

Detecting Early Stage Lung Cancer using a Neural Network Trained with Patches from Synthetically Generated X-Rays

Alex Chang and Abhishek Moturu

Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

January 4, 2019

Abstract

The aim of this research is to train a neural network to detect early stage lung cancer with high accuracy. Since X-rays are a relatively cheap and quick procedure that provide a preliminary look into a patient's lungs and because real X-rays are often difficult to obtain due to privacy concerns, a neural network can be trained using patches from synthetically generated frontal chest X-rays using Beer's Law on chest X-ray Computed Tomography (CT) scans with and without randomly inserted lung nodules. This provides a large, diverse training dataset that can be used to train a neural network to classify each patch of a given X-ray based on whether or not it contains a nodule. This research project involves bone removal to obtain soft-tissue dual-energy X-rays, lung segmentation to separate lungs within CT scans and randomize nodule placement, nodule generation to grow nodules of various shapes, sizes, and radiodensities, X-ray generation using Beer's Law, image processing to produce realistic and effective X-rays, preliminary neural network exploration to study the feasibility of detecting nodules, sliding windows to create heatmap visualizations of X-rays to highlight areas of high nodule probability, and analyzing these various methods and the results of the neural network to improve accuracy while reducing space and time complexity. This research may be helpful in detecting lung cancer at a very early stage. This report is an extension of our previous report, "Creation of Synthetic X-Rays to Train a Neural Network to Detect Lung Cancer" [1].

1 Introduction

Lung cancer has the highest projected incidence and mortality rate of all cancers and also has one of the lowest 5-year net survival rates [2]. Starting as a group of malignant cells in the lungs, lung cancer grows into and destroys nearby tissues [3]. Groups of these malignant cells form cancerous tumours. Detecting lung cancer early is crucial for a good prognosis since most diagnoses occur in the later stages of the disease, which have higher mortality rates [4]. A solitary pulmonary nodule (SPN) is a lesion in the lungs that is less than 3 cm in diameter. An SPN may be found in up to 0.2% of chest X-rays and may be benign or malignant. As a result, early detection of small (SPN-sized) nodules may reduce the overall mortality rate of lung cancer [5].

Lung cancer can be diagnosed using various medical imaging techniques such as X-rays, X-ray Computed Tomography (CT) scans, PET scans, ultrasounds, and MRI scans [6]. However, this project focuses on X-rays due to their relative affordability and accessibility. X-rays are grayscale images of a certain part of the body that are dark in regions of low radioden-

sity (such as the lungs) and bright in areas of high radiodensity (such as bones) [7]. Since the lungs appear dark in the X-rays, nodules, which are soft tissues with much higher radiodensity, appear visibly brighter, even if only slightly, within the lungs. CT scans are used as manipulable 3-dimensional models of the human body which allow for bone removal, nodule placement, and nodule generation to create synthetic X-rays which can then be enhanced using image processing techniques.

Convolutional Neural Networks (CNNs) have taken off in the past few years due to their immense applicability in various fields, especially Computer Vision. Well-trained CNNs may be able to pick up details that human experts may have trouble discerning and are therefore very useful in the field of medical imaging [8]. Finding a suitable architecture for this particular problem and fine-tuning hyperparameters is crucial in obtaining good results. In previous research, the GoogLeNet CNN ([9]), which was trained to detect 5 different lung abnormalities, achieved an overall accuracy of 91% and the CheXNet CNN [10], which was trained to detect 14 different lung abnormalities, achieved an overall accuracy of 85% (See **Section 3**).

Different aspects of this project have been implemented in the most appropriate programming languages. Converting the DICOM files, which contain the CT scans provided by Dr. Barfett, to manipulable arrays was done in MatLab along with bone removal, nodule insertion, and X-ray generation. MatLab's Image Processing Toolbox was used to enhance contrast, resize, invert, flip, rotate, and save the X-rays. Then, Python's NumPy was used to segment the lungs within the CT scans to randomize nodule placement and generate nodules. Finally, Python's TensorFlow, Keras, and OpenCV were used to create a small neural network and train it to identify whether or not the patches of a given X-ray contained nodules.

2 Synthetic X-Ray Creation

To produce the synthetic X-rays, we employ the techniques described in our initial report [1].

CT scans are used as 3D models of patient bodies within which to simulate nodule growth. To simulate soft-tissue dual-energy X-rays and remove bone obstructions, bone voxels are replaced with water voxels using a thresholding method.

We segment each CT scan using our previously described technique to obtain the positions of lung voxels. Each nodule is then centered at a random location within the lung. To grow the nodule, a radiodensity is randomly picked between 30-150HU and a diameter is randomly selected between 0.5-3cm [11, 12]. Lung nodules start to become visible when they reach a diameter of around 0.5cm and most are clearly visible when they reach a diameter of around 1cm [13, 14], so a diameter of 1-3cm with radiodensities of 30-150HU, 350HU, 500HU, and 1000HU were also explored. The nodule is generated using the lobulated model. All random choices are picked from a uniform random distribution.

The nodule is then inserted into the CT scan at the chosen location by replacing the radiodensity of the voxels with the chosen radiodensity.

Finally, using the Beer-Lambert law [15], rays are traced with the parallel-ray approximation of X-ray generation (point-source is infinitely far away from the patient) to create the X-rays. However, these X-rays appear faint.

To generate synthetic X-rays that appear realistic, multiple image processing techniques were employed. Inverting the colours to obtain the X-ray look (1-img), flipping the image along

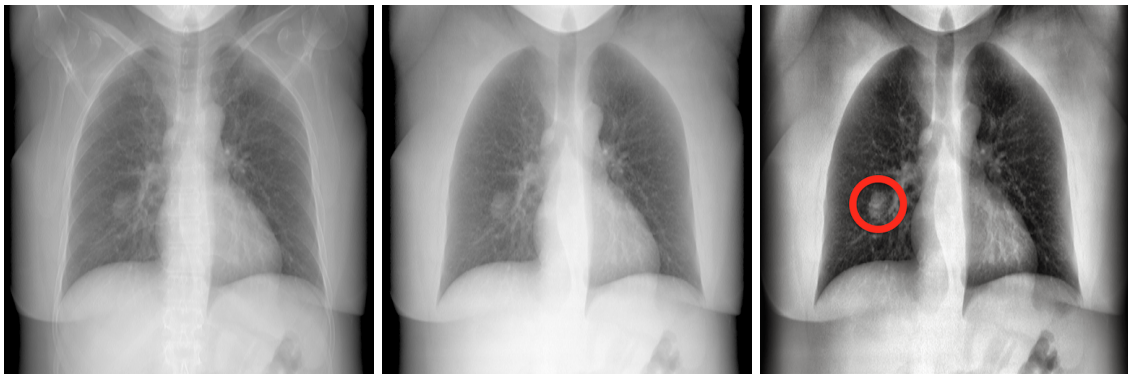


Figure 1: The following are three synthetic X-ray images of the same lungs: standard X-ray (left), soft-tissue dual-energy X-ray (middle), and gamma corrected and contrast-limited adaptive histogram equalized X-ray (right). Note that the synthetic nodule is circled in red.



Figure 2: The following are three real X-ray images of the same lungs: standard X-ray (left), soft-tissue dual-energy X-ray (middle), and gamma corrected and contrast-limited adaptive histogram equalized X-ray (right). Note that the real nodule is circled in red.

the vertical axis so that the heart is on the correct side (flip), rotating the image so the lungs are the right side up (rot90), and resizing the image to 512×512 pixels (imresize, bilinear interpolation) was straightforward.

Uniformly enhancing the contrast of the X-rays such that the lungs appear dark and the tumours appear bright enough to be visible (once pointed out) took multiple experiments. Initially, the X-rays from each CT were adjusted using handpicked values (imadjust) [16]. As this was inefficient, histogram equalization (histeq) [17] was considered to make the contrast enhancement more uniform across all X-rays. However, histogram equalization equalizes globally, which may cause problems when there is a lot of dark space in the X-rays (common for smaller patients) [18]. This issue is solved by using contrast-limited adaptive histogram equalization, or CLAHE (adapthisteq) [19], which performs histogram equalization locally and globally making it more robust to noise. Both the histogram equalization methods highlight faint nodules well, however, to brighten faint nodules even more before equalization, gamma correction (imadjust with gamma parameter) [20] can be used. When $\gamma > 1$, the mapping is weighed towards brighter values, making the lungs darker and the nodules comparatively brighter. See **Figure 1** and **Figure 2** for examples of synthetic and real X-rays.

3 Full X-Ray Classification

Our first approach to classify nodule presence in X-rays was to input the full image to a convolutional neural network. After obtaining poor results from augmenting real datasets with our artificial images, we hypothesized that the lack of diversity from our synthetic X-rays could be the source of the classification inaccuracy. Since several images are produced from the same patient’s CT scan (around 1000 per CT), we speculated that the change in the source point and nodule placement locations would not introduce sufficient variation for the neural network to focus on nodule presence in real X-rays. The classifier seemed to be learning to recognize patient bodies instead of the nodules.

To test our hypothesis, we trained a classifier using only synthetic images. A dataset consisting of 33556 images of dimensions 256×256 , produced from 35 CT scans using the point-source method [1], was used to train a neural architecture similar to AlexNet, a very influential convolutional neural network in the field of Computer Vision. The dataset contained equal numbers of positive and negative examples, and was partitioned into a 70:30 split for training

and validation, respectively. After 25 epochs of training with a batch size of 128, the validation accuracy remained at 50%, which suggested that our method of training was inappropriate for this problem.

Rajpurkar et al. [10] have shown promising results for a similar multiclass classification problem using the ChestX-ray14 dataset. Their model, trained from a DenseNet contains over 5 times the number of layers in AlexNet. However, DenseNets are shown to exploit the advantages of a deep learning while maintaining computational efficiency; their 121-layer model contains under 10 million parameters, versus 60 million in AlexNet. Other subtleties the group from Cornell carefully implemented, which will be explored in the future, include initializing the parameters using a pretrained model, gradually changing the learning rate, and picking optimal choices of training batch size.

4 Creation of X-Ray Patches

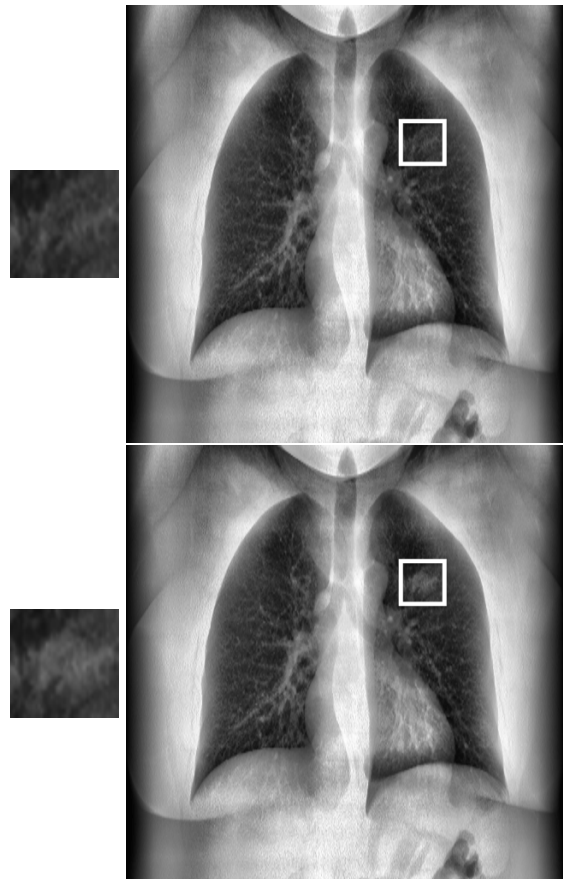


Figure 3: Top: Patch of a synthetic X-ray without any nodules. Bottom: Patch at the same location with a nodule inserted.

Since the full X-ray method was unsuccessful in detecting nodules, we started exploring the idea of classifying patches of the X-ray image based on whether or not they contained nod-

ules. To generate a balanced training data with an equal number of positive and negative examples, a position is first randomly picked within the CT scan to place a nodule. Then, a 64×64 X-ray patch is cropped around that position without nodule insertion. This generates a patch that is a negative example since it does not contain a nodule. Then, the nodule is inserted in the CT in the chosen position and another 64×64 X-ray patch is cropped around that same position. This generates a patch that is a positive example since it contains a nodule. Repeating this process 10,000 times, a total of 20,000 training image patches were created to train the neural network with 50% positive and 50% negative examples. The full X-rays are also saved (See **Section 2**) for testing and reference purposes. See **Figure 3** for examples of patches and the corresponding full X-rays.

5 Sliding Window Detection

As discussed, with poor results from the classification of full X-rays, we explored the possibility of reducing the size of the problem by classifying smaller image patches. Using the parallel-ray tracing method, patches of size 64×64 were cropped from original 512×512 images to produce a large, diverse dataset.

The well-known sliding window method in object detection [21] consists of inputting the patches centered at each pixel of an image into a classifier, hence "sliding" the classification window across the image. When appropriate, the sliding window may skip rows or columns to reduce the number of operations. This allows for a thorough search by determining object presence at almost every location in the image [22].

Although its use is often limited in real-time application due to its computationally inefficient brute-force approach, this method executes in a reasonable time in the context of this project (approx. 5 minutes per image). With the probability of nodule presence of each patch in the image, we build a confidence heatmap, which could assist radiologists in their analysis.

6 Heatmap Visualization

To visualize the results of the neural network, each 64×64 patch of the X-ray (3×3 stride) is traversed and classified using the neural network and assigned a probability of nodule presence.

In **Figures 4-8**, each row contains the results of the neural network trained with X-rays of certain specifications. Each of the three images contains two X-rays: the X-ray on the left contains a nodule (circled in red for clarity) and the X-ray on the right is a heatmap visualization

of the same X-ray with red dots at the centre of every 64×64 patch where there is above a 95% chance of a nodule. Of the three pairs of X-rays in each row, the X-ray on the left was synthetically generated using the CT scan that was used to generate the training data, the X-ray in the middle was synthetically generated using another CT scan, and the X-ray on the right is a real X-ray provided by Dr. Barfett. Note that some classifiers correctly identify the nodule region, but all the classifiers identify several false positives, especially near the edges of the lungs and the heart and near nodule-like shapes.

To get rid of some of the false positives, the intensity of the region was taken into account. Patches that did not fall in a certain average, maximum, and minimum intensity threshold were not added to the heatmap visualization as they likely lie outside the lungs. These thresholds were found through experimentation.

The training data that seemed to provide the best results, in terms of finding the nodules (in real and synthetic X-rays) and not identifying too many false positives, were the one containing nodules of radiodensity 30-150 and diameter 1-3cm (See **Figure 5**) and the one containing nodules of radiodensity 350HU and diameter 1-3cm (See **Figure 6**). The classifiers trained with the other training data failed to properly identify a nodule in the real X-ray that was visible even to untrained human eyes.

7 Neural Network

The learning models were trained using the Keras deep learning framework on Google's servers using Colab. The datasets were shuffled, split 70:30 into training and validation sets, and directly inputted to the neural network without any prior transformations for the purpose of the exploratory phase. Further investigation in data augmentation techniques is required to introduce additional variation.

7.1 Architecture

The simple convolutional neural network architecture that was used to classify patches is shown in **Figure 9, 10**. Over 4 million of the approx. 5 million model parameters are attributed to the connection of the last convolutional layer and first fully connected layer. To reduce this number, fewer filters are used in the last convolutional layer. Additionally, this layer performs convolutions with a stride of 4 along each axis.

With each dataset, the initialized model was trained for 25 epochs, with batch sizes of 128. During the training with each dataset, the model resulting from the epoch with the lowest validation loss (sparse categorical cross-entropy loss)

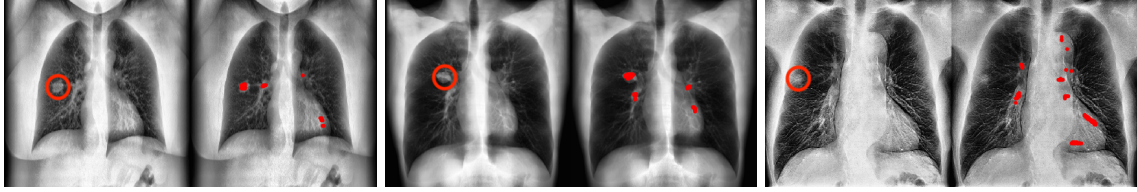


Figure 4: Visualization of the results of the neural network trained with patches containing nodules of radiodensity **30-150HU** and diameter **0.5-3cm**. Left: nodule detected, some false positives. Middle: nodule detected, some false positives. Right: Nodule undetected, several false positives.

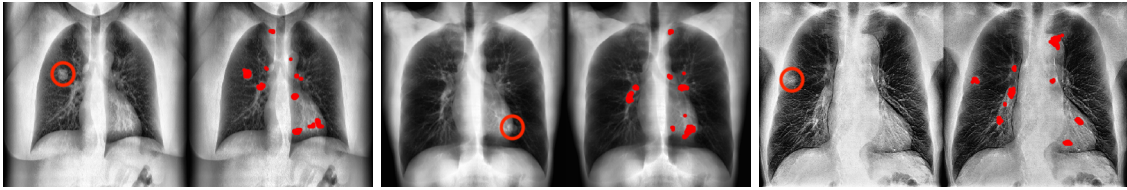


Figure 5: Visualization of the results of the neural network trained with patches containing nodules of radiodensity **30-150HU** and diameter **1-3cm**. Left: nodule detected, several false positives. Middle: nodule detected, several false positives. Right: Nodule detected, several false positives.

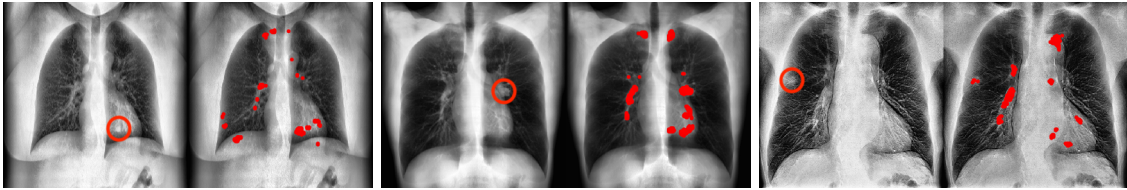


Figure 6: Visualization of the results of the neural network trained with patches containing nodules of radiodensity **350HU** and diameter **1-3cm**. Left: nodule detected, many false positives. Middle: nodule detected, many false positives. Right: Nodule detected, many false positives.

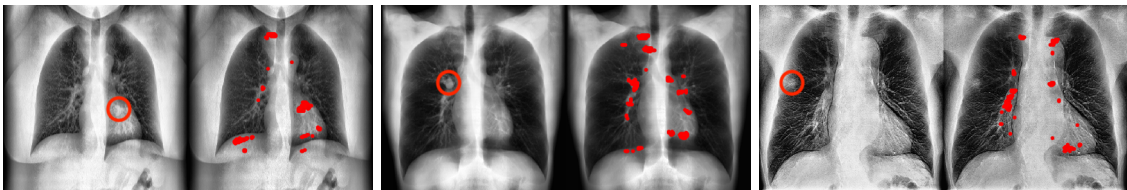


Figure 7: Visualization of the results of the neural network trained with patches containing nodules of radiodensity **500HU** and diameter **1-3cm**. Left: nodule detected, many false positives. Middle: nodule detected, many false positives. Right: Nodule undetected, many false positives.

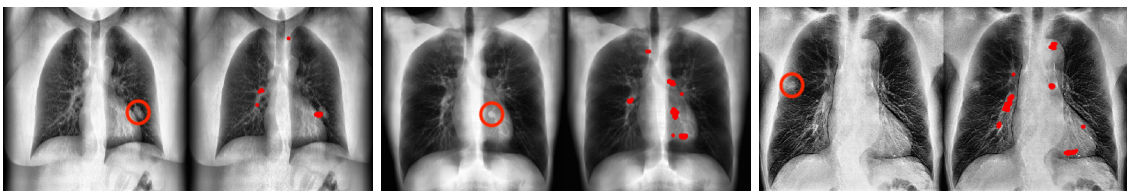


Figure 8: Visualization of the results of the neural network trained with patches containing nodules of radiodensity **1000HU** and diameter **1-3cm**. Left: nodule detected, some false positives. Middle: nodule detected, some false positives. Right: Nodule undetected, several false positives.

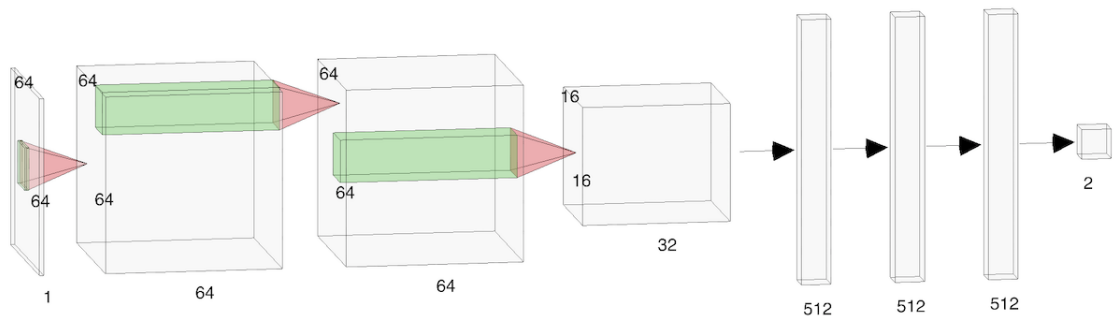


Figure 9: The simple preliminary convolutional neural network used to classify patches. From left to right, the network takes a single-channel (grayscale) image as input and contains 3 convolutional layers, 3 fully connected layers, and a binary output layer. The size of all the filters is 5×5 . Generated using [23].

was saved and its corresponding validation accuracy (percentage of patches correctly classified) is reported in **Table 1**. As seen in **Figure 11**, 25 epochs of training appear to suffice for the model to reach its point of minimal validation loss (approx. around epoch 12). Training for more epochs leads to overfitting.

7.2 Results

As shown in **Table 1**, an increase in validation accuracy is observed with an increase in nodule radiodensity and diameter. However, the validation accuracy does not seem to be a useful statistic when analyzing the number of false positives. In many cases of obvious nodules, the classifier trained with high radiodensity nodules fails to predict real nodule presence in certain patches (false negatives), suggesting the need to

work with low radiodensity nodules. This approach seems to be able to recognize real nodule patches even if training was done only on synthetic nodule patches. Further tests on real X-rays with nodules will be performed on the NIHCC ChestX-ray14 dataset.

Training Data Specifications (radiodensity, diameter)	Validation Accuracy (%)
30-150HU, 0.5-3cm	70.55%
30-150HU, 1-3cm	74.17%
350HU, 1-3cm	88.70%
500HU, 1-3cm	93.08%
1000HU, 1-3cm	97.45%

Table 1: Validation accuracy of the neural network trained and validated using the respective radiodensity and diameter specifications.

Layer (type)	Output Shape	Param #
conv2d_3 (Conv2D)	(None, 64, 64, 64)	1664
conv2d_4 (Conv2D)	(None, 64, 64, 64)	102464
conv2d_5 (Conv2D)	(None, 16, 16, 32)	51232
flatten_1 (Flatten)	(None, 8192)	0
dense_4 (Dense)	(None, 512)	4194816
dense_5 (Dense)	(None, 512)	262656
dense_6 (Dense)	(None, 512)	262656
dense_7 (Dense)	(None, 2)	1026

Total params: 4,876,514
 Trainable params: 4,876,514
 Non-trainable params: 0

Figure 10: A description of the model with the output shape and number of parameters in each layer.



Figure 11: Training (dashed) and validation (solid) accuracy (left) and loss (right) metrics of neural network over 25 epochs trained using patches with nodules of radiodensity and diameter: (1) **30-150HU, 0.5-3cm** (2) **30-150HU, 1-3cm** (3) **350HU, 1-3cm** (4) **500HU, 1-3cm**, and (5) **1000HU, 1-3cm**.

8 Future Work

There are several topics that could be further pursued in the future. Primarily, finding the ideal neural network architecture and corresponding hyperparameters will require many experiments to improve accuracy and reduce run-time during classification.

To increase further variability in the data, we intend on applying data augmentations such as affine transformations, sharpening, Gaussian noise, scaling, salt and pepper methods, flipping, contrast enhancement methods, and more [24]. It may also be helpful to change the ratio of positive to negative examples in the training data by balancing the cases realistically.

In addition, tweaking the size of the patches and sliding windows may help reduce the number of false positives in the heatmap visualizations. Working with an image pyramid, rather than the single image, can also help weed out some of the false positives.

We will run some more experiments to find the least diameter (currently 1-3cm) and radio-density (currently 30-350HU) with which the neural network can be trained and still be able to detect nodules in real X-rays with high accuracy. We hope to a run the classifiers on many more real X-rays to ensure that neural networks trained with synthetic X-rays perform well in detecting nodules in real X-rays.

With regards to image processing, gamma correction and contrast-limited adaptive histogram equalization using MatLab's `imadjust` (with a gamma parameter) and `adaphisteq` provide good results. We will fine tune these parameters further to find the best contrast enhancement that makes the nodules most visible [25] in various resolutions.

Also, we can generate nodules of various shapes and calcification patterns that may also give information on the possible malignancy of a given nodule [26]. Furthermore, other shape-growing algorithms may be helpful in detecting other types of lung abnormalities.

Finally, we hope to analyze the results of the neural network classifier further using metrics such as recall, precision, and F1-score, for a better insight of the classification accuracy [27] and try different loss functions.

Contact: Alex & Abhishek
(le.chang & abhishek.moturu [at] mail.utoronto.ca).

Acknowledgments

We would like to thank those who consented to provide their chest CT scans for research. Thanks also to Hui-Ming Lin for exporting the CT scans. Many thanks to Prof. Ken Jackson & Dr. Joseph Barfett for their continued supervision and support.

References

- [1] A. Moturu and A. Chang, "Creation of synthetic x-rays to train a neural network to detect lung cancer." http://www.cs.toronto.edu/pub/reports/na/Project_Report_Moturu_Chang_1.pdf, 2018. (Accessed on 01/04/2019).
- [2] "Lung cancer - non-small cell: Stages | cancer.net." <https://www.cancer.net/cancer-types/lung-cancer-non-small-cell/stages>. (Accessed on 01/04/2019).
- [3] "What is lung cancer? - canadian cancer society." <http://www.cancer.ca/en/cancer-information/cancer-type/lung/lung-cancer/?region=on>. (Accessed on 01/04/2019).
- [4] "Canadian-cancer-statistics-2018-en.pdf." <http://www.cancer.ca/~media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2018-EN.pdf>. (Accessed on 01/04/2019).
- [5] "The solitary pulmonary nodule. - pubmed - ncbi." <https://www.ncbi.nlm.nih.gov/pubmed/16567482>. (Accessed on 01/04/2019).
- [6] "Diagnosis of lung cancer - canadian cancer society." <http://www.cancer.ca/en/cancer-information/cancer-type/lung/diagnosis/?region=on>. (Accessed on 01/04/2019).
- [7] "X-ray - canadian cancer society." <http://www.cancer.ca/en/cancer-information/diagnosis-and-treatment/tests-and-procedures/x-ray/?region=on>. (Accessed on 01/04/2019).
- [8] J. L. Patel and R. K. Goyal, "Applications of artificial neural networks in medical science," *Current clinical pharmacology*, vol. 2, no. 3, pp. 217–226, 2007.
- [9] M. Cicero, A. Bilbily, E. Colak, T. Dowdell, B. Gray, K. Perampaladas, and J. Barfett, "Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs," *Investigative radiology*, vol. 52, no. 5, pp. 281–287, 2017.
- [10] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *CoRR*, vol. abs/1711.05225, 2017.
- [11] "Hounsfield unit | radiology reference article | radiopaedia.org." <https://radiopaedia.org/articles/hounsfield-unit>. (Accessed on 01/04/2019).
- [12] "Ct basics and pet attenuation." <http://www.people.vcu.edu/~mhcrosthwait/crs322/petctdata.html>. (Accessed on 01/04/2019).
- [13] "Solitary lung nodule symptoms, causes, and treatments." <https://www.webmd.com/lung/solitary-pulmonary-nodule#1>. (Accessed on 01/04/2019).
- [14] "Solitary pulmonary nodule: Overview, types of benign pulmonary tumors, etiology." <https://emedicine.medscape.com/article/2139920-overview>. (Accessed on 01/04/2019).
- [15] T. G. Feeman, *The Mathematics of Medical Imaging*. Springer International Publishing, 2015.
- [16] "Adjust image intensity values or colormap - matlab imadjust." <https://www.mathworks.com/help/images/ref/ijmadjust.html>. (Accessed on 01/04/2019).

- [17] "Enhance contrast using histogram equalization - matlab histeq." <https://www.mathworks.com/help/images/ref/histeq.html>. (Accessed on 01/04/2019).
- [18] "Contrast stretching (image processing)." <https://goo.gl/xK1bo1>. (Accessed on 01/04/2019).
- [19] "Contrast-limited adaptive histogram equalization (clahe) - matlab adapthisteq." <https://www.mathworks.com/help/images/ref/adapthisteq.html>. (Accessed on 01/04/2019).
- [20] "Gamma correction - matlab & simulink." <https://www.mathworks.com/help/images/gamma-correction.html>. (Accessed on 01/04/2019).
- [21] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *NIPS*, 2013.
- [22] "Machine learning is fun! part 3: Deep learning and convolutional neural networks." <https://goo.gl/RgVt3h>. (Accessed on 01/04/2019).
- [23] "Nn svg." <http://alexlenail.me/NN-SVG/AlexNet.html>. (Accessed on 01/04/2019).
- [24] "Overview of augmenters — imgaug 0.2.7 documentation." <https://imgaug.readthedocs.io/en/latest/source/augmenters.html>. (Accessed on 01/04/2019).
- [25] S. Karuppanagounder and K. Palanisamy, "Medical image contrast enhancement based on gamma correction," *International Journal of Knowledge Management and e-Learning*, vol. 3, pp. 15–18, 02 2011.
- [26] "Evaluation of the solitary pulmonary nodule - american family physician." <https://www.aafp.org/afp/2015/1215/p1084.html>. (Accessed on 01/04/2019).
- [27] "Precision vs recall – towards data science." <https://towardsdatascience.com/precision-vs-recall-386cf9f89488>. (Accessed on 01/04/2019).