CLASSIFY X-RAY IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

Federica Crobu[1] and Agostino Di Ciaccio[2]

[1] Sapienza Università di Roma, (e-mail: federicacrobu@gmail.com)

[2] Department of Statistics, Sapienza Università di Roma,
(e-mail: agostino.diciaccio@uniroma1.it)

ABSTRACT: In recent years, computer-assisted diagnostic systems have gained increasing interest through the use of deep learning techniques. In this work we show how it is possible to classify X-ray images through a multi-input convolutional neural network. The use of clinical information together with the images allowed to obtain better results than those present in the literature on the same data.

KEYWORDS: deep learning, convolutional neural networks, chest x-ray, multi-input neural networks.

1 Introduction

In recent years, the deep neural networks (DNN) and in particular the deep convolutional neural networks (DCNN) have attracted the attention of the researchers for their great ability to analyse images. One of the most fascinating and advantageous branches for the application of these models is medicine. Thanks to them we can now imagine a future in which doctors are helped by computers to recognize diseases and make diagnoses. Furthermore, it could be a drastic improvement in the underdeveloped countries where the availability of doctors is problematic and pathologies such as pneumonia are still one of the main causes of death.

In the classical context of image recognition, the goal is to classify what is contained in an image, however, in the analysis of medical images, the challenge is quite different. In fact, to emulate the role of the doctor, the model needs much more information that cannot be deduced from the analysis of radiographic images only. Therefore, it is also necessary to consider many other information collected on the patients such as clinical and demographic details.

The correlation of certain pathologies with age or smoking is well known, for example. Other diseases may have genetic predispositions and many diseases can be related to each other. Usually, doctors can obtain and use all this information and it is advantageous to provide them also to the predictive model.

From the technical point of view, the goal of including more inputs of different nature, images and numerical values, has been achieved using a multi-input neural

network architecture. Through this model we were able to obtain a very accurate classification, as it is shown in the following sections.

2 The X-ray data and the previous works

The availability of large medical databases containing both images and clinical information is scarce. Currently, the largest database is the ChestX-ray14, chosen for this application. It was released by the United States National Institutes of Health (NIH) and contains over 112,000 radiographic frontal chest images of 30,805 patients. Each of them can be healthy or sick, with one or more of the following 14 diseases: Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Fibrosis, Hernia, Infiltration, Mass, Nodule, Pleural thickening, Pneumonia, Pneumothorax. Furthermore, a "no finding" category represents the images in which none of the previously mentioned diseases have been detected.

As can be understood by analysing the database, for many patients are available multiple results of the tests, which can be useful for capturing the progress of diseases over time.

The labels, corresponding to the pathologies identified in each image, were extracted from radiological reports using natural language processing techniques with an accuracy that is declared by authors over 90% (Rajpurkar et al. 2017). Therefore, we cannot fully trust the labelling process and, furthermore, some researchers have raised many doubts about the correctness of the labels. The criticism of the radiologist Luke Oakden-Rayner (2017) that, after observing the images, states that there are incorrect labels. Finally, it should be noted that many diagnoses present more concomitant diseases.

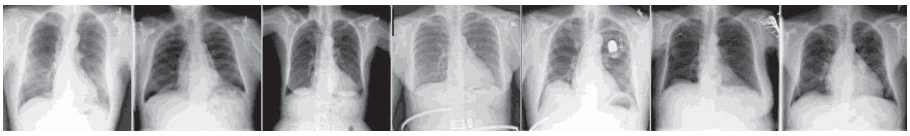

Fig.1 – Some images of the database ChestX-ray14

This dataset has been already used by many other researchers. Surely, the best-known work was made by a Stanford's team (Rajpurkar et al. 2017). They proposed an architecture called *CheXNet* based on the usage of the DCNN architecture called *DenseNet121* (Huang et al. 2017). This work represents, at this moment, the state-of-the-art results in terms of AUC scores.

Other important works are the one of Yao et al. (2017) and the one of Wang et al. (2017). The first is mainly based on an architecture consisting of a DenseNet as encoder and on a recurrent neural network as decoder. Wang tries to apply some of the most famous CNN architectures (excluding DenseNet), achieving the best results with ResNet-50. Other interesting and more recent works are the ones of Baltrushat (2018) and Guendel (2018). Baltrushat based his work on a ResNet-50 to analyze the images, supported by the use of age, gender and view position.

3 The proposal and the results

Inspired by the work of the Stanford team, we decided to improve the model by exploiting the few clinical and demographic information available with these images. We have considered age, sex, sight position and 14 new variables containing the patient's information obtained from previous pathological history present in the same data.

The goal was therefore to improve the DenseNet121 model with another parallel neural network, with two small dense layers (32 and 16 neurons), which processes the non-image characteristics. The two independent networks are then concatenated and connected to the output layer based on 14 neurons with sigmoid activation function, whose task is to estimate the probability of the presence of each disease in the X-ray image. The final network has a complex structure with 123 'main' layers and 7,053,182 parameters. We used the pretrained weights of DenseNet121 on Imagenet as initialization of the network.

To solve this multi-input multi-class problem, we have employed a weighted binary cross-entropy loss function with data augmentation.

Our results provide an interesting improvement of the state-of-the-art, confirming our intuition of the architecture's power. Following the literature, we have adopted the AUC index as main tool to evaluate the quality of the predictions. In the table 1 we can see the comparison of the performances of our model with the best results obtained by other researchers in terms of the mean AUC scores.

	Wang et al. (2017)	***Yao et al. (2018)***	***CheXNet (2017)***	***Our Multi-input***
Official split	Yes	No	No	Yes
Atelectasis	0.716	0.772	0.809	**0.816**
Cardiomegaly	0.807	0.904	0.925	**0.925**
Effusion	0.784	0.859	0.864	**0.867**
Infiltration	0.609	0.695	0.735	0.731
Mass	0.706	0.792	0.868	**0.897**
Nodule	0.671	0.717	0.780	**0.827**
Pneumonia	0.633	0.713	0.768	**0.776**
Pneumothorax	0.806	0.841	0.889	**0.927**
Consolidation	0.708	0.788	0.790	**0.801**
Edema	0.835	0.882	0.888	**0.893**
Emphysema	0.815	0.829	0.937	**0.946**
Fibrosis	0.769	0.767	0.805	**0.881**
Pleural Thickening	0.708	0.765	0.806	**0.827**
Hernia	0.767	0.914	0.916	**0.963**
Average	0.738	0.803	0.841	**0.863**

Table 1. Comparison of the AUC on test data.

We have chosen the subdivision suggested by the data authors, and we have also verified that the previously proposed approaches with different splits still obtain the same results with this subdivision. The size of the test-set on which the AUC was measured has dimensions greater than 25,000 and therefore guarantees a great stability of the results with respect to the possible subdivisions. It is evident in the table that the average AUC has been significantly improved by our approach and, for most classes, we have clearly outperformed previous jobs.

4 Conclusions

The results of this application have confirmed the validity of our approach: a multi-input neural network architecture can significantly improve predictions. Clearly, the idea of combining different sources of heterogeneous information can be applied in other fields of medicine, as in the analysis of MRI scans. Whenever the patient's clinical and/or demographic information is available, it is possible and fruitful to apply this approach. Similarly, this technique can be used in other application areas.

References

BALTRUSCHAT, I.M., NICKISCH, H., GRASS, M., KNOPP, T., SAALBACH, A. 2018. Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification. *arXiv:1803.02315*.

GUENDEL, S., GRBIC, S., GEORGESCU, B., ZHOU, K., RITSCHL, L., MEIER, A., COMANICIU, D. 2018. Learning to recognize abnormalities in chest x-rays with location-aware dense networks. *arXiv preprintarXiv:1803.04565*.

HUANG, G., LIU, Z., VAN DER MAATEN, L., WEINBERGER, K. Q. 2017. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, pp. 2261-2269.

OAKDEN-RAYNER, L. 2017. *Exploring the ChestXray14 dataset: problems*, https://lukeoakdenrayner.wordpress.com.

RAJPURKAR, P.,IRVIN, J., ZHU, K., YANG, B., MEHTA, H., DUAN, T., DING, D., BAGUL, A., BALL, R. L., LANGLOTZ, C., SHPANSKAYA, K., LUNGREN, M. P., NG, A. 2017. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning, *arXiv preprint arXiv:1711.05225*.

WANG, X., PENG, Y., LU, L., LU, Z., BAGHERI, M., SUMMERS, R. M. 2017. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 3462-3471.

YAO, L., POBLENZ, E., DAGUNTS, D., COVINGTON, B., BERNARD, D., LYMAN, K. 2017. Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv preprint arXiv:1710.10501*.