Designing Chest X-ray Datasets for Improving Lung Nodules Detection Through Convolutional Neural Networks

Raúl ACEÑERO EIXARCH [a,1], Raúl DÍAZ-USECHI LAPLAZA [b] and
Rafael BERLANGA [a]

[a] *Universitat Jaume I, Castellón, Spain*
[b] *Hospital General de Castellón, Spain*

Abstract. In this paper, we propose a method for building alternative training datasets for lung nodule detection from plain chest X-ray images. Our aim is to improve the classification quality of a state-of-the-art CNN by just selecting appropriate samples from the existing datasets. The hypothesis of this research is that high quality models need to learn by contrasting very clean images with those containing nodules, specially those difficult to identify by non-expert clinicians. Current chest X-ray datasets mostly include images where more than one pathology exist and/or contain devices like catheters. This is because most samples come from old people which are the usual patients subject to X-ray examinations. In this paper, we evaluate several combinations of samples from existing datasets in the literature. Results show a great gain in performance for some of the evaluated combinations, confirming our hypothesis. The achieved performance of these models allows a considerable speed-up in the screening of patients by radiologist.

Keywords. Convolutional Neural Networks, X-ray images, Lung Nodules Detection

Introduction

Radio-diagnosis is a low-cost and universally widespread method based on the analysis of X-ray images. Its main drawback is that it must be carried out by highly qualified people (i.e., radiologists) which are scarce in the public health systems. Thus, most X-ray images are directly delivered to doctors without supervision of radiologists. This motivates the use of automatic screening tools able to detect and send suspicious cases to radiologists. More specifically, these screening system must target to diseases like lung cancer because its early detection is crucial for patients survival.

The initial stage of the neoformative process of lung cancer, before becoming a lung mass, is represented on the chest X-ray as a pulmonary nodule, in most cases in a very subtle way. Even trained eyes can miss these findings, and once the pathology is diagnosed, it is usual that clues become more evident. Lung tumors present different cell lines, and present in different ways in the lung parenchyma. They may have a central

---

[1]Corresponding Author.

arrangement, being located in perihilar or even endobronchial topographies, as well as a more peripheral distribution. Initially, they will be represented as small pulmonary nodules, always smaller than 3cm, and once that figure is exceeded, they will become pulmonary masses. Densotomographically, they can present densities similar to those of soft tissue tissues. Nodules have poorly defined borders, with a poor definition with respect to the adjacent structures, adopting a morphology of "spiculate borders", which as the disease progresses, can condition the invasion of organs and structures adjacent to the tumor. Given its variability in morphologies and locations where the pulmonary nodule can settle, it is important to recognize its presentation patterns both to identify them and to differentiate them from other pathologies (metastases, abscesses, interstitial diseases and much more).

Currently, there are no automatic methods for early detection of lung tumors in X-ray images, as there are for other neoplasms (breast, prostate, colorectal, etc.) Most methods for nodule detection and classification are defined for computed tomography (CT) scans, which offer greater sensitivity and specificity than X-ray. However, CT scans require high doses of radiation, making the risk/benefit in a large population unfavorable.

1. Related Work

In this section we revise the main approach in dataset generation for automatic radio-diagnosis for lung nodule detection. Most of these approaches rely on state-of-the-art Convolutional Neural Networks (CNN).

In the literature we can find some contributions aimed at generating datasets from existing RIS/PACS systems. Most of them take profit from existing reports attached to the images to automatically generate labels for findings and diseases. For example, ChestX-ray8 [1] provides 108,948 frontal view X-ray images of 32,717 unique patients automatically labeled with eight diseases. Recently, this dataset was extended up to 14 diseases (ChestX-ray14).

ChestX-ray14 was used to train a specialized 121-layers CNN called CheXNet [2]. This model was mainly targeted at detecting pneumonia, but also performed well in the other diseases. Unfortunately, this dataset does not include lung nodules samples.

PadChest[3] provides an even larger dataset with 160,000 images from 67,000 patients labeled with a very large number of concepts automatically extracted from the attached reports. In this case, labels cover 174 findings, 19 differential diagnoses and 104 anatomical locations, which are mapped to the Unified Medical Language System (UMLS). The extraction process is performed by a recurrent neural network trained with 27% of manually annotated images.

As far as we know, there is no work reporting results of a CNN trained with PadChest. Perhaps, an issue for this is to find a method that allow us to get a proper training dataset for the target disease. For example, PadChest contains an acceptable number of samples with lung nodules. However, most samples belong to old people often presenting other pathologies, leading to noisy images. Besides, errors in the image labels can produce further noise in the training datasets.

Compared to ImageNet [6], current chest X-ray datasets do not achieve a critical mass for training CNN models. Most approaches in the literature train and test on the same (small) dataset, without validating the model with other different datasets. More-

CNN model	Dataset	Samples	Classes
ChestX-ray8[1]	NCCIH	112,120	8 diseases (nodules ≈ 6,323)
ChestX-ray14[8]	NCCIH	112,120	14 diseases (nodules ≈ 6,323)
CheXNet [2]	NCCIH	112,120	14 diseases (nodules ≈ 6,323)
Genetic DL[9]	PLCO	185,421	Lung cancer
Faster R-CNN[10]	LIDC-IDRI	1,018	Lung cancer
DENSENET[11]	JSRT	247	Lung nodules
Fastai[12]	Chest X-ray Pneumonia	1,341	Pneumonia/Normal

Table 1. Main CNN approaches and training datasets for nodule detection

over, these approaches show very disparate results, which could indicate the lack of a critical mass in the existing datasets. Thus, current models seem not to generalize well and are not ready yet for production [4]. Authors in [5] discuss the unintended consequences of these drawbacks.

In this short paper we present some preliminary work showing these issues, and how the selection of different samples from different datasets can improve notably the performance of the learned models.

2. Material and Methods

For the screening experiments we mainly use two large datasets of chest X-ray images associated where lung nodules are present. More specifically, we use Chest-X-ray8[7] and PadChest[3]. Table 1 describe these datasets, as well others related to X-ray for extracting samples for the normality class.

Our ultimate goal is the creation of a large and balanced lung cancer dataset that contains X-ray images with nodules difficult to detect by a non-expert people. In this case, images must have associated reports where lung cancer is diagnosed with CT. In addition, X-ray images are requested before the disease is declared, since they will be used to look for patterns in images where the findings are too subtle to be detected by the human eye. For this purpose we added X-ray images from the Hospital of Castelln fulfilling these constraints.

Table 1 shows different CNN approaches along with the datasets with which they were trained on. As earlier mentioned, the choice of these datasets is critical as quality results can vary greatly.

The Tensorflow and Pytorch frameworks have links to their standard neural network zoo models already created along with their precision scores on the Imaginet dataset. One of the CNNs with the highest accuracy rate is Resnet-152, which has been chosen for the experiments.

Resnet-152 was used in two different ways: (1) using a pre-trained model with Imagenet, and (2) training from scratch. In both cases precision results were very similar. However, they notably differ when testing with NCCIH images not seen during the training phase, achieving the models trained from scratch better results. Even so, results were not good enough for screening.

After analyzing the images by a radiologist, we realized that images representing normal cases were not so. Therefore, we decided to look for a series of images completely free from abnormalities. Thus, we selected a Kaggle dataset for pneumonia in

children [12]. Training from scratch with a combination of images from NCCIH for nodules and these samples achieved good results for the NCCIH test dataset (100% for nodules and 50% for normal cases).

We tested the new trained model over the PadChest dataset, obtaining good results for the nodule detection but some false positives from images tagged as normal. However, these normal cases belong to old people and contain many noisy elements like electrodes, catheters and so on, which are not reflected in the reports.

3. Conclusions

In this paper we present a preliminary study about the impact of some decisions when training a CNN for classifying X-ray images. Current datasets provided in the literature are either too small or non comprehensive enough to train a deep CNN. Large datasets are provided with tags inferred from the reports, but they do not reflect well the quality and discrimination power of the corresponding images. In our preliminary experiments we show that combining good examples from different datasets can achieve a substantial improvement. In the next steps we aim at building a very large dataset by combining existing ones as well high quality data from hospitals directly assessed by radiologists. The idea is to give the model the necessary elements that allow the model to discern the target pathology at earlier stages. Future work also regards how to automatically discern good from bad examples to train these models.

References

[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri and R. M. Summers: ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In Proc. of CVPR 2017, pp. 3462-3471 (2017)

[2] O. Gozes and H. Greenspan: Deep Feature Learning from a Hospital-Scale Chest X-ray Dataset with Application to TB Detection on a Small-Scale Dataset. In Proc. of EMBC 2019, pp. 4076-4079.

[3] A. Bustos, A. Pertusa, J.M. Salinas, M. de la Iglesia-Vay: PadChest: A large chest x-ray image dataset with multi-label annotated reports. Medical Image Analysis, Vol. 66 (2020)

[4] J. R. Zech, et al: Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLOS Medicine, 15(11):e1002683, 2018.

[5] Federico Cabitza, Raffaele Rasoini, and Gian Franco Gensini. Unintended consequences of machine learning in medicine. Jama, 318(6):517518, 2017.

[6] IMAGENET Dataset, `http://www.image-net.org/`[Last visited: 24/04/2021]

[7] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. M. Summers: ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. IEEE CVPR, pp. 3462-3471 (2017)

[8] Wang H, Jia H, Lu L, Xia Y.: Thorax-Net: An Attention Regularized Deep Neural Network for Classification of Thoracic Diseases on Chest Radiography. IEEE J Biomed Health Inform. Vol. 24(2):475-485 (2020)

[9] A Genome Wide Scan of Lung Cancer and Smoking Dataset. [Last Visited: 20/04/2021] `https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000093.v2.p2/`

[10] A. Bhandary et al.: Deep-learning framework to detect lung abnormality A study with chest X-Ray and lung CT scan images. Pattern Recognition Letters, Vol. 129: 271-278 (2020).

[11] X. Li, et al: Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection. Artificial Intelligence in Medicine, Vol. 103 (2020)

[12] Chest X-Ray Images (Pneumonia), [Last visited: 24/04/2021] `https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia`