

# Using Deep Learning for Classification of Early Lung Nodules on Computed Tomography Images

Lucas Lima<sup>1</sup>, Marcelo Oliveira<sup>1</sup>

<sup>1</sup>Institute of Computing – Federal University of Alagoas (UFAL) – Maceió, AL, Brazil

{l11,olveiramc}@ic.ufal.br

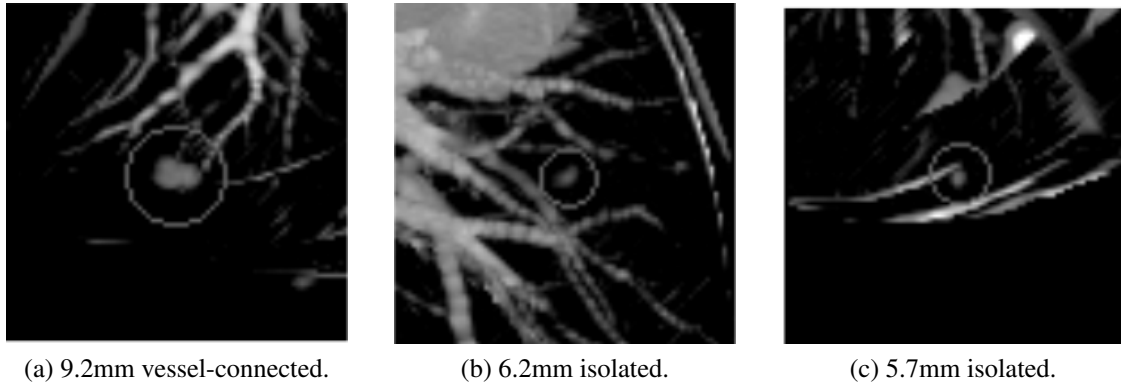
**Abstract.** Lung cancer is the leading cause of cancer mortality, accounting for approximately 20% of all cancer-related deaths. Nevertheless, despite the development of new therapeutic agents and technologies, only 16% of lung cancer patients are diagnosed at early stages. Therefore, to diagnose in early stages, when the nodules are very small, is a complex task for specialists and presents some challenges. To assist the specialists, the main purpose of this work is to propose the use of Deep Learning to classify 25,200 small pulmonary nodules balanced with diameter 5-10mm. The result was of 0.992 (+/- 0.001) of area under ROC curve using 10-fold cross validation. The proposed method showed to be promising to assist the specialists in classification of small lung nodules.

## 1. Introduction

Cancer is characterized by a disease that causes the disordered growth of abnormal cells which spreads through neighboring tissues destroying them. Lung cancer is the leading cause of cancer-related deaths in the world, accounting for approximately 20% of all cancer-related deaths [Tammemagi and Lam 2014]. Half of the patients die within one year, and the 5-year survival rate is about 17.8% when diagnosed [Zappa and Mousa 2016]. Patients diagnosed in early stages have a 1-year survival rate of 81-85% while in an advanced stage have 15-19% [Blandin Knight et al. 2017]. Therefore, malignant lung nodules need to be detected and diagnosed as soon as possible to enable the early treatments. But the process to identify lung nodules in early stages, those less than 10mm in diameter, is a challenging task even for the experienced radiologist. Besides small size, the early nodules have low contrast in comparison to the lung tissue and can be attached to other complex lung structures (Figure 1) [Hua K-L and Y-J 2015] [Alilou et al. 2014].

With the objective to assist the radiologists in the lung cancer classification, Computer-Aided Diagnosis (CAD) systems have been used to determine the probability of malignancy of a lung nodule [Gillies et al. 2016]. Until recently, the effectiveness of CAD systems and many other pattern recognition applications depended on meticulously handcrafted features, topped off with a learning algorithm to map it to a decision variable. Medical image analysis and interpretation are fundamental cognitive tasks in the radiologist's diagnosis. During the process of designing features, and features such as contrast of the lesion, spiculation patterns and the sharpness of the border it is often necessary to consult the experts. These feature transformations provide a platform to instill task-specific, a priori knowledge but cause a large bias towards how we humans think the task is performed [Chartrand et al. 2017] [Kooi et al. 2017].

Figure 1. Lung nodules (circulated in gray) with their sizes and anatomical structures connected to them, respectively [Alilou et al. 2014].



A technique of machine learning called “Deep Learning” (DL) have attracted the attention of the scientific community to resolve many problems of computer vision such as image classification and image retrieval [Guo et al. 2016]. DL has the characteristic of grouping representation-learning methods in which the algorithm learns a composition of features from multiple levels of representation in a hierarchical way starting at one level and reaching progressively more abstract levels [LeCun et al. 2015, Chartrand et al. 2017]. Among the architectures presenting in DL there is the Convolutional Neural Network (CNN). CNN architecture has been the most effective approach to classify images since 2012 when Krizhevsky et al. [Krizhevsky et al. 2012] won the ILSVRC2012 competition. CNN is an interleaved set of feed-forward layers composed of convolutional filters followed by reduction, rectification or pooling layers, where each layer extracts high-level abstract features from the images [Ravi et al. 2017]. It has been already applied to classify small lung nodules [Anirudh et al. 2016].

The main goal of this work was to assess the Deep Learning capability in classify early lung nodules with a diameter of 5-10mm.

The remainder of this paper is organized as follows. First, we present the related works (Section 2). Next, we show the public database used, how the early nodules were selected, the techniques used to augment the samples of our database, the CNN’s hyperparameters used, and how we evaluated our results (Section 3). Next, we presents the main results obtained by the methodology proposed and the discussion of this work (Section 4). Lastly, we conclude this paper and presents our main contribution to the state-of-art (Section 5).

## 2. Related Works

Many works have been done aiming do classify lung nodules, but most focus on both large and small nodules, and few focus on small nodules. In this paper, we separated the works made by Felix et al. [Felix et al. 2016], Dilger et al. [Dilger et al. 2015] and Anirudh et al. [Anirudh et al. 2016].

Felix et al. [Felix et al. 2016] used selected features extracted from 274 small lung nodules of the Lung Image Database Consortium (LIDC) with a diameter between 3-10mm. The features extracted were Texture Features (TF) and Margin Sharpness Fea-

tures (MSF) and they used the technique called Evolutionary Genetic Algorithm (EGA) to select the features on the integration of TF and MSF. The best classifier used was Multilayer Perceptron (MLP) and achieved the area under ROC curve (AUC) of 0.820 ( $\pm 0.053$ ) using 10-fold cross validation. They used nodules with up to 10mm and got good results, but they used many steps: extraction and selection of the most relevant attributes out of a total of 48 through the technique EGA to classify the small lung nodules finally.

Dilger et al. [Dilger et al. 2015] used an Artificial Neural Network (ANN) to classify 50 nodules (22 malignant and 28 benign) from high-resolution computed tomography scans with diameter between 4-30mm and a set of 47 selected features (intensity, shape, size, and texture), extracted from both the pulmonary nodules and the tissue of the region of lung parenchyma, and achieved the highest AUC of 0.935 ( $\pm 0.0096$ ) using leave-one-case-out cross validation. The nodules were acquired from the National Institutes of Health (NIH). They got excellent results, but they used nodules bigger than 10mm reaching 30mm, which includes nodules in advanced stages.

Anirudh et al. [Anirudh et al. 2016] employed a 3D Convolutional Neural Network (CNN), where 20 and 47 scans were used for training and test, respectively. The diameter was between 3-20mm and achieved sensitivities of 80% for ten false positives per scan. Although they used nodules smaller than 3mm, they also used nodules bigger than 10mm reaching 20mm, so they got nodules more advanced.

### 3. Materials and methods

The overview schema of this work can be view in Figure 2. First, we got a database from medical imaging from a public repository (Section 3.1). Next, we measured the nodules' sizes and selected only small nodules (Sectio 3.2), then the database was balanced for the experiments in an equal number of benign and malignant nodules (Section 3.3), we created a Convolutional Neural Network (CNN) architecture (Section 3.4) and applied it in the database with and without data augmentation to generate the results.

#### 3.1. Lung Database Acquisition

The database used in this work was the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) [Samuel G. Armato et al. 2011]. This base consists of Computed Tomography (CT) scans with marked-up annotated lesions identified and classified by four experienced radiologists. It is now composed of 1,010 CT scans of the patients, 1,308 studies, and 244,529 images. The radiologists identified the nodules according to subjective characteristics, and one of them is the likelihood of malignancy on a scale from 1 to 5, following the conditions:

- Malignancy 1: high probability of being benign;
- Malignancy 2: a moderate chance of being benign;
- Malignancy 3: indeterminate probability;
- Malignancy 4: a moderate probability of being malignant;
- Malignancy 5: high probability of being malignant.

The LIDC-IDRI is a collection where there is no correlation between images, exams data and classification of the nodules by radiologists. Furthermore, information about nodule size or image features does not exist. So, aiming to integrate information about the

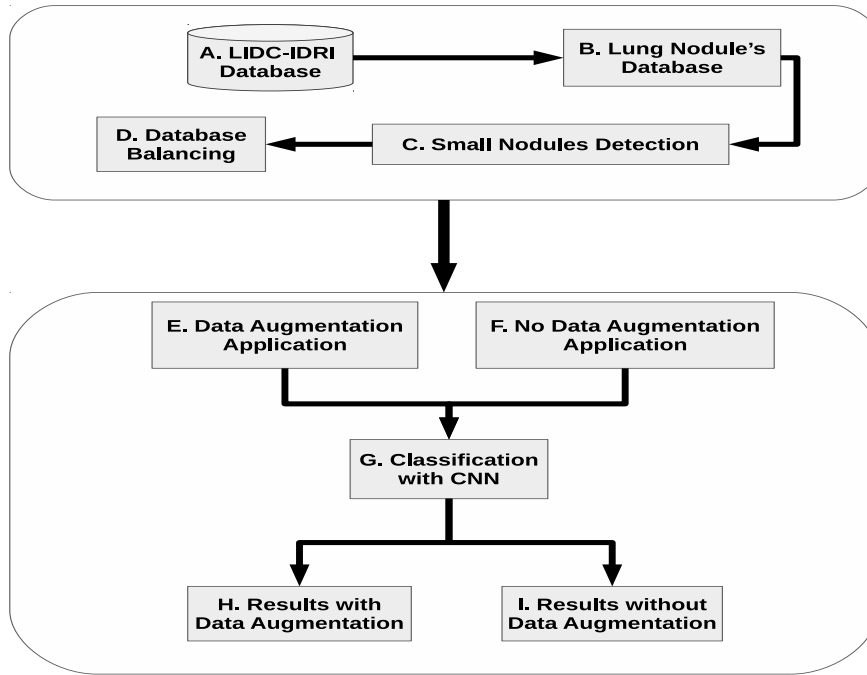


Figure 2. Schema overview applied in this work.

nodules in LIDC and images features we used the database developed by Ferreira Junior et al. [Ferreira Junior et al. 2016]. All images lesions were manually segmented using the radiologist's marks and later placed into the respective database, making access to the information of the injuries practicable.

Nodules with the likelihood of malignancy three were discarded in this work because of their probability of malignancy on them is indeterminate.

### 3.2. Small Lung Nodule Selection

To perform the selection of the nodules according to their diameters, we initially estimated their sizes which were then attached to our database, as the LIDC-IDRC original database did not contain. The size of a nodule was calculated as a simple 2D measure of the greatest diameter, which can be realized in the axial plane along the axis of the longest diameter [Bartholmai et al. 2015]. These approximations consisted of calculating the Euclidean distance between the minimum and maximum coordinates in the respective  $x$  and  $y$  axes of all slices of a nodule.

According to the literature [Yankelevitz et al. 1999] [Reeves and Kostis. 2000], a nodule can be as a small nodule when it has a diameter of less than 10mm. However, Bartholmai et al. [Bartholmai et al. 2015] affirm that tests have already shown that nodules with a diameter less than 5mm are probably benign, even in cases of high-risk patients, which makes it uninteresting to work with nodules predisposed to be benign. So, the threshold chosen for this work was nodules greater equal to 5mm and less than 11mm.

### 3.3. Balancing Database

Counting by slices of all small nodules, we got 3,513 slices, where 2,582 were benign, and 931 were of malignant nodules. In this work, our focus was early solid nodules, so

nonsolid and ground-glass opacity nodules are out of the scope of this paper. So, our final number of slices were of 2,172 and 603 of benign and malignant slices of solid nodules, respectively. To ensure data balancing in all phases of the experiments and to perform a fair classification we used only 600 slices of benign and 600 slices of small malignant nodules, totalizing 1200 slices of nodules in early stage (see Table 1).

Table 1. Numbers of small nodules between 5-10mm used from our database balanced in 600 for benign and malignant.

	<b>Benign</b>		<b>Malignant</b>		
<b>Probability of Malignancy</b>	1	2	4	5	<b>Total</b>
<b>Number of nodules</b>	169	431	492	108	1200
<b>Sum</b>	600		600		

### 3.4. Convolutional Neural Network

In the 1960s a Convolutional Neural Network (CNN) is a model inspired by the principles of Deep Learning was proposed, which is a multi-layer neural network that it can extract multi-level hierarchies of features presented in two-dimensional images and is widely used in the field of computer vision and offers the advantages of being easy to train and have fewer parameters than fully connected networks with the same number of hidden units [Song et al. 2017].

The CNN can have several layers composed of three main neural layers: convolutional layers, pooling layers and fully connected layers [Guo et al. 2016]. CNN receives as input an  $n \times n \times m$  image where  $n$  is the height and width and  $m$  the number of channels of the image. If  $m$  has value one the image is in grayscale if  $m$  is three the image is in RGB mode.  $K$  convolutional filters of size  $a \times a$  are applied in convolutional layers, where  $a < n$  to extract the features in the image received. The pooling layers are non-linear down-sampling layers where the operations can be of max-pooling or average-pooling on the matrices they receive. Max-pooling calculates the maximum and average-pooling calculates the average in a neighborhood of feature points, in which this neighborhood or filter is usually a matrix of  $2 \times 2$ . Pooling layers can be applied optionally after each convolutional layer. The fully connected layers have the role of classifying the input images through the patterns learned by the CNN, which is usually a Multilayer Perceptron (MLP). During the process of training in each layer is applied some activation function to normalize the output in the neurons which are non-linear and derivable, and can be for example Softmax and Rectified Linear Unit (ReLU) [Sonoda and Murata 2015] [Litjens et al. 2017].

The dropout and data augmentation techniques were used to improve the performances and avoid the overfitting problem. Dropout was developed and proposed by Srivastava, N. et al. [Srivastava et al. 2014], and besides reducing overfitting, it still gives a significant improvement over other regularization methods. In dropout, there is a parameter  $p$  or “dropout rate” which is a value between zero and one and which excludes the features proportionally in this fraction put. CNN has the disadvantage to require a large database of labeled images to be efficient in the task of classification, how we had to this work few samples, 1200 samples, so, the application of data augmentation was essential to make progress in the accuracy, and still control the overfitting problem.

CNN can execute its training in supervised training using some Gradient Descent Algorithm (GDA), among them there is the Adadelta algorithm which has the advantage to alleviate the job of choosing a learning rate, a task that we face in Stochastic Gradient Descent (SGD) algorithm [Zeiler 2012].

The CNN architecture used in this work and its parameters values and outputs in each layer is shown (Figure 3) and is detailed (see Table 2).

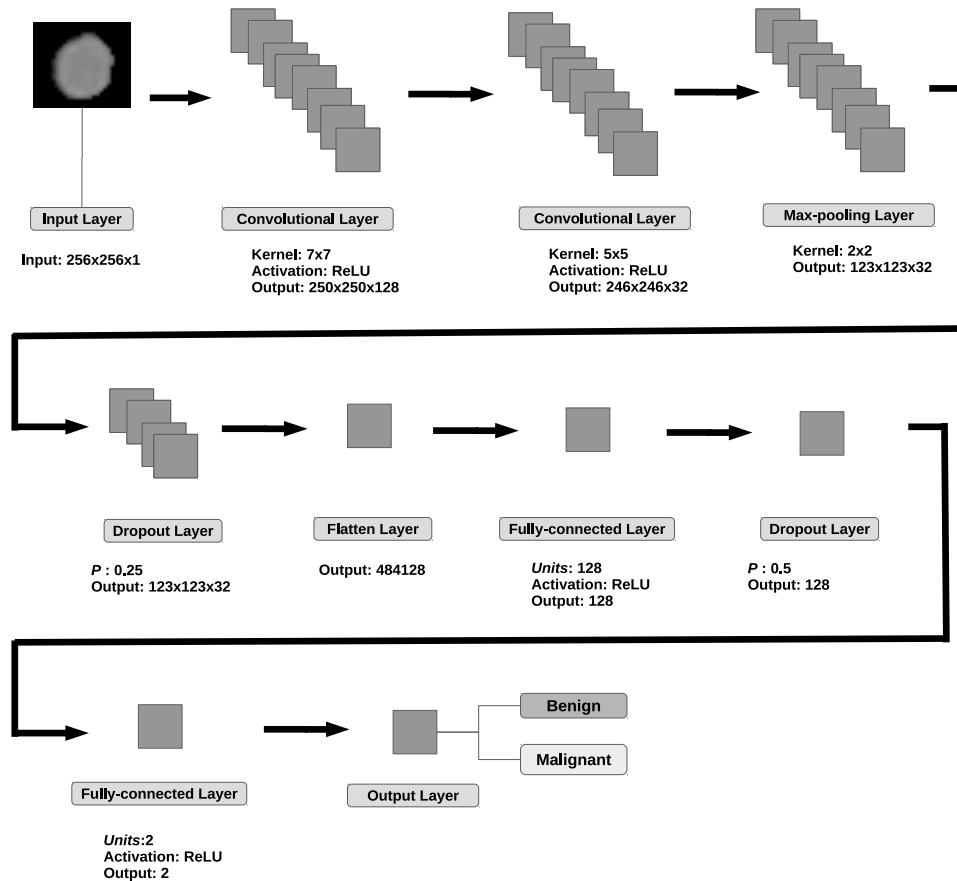


Figure 3. The architecture of the proposed CNN with the parameters and the output values.

## 4. Experiments and Results

### 4.1. Experiment Setup

All the experiments were done on a graphics processing unit (GPU) NVidia Titan X with 12 Gigabytes of RAM, 3584 cores, the speed of 1,5 GHz and Pascal Architecture. The experiments were done using Keras (v.2.0.9) with Tensorflow version (v.1.3.0) as backend [Chollet et al. 2015] [Abadi et al. 2015].

The operations applied using the data augmentation technique were: rotation, flip from left to right and from top to bottom. We applied 21 new samples to each image, making our database jump from 1,200 to 25,200 (12,600 benign and 12,600 malignant) samples.

Table 2. CNN architecture and parameters.

Layer	Type	Input	Output
1	Convolutional	256x256x1	250x250x128
2	Convolutional	250x250x128	246x246x32
3	Max-pooling	250x250x128	123x123x32
4	Dropout	123x123x32	123x123x32
5	Flatten	123x123x32	484128
6	Fully-connected	484128	128
7	Dropout	128	128
8	Fully-connected	128	2

Table 3. Comparison of filters and kernel values in CNN in layer 1.

Filter	Kernel		
	3x3	5x5	7x7
<b>16</b>	69.92% ( $\pm$ 3.60%)	70.33% ( $\pm$ 3.62%)	71.25% ( $\pm$ 3.15%)
<b>32</b>	70.50% ( $\pm$ 3.08%)	71.75% ( $\pm$ 3.22%)	71.83% ( $\pm$ 3.96%)
<b>64</b>	69.58% ( $\pm$ 3.06%)	71.92% ( $\pm$ 4.17%)	71.42% ( $\pm$ 3.05%)
<b>128</b>	70.92% ( $\pm$ 3.52%)	71.58% ( $\pm$ 4.26%)	<b>72.25% (<math>\pm</math> 4.00%)</b>

We did the experiments in three stages: training, validation, and test, using 80%, 10% and 10% of the samples, respectively. All steps were balanced, and it was used 10-fold cross-validation to reduce the impact of data bias. We generated two kinds of results, one using data augmentation, with 25,200 samples, and the other not using, with 1,200 samples.

All images had a resolution of 256x256 in grayscale mode. The only pre-processing executed was to normalize the pixels between zero and one before the application of the CNN.

#### 4.2. Construction of the CNN

The CNN used in this work was very similar applied in mnist with keras [Fchollet et al. 2018]. The difference is in the first two layers, the convolutional layers. It contains two convolutional layers, in the first, was made a comparison between the parameters values, kernel, and filter, where we vary the values of the kernel from 3 to 7 from 16 to 128, respectively. The best result was the combination of the kernel 7x7 and the filter 128, as we can see in Table 3. The second layer had a kernel of size 5x5 and 32 filters; followed by one layer of Max-pooling, with a kernel of size 2x2; next, we applied a dropout layer with  $p = 0.25$ . Next, a flatten layer was applied, where this layer only took the input in 3D and turned it in 1D without affecting the neurons. Next comes a Fully-connected layer, followed by another dropout layer, with  $p = 0.5$ , and to finish other Fully-connected layer to output the classification of the image in benign or malignant, totalizing our architecture in eight layers. The convolutionals and the first fully-connected layers used the ReLU activation function, and the last fully-connected layer used the Softmax activation function. The batch size and epochs were set with the value of 50 for both go get the best results in a short time. The GDA used was the Adadelta algorithm.

The rest of the architecture was followed by the example made with the Mnist

database using Keras that is in a grayscale mode like ours [Fchollet et al. 2018].

### 4.3. Results and Analysis

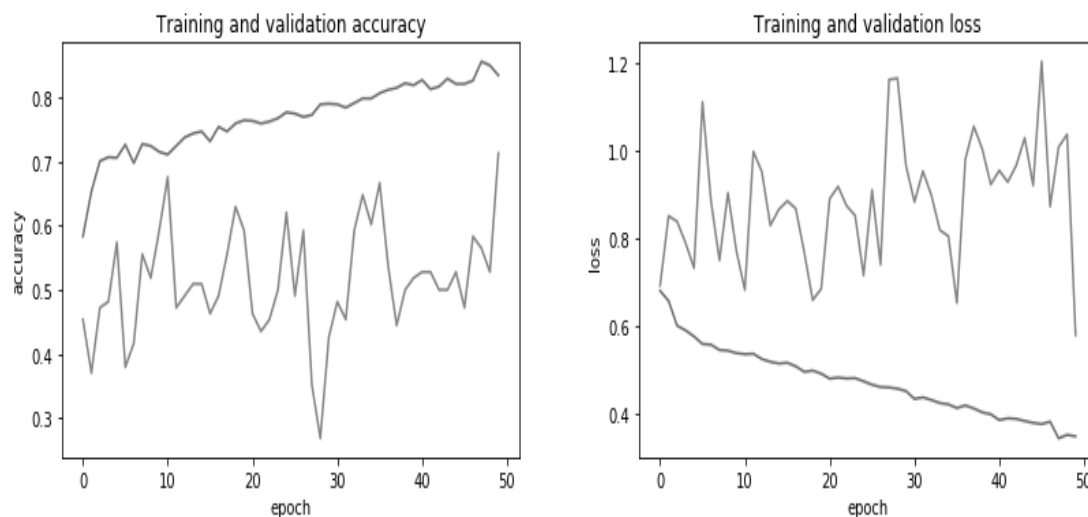
In this section is presented the results acquired with the CNN architecture showed in the last part applied in the same database without and with data augmentation. To evaluate the performances alleviating the impact of data bias, we used 10-fold cross validation, where training, validation and test sets balanced in all iterations.

The results obtained in the two conditions (without and with data augmentation), are showed (see Table 4), through the metrics Sensitivity (SEN), Specificity (SPE), Accuracy (ACC) and area under ROC curve (AUC), with mean followed by  $\pm$  standard deviation generated by the 10-fold cross validation.

Table 4. Results achieved without and with data augmentation.

	<b>Without Data Augmentation</b>	<b>With Data Augmentation</b>
<b>Sentivity (%)</b>	67.33% ( $\pm$ 8.95%)	95.87% ( $\pm$ 1.73%)
<b>Specificity (%)</b>	80.33% ( $\pm$ 9.57%)	97.39% ( $\pm$ 0.88%)
<b>Accuracy (%)</b>	73.83% ( $\pm$ 3.50%)	96.63% ( $\pm$ 0.55%)
<b>AUC (%)</b>	0.808 ( $\pm$ 0.048)	0.992 ( $\pm$ 0.001)

Here, is showed the training and validation accuracy, and training and validation loss acquired in the first iteration of the experiment without data augmentation in the figures 4a and 4b, respectively.



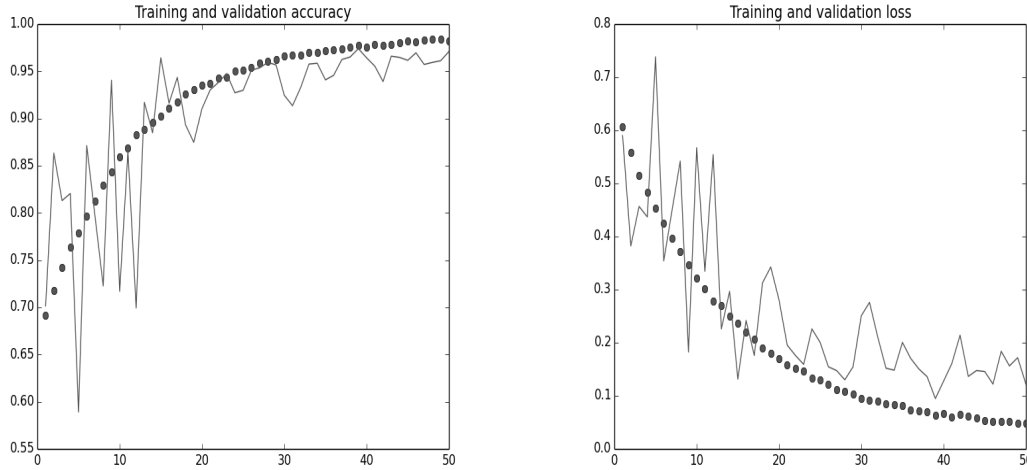
(a) Training (line above) and validation accuracy (line below). (b) Training (line above) and validation (line below) loss.

Figure 4. Loss and validation accuracy in training without data augmentation.

Next, is presented the training and validation accuracy, and training and validation loss acquired in the first iteration of the experiment without data augmentation in the



figures 5a and 5b, respectively. Here we can see that the curves are more closed, breaking the problem of overfitting and improving the validation accuracy.

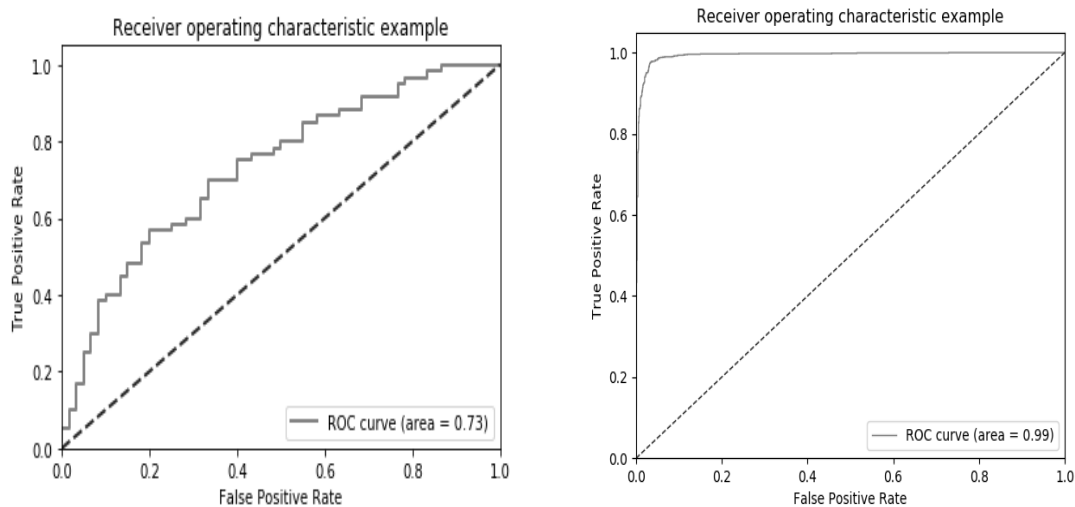


(a) Training (circles) and validation accuracy (line). (b) Training (line) and validation (circles) loss.

Figure 5. Loss and validation accuracy in training with data augmentation.

The CNN applied in data augmentation’s group achieved the best results, a sensitivity of 95.87%, specificity of 97.39%, the accuracy of 96.63% and AUC of 0.992.

To compare the performance between CNN with and without data augmentation is presented the ROC curve between both in the first iteration of the experiment in Figure 6, from which we can view that the use of the CNN with data augmentation is better than using CNN without data augmentation. AUC was of 0.73 without and 0.99 with data augmentation, respectively.



(a) ROC curve without data augmentation. (b) ROC curve with data augmentation.

Figure 6. AUC results.

The CNN proposed enabled us to achieve excellent results in the metrics evaluated because mainly the CNN can extract essential features from the images like texture, shape, intensity, and even features so deep that is unknown of two different dimensions. The number of layers in our CNN architecture was relatively small because of the limitations of the database used, and it was not necessary to modify anything in the CNN architecture due to the increase of the database.

Next, is showed the comparison between our the best result using CNN with the results obtained by [Felix et al. 2016], [Dilger et al. 2015], and our results were better (see Table 5). How Anirudh et al. [Anirudh et al. 2016] evaluated their experiments using free receiver operating characteristic (FROC), a metric not applied in this work, we didn't compare our results with theirs.

Table 5. Comparison with related works

Work	Database (samples)	AUC	Classifier
Felix et al [Felix et al. 2016]	LIDC (274)	0.820 ( $\pm$ 0.053)	MLP
Dilger et al [Dilger et al. 2015]	NIH (50)	0.935 ( $\pm$ 0.0096)	ANN
Proposed paper	LIDC (22,500)	<b>0.992 (<math>\pm</math> 0.001)</b>	CNN

## 5. Conclusion

In this paper, we proposed the use of Convolutional Neural Network (CNN) with data augmentation technique with the objective to improve the accuracy in the classification of lung nodules in early stages with diameter smaller than 11mm using 10-fold cross validation balancing the database during all stages of the experiments. Only the operations (rotation, flip from left to right and flip from top to bottom) were needed to augment our database in a sufficient quantity without distorting the images.

The CNN architecture used achieved an excellent AUC of 0.992, reaching almost the perfection of 1. And our results were better than the related works that also used the AUC as metric to show the results. The main contribution of this work was to show the efficiency of the CNN integrated with data augmentation in the classification of small pulmonary nodules.

As future work, we aim to explore the parenchyma, which is a tissue around the lung nodules, analyze transfer learning techniques, and expand our analysis to semisolid and ground glass opacity nodules beyond the solid nodules to improve the classification of small lung nodules.

The authors would like to thank the financial supporting provided by the Brazilian institution Coordination for the Improvement of Higher Education Personnel (CAPES).

## References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals,

- O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems.
- Alilou, M., Kovalev, V., Snezhko, E., and Taimouri, V. (2014). A comprehensive framework for automatic detection of pulmonary nodules in lung ct images. *Image Analysis Stereology*, 33(1).
- Anirudh, R., Thiagarajan, J. J., Bremer, T., and Kim, H. (2016). Lung nodule detection using 3d convolutional neural networks trained on weakly labeled data. In *Medical Imaging: Computer-Aided Diagnosis*.
- Bartholmai, B., Koo, C., Johnson, G., White, D., Raghunath, S., Rajagopalan, S., Moynagh, M., Lindell, R., and Hartman, T. (2015). Pulmonary nodule characterization, including computer analysis and quantitative features. *Journal of Thoracic Imaging*, 30(2):139–156.
- Blandin Knight, S., Crosbie, P. A., Balata, H., Chudziak, J., Hussell, T., and Dive, C. (2017). Progress and prospects of early detection in lung cancer. *Open Biology*, 7(9).
- Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdal, M., Turcotte, S., Pal, C. J., Kadoury, S., and Tang, A. (2017). Deep learning: A primer for radiologists. *RadioGraphics*, 37(7):2113–2131. PMID: 29131760.
- Chollet, F. et al. (2015). Keras. <https://keras.io>.
- Dilger, S. K. N., Uthoff, J., Judisch, A., Hammond, E., Smith, S. L. M. B. J., John D. Newell, J., Hoffman, E. A., and Sieren, J. C. (2015). Improved pulmonary nodule classification utilizing quantitative lung parenchyma features. *Journal of Medical Imaging*, 2(4):041004–1–041004–10.
- Fchollet, Matsuyamax, Smerity, kemaswill, and Treszkai (2018). Keras examples mnist. [https://github.com/keras-team/keras/blob/master/examples/mnist\\\_cnn.py/](https://github.com/keras-team/keras/blob/master/examples/mnist\_cnn.py/).
- Felix, A. d. L. F., Ferreira Junior, J. R., Oliveira, M. C., and Machado, A. P. (2016). Using 3d texture and margin sharpness features on classification of small pulmonary nodules. In *Proceedings... Conference on Graphics, Patterns and Images*, 29. (SIB-GRAPI), IEEE Computer Society's Conference Publishing Services.
- Ferreira Junior, J. R., Oliveira, M. C., and de Azevedo-Marques, P. M. (2016). Cloud-based nosql open database of pulmonary nodules for computer-aided lung cancer diagnosis and reproducible research. *Journal of Digital Imaging*, 29(6):716–729.
- Gillies, R. J., Kinahan, P. E., and Hricak, H. (2016). Radiomics: Images are more than pictures, they are data. *Radiology*, 278(2):563–577. PMID: 26579733.
- Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., and Lew, M. S. (2016). Deep learning for visual understanding. *Neurocomput.*, 187(C):27–48.
- Hua K-L, Hsu C-H, H. S. C. W.-H. and Y-J, C. (2015). Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoTargets and therapy*, 8:2015–2022.
- Kooi, T., Litjens, G., Ginneken, B., A. Gubern-Mérida, C., Sánchez, Mann, R., Den, H., and Karssemeijer, N. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Medical Image Analysis*, 35:303–312.

- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521:436–444.
- Litjens, G. J. S., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciampi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *CoRR*, abs/1702.05747.
- Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Pérez, J. A., Lo, B. P. L., and Yang, G.-Z. (2017). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, 21:4–21.
- Reeves, A. and Kostis, W. (2000). Computer-aided diagnosis of small pulmonary nodules. *Seminars in Ultrasound, CT and MRI*, 21(2):116–128.
- Samuel G. Armato, Geoffrey McLennan, L. B. M. F. M., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., Max, D., Pais, R. C., Qing, D. P., Roberts, R. Y., Smith, A. R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G. W., Jude, C. M., Munden, R. F., Petkovska, I., Quint, L. E., Schwartz, L. H., Sundaram, B., Dodd, L. E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Castele, A. V., Gupte, S., Sallam, M., Heath, M. D., Kuhn, M. H., Dharaiya, E., Burns, R., Fryd, D. S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B. Y., and Clarke, L. P. (2011). The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans. *Medical Physics*, 38(2):915–931.
- Song, Q., Zhao, L., Luo, X., and Dou, X. (2017). Using deep learning for classification of lung nodules on computed tomography images. In *Journal of healthcare engineering*.
- Sonoda, S. and Murata, N. (2015). Neural network with unbounded activations is universal approximator. *CoRR*, abs/1505.03654.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958.
- Tammemagi, M. C. and Lam, S. (2014). Screening for lung cancer using low dose computed tomography. *BMJ*, 348. PMID: 24865600.
- Yankelevitz, D., Gupta, R., Zhao, B., and Henschke, C. (1999). Pulmonary nodule characterization, including computer analysis and quantitative features. *Radiology*, 212(2):561–566. PMID: 10429718.
- Zappa, C. and Mousa, S. A. (2016). Non-small cell lung cancer: current treatment and future advances. *Translational Lung Cancer Research*, 5(3).
- Zeiler, M. D. (2012). ADADELTA: an adaptive learning rate method. *CoRR*, abs/1212.5701.