Classification of malignancy of lung nodules in CT images using Convolutional Neural Network

Giovanni L. F. da Silva[1]**, Aristófanes C. Silva**[1]**, Anselmo C. de Paiva**[1]**,
Marcelo Gattass**[2]

[1]Núcleo de Computação Aplicada – Federal University of Maranhão (UFMA)
Caixa Postal 65085-580 - São Luís - MA - Brazil

[2]Pontifical Catholic University of Rio de Janeiro (PUC-Rio)
Caixa Postal 22453-900 - Rio de Janeiro - RJ - Brazil

{gioh.lucca, aricsilva, anselmo.c.paiva}@gmail.com

mgattass@tecgraf.puc-rio.br

Abstract. *Lung cancer presents the highest mortality rate, besides being one of the smallest survival rates after diagnosis. Thereby, early detection is extremely important for the diagnosis and treatment. This paper proposes three different architectures of Convolutional Neural Network (CNN), which is a deep learning technique, for classification of malignancy of lung nodules without computing the morphology and texture features. The methodology was tested onto the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), with the best accuracy of 82.3%, sensitivity of 79.4% and specificity 83.8%.*

1. Introduction

Lung cancer is the most frequent malignant tumor and has an increase of 2% per year in its worldwide incidence. In 90% of cases, lung cancer is associated with the consumption of tobacco products. In Brazil, it is estimated an average of 28,220 cases of lung cancer in 2016, where 17,330 male and 10,890 female [INCA 2016].

One of the best opportunities to diagnose lung cancer is when an asymptomatic patient, normally a smoker, undergoes a computerized tomography (CT) exam [Srichai et al. 2007]. The detection of such nodules using CT is not a simple task, since they can have contrasts similar to other structures, low density, and a small size in an area of complex anatomy (can be connected to blood vessels or in the borders of the lung), among other issues [Leef and Klein 2002].

A variety of computer-aided detection techniques have been proposed for the detection and characterization of tumors. The development of such techniques involves two main categories: as computer-aided detection (CAD) and computer-aided diagnosis (CADx). The construction of a CADx system would reduce the number of unnecessary biopsies in patients with benign tumors, preventing physical and mental depression in patients. Thus, CADx acts as a second reader and aids experts for accurate and efficient diagnosis of cancer cells in the earlier stages of disease [Parveen and Kavitha 2014].

Deep learning techniques, in recent years, have highlighted the possibility of automatically uncovering features from the training images and exploiting the interaction,

even hierarchy, among features within the deep structure of a neural network. Accordingly, the issues of feature computing, selection, and integration can potentially be addressed by this new learning framework without a complicated pipeline of image processing and pattern recognition steps [Hua et al. 2015]. Thus, this paper proposes a methodology for classifying lung nodules in whether benign or malignant, using Convolutional Neural Network as features extractor and classifier, which can represent a second opinion to the expert.

The paper is organized as follows. Section 2 presents the related work to this research. Section 3 presents the proposed methodology for the classification of lung nodules. Section 4 presents and discusses the results. Finally, in Section 5 presents final remarks about this paper.

2. Related Work

Various initiatives are frequently developed aiming at increasing the accuracy of lung cancer diagnosis using CADx systems. Hua et al. (2015) proposed a methodology using deep learning techniques, Deep Belief Network (DBN) and Convolutional Neural Network (CNN), with a sensitivity rate of 73.40% and specificity rate of 82.20% using DBN and a sensitivity rate of 73.30% and specificity rate of 78.70% using CNN. Kumar et al. (2015) proposed a methodology using Stacked Autoencoder (SAE), deep learning technique, with an accuracy rate of 75.01%.

The work published by Kuruvilla and Gunavathi (2014) proposed a methodology based on texture features using Artificial Neural Network (ANN), with an accuracy rate of 93.30%. In their part, Gupta and Tiwari (2014) proposed a methodology based on shape features using ANN, with an accuracy rate of 90.00%. Dandil et al. (2014) proposed a methodology based on texture features using Principal Component Analysis (PCA) and ANN, with an accuracy rate of 90.63%. Parveen and Kavitha (2014) proposed a methodology based on texture features using Support Vector Machine (SVM), with a sensitivity rate of 91.38% and specificity rate of 89.56%.

The methodology proposed by Nascimento et al. (2012) based on texture features, uses Linear Discriminant Analysis (LDA) and SVM, with an accuracy rate of 92.78%. Orozco et al. (2013) proposed a methodology based on texture features using Correlation-based Feature Subset Selection, k-Nearest Neighbor (KNN) and SVM, with an accuracy rate of 82.66% and Krewer et al. (2013) proposed a methodology based in a combination of texture and shape features using Correlation-based Feature Subset Selection and KNN, with an accuracy rate of 90.91%.

3. Materials and Methods

The proposed methodology is divided into three steps as described in Figure 1. In summary, the first step is the image acquisition from the images database of CT exams in the LIDC-IDRI database. In the second step, the nodules segmentation is conducted through markings of experts. Finally, the diagnostic is completed using Convolutional Neural Network following of the results evaluation.

3.1. Image Acquisition

The image database used in this work is the LIDC-IDRI [Armato III et al. 2011], which is available online as a result of an association between the Lung Image Database Consor-
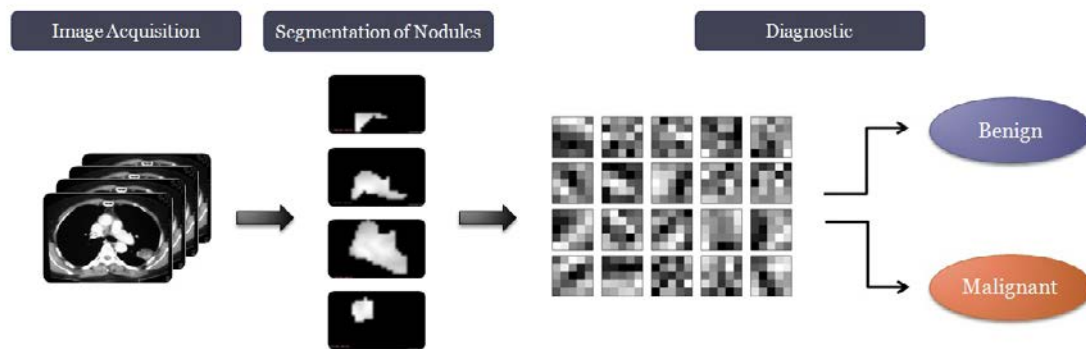
Figure 1. Proposed Methodology.

tium and the Image Database Resource Initiative, including 1,018 CT exams. However, two factors made some of them (185 exams) inappropriate for this methodology. The first factor is related to exams that do not present nodules equal to or larger than 3 mm. The second factor is the divergence of information found in the marking file of an exam versus the information present in the DICOM header of the same exam, which invalidates the marking [Carvalho Filho et al. 2014]. Therefore, the proposed methodology was applied to 833 exams.

The database supplies an XML file with contour information for the slices, and several features including calcification, texture and malignancy with values ranging from 1 to 5 for lung nodules larger than 3 mm. The process of annotating the nodules of the LIDC-IDRI database was performed by four experts in two stages. In the first stage, each expert analyzed the exams individually. In the second stage, the results of the four analyses of the first stage were presented together to the four experts. During this stage, each of the experts re-analyzed the exams and remade their annotations independently [Nascimento et al. 2012].

There is no consensus imposition, all nodules indicated by the revision of experts are taken into account and recorded. Thus, it is possible to have a different diagnosis for the same nodule. This paper considers only one instance per nodule, with the objective of minimizing the impact of subjectivity in exams. The classification of malignant or benign is obtained first with computation as presented in Jabon et al. (2009), which summarizes the features of each nodule as determined by the four experts by computing the mode or the median into a single value. According to the result of this summary, this paper considers that malignant nodules are those cases that present malignancy semantic values of moderately suspicious or highly suspicious and benign nodules are those cases that present characteristics that are highly or moderately indicative of a benign tumor. Regarding contour, the value that contains larger bounds among the four markings made during the annotating process was used. As a total, 1,405 nodules three-dimensional (1,011 benign and 394 malignant) were obtained.

3.2. Nodules Segmentation

With respect to nodules segmentation, information was obtained from a XML file containing the coordinates of the nodules with analysis criterion of each expert. However, the segmentation used in this work is summarized and presented in the Subsection 3.1, in

which only the larger bound was used to represent the instance of the nodules described by up to four markings by experts.

3.3. Diagnostic of Lung Nodules

According to Hua et al. (2015), the physical resolution of CT images on the x and y axes is higher, and three-dimensional features may sometimes be inaccurate due to low resolution on the z axis. Thus, each two-dimensional CT slice is served as individual training sample, resulting in 8,296 nodules (4,329 benign and 3,967 malignant).

To facilitate the training on the CNN architecture, each CT slice is resized to 28 x 28, as illustrated in Figure 2. Nodules with smaller dimensions, Figure 2 (a), were extracted and put into background images of 28 x 28 and nodules with larger dimensions, Figure 2 (b), were rescaled by setting the higher axis in the dimension 28 and reducing the less axis proportionally, finally, they were extracted and put into background images of 28 x 28. Thereon, each CT slice is submitted to the CNN for classification of the malignant or benign nature of lung nodules without computing the morphology and texture features.
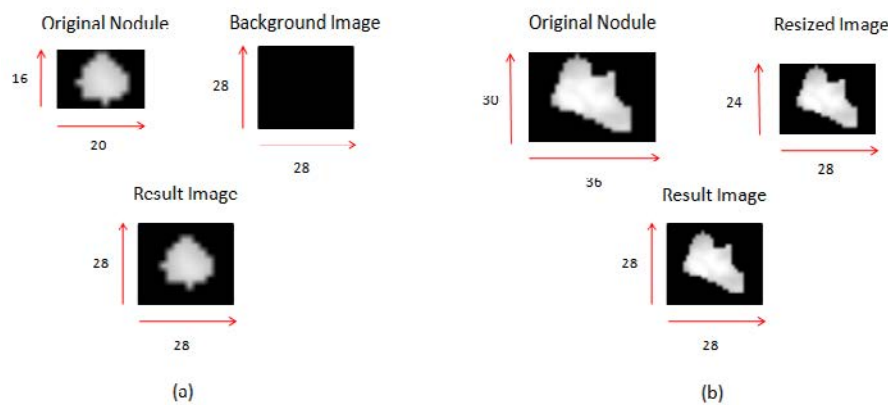

Figure 2. Image Resizing.

3.3.1. Convolutional Neural Network

Convolutional Neural Networks are a biologically-inspired trainable architecture that can learn multi-level hierarchies of features [LeCun et al. 2010]. The network extracts implicitly features of visual patterns presented as input and classify patterns from the extracted features. Typical CNN usually consist of convolutional layers, pooling layers, neuron layers and fully-connected layers [Kang and Wang 2014].

Convolutional layers have trainable filters (also called features maps) that are applied across the entire input [LeCun et al. 1989]. Each filter detects a particular feature that occurs in any part of the input [LeCun et al. 2010]. Pooling layers are non-linear down-sampling layers that yield maximum or average values in each sub-region of input image or feature maps. Pooling layers increase the robustness of translation and reduce the number of network parameters [Kang and Wang 2014]. Neuron layers apply non-linear activations on input neurons. Common activations are softmax function, sigmoid function, rectified linear unit, etc. Fully-connected layers are responsible for classifying patterns presented to CNN.

Supervised training is performed using a form of stochastic gradient descent to minimize the discrepancy between the desired output and the current output of the network [LeCun et al. 2010]. This paper uses cross entropy as the loss in the objective function [Rubinstein and Kroese 2013]. All the coefficient of all the filters in all the layers are updated simultaneously with the back-propagation method [Hecht-Nielsen 1989].

Concerning classification of the malignant or benign nature of lung nodules, three different architectures were built with the purpose of analyzing the performance of each one over the same parameters such as number of epochs, learning rate, filter size, training and testing samples, etc. One used a simple annotation to indicate the layer parameters: C(N) indicates convolutional layer with N filters, P(S) denotes pooling layer with kernel size S, CC(K) indicates fully-connected layer with K neurons, R and S represent rectified linear unit (ReLU) and softmax function, and D indicates dropout layer to help preventing overfitting, one of the problems that occur during neural network training.

Figure 3 shows architecture 1 consisting of a convolutional layer with 32 filters and ReLU activation, another convolutional layer with 16 filters and ReLU activation, a pooling layer with kernel size of 2 and a dropout layer, followed by a fully-connected layer with 16 neurons, a ReLU activation and a dropout layer, on top, a fully-connected layer with 2 neurons and a softmax function.
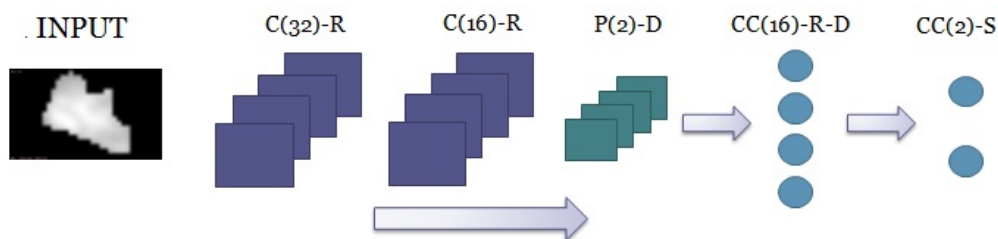

Figure 3. Architecture 1 of CNN.

Architecture 2, described in Figure 4, consists of a convolutional layer with 16 filters and ReLU activation, a pooling layer with kernel size of 2, another convolutional layer with 32 filters and ReLU activation, a pooling layer with kernel size of 2 and a dropout layer, followed by fully-connected layer with 16 neurons, a ReLU activation and a dropout layer, on top, a fully-connected layer with 2 neurons and a softmax function.
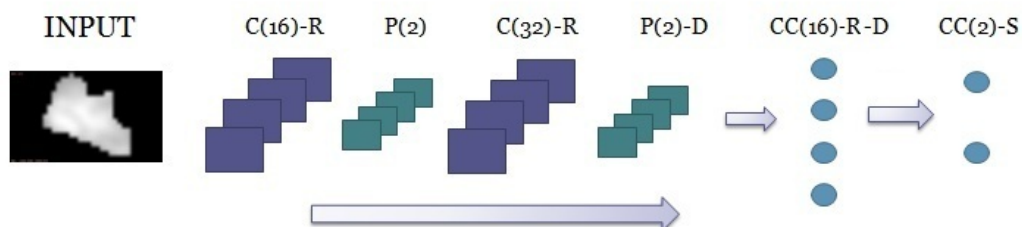

Figure 4. Architecture 2 of CNN.

Finally, architecture 3, presented in Figure 5, consists of a convolutional layer with 16 filters and ReLU activation, a pooling layer with kernel size of 2, another convolutional

layer with 32 and ReLU activation, a convolution layer with 8 filters and ReLU activation, followed by fully-connected layer with 16 neurons, a ReLU activation and a dropout layer, another fully-connected layer with 32 neurons and ReLU activation, on top, a fully-connected layer with 2 neurons and a softmax function.
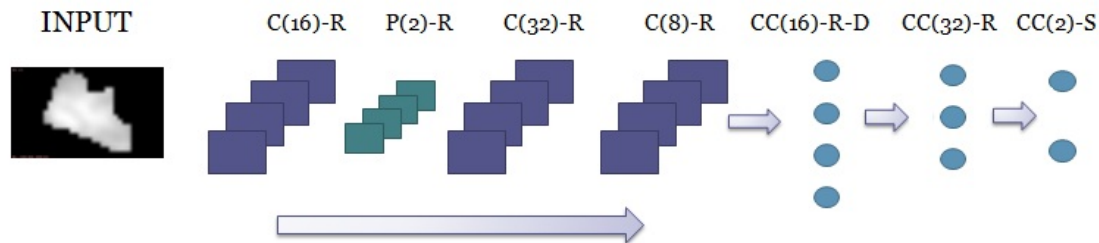

Figure 5. Architecture 3 of CNN.

4. Results and Discussion

In order to evaluate the proposed architectures, one used 7,000 nodules (3,500 benign and 3,500 malignant) for training and 1,296 nodules (829 benign and 467 malignant) for testing. All architectures were trained and tested with the same data set, with learning rate of 0.01, kernel size of 5 in convolutional layer and kernel size of 2 in pooling layer, with training in batch of 30 by 30, for 30 epochs. Table 1 summarizes the performance of accuracy, sensitivity and specificity for each architecture.

Table 1. Results for all architectures.

Models	Accuracy (%)	Sensitivity (%)	Specificity (%)
Architecture 1	80.2	75.2	83.1
Architecture 2	80.2	82.9	78.6
Architecture 3	**82.3**	**79.4**	**83.8**

Architecture 1 obtained accuracy of 80.2%, sensitivity of 75.2% and specificity of 83.1%, architecture 2 achieved the same accuracy of architecture 1, sensitivity of 82.9% and specificity of 78.6%. The best accuracy was obtained by architecture 3 with 82.3%, sensitivity of 79.4% and specificity of 83.8%, we can attribute this to the fact contain more convolution layer, thus generating more features maps, and the presence of a hidden layer in the fully-connected layer.

Regarding sensitivity, success rate of malignant nodules, the architecture 2 is better than other architectures with rate of 82.9%, showing the importance of pooling layer after each convolutional layer. According to Browne and Ghidary (2003), a CNN with convolutional layer followed by pooling layer, the outcome is a gradual increase in the number of features used to describe the image, combined with a gradual decrease in spatial resolution.

In the Figure 6 are presented cases in which all the three architectures of CNN succeeded and failed to the classification. Figure 6 (a) presents a benign nodule that networks classified as benign, Figure 6 (b) presents a benign nodule that networks classified

as malign, Figure 6 (c) presents a malignant nodule that networks classified as malignant and Figure 6 (d) presents a malignant nodule that networks classified as benign. It shows that all the architectures have learned how to classify the nodules in malignant and benign considering most the morphology instead of the texture, thereby causing some false positives and false negatives in the classification.

Figure 6. Examples of Nodules.

Table 2 shows a brief comparison between the results found on this paper and some of the related works. It is important to emphasize that to perform a reliable comparison with these previous works, it should be necessary to use the same image database, same training and testing set, and same settings for the classifiers, among other parameters. Notwithstanding, comparing with the works that used deep learning techniques such as Hua et al. (2015) and Kumar et al. (2015), the present work presents better results. So, it can be attributed to the large amount of images used in training, this is, 7,000 nodules (3,500 benign and 3,500 malignant).

Table 2. Comparison with other publications with respect to the classification of lung nodules in benign and malignant. Accuracy (Acc), Sensitivity (Sen) and Specificity (Spe).

Work	Database (Samples)	Acc (%)	Sen (%)	Spe (%)
Nascimento, 2012	LIDC (73)	92.78	85.64	97.89
Orozco et al., 2012	NBIA-ELCAP (113)	-	96.15	52.17
Krewer et al., 2013	LIDC-IDRI (33)	90.91	85.71	94.74
Dandil et al., 2014	Private (128)	90.63	92.30	89.47
Parveen and Kavitha, 2014	Private (3,278)	-	91.38	89.56
Kuruvilla and Gunavathi, 2014	LIDC (110)	93.30	91.40	100
Gupta and Tiwari, 2014	Private (120)	90	86.66	93.33
Hua et al., 2015	LIDC (2,545)	-	73.30	78.70
Kumar et al., 2015	LIDC (4,323)	75.01	83.35	-
Proposed Architecture 1	**LIDC-IDRI (8,296)**	**80.2**	**75.2**	**83.1**
Proposed Architecture 2	**LIDC-IDRI (8,296)**	**80.2**	**82.9**	**78.6**
Proposed Architecture 3	**LIDC-IDRI (8,296)**	**82.3**	**79.4**	**83.8**

5. Conclusion

Lung cancer remains the leading cause of cancer mortality worldwide, with one of the lowest survival rates after diagnosis. Therefore, early detection greatly increases the chances of improving patient survival. The major contribution of this paper lies in the

application and analysis of three different architectures of CNN, a deep learning technique, in the classification of lung nodules as benign or malignant without computing the morphology and texture features.

The methodology was tested on the LIDC-IDRI database and the best result was achieved by means of the architecture 3 with accuracy of 82.3%, sensitivity of 79.4% and specificity of 83.8%, outperforming the results obtained by other deep learning techniques. Finally, it is important to highlight the possibility of automatically uncovering features from the training images using deep learning technique and still maintain promising results, encouraging new applications of deep learning technique in CADx systems.

References

Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., et al. (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.

Browne, M. and Ghidary, S. S. (2003). Convolutional neural networks for image processing: an application in robot vision. In *AI 2003: Advances in Artificial Intelligence*, pages 641–652. Springer.

Carvalho Filho, A. O. d., de Sampaio, W. B., Silva, A. C., de Paiva, A. C., Nunes, R. A., and Gattass, M. (2014). Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index. *Artificial intelligence in medicine*, 60(3):165–177.

Dandil, E., Cakiroglu, M., Eksi, Z., Ozkan, M., Kurt, O. K., and Canan, A. (2014). Artificial neural network-based classification system for lung nodules on computed tomography scans. In *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*, pages 382–386. IEEE.

Gupta, B. and Tiwari, S. (2014). Lung cancer detection using curvelet transform and neural network. *International Journal of Computer Applications*, 86(1).

Hecht-Nielsen, R. (1989). Theory of the backpropagation neural network. In *Neural Networks, 1989. IJCNN., International Joint Conference on*, pages 593–605. IEEE.

Hua, K.-L., Hsu, C.-H., Hidayati, S. C., Cheng, W.-H., and Chen, Y.-J. (2015). Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoTargets and therapy*, 8:2015–2022.

INCA (2016). Instituto nacional do câncer. tipos de câncer: Pulmão. http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pulmao. Accessed: 2016-02-29.

Jabon, S. A., Raicu, D. S., and Furst, J. D. (2009). Content-based versus semantic-based retrieval: an lidc case study. In *SPIE Medical Imaging*, pages 72631L–72631L. International Society for Optics and Photonics.

Kang, K. and Wang, X. (2014). Fully convolutional neural networks for crowd segmentation. *arXiv preprint arXiv:1411.4464*.

Krewer, H., Geiger, B., Hall, L. O., Goldgof, D. B., Gu, Y., Tockman, M., and Gillies, R. J. (2013). Effect of texture features in computer aided diagnosis of pulmonary nodules in low-dose computed tomography. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 3887–3891. IEEE.

Kumar, D., Wong, A., and Clausi, D. A. (2015). Lung nodule classification using deep features in ct images. In *Computer and Robot Vision (CRV), 2015 12th Conference on*, pages 133–138. IEEE.

Kuruvilla, J. and Gunavathi, K. (2014). Lung cancer classification using neural networks for ct images. *Computer methods and programs in biomedicine*, 113(1):202–209.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.

LeCun, Y., Kavukcuoglu, K., Farabet, C., et al. (2010). Convolutional networks and applications in vision. In *ISCAS*, pages 253–256.

Leef, J. L. and Klein, J. S. (2002). The solitary pulmonary nodule. *Radiologic Clinics of North America*, 40(1):123–143.

Nascimento, L. B., de Paiva, A. C., and Silva, A. C. (2012). Lung nodules classification in ct images using shannon and simpson diversity indices and svm. In *Machine Learning and Data Mining in Pattern Recognition*, pages 454–466. Springer.

Orozco, H. M., Villegas, O. O. V., Dominguez, H. d. J. O., and Sanchez, V. G. C. (2013). Lung nodule classification in ct thorax images using support vector machines. *The 11th International Conference on Information Sciences, Signal Processing and their Applications: Main Tracks*, pages 277–283.

Parveen, S. S. and Kavitha, C. (2014). Classification of lung cancer nodules using svm kernels. *International Journal of Computer Applications*, 95(25).

Rubinstein, R. Y. and Kroese, D. P. (2013). *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*. Springer Science & Business Media.

Srichai, M. B., Naidich, D. P., Müller, N. L., and Webb, W. R. (2007). *Computed tomography and magnetic resonance of the thorax*. Lippincott Williams & Wilkins.