%	0.78
macro avg	77.50%	77.14%	77.26%	660.00
weighted avg	77.55%	77.58%	77.50%	660.00

To conclude, we go on to analyze

the accuracy obtained using the Naive Bayes classifier.

In this case we have a score of 74%, which is lower than both previous models.

According to the confusion matrix, the Naive Bayes model made a total of 255 and 231 correct predictions for class 0 and class 1, respectively.



The real difference of this model from the other two can be seen by looking at the *recall*:

	precision	recall	f1-score	support
0	78.70%	70.83%	74.56%	360.00
1	68.75%	77.00%	72.64%	300.00
accuracy	73.64%	73.64%	73.64%	0.74
macro avg	73.73%	73.92%	73.60%	660.00
weighted avg	74.18%	73.64%	73.69%	660.00

In fact, the Naive Bayes classifier is the only one that has a higher percentage of malignant tumors classified correctly (as opposed to CNN and SVM, which classify benign ones better).

It's not uncommon for different machine learning models to have different strengths and weaknesses, and this can sometimes lead to different models

performing better or worse on different tasks.

To improve the performance of Naive Bayes classifier on benign tumors and of CNN or SVM on malignant tumors, various approaches can be attempted, such as refining the hyperparameters of each model or combining the strengths of both models using a hybrid approach, such as, for example, using the Naive Bayes classifier as a feature extractor for CNN.

So, we concluded that, in general, CNN remains the best method to accomplish our goal of skin cancer classification.

However, why in image classification does the accuracy of CNN come higher than one of SVM model? And why is the accuracy of the Naive Bayes classifier also lower than the SVM?

Convolutional neural networks are a type of deep learning neural network that are particularly well-suited for image classification tasks.

They are designed to process data with a grid-like topology, such as an image, and are able to learn hierarchies of features through the use of convolutional layers.

This allows them to learn more com-

plex features of the input data and can lead to higher accuracy in image classification tasks compared to other machine learning models, such as support vector machines (SVMs) or Naive Bayes classifiers.

On the other hand, SVMs work by finding the hyperplane in a high-dimensional space that maximally separates different classes. While they can be effective for certain types of classification tasks, they may not be as effective as CNNs for image classification tasks because they do not have the ability to learn hierarchical features from the data.

As for the naive Bayes classifier, it is a probabilistic classifier based on the Bayes theorem with the assumption of independence between features.

It is a relatively simple and fast model that is often used in classification tasks, but it tends to have lower accuracy compared to more complex models such as CNNs or SVMs. This is because the assumption of independence between features is often not valid in real-world data, and the model may not be able to capture the complex relationships between features that can affect the classification.

5 Conclusions

The results obtained are proof of how possible cancer recognition is through machine learning although there are still many challenges, however, to make these algorithms a complete and reliable solution. One of the main problems is that the currently available artificial intelligence models are mainly based only on pa-

tients' imaging data, ignoring other important information such as ethnicity, skin, disease, existing sun damage, lifestyle habits (such as sun exposure, smoking and alcohol intake) and patients' medical history. To get to a point where artificial intelligence can "replace" (or be of great help to) the physician in making such a diag-

nosis, it will be important to work to improve existing artificial intelligence solutions and increase the diagnostic accuracy of the methods used to diagnose skin cancer. If we can overcome these challenges, artificial intelligence can become an important tool for providing accurate diagnosis of skin cancer in a cost-effective and accessible way, even remotely.

6 Project Link

This is the link to our colab notebook:

<https://colab.research.google.com/drive/1MSN3LDGG2cJhLy9je1FSdBPphX-ybULc?usp=sharing>

References

- [1] <https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign>
- [2] The lancet Regional Health – Europe, “Changing epidemiology and age-specific incidence of cutaneous malignant melanoma in England: An analysis of national cancer registration data by age, sex, and anatomical site, 1981-2018”, Volume 2, 100024, March 01, 2021.
- [3] Carlos Ricotti, MDa , Navid Bouzari, MDb , AmarAgadi, MDc , Clay J. Cockerell, “Malignant skin Neoplasms”, 2009 Elsevier Inc.
- [4] Istituto Superiore di Sanità, “Melanoma”, <https://www.epicentro.iss.it/melanoma/>
- [5] Ann Pietrangelo in Healthline, “What is the ABCDE rule for detecting skin cancer?”, February 9, 2022.
- [6] Hekler, A., Utikal, J.S., Enk, A.H., Hauschild, A., Weichenthal, M., Maron, R.C., Berking, C., Haferkamp, S., Klode, J., Schadendorf, D., “Superior skin cancer classification by the combination of human and artificial intelligence.” Eur. J. Cancer 2019, 120, 114–121.
[https://www.ejancer.com/article/S0959-8049\(19\)30427-7/fulltext](https://www.ejancer.com/article/S0959-8049(19)30427-7/fulltext)
- [7] Humayun, M., Sujatha, R., Almuayqil, S.N., Jhanjhi, N.Z., “A Transfer Learning Approach with a Convolutional Neural Network for the Classification of Lung Carcinoma.” Healthcare 2022, 10, 1058.
<https://www.mdpi.com/2227-9032/10/6/1058>
- [8] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018), 22 June 2018
- [9] Wang M, Li C, Zhang W, Wang Y, Feng Y, Liang Y, Wei J, Zhang X, Li X and Chen R, ”Support Vector Machine for Analyzing Contributions of Brain Regions During Task-State fMRI”, march 06, 2019.