# Lung Nodule Classification Based on Deep Convolutional Neural Networks

Julio Cesar Mendoza Bobadilla and Helio Pedrini[✉]

Institute of Computing, University of Campinas,
Campinas, SP 13083-852, Brazil
`helio@ic.unicamp.br`

**Abstract.** Lung nodule classification is one of the main topics on computer-aided diagnosis (CAD) systems for detecting nodules. Although convolutional neural networks (CNN) have been demonstrated to perform well on many tasks, there are few explorations of their use for classifying lung nodules. In this work, we present a method for classifying lung nodules based on CNNs. Training is performed by balancing the mini-batches on each stochastic gradient descent (SGD) iteration to address the lack of nodule samples compared to background samples. We show that our method outperforms a base feature-engineering method using the same techniques for other stages of lung nodule detection, and show that CNNs obtain competitive results when compared to state-of-the-art methods evaluated on Japanese Society of Radiological Technology (JSRT) dataset [13].

**Keywords:** Lung cancer · Chest imaging · Image classification · Convolutional Neural Networks

## 1 Introduction

Lung cancer affects men and women around the world and can be characterized by abnormal cell growth in the lung tissues. A chest radiography (CXR) or a computerized tomography (CT) is typically one of the first diagnostic steps performed by radiologists [2]. Lung nodules are small masses of tissue, which appear as white shadows on a CXR image or CT scan. Benign nodules have approximately 5–30 millimeters, whereas larger nodules are more likely to be malignant [1].

The detection of lung nodules on an imaging test is a challenging task due to low contrast of the lesions, possible overlap with ribs and large pulmonary vessels, as well as their irregular size, shape and opacity [8]. Pulmonary radiologists usually detect lung nodules by considering brightness and shape of circular objects within the thoracic cavity. More conclusive diagnosis of lung cancer can be conducted through histological examination of the tissue [7].

Advances in imaging modalities and Computer-Aided Diagnosis (CAD) systems have helped radiologists identify lung nodules more quickly and accurately.

Potential locations of nodules can be automatically detected by considering bright masses of expected size, shape and texture of a lung nodule. Techniques available for detecting lung nodules generally involve three main stages: lung segmentation, candidate detection, and nodule classification.

The main contribution of this work is the evaluation of Convolutional Neural Networks (CNN) for lung nodule classification. Our CNN is trained through stochastic gradient descent (SGD) algorithm. The input of the CNN is a set of candidate regions. The amount of false positive candidates is large when compared to the number of true positives, leading to an unbalanced classification problem. We address this issue by balancing the mini-batches on each SGD iteration. The proposed method is evaluated on the Japanese Society of Radiological Technology (JSRT) dataset [13]. Our method outperforms the results obtained with a feature-engineering approach [6] used as baseline, which uses the same algorithms for previous stages to nodule classification, and obtains competitive results when compared to other methods of the literature.

The remainder of the paper is organized as follows. Section 2 briefly describes background information on previous work. Section 3 presents the methodology proposed in this work. Section 4 describes and discusses our experimental results. Section 5 concludes the paper with final remarks and directions for future work.

## 2   Background

The lung nodule detection typically involves three sequential steps. The first one consists in segmenting the lung area to prevent the detection of structures similar to nodules located outside the lung using methods such as active contour models, active shape models, regression, among others. The second step is the detection of nodule candidates through appropriate filtering strategies, such as convergence index-based filters, Laplacian of Gaussian filters, pixel classification, among others. The third stage is the lung nodule classification. Most of the available methods follow the classical pattern recognition pipeline: feature extraction, feature selection, and classification. On the other hand, deep learning approaches have been rarely used for lung nodule classification.

### 2.1   Lung Nodule Classification

A method proposed by Coppini et al. [3] uses a multi-scale neural network architecture to distinguish true nodules from background patterns. A set of Gabor filters and a Laplacian of Gaussian filter are used to extract features from the input regions and to feed a 3-layer neural network. Then, an optimal linear combiner determines the nodule probability for each candidate.

Most approaches segment nodule regions to capture features from particular regions of the nodules. Schilham et al. [11] use multi-scale techniques to segment nodule candidates. Then, they convolve the input regions with a set of Gaussian filters to extract statistics from the inner and band regions of the nodule. A two-step classification is performed using approximate $k$-nearest neighbor algorithm.

Some methods perform transformations on the input regions. Wei et al. [15] transform the candidate regions using Adaptive Ring filter, Iris filter, and Sobel filter. They segment the nodule candidates using active contour models, and extract geometric features, contrast features, first and second order statistics from the inner and outer regions on the original and transformed candidates. The best features are selected using the forward stepwise selection algorithm. Classification is performed by using a statistical method based on the Mahalanobis distance measure.

Shiraishi et al. [14] proposed a method that describes nodules using geometric, intensity, background and edge gradient features from the original and density corrected, nodule enhanced, and contralateral subtracted images of each suspicious nodule. Classification is performed using artificial neural networks.

Hardie et al. [6] proposed an adaptive distance-based thresholding approach for nodule segmentation. Then, a set of geometric, gradient and intensity features are extracted from the local enhanced, normalized, nodule-mask and original versions of each nodule candidate. A Fisher Linear Discriminant is used for classification. Similar to [15], Hardie et al. use a sequential forward selection algorithm to find a proper subset of features.

A method proposed by Chen and Suzuki [2] uses a regression model to suppress rib-like structures on the images. Nodule candidates are segmented using dilatation and erosion followed by a watershed-based segmentation. Morphological and gray-level-based features are extracted from each nodule candidate in the original and rib-suppressed images. An SVM classifier with Gaussian kernel is used to classify each nodule.

## 2.2 Deep Learning

Several knowledge domains have benefited from deep learning research [9,12], such as image and video classification, audio and speech recognition, bioinformatics, natural language processing, and data mining. Different deep learning architectures have been developed to address various tasks, including CNNs, belief networks, Boltzmann machines, and recurrent neural networks.

Methods based on deep learning can be applied to supervised and unsupervised learning problems. In supervised tasks, the algorithms avoid the process of constructing feature descriptors by transforming the data into compact intermediate representations and deriving layered structures [4]. In unsupervised tasks, the algorithms explore unlabeled data, which are typically more available than labeled data.

## 3 Proposed Method

Our lung nodule classification method performs the candidate analysis directly on the pixels of the images using a CNN, instead of extracting features and using classifiers. The initial stages of lung segmentation and candidate detection follow the approach developed by Hardie et al. [6].

The classification starts by preprocessing and augmenting the candidate regions obtained in the segmentation process. A CNN is then trained with backpropagation using the augmented dataset. Finally, we reduce overlapping detections through an adjacent candidate rejection rule.

### 3.1   Preprocessing and Data Augmentation

We applied a $z$-score normalization to the pixel intensities over the dataset to have zero mean and unit variance. Candidate detection stage produces approximately 150 negative samples for each positive one. To balance the dataset and prevent overfitting, we augment the data available by increasing the amount of positive samples and applying the following transformations to all candidates, including negative samples: (i) horizontal/vertical random shift with a factor of each dimension of the region of interest, selected over a uniform distribution between 0 and 0.1; (ii) random zoom with a scale factor selected over a logarithmic distribution of 1.0 and 1.2; (iii) horizontal flip: yes or no with a probability of 0.5; (iv) random rotation with an angle between $-5°$ and $5°$.

### 3.2   Architecture

Inspired by the work developed by Graham [5], we explored a family of CNNs known as DeepCNets. DeepCNets are composed of alternating convolutional and max-pooling layers with a linearly increasing number of filters. We evaluated architectures with 3 to 6 convolutional layers, achieving better results with architectures with 6 convolutional layers. Figure 1 depicts our architecture.
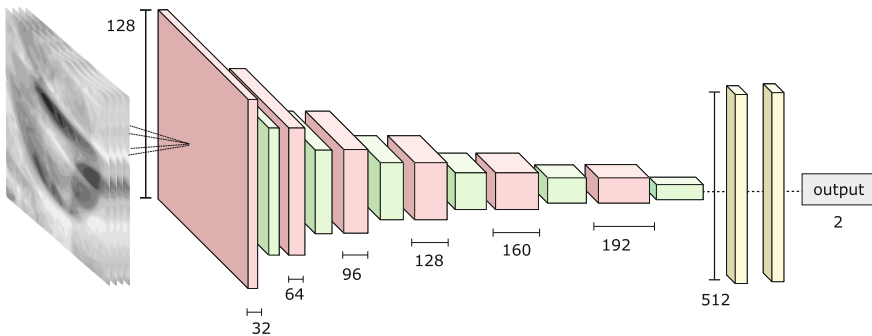


**Fig. 1.** Our CNN architecture. Nodule candidates are fed to the network composed of six alternating convolution and max-pooling layers, and two fully connected layers.

The network has 6 convolutional layers, 6 max-pooling layers, and 2 fully connected layers. Filters used on convolutional layers have a $3 \times 3$ receptive field. Max-pooling is performed over windows of $2 \times 2$, with stride 2. We use Leaky Rectifier Linear Units with alpha 0.333 for all convolutional and fully connected

layers. We use dropout regularization with a drop ratio 0.25 on convolutional layers and 0.5 on fully connected layers. The output of the last layer is connected to a softmax layer that produces a distribution over the two classes.

### 3.3   Learning

For the learning stage, we use the Stochastic Gradient Descent (SGD) with Nesterov momentum. We initialized the weights on each layer with the orthogonal random matrix initialization procedure proposed by Saxe et al. [10]. The learning rate is decreased when the error on the test set stops improving. The learning rate is decreased three times. We experimented with weight decay regularization, however, it did not improve our results. Therefore, the final model does not use weight decay regularization.

Motivated by the work proposed by Yan et al. [16], we balance the amount of positives and negatives samples for each batch used to feed the CNN in the training step. For positives, we select half of the batch size samples randomly. For negatives, we select half of the batch size samples iteratively. Positive and negative samples are perturbed before they feed the CNN on each SGD iteration.

### 3.4   Adjacent Rejection

The trained CNN is used to measure the probability of a candidate being a nodule. Since more than one candidate can be related to a positive sample, we use the same adjacency rejection rule explained in [6]. This procedure reduces the number of final detections by preserving the candidates that have the maximum detection probability on their radial neighborhood of 22 mm.

## 4   Experimental Results

This section describes the dataset used in our experiments, the results obtained with our method, as well as a comparison of the results to other methods available in the literature.

### 4.1   Dataset

We use the Japanese Society of Radiological Technology (JSRT) database [13] composed of 154 nodule and 93 non-nodule chest radiographs. The images have $2048 \times 2048$ pixels in size and 4096 gray levels. Nodules are classified as malignant or benign and graded as to their visual subtlety: extremely subtle, very subtle, subtle, relative obvious and obvious. The images were annotated by three radiologists and validated using CT images to provide groundtruth information.

We validate our method with 10-fold cross-validation. We use sensitivity and false positive per image (FPPI) measures to compare our method to others in the literature. Sensitivity and FPPI are meaningful on lung nodule detection, in the sense that methods with high sensitivity tend to produce a large amount of false positives and, consequently, demanding unavailing attention of the physicians in the analysis of samples with low probability of being a nodule.

## 4.2   Results

Our CNN was trained with SGD optimization with Nesterov momentum of 0.9, a batch size of 32 samples, learning rate of 0.1 on 55 epochs, decreasing the learning rate by 10 on 15, 25 and 40 epochs. Our model takes approximately 10 h to be trained on a Tesla K40c GPU device.

Due to the variability in the validation, labeling and scoring procedures, it is not simple to make a precise comparison of the methods that use the JSRT database. In our experiments, we use the validation, labeling and scoring procedures as described in [6]. Our CNN was trained by excluding 14 images containing nodules on the opaque regions of the lung. In Fig. 2(a), we compare the performance of our method to the one proposed by Hardie et al. [6]. A comparison with other methods available in the literature is provided in Fig. 2(b). Since some methods report their results with the entire database, we adjust the sensitivity values by considering excluded cases as missed.
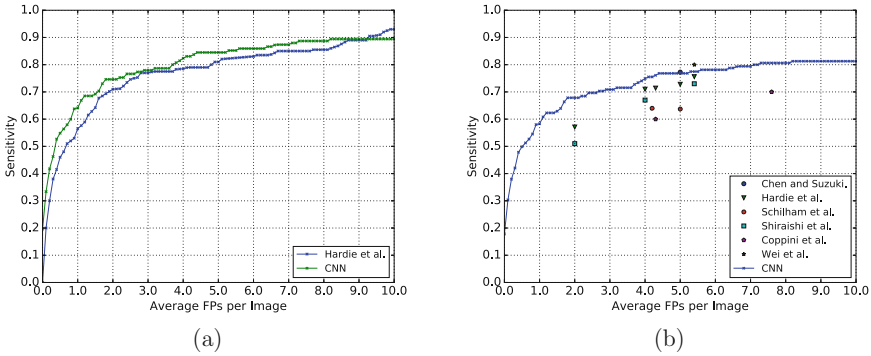


**Fig. 2.** (a) Comparison of the CAD performance between our method and the one proposed by Hardie et al. [6] on the JRST database. Opaque cases are excluded; (b) Comparison of the CAD performance on the JRST database. Sensitivity values are adjusted by considering opaque cases as missed.

Table 1 shows the sensitivity values obtained with our CNN method and state-of-the-art approaches at the same FPPI values. Chen and Suzuki [2] reported a slightly higher sensitivity. Their systems use dual-energy radiographs for rib-suppression on the training stage. The method proposed in [15] extracts a large set of features from the images at different scales. The methods described in [3,11,15] employ the entire JSRT database.

Figure 3 shows some samples of true positives and false positives detected with the proposed method. The samples are shown with respect to their probability values.

**Table 1.** CAD system performance comparison.

| Average FPPI | Method | Reported sensitivity (%) | CNN sensitivity (%) |
|---|---|---|---|
| 5.0 | Chen and Suzuki [2][a,c] | **85** | 84.5 |
| 5.0 | Hardie et al. [6][a] | 80.1 | **84.5** |
| 2.0 | Schilham et al. [11] | 51 | **67.8** |
| 4.0 | Schilham et al. [11] | 71 | **74.6** |
| 5.0 | Shiraishi et al. [14][a,b] | 70.1 | **84.5** |
| 4.3 | Coppini et al. [3] | 60 | **76.1** |
| 5.4 | Wei et al. [15] | **80** | 77.7 |

[a]Results reported in this row exclude opaque cases.
[b]Results based on 924 chest radiographs that include the JRST cases.
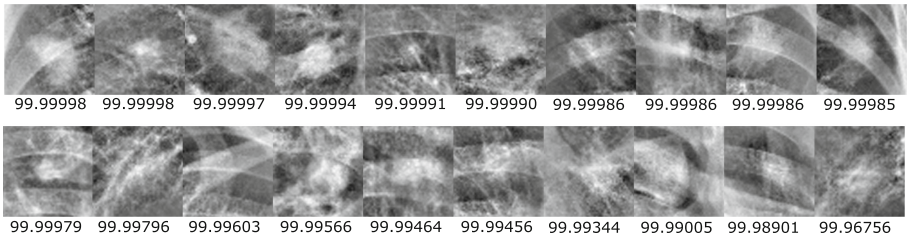[c]A private database was used for training and the JSRT for testing.



**Fig. 3.** Top detection results with their respective probability values (sorted in descending order). The first row shows the true positive samples, whereas the second row shows the false positive samples.

## 5   Conclusions

We proposed a method based on CNNs to classify lung nodules in chest radiographs. Our method outperforms the baseline work without the need of extracting several features from the images.

Experimental results show that CNNs can operate effectively on lung nodule classification through data augmentation and dropout regularization. By balancing the number of positive and negative samples on the batches, it was possible to train a CNN with only 140 positive samples.

# References

1. Brant, W.E., Helms, C.A.: Fundamentals of Diagnostic Radiology. Lippincott Williams & Wilkins, Philadelphia (2012)
2. Chen, S., Suzuki, K.: Computerized detection of lung nodules by means of "virtual dual-energy" radiography. IEEE Trans. Biomed. Eng. **60**(2), 369–378 (2013)
3. Coppini, G., Diciotti, S., Falchini, M., Villari, N., Valli, G.: Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms. IEEE Trans. Inf. Technol. Biomed. **7**(4), 344–357 (2003)
4. Deng, L., Yu, D.: Deep learning: methods and applications. Found. Trends Sig. Process. **7**(3–4), 197–387 (2014)
5. Graham, B.: Spatially-Sparse Convolutional Neural Networks. arXiv preprint arXiv:1409.6070 (2014)
6. Hardie, R.C., Rogers, S.K., Wilson, T., Rogers, A.: Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. Med. Image Anal. **12**(3), 240–258 (2008)
7. Hauser, S., Josephson, S.A.: Harrison's Neurology in Clinical Medicine. McGraw Hill Professional, New York (2013)
8. Katsuragawa, S., Doi, K.: Computer-aided diagnosis in chest radiography. Comput. Med. Imaging Graph. **31**(4), 212–223 (2007)
9. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
10. Saxe, A.M., McClelland, J.L., Ganguli, S.: Exact Solutions to the Nonlinear Dynamics of Learning in Deep Linear Neural Networks. arXiv preprint arXiv:1312.6120 (2013)
11. Schilham, A.M., van Ginneken, B., Loog, M.: A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database. Med. Image Anal. **10**(2), 247–258 (2006)
12. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
13. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K.I., Matsui, M., Fujita, H., Kodera, Y., Doi, K.: Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. Am. J. Roentgenol. **174**(1), 71–74 (2000)
14. Shiraishi, J., Li, Q., Suzuki, K., Engelmann, R., Doi, K.: Computer-aided diagnostic scheme for the detection of lung nodules on chest radiographs: localized search method based on anatomical classification. Med. Phys. **33**(7), 2642–2653 (2006)
15. Wei, J., Hagihara, Y., Shimizu, A., Kobatake, H.: Optimal image feature set for detecting lung nodules on chest X-ray images. In: Lemke, H.U., Inamura, K., Doi, K., Vannier, M.W., Farman, A.G., Reiber, J.H.C. (eds.) Computer Assisted Radiology and Surgery, pp. 706–711. Springer, Heidelberg (2002)
16. Yan, Y., Chen, M., Shyu, M.L., Chen, S.C.: Deep learning for imbalanced multimedia data classification. In: IEEE International Symposium on Multimedia, pp. 483–488, December 2015