Convolutional Neural Networks for mammogram classification in BIRADS standard: development and preliminary tests

Francesca Lizzi[1,2], **Paolo Bosco**[1], **Davide Caramella**[2,3], **Carolina Marini**[3],

Alessandra Retico[1], **Antonio Traino**[3], **Maria Evelina Fantacci**[1,2]
[1] National Institute for Nuclear Physics (INFN), Pisa, Italy
[2] University of Pisa, Pisa, Italy
[3] Azienda Ospedaliero-Universitaria Pisana (AOUP), Pisa, Italy
francescalizzi1@gmail.com

Abstract

The aim of our project is to transpose the European Directive 59/2013/EURATOM which regulates the use of ionizing radiation in medicine and the information provided to patients on radiation dose. Controlling the dose is a really important aspect when we deal with breast cancer screening because breast is made of a radio-sensitive tissue and screening programs expose to radiation a lot of healthy women. In order to personalize the dose index, it is necessary to take into account breast density because to have a sufficient sensitivity a higher dose is given to the patient. For this reason we propose to classify mammograms with a Convolutional Neural Network (CNN) in BIRADS standard. To assess density in screening program, an automatic algorithm is necessary because of the large number of women who partecipates.

1 Introduction

Breast cancer is one of the most diagnosed and fatal cancer all over the world. The strongest weapons we have against it are prevention and early diagnosis. Mammography is still the most used imaging system in screening programs but it suffers of some difficulties, especially when we deal with dense breast [1]. Mammographic density is the amount of white radiographic appearance of dense tissue as seen on a mammogram [2]. In the Fifth Edition of BIRADS (Breast Imaging-Reporting and Data System), four density classes are defined by guideline based on text descriptions [3]. The RADIOMA project has the goal of developing and implementing a new personalized dose index according to European Directive 59/2013/EURATOM [4]. In order to achieve this goal the assessment of density is necessary. We chose as main instrument a convolutional neural network (CNN). CNNs are usually used to analyze natural images but in the very last year their application on medical images has been very promising [5]. To have enough data, with the AOUP's team, we are collecting a large (8000 images) dataset of negative mammograms. To perform preliminary analysis, we used a public research database named Mammogrid [6], which contains digital images. Subsequently, a CNN for density classification has been developed and preliminary results are reported in this abstract.

2 Data and methods

CNNs have been trained in order to solve a simple classification problem on four classes on Mammogrid database [6] and two of the density classes on the AOUP daataset. The chosen classes are

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

defined by the laterality and the projection of mammograms. So the four classes are: cranio-caudal projections on left and right breast and medio-lateral oblique projections on left and right breast, as shown in Figure 1.

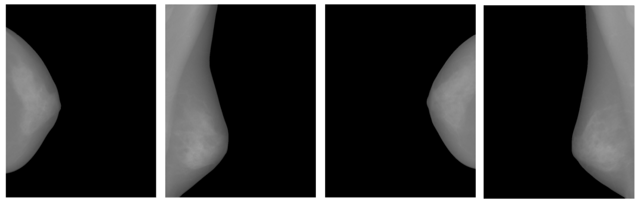

Figure 1: Samples of the training set for each class

108 images of 28 subjects were available. Among the 108 images, we put 88 of them in training and the others in validation. The CNN has been trained on a Toshiba Satellite on CPU (4th Generation Intel® Core™ i7-4720HQ Processor). The low number of images and the limited computing power initially available strongly influenced the architecture of the network and the size of input images. The network has been implemented with Keras 2.0, using Python 2.7 with Tensorflow on backend. The chosen architecture for this CNN is inspired by the VGG19 one [7]. The advancement brought by this CNN is the stacking of convolutional layers. In fact, in VGG19 three convolutional layers with a little receptive field (3x3) are stacked in order to have a larger effective receptive field (7x7) with less parameters. The loss function is a categorical crossentropy and the network was trained over 50 epochs. The optimizer is "rmsprop" and it is an adaptive learning rate method proposed by Geoffrey Hinton [8]. As regularization method, data augmentation and dropout have been chosen. Dropout consists in randomly setting a fraction rate of input units to 0 at each update during training time in order to help in preventing overfitting. Learning rate of the optimizers has been set at 0.01 and it decreases over epochs with a decay rate equal to 0.9. Furthermore, if the validation loss assumes the same value for three consecutive epochs, learning rate will be further decreased by a factor 5 until a minimum of 0.0001. The architecture is composed of: two 2D-convolutional layers with 32 kernels of 3x3 dimension with stride 1 and the input size was fixed to 50x50x1. The activation functions are LeakyReLu. The following layer is a dropout layer with a dropout rate of 0.1, since the low number of images. It is followed by a max pooling layer on matrices of size 2x2. The pooling layer is followed by two 2D-convolutional layers with 32 kernels of 3x3 size, a dropout layer and another max pooling. The activation functions are ReLu. The flatten layer flattens the input into a vector and it is followed by the first fully connected layer. The last layer is fully connected and its number of neurons is equal to the number of classes, i.e. four. The activation function for the last layer is a sigmoid.
The dataset used for density classification problem consists in 73 right cranio-caudal mammograms belonging to density class A and 57 right cranio-caudal mammograms belonging to density class A extracted from the first mammograms made available of the dataset we are collecting. Since calculus limitation, we used only one of the four projections, i.e. the right cranio-caudal, and only A and D classes to train the CNN. Among them, we put 63 mammograms classified "A" and 47 mammograms classified "D" in the training set. The others 20 images were moved in validation set. Since the first CNN worked enough well, the CNN for density classification has been built equal to the first one except for the number of neurons of the last layer, which is equal to two.

3 Results

The trend of training loss and training accuracy for projection classification is reported in Figure 2. The red line represents the loss function and the value at last epoch is a minimum and it is equal to 0.0107. The green line represents instead the accuracy, i.e. how well the classifier works on training set. Accuracy goes to 1 during the 37th epoch and this means that this CNN works with a precision of 100% both on training and validation set. The training requires about 8 seconds per epoch.

In Figure 3, the trend of training loss and training accuracy for density classification is shown. The red line represents the loss function: the value at last epoch is a minimum and it is equal to 0.2732. The green line represents the accuracy: accuracy goes to 0.8852 during the 37th epoch and this means that this CNN works with a precision of 88.52% on the training set. The mean duration of a epoch is about 12 seconds.

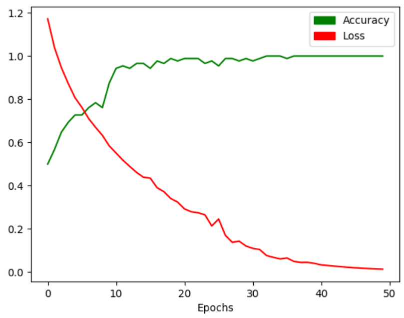

Figure 2: Loss and accuracy over epochs for projection classification

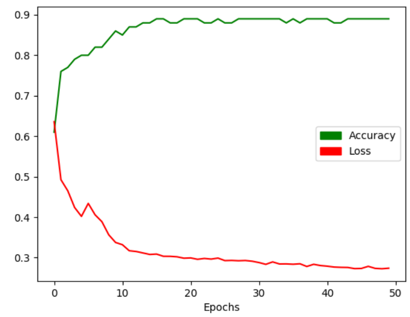

Figure 3: Loss and accuracy over epochs for density classification

4 Discussion

The CNN used to classify the projections works in an excellent way and it classifies perfectly both training and validation set. This very simple problem is not only solved with a high accuracy but also in a very little time, i.e. 400 seconds. The second classifier works well but with an accuracy equal to 88% on training set and its training over the same number of epochs requires more time than the first one. This is because the assessment of density is not a so simple problem to solve. The strong downsampling of the images, the use of only one of the four projection and the architecture choice are important constraints that can be overpass easly with a higher calculus power. Finally, the use of a more powerful hardware and of the whole dataset, that we are currently doing, could led to a well-working classifier even on the BIRADS classification problem.

References

[1] Magnus Løberg, Mette Lise Lousdal, Michael Bretthauer, and Mette Kalager. Benefits and harms of mammography screening. *Breast Cancer Research*, 17(1), December 2015.

[2] Kavitha Krishnan, Laura Baglietto, Jennifer Stone, Julie A. Simpson, Gianluca Severi, Christopher F. Evans, Robert J. MacInnis, Graham G. Giles, Carmel Apicella, and John L. Hopper. Longitudinal Study of Mammographic Density Measures That Predict Breast Cancer Risk. *Cancer Epidemiology Biomarkers & Prevention*, 26(4):651–660, April 2017.

[3] EA Sickles, CJ D'Orsi, LW Bassett, and et al. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System, 2013.

[4] Council Directive 59/2013/EURATOM of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionizing radiation, January 2014.

[5] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, December 2017.

[6] R. Warren, A.E. Solomonides, C. del Frate, I. Warsi, J. Ding, M. Odeh, R. McClatchey, C. Tromans, M. Brady, R. Highnam, M. Cordell, F. Estrella, M. Bazzocchi, and S.R. Amendolia. MammoGrid — a prototype distributed mammographic database for Europe. *Clinical Radiology*, 62(11):1044–1051, November 2007.

[7] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[8] Geoffrey E. Hinton. http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.