

Skin Cancer Classification Using Inception Network and Transfer Learning

Priscilla Benedetti[1(✉)], Damiano Perri[2(✉)], Marco Simonetti[2(✉)], Osvaldo Gervasi[3(✉)], Gianluca Reali[1(✉)], and Mauro Femminella[1(✉)]

[1] Department of Engineering, University of Perugia, Perugia, Italy
pris.benedetti92@gmail.com,{gianluca.reali,mauro.femminella}@unipg.it
[2] Department of Mathematics and Computer Science, University of Florence, Florence, Italy
damiano.perri@gmail.com,marco.simonetti@libero.it
[3] Department of Mathematics and Computer Science, University of Perugia, Perugia, Italy
osvaldo.gervasi@unipg.it

Abstract. Medical data classification is typically a challenging task due to imbalance between classes. In this paper, we propose an approach to classify dermatoscopic images from HAM10000 (Human Against Machine with 10000 training images) dataset, consisting of seven imbalanced types of skin lesions, with good precision and low resources requirements. Classification is done by using a pretrained convolutional neural network. We evaluate the accuracy and performance of the proposal and illustrate possible extensions.

Keywords: Machine learning · Convolutional neural network · Keras · TensorFlow

1 Introduction

Training of neural networks for automated diagnosis of pigmented skin lesions can be a difficult process due to the small size and lack of diversity of available datasets of dermatoscopic images. The HAM10000 ("Human Against Machine with 10000 training images") dataset is a collection of dermatoscopic images from different populations acquired and stored by different modalities. We used the benchmark dataset, with a small number of images and a strong imbalance among the 7 different types of lesions, to prove the validity of our approach, which is characterized by good results and light usage of resources.

Exploiting a highly engineered convolutional neural network with transfer learning, customized data augmentation and a non-adaptive optimization algorithm, we show the possibility of obtaining a final model able to precisely recognize multiple categories, although scarcely represented in the dataset. The whole training process has a limited impact on computational resources, requiring no more than 20 GB of RAM space. The rest of paper is structured as

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12249, pp. 536–545, 2020.
https://doi.org/10.1007/978-3-030-58799-4_39

follows: Sect. 2 describes the related work in the field of medical images processing. Section 3 illustrates the dataset of interest. Section 4 gives an overview of the model architecture. Section 5 includes the training process and shows experimental results. Finally, some final comments and future research directions are reported in Sect. 6.

2 Related Work

Processing of biomedical images has always been a field strongly beaten by CNN pioneers. The first related papers date back to 1991 [1], with a strong impulse in the following years in the search for methods for automating the classification of pathologies and related diagnosis [2,3].

Nowadays, almost thirty years later, reliability of networks reached a rather high level, as well as intrinsic complexity. This reliability allowed a wide diffusion of the approach of subjecting diagnostic images to automatic classification systems, from evolutionary algorithms [4–7] to deep networks [8–11], being them either convolutive or not. Even in the medical sector of dermatology, automatic image recognition and classification was used for decades to detect tumor skin lesions [12,13].

Recent and promising research has highlighted the possibility that properly trained machines can exceed the human recognition and classification capability to recognize skin cancers. The scores obtained are very encouraging [14] and we are confident that in the near future the recognition capacity of these forms of pathologies will become almost total.

Today CNNs are used for image feature extraction. Features are used for image classification [15,16].

3 The Dataset

Dermatoscopy is often used to get better diagnoses of pigmented skin lesions, either benign or malignant. With dermatoscopic images is also possible to train artificial neural networks to recognize pigmented skin lesions automatically. Nevertheless, training requires the usage of a large number of samples, although the number of high quality images with reliable labels is either limited or restricted to only a few classes of diseases, often unbalanced.

Due to these limitations, some previous research activities focused on melanocytic lesions (in order to differentiate between a benign and malignant sample) and disregarded non-melanocytic pigmented lesions, even if very common. In order to boost research on automated diagnosis of dermatoscopic images, HAM10000 has been providing the participant of the ISIC 2018 classification challenge, hosted by the annual MICCAI conference in Granada, Spain [17], specific images.

The set of 10015 8-bit RGB color images were collected in 20 years from populations from two different sites, specifically the Department of Dermatology of the Medical University of Vienna, and the skin cancer practice of Cliff

Rosendahl in Queensland. Relevant cases include a representative collection of all important diagnostic categories of pigmented lesions [17]:

- **akiec:** Actinic Keratoses and Intraepithelial Carcinoma, common noninvasive variants of squamous cell carcinoma that can be treated locally without surgery. [**327 images**]
- **bcc:** Basal cell carcinoma, a cancer that rarely metastasizes but grows destructively if untreated. [**514 images**]
- **bkl:** Generic label that includes seborrheic keratoses, solar lentigo and lichen-planus like keratoses (LPLK), which corresponds to a seborrheic keratosis or a solar lentigo with inflammation and regression, often mistaken for melanoma. [**1099 images**]
- **df:** Dermatofibroma, a benign skin lesion. [**115 images**]
- **nv:** Melanocytic nevi are benign neoplasms of melanocytes. [**6705 images**]
- **mel:** Melanoma, if diagnosed in an early stage, it can be cured by simple surgical excision. [**1113 images**]
- **vasc:** Vascular skin lesions in the dataset range from cherry angiomas to angiokeratomas and pyogenic granulomas. [**142 images**]

More than 50% of lesions are confirmed through histopathology (histo), the ground truth for the rest of the cases is either follow-up examination (followup), expert consensus (consensus), or confirmation by in-vivo confocal microscopy (confocal). Other features in the individual dataset include age, gender and body-site of lesion (localization) [17].

4 Model Architecture

Since their first appearance in Le Cun *et al.* publication [18], Convolutional Neural Networks (CNN) have been widely applied to data that have a known, grid-like structure. Possible set of interests are time-series data, which can be modeled as a 1D grid taking samples at regular time intervals, and image data, which can be thought of as a 2D grid of pixels. The foundational layer of a convolutional network consists of three stages. In the first stage, the layer performs several parallel convolutions to produce a set of linear activations. In the second stage, each convolution output is run through a nonlinear activation function, such as the rectified linear activation function (ReLU). In the third stage, a pooling function is used to modify the output of the layer further [19]. The max pooling operation, used in this work, reports the maximum output within a rectangular neighborhood. Since the objective images of skin lesions present great variation in size, we decided to use a network made by inception modules, which make use of filters of different size operating at the same level. This usage of wide modules with multiple cheap convolutional operations entails a reduced computational complexity with respect to a deep network with large convolutional layers. In the specific model we used, based on Inception-ResNet-v2, another point of speed improvement is the introduction of residual connections, which replace pooling operations within the main inception modules. However, the previously

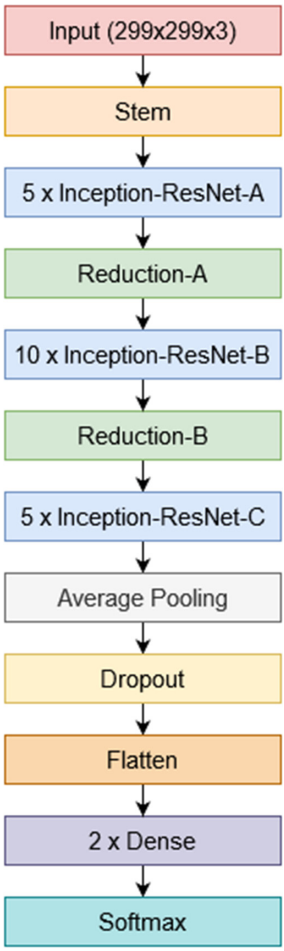

Fig. 1. General schema for the used network, an Inception-ResNet-v2 architecture with the addition of a flattening layer, 2 fully-connected layers and a final softmax activation.

mentioned max pooling operations are still present in the reduction blocks. The structure of the network used in this work is shown on Fig.1

The original Inception-ResNet-v2 architecture [20] has a stem block consisting of the concatenation of multiple convolutional and pooling layers, while Inception-ResNet blocks (A, B and C) contain a set of convolutional filters with an average pooling layer. As prevously mentioned, reduction blocks (A, B) replace the average pooling operation with a max pooling one. This structure has been extended with a final module consisting of a flattening step, two fully-connected layers of 64 units each, and the softmax classifier. The overall module is trainable on a single GPU with reduced memory consumption.

5 Training Process and Experimental Results

This work consists of two training rounds, after a step of data processing in order to deal with the strong imbalance of the dataset:

- A first classification training process using class weights.
- Rollback of previous obtained best model to improve classification performance with a second training phase.

5.1 Data Processing

In the first stage of data processing, after the creation of a new column with a more readable definition of labels, each class was translated into a numerical code using *pandas.Categorical.codes*. Afterwards, missing values in "age" column was filled with column mean value. Figure 2 and Fig. 3 show the HAM10000 data distribution.

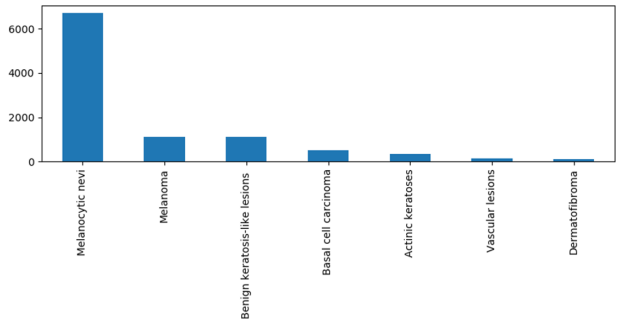

Fig. 2. Plotting the frequency of each class, the imbalance between Melanocytic Nevi and the rest of the possible categories is manifest.

Finally, images are loaded and resized from 450×600 to 299×299 in order to be correctly processed by the network. After a normalization step on RGB arrays, we split the dataset into a training and validation set with 80:20 ratio (Fig. 4).

In order to re-balance the dataset, we chose to shrink the amount of images for each class to an equal maximum dimension of 450 samples. This significant decrease of available images is then mitigated by applying a step of data augmentation. Training set expansion is made by altering images with small transformations to reproduce some variations, such as horizontal flips, vertical flips, translations, rotations and shearing transformations.

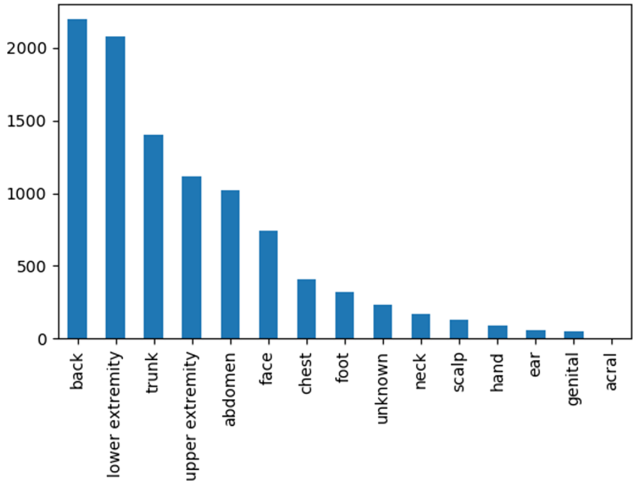

Fig. 3. Plot of samples' body locations.

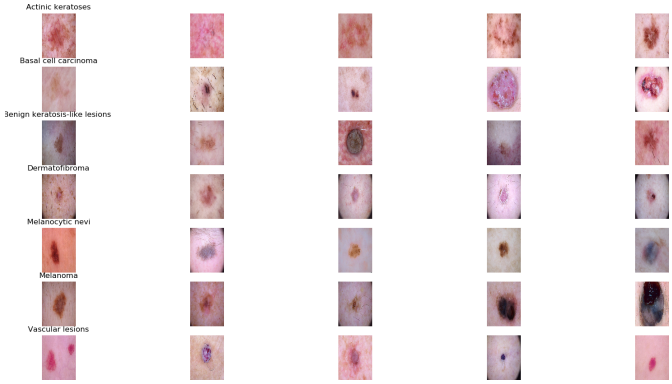

Fig. 4. Some HAM10000 images resized with OpenCV.

5.2 Baseline

Due to the limited number of samples for the training process, we decided to take advantage of transfer learning, utilizing Inception-ResNet-v2 pre-trained on ImageNet [21] and Tensorflow, a deep learning framework developed by Google, for fine-tuning of the last 40 layers. Keras library offers a wide range of optimizers: Adaptive optimization methods such as AdaGrad, RMSProp, and Adam are widely used for deep neural networks training due to their fast convergence times. However, as described in [22], when the number of parameters exceeds the number of data points these optimizers often determine a worse generalization capability compared with non-adaptive methods. In this work we used a stochastic gradient descent optimizer (SGD), with learning rate set to 0.0006

and usage of momentum and Nesterov Accelerated Gradient in order to adapt updates to the slope of the loss function (categorical crossentropy) and speed up the training process. The total number of epochs was set to 100, using a small batch size 10. A set of class weight was introduced in the training process to get more emphasis on minority class recognition. A maximum patience of 15 epochs was set to the early stopping callback in order to mitigate the overfitting visible in Fig. 5, which shows the history of training and validation process.

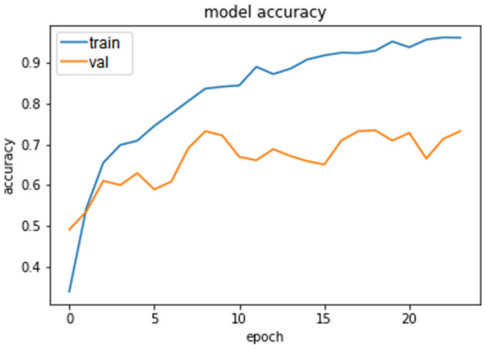

Fig. 5. Accuracy and Loss for each epoch.

Finally, the model achieves an accuracy of 73.4% on the validation set, using weights from the best epoch. Figure 6 shows the confusion matrix for the model on the validation set.

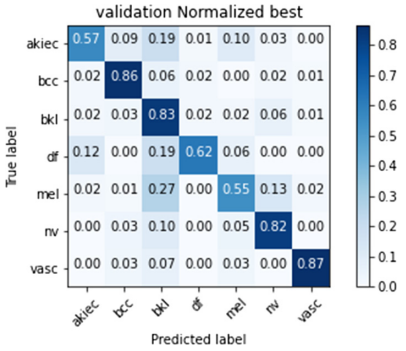

Fig. 6. Two of the minority classes, Actinic Keratoses (akiec) and Dermatofibroma (df), are not properly recognized. Melanoma (mel) is often mistaken for generic keratoses (bkl), as already mentioned in Fig. 3

5.3 Resuming Training from the Best Epoch

In order to improve classification performance, specially on minority classes, we loaded the best model obtained in the first round to extend the training phase and explore other potential local minimum points of the loss function, by using an additional amount of 20 epochs. This second step led to an enhancement in overall predictions, reaching the maximum accuracy value of 78.9%.

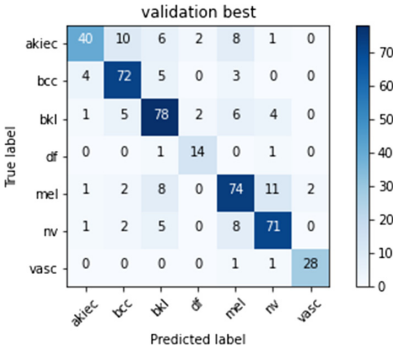

Fig. 7. Final confusion matrix.

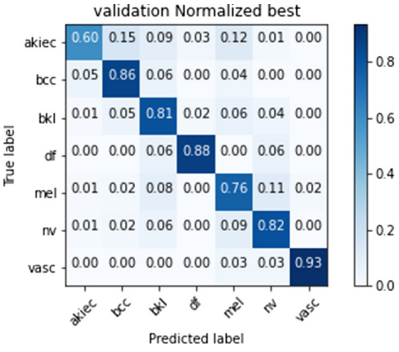

Fig. 8. Normalized final confusion matrix.

Figure. 7 shows the normalized confusion matrix on the validation set for the final fine-tuned model. In this case, 6 out of 7 categories are classified with a total ratio of True Positives higher than 75%, even in presence of extremely limited sample set, as vascular lesions (vasc), 30 samples, and dermatofibroma (df), 16 samples. The whole process of training has required less than four hours on Google Colab cloud's GPU, for an overall RAM utilization below 20 GB (Fig. 8).

6 Conclusions and Future Works

In conclusion, in this paper we investigate the possibility of obtaining improved performances in the classification of 7 significantly unbalanced different types of skin diseases, with a small amount of available images. With use of a fine-tuned deep inception network, data augmentation and class weights, the model can achieve a good final diagnostic accuracy. The described training process has a light resource usage, requiring less than 20 GB of RAM space, and it can be executed in a Google Colab notebook. For future improvements larger datasets of dermatoscopic images are needed. The model shown in this paper can be regarded as a starting point to implement a lightweight diagnostic support system for dermatologists, for example in the Web as well as through a mobile application.

References

1. Zhang, W., Hasegawa, A., Itoh, K., Ichioka, Y.: Image processing of human corneal endothelium based on a learning network. Appl. Opt. **30**, 4211–4217 (1991)
2. Zhang, W., Doi, K., Giger, M.L., Wu, Y., Nishikawa, R.M., Schmidt, R.A.: Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network. Med. Phys. **21**(4), 517–524 (1994)
3. Vella, F., Neri, I., Gervasi, O., Tasso, S.: A simulation framework for scheduling performance evaluation on CPU-GPU heterogeneous system. In: Murgante, B., et al. (eds.) ICCSA 2012. LNCS, vol. 7336, pp. 457–469. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-31128-4_34
4. Chen, J.-C., Yeh, C.-M., Tzeng, J.-E.: Pattern differentiation of glandular cancerous cells and normal cells with cellular automata and evolutionary learning. Exp. Syst. Appl. **34**(1), 337–346 (2008)
5. Guo, P.F., Bhattacharya, P.: An evolutionary approach to feature function generation in application to biomedical image patterns (2009)
6. Gervasi, O., Russo, D., Vella, F.: The AES implantation based on OpenCL for multi/many core architecture. In: 2010 International Conference on Computational Science and Its Applications, Fukuoka, ICCSA 2010, pp. 129–134. IEEE Computer Society, Washington (2010). https://doi.org/10.1109/ICCSA.2010.44
7. Mariotti, M., Gervasi, O., Vella, F., Cuzzocrea, A., Costantini, A.: Strategies and systems towards grids and clouds integration: a DBMS-based solution. Fut. Gener. Comput. Syst. **88**, 718–729 (2018). https://doi.org/10.1016/j.future.2017.02.047
8. Pang, S., Yu, Z., Orgun, M.A.: A novel end-to-end classifier using domain transferred deep convolutional neural networks for biomedical images. Comput. Methods Prog. Biomed. **140**, 283–293 (2017)
9. Zhou, Z., Shin, J., Zhang, L., Gurudu, S., Gotway, M., Liang, J.: Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally (2017)
10. Gervasi, O., et al. (eds.): ICCSA 2016. LNCS, vol. 9787. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42108-7
11. Gervasi, O., et al. (eds.): ICCSA 2017. LNCS, vol. 10406. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62398-6

12. Lau, H.T., Al-Jumaily, A.: Automatically early detection of skin cancer: study based on nueral netwok classification(2009)
13. Dorj, U.-O., Lee, K.-K., Choi, J.-Y., Lee, M.: The skin cancer classification using deep convolutional neural network. Multimedia Tools Appl. **77**(8), 9909–9924 (2018). https://doi.org/10.1007/s11042-018-5714-1
14. Maron, R.C., et al.: Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. Eur. J. Cancer **119**, 57–65 (2019)
15. Biondi, G., Franzoni, V., Gervasi, O., Perri, D.: An approach for improving automatic mouth emotion recognition. In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11619, pp. 649–664. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24289-3_48
16. Perri, D., Sylos Labini, P., Gervasi, O., Tasso, S., Vella, F.: Towards a learning-based performance modeling for accelerating deep neural networks. In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11619, pp. 665–676. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24289-3_49
17. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data. **5**, 180161 (2018). https://doi.org/10.1038/sdata.2018.161
18. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998). https://doi.org/10.1109/5.726791
19. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT press, Cambridge (2016)
20. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning (2016)
21. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: 25th International Conference on Advance in Neural Information Processing System, pp. 1106–1114 (2012)
22. Wilson, A., Roelofs, R., Stern, M., Srebro, N., Rech, B.: The marginal value of adaptive gradient methods in machine learning. In: NIPS 2017 (2017)