Automated Detection of Abnormalities in Chest X-Ray Images Using Convolutional Neural Networks

[1]K. Aravinth Raaj, [2]B. Joel Sherwin, [3]S. Sabeetha Saraswathi

[3]Assistant professor

Department of Computer Science and Engineering, TRP Engineering College, Tiruchirappalli, India.

Abstract — *Heart and lung diseases account for 11% of total deaths in India, with the cumulative count of all chest related deaths accounting for almost 30% of the world's mortality rate. A computer-aided diagnosis system is proposed that utilizes Convolutional Neural Networks (CNN) to classify images of chest x-rays into normal or abnormal labels, and if abnormal, attempts to identify the kind of abnormality/disease further. We trained Convolutional Neural Network (CNN) to classify tuberculosis, pneumonia, and cardiovascular abnormalities. The pre-trained neural network GoogLeNet is tuned using a dataset of labeled x-rays, and the model is used to classify an x-ray without prior knowledge in medicine. We extend our work further to classify all 11 diseases.*

Keywords — *neural networks; deep learning; medical imaging; computer-aided diagnosis.*

I. INTRODUCTION

Chest Radiography (chest x-ray) is one of the most widely used medical imaging technologies available today. X-rays are easy to obtain and are far more economical than other means of medical imaging. Radiologists use chest x-ray images to diagnose many diseases, including pneumonia, heart failure, pneumothorax, bone fracture, hiatal hernia, and lung diseases.

Chest x-ray photographing is a non-invasive process that takes only half an hour. Even in undeveloped regions, it is affordable to get modern digital radiography (MR) machines. Despite how easy it is to obtain an x-ray, the challenge lies in diagnosing it. It might often take a trained radiologist several minutes to diagnose a single x-ray and write a detailed report. On top of that, the possibility for misdiagnosis of x-rays is high, with studies saying that as high as 50% of lung nodules are missed. Developing a computer-aided diagnosis system can help in the early identification of fatal diseases. In this work, a model is developed that can automatically detect anomalies in chest X-rays.

Due to the high patient burden and low availability of skilled radiologists, there is often a high human error rate in analyzing the chest x-rays. For an experienced radiologist, it might take several minutes to review an x-ray and write a report. Many of them have to work overtime, increasing misdiagnosis due to exhaustion. A quick and accurate CAD system can help under-developed and developing countries where the number of patients is high and medical services cannot be provided actively.

The recent development of artificial intelligence has opened up new opportunities to build AI-based computer-aided diagnosis systems to automatically analyze medical problems, like automatic vertebrae detection, automatic coronary calcium scoring, lymph nodule (LN) detection interstitial lung disease (ILD) classification. Though these are achieved using high-resolution Magnetic Resonance Imaging (MRI) or Computer Tomography (CT) images, these images provide more details, but plain X-Ray images are widely used.

The challenging task for any CAD system is to analyze the plain X-Ray images. In contrast, the MRI and CT images are layered images that provide more details and produce higher noise ratio levels to make the process easier. Many X-ray image datasets are available openly, but the practical problem is only a few datasets labeled as MRI or CT images. Traditional Computer-aided diagnosis systems are based on handcrafted images and have general processing like enhancement and subtraction techniques, segmentation of lung fields, ribcages, and other structures. They analyze the size, lung nodules, and texture. These CAD systems depend on manually extracted features. It takes a long time for researchers to develop a good set of features, especially for complicated images like x-rays.

Deep convolutional neural networks have gained more interest and popularity for their excellent image recognition, image classification, and semantic segmentation. Convolutional neural networks are applied in many medical image processing tasks.

CNN is an end-to-end network that is easy to train, so they are good for analyzing X-ray images. They do not require any manual feature engineering. Instead, they train on the classification labels and raw images. A computer scientist can build these models without any medical knowledge, with the only requirement being large amounts of training data.

II. EXISTING MODELS

Detection of abnormalities in chest X-ray is difficult and a tedious process. Many researchers have trained a neural network and developed some algorithms to detect chest X-ray abnormalities like Tuberculosis, Pneumonia, Nodules, Cardiopulmonary problems, etc. In most research and studies, chest X-ray images are classified using handcrafted features and classifiers. For detecting the patterns on a chest x-ray, textural and geometrical features are used.

Ginneken et al. [1] were the first to propose a good method to detect tuberculosis automatically. They use chest x-Ray images to analyze the different abnormalities. Subtraction techniques are used to remove normal structures in chest x-Ray images, so the abnormalities can be viewed clearly by a radiologist or to the computer to detect. Hogeweg et al. [2] also proposed a method by combining scores. These lung fields are divided into several regions and texture. For performing classification and calculating the weight K-nearest algorithm is used. Chest x-Ray images are divided into small patches, and these patches are used to obtain the pixel level score by Linear Discriminant Analysis (LDA). The shape abnormality score is obtained by combining Mahalanobis distance at image level and pixel score. This is further improved in which focal, textural, and shape abnormalities are separately analyzed and combined with the TB score [3]. Hooda et al. [4] developed a CAD system for TB detection with 7 convolutional layers and 3 fully connected layers. They use Adam optimizer, which gives higher accuracy, and the model was evaluated using Shenzhen and Montgomery datasets, which are available publicly.

Pranav Rajpukar et al. [4] developed an algorithm that can detect pneumonia from chest x-ray images. CheXNet uses 121-layer convolutional neural networks which trained on frontal chest x-ray images available publically. Only frontal view x-rays are used in this model; accuracy can be increased by up to 15% by including the lateral view [5]. They split databases randomly by the following works of Wang et al. and Yao et al. They compare the per-class AUROC of the model against 13 classes of Yao et al. [5] and 1 class of Wang et al. [6].

Jaeger et al. [10] proposed detecting tuberculosis in which intensity mask, Log Gabor mask, and lung model mask are used for lung segmentation. To find the pathological patterns in chest x-ray images, different shapes and texture descriptions are used. Histogram bins are used for each descriptor to represent its distribution, and the value for every descriptor is considered a feature. CXR images are classified into normal, and abnormal using a classifier called a linear support vector machine (SVM). And also, these authors proposed a method which has two separate features, namely CBIR (Content-based image retrieval) and objects detection based features are used after lung boundary segmentation by graph cut segmentation method. Finally, SVM is used to classify CXR.

Identifying pathologic abnormalities in high-resolution images is quite difficult due to variations in normal anatomic, prior medical interventions such as a pacemaker, pathology, patient orientation, or different imaging techniques. The initial task is to identify the objects in an image; machine learning is well suited for classification. Few algorithms for the CAD system have been incorporated into widespread clinical use [7]. Rajkomar et al. [8] used GoogLeNet with image augmentation and Imagenet to classify images into either frontal or lateral classes with 100 percent accuracy.

III. INTRODUCTION TO DEEP LEARNING

Deep learning involves multiple processing layers of computational models to learn representation data with multiple levels of abstraction. The fields of visual object recognition, object detection, speech recognition, and many other domains such as genomics and drug discovery have significantly benefited. Using backpropagation, deep learning discovers detailed structure in large data sets to show how the machine should compute each layer's representation by changing its internal parameters. Deep convolutional networks have brought breakthroughs in video, image, speech, and audio. Neural networks are widely utilized today by a number of companies for purposes as varied as tagging algorithms by Facebook, reverse image search by Google, and product recommendations by Amazon. Neural networks have gained popularity due to companies having access to an increasingly large amount of data. The more data available, the better network can be trained.

Neural networks are modeled based on the human brain. The basic blocks of a neural network are neurons. In biological neurons, there are a number of inputs (dendrites), processor (nucleus), and an output (axon). When the neuron is triggered, it accumulates all its input, and it fires a signal through the axon when the threshold value goes over. An artificial neuron is based on the same principles; they are activated when the activation function's value goes above a certain threshold value.

$$f(x)=max(0,x)$$

ReLU

$$f(x)=\frac{1-e^{-2x}}{1+e^{-2x}}$$

Hyperbolic

$$f(x)=\frac{1}{1+e^{x}}$$

Sigmoid:

Examples of activation functions

Backpropagation algorithms are one of the more popular methods to train a neural network. The input is given to a network that will produce an output, and then the network is given with correct output for the given input (ideal output). The network takes these ideal outputs to produce more accurate output next time by adjusting the weights, going backward to reach the output layer's input layer. The next time the same output is given to the network, it will produce an output closer to the ideal one. The training process is repeated

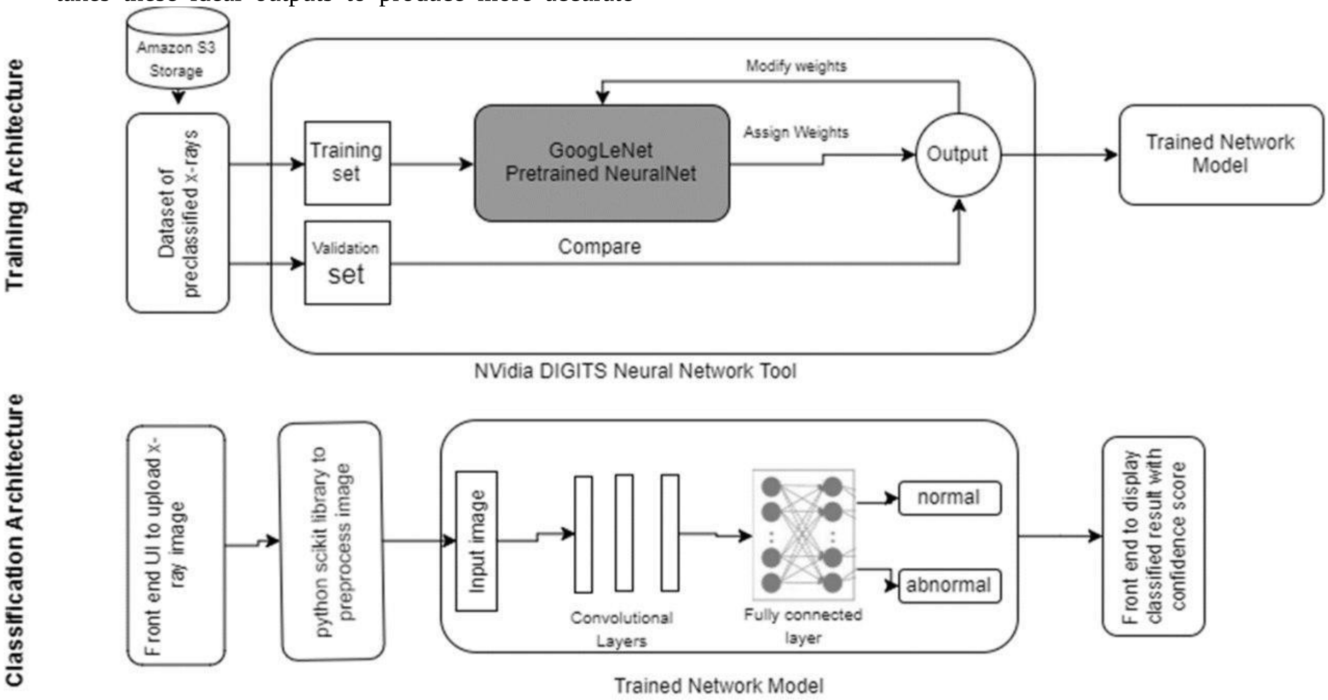
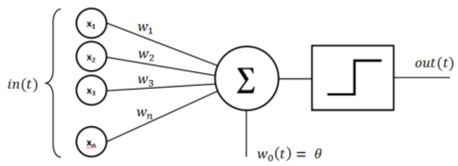


Fig. 1: Architecture of the CAD system

Over many epochs until the error between the ideal output and the network's output are small enough.



Representation of a neuron

Neurons produce output using input from many other nodes or external sources. Every input has its weight (w), which is assigned based on relative importance to the inputs. The block of nodes is called a layer. The input layer just passes information to the hidden layers without performing any computational process. The intermediate and computation processes are done on the hidden layer, and the weight signals or information are transferred. Finally, the output layer uses an activation function that maps to the desired output format. A rule called learning rule modifies the parameters of the network. This process typically amounts to modifying the weights and thresholds

Convolutional Neural Networks are similar to ordinary neural networks because they also consist of neurons that also can learn weights and biases. The biological visual cortex's architecture inspires the architecture of neurons in a convolutional neural network. This has numerous advantages over traditional neural networks in computer vision, as CNN's consistently outperform other networks in jobs relating to image classification and object recognition. The neurons in a CNN respond to desirable features or stimuli in a restricted region of the image known as the receptive field. Receptive fields partially overlap, over-covering the entire visual field. In a convolutional layer, a filter moves over the receptive fields of the images one by one. A unit response can be approximated mathematically by a convolution operation. CNN's require large amounts of data to train on. Aside from computer vision, CNNs have a wide range of applications, ranging from recommender systems to natural language processing.

Deep learning came into existence due to the over-dependence of handcrafted features. Deep learning first evolved from the biological experiment about the study of the cortex of a cat. In 1960 Hubel and Wiesel observed that the cat's brain is activated by the rectangular topology of cells for a particular draw line orientation. Subsequently, they spring up with the idea of hierarchical patterns where greater details are composed to lower. Later, Fukushima proposed a layered network based on unsupervised learning. In 1998 Lecun [9] proposed LeNet architecture for recognizing handwriting characters using supervised learning techniques. The potential of CNN architecture cannot be completely explored because LeNet is a small network. After graphical processing units (GPU) became affordable, the true potential of deep learning is realized. Krizhevsky et al. [9] in 2012 worked on Imagenet datasets, which first explored the depth of deep learning. AlexNet architecture was proposed by Krizhevsky, which is trained over millions of datasets of Imagenet to classify them into thousand categories.

CNN is the most commonly used in deep learning techniques, and it is mainly used for image classification. CNN is based on the idea of sharing weights between different layers and local spatial connectivity. From the base of the network of the convolutional operator, they derive their names. These operators are used to extract the features of the images. Every image is applied with multiple filters in each convolutional layer. Feature maps obtained by matrices filters or kernel over the complete image are considered the convolutional layer's output. Each convolutional layer's output depends on the number of filters, padding, stride, and filters' size. This layer preserves the linear and spatial relationships between pixels.

ReLu operator is applied after each convolutional layer to include nonlinearity in the network. Negative values in a feature map are replaced with zero by an element-wise operator. There are alternative nonlinear functions like sigmoid and tanh, but the ReLu function is better than others in most situations. The network pooling step is performed after the convolutional steps to reduce each feature map's size. Max pooling and average polling functions are most commonly used. The largest element is selected from the window in max pooling, and the average of all elements is taken in average pooling. Reducing the number of parameters and computation in the network avoids overfitting by pooling. This also helps to detect objects of any size. Fully connected layers are used at the end of the network, similar to the traditional neural network. In this layer, every neuron in the previous layer is connected to the current layer. The purpose of this layer is to classify the input images into various classes. After these processes, softmax is used to obtain the final output.

IV. PROPOSED METHOD

A dataset consisting of 2706 x-rays was sourced from NIH, of which 1086 were normal, and 1860 were abnormal. The label of each x-ray was extracted from its metadata through API calls to the NIH website. Once downloaded, the images were split into folders based on their labels for training purposes. As training, a neural network from scratch is a time and compute-intensive process. It was decided to use a pre-trained neural network and perform transfer learning by continuing the back propagation method.

The trained neural network GoogLeNet was chosen to be fine-tuned using the sourced dataset. GoogLeNet has 22 layers, including 9 Inception modules, 2 convolutional layers, and three pooling layers. This was one of the first CNN architectures that strayed from the general approach of simply stacking convolution and pooling layers on top of each other in a sequential structure. GoogLeNet places notable consideration on memory and power usage. The network is 22 layers deep when counting only layers with Parameters. The overall number of layers used for the construction of the network is about 100. However, this number depends on the machine learning infrastructure system used.

The training was performed using the Caffe deep learning framework with the NVIDIA DIGITS Interactive Deep Learning System as a front end. Due to speed and time issues on the initial training, which was done on the local system, the training process was migrated over to an AWS EC2 g2.2xlarge instance, with an Intel Xeon E5-2670 CPU and an NVIDIA Kepler GK104 GPU and 15GB of RAM. This led to drastic improvements in training speed. Once the dataset was uploaded and training was completed, we were able to achieve an accuracy of 0.73 without any image pre-processing or manual modification to the network

type	patch size/ stride	output size	depth
convolution	$7\times7/2$	$112\times112\times64$	1
max pool	$3\times3/2$	$56\times56\times64$	0
convolution	$3\times3/1$	$56\times56\times192$	2
max pool	$3\times3/2$	$28\times28\times192$	0
inception (3a)		$28\times28\times256$	2
inception (3b)		$28\times28\times480$	2
max pool	$3\times3/2$	$14\times14\times480$	0
inception (4a)		$14\times14\times512$	2
inception (4b)		$14\times14\times512$	2
inception (4c)		$14\times14\times512$	2
inception (4d)		$14\times14\times528$	2
inception (4e)		$14\times14\times832$	2
max pool	$3\times3/2$	$7\times7\times832$	0
inception (5a)		$7\times7\times832$	2
inception (5b)		$7\times7\times1024$	2
avg pool	$7\times7/1$	$1\times1\times1024$	0
dropout (40%)		$1\times1\times1024$	0
linear		$1\times1\times1000$	1
softmax		$1\times1\times1000$	0

Fig. 2: GoogLeNet incarnation of the Inception architecture

GoogLeNet has twenty-two layers: two convolutional layers, four max pool layers, nine inception layers, and one average pool layer. The convolutional layer is the building block of a convolutional neural network. These layers comprise a set of independent filters. These filters are independently convolved with image producing feature maps. The pooling layer is another building block of the convolutional neural network. It operates each feature map individually that reduces the representation's spatial size to reduce the number of parameters and computation in the network. The fully connected layer is the final layer of the convolutional neural network, in which all the activations in the previous layer are connected to neurons.

Neurons are mathematical functions. These neurons' function is to collect the input and multiply each input value by value to calculate the weight. When the threshold value of the input signal limit is exceeded, it fires an output signal. Then the output fed to other preceptors.

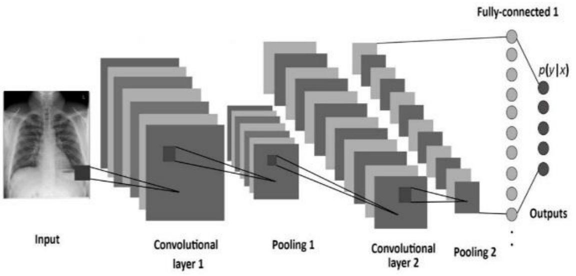


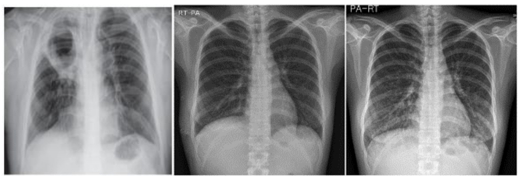
Fig. 3 Visualization of working of GoogLeNet

CNN's have had the standard architecture of stacked convolutional layers, sometimes followed by max-pooling and contrast normalization layer, and one or more fully connected layers, ever since Lenet-5. For larger datasets, the solution has been to increase the number of layers. Despite the argument that the max-pooling layer may result in loss of accuracy in terms of spatial information, the same architecture has been employed for a number of applications, ranging from object detection to human pose estimation. GoogLeNet takes a different approach to the architecture by taking cues from the primate visual cortex. The network comprises a number of "inception modules," each having a 1x1, 3x3, and 5x5 convolutions, along with max pooling. This approach was proved to be highly effective as it won the classification and object recognition challenges in the ImageNet Large Scale Visual Recognition Competition 2014.

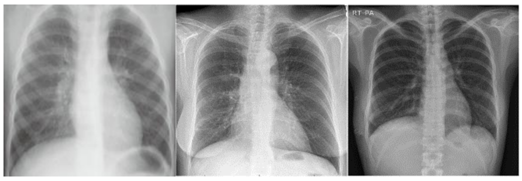
V. EXPERIMENTAL RESULTS

Results are evaluated by training and testing the proposed architecture on an openly accessible repository of labeled, anonymous x-rays images, made available by the National Institute of Health (NIH). 2706 random x-rays were chosen for training, with 1086 being normal and 1860 of the labeled with some abnormal diagnosis. A dataset of 275 x-rays of patients afflicted by tuberculosis was obtained from the Shenzhen Hospital X-ray dataset.

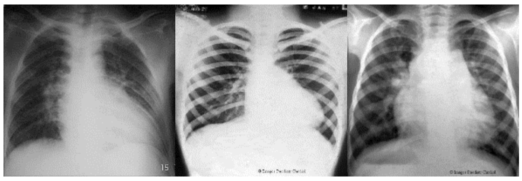
a) Normal x-rays



b) X-rays with tubercluoses



c) X-rays with pneumonia



d) X-rays with heart disease

Fig.4 Sample datasets

Out of 3257 images, 2442 CXRs (75%), which are randomly chosen, are used for training while the remaining 815 images (25%) are used for validation purposes. Accuracy is the ratio of the number of correctly classified CXRs s to the total number of CXRs in the validation set. The training images are divided into batches of size 50. The architecture is trained for 30 epochs. Weights in each layer are initialized randomly using zero-mean with a standard deviation of 0.01, and the initial biases are also random. The learning rate was set to 0.01. The training process was carried out on an AWS EC2 g2.2xlarge instance.

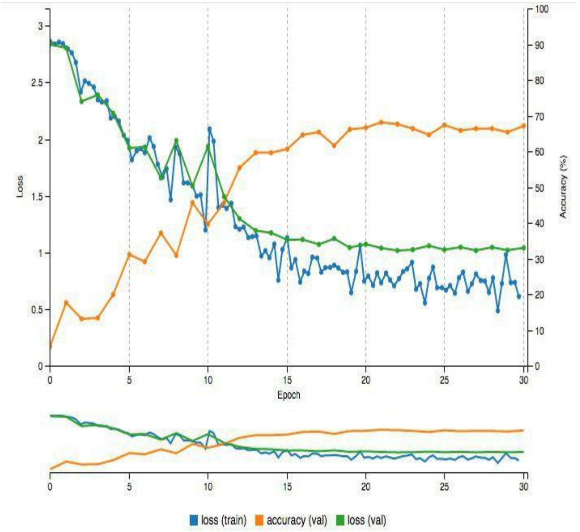


Fig: 5 Accuracy over multiple epoch

Without any alterations made to the network for this specific job and only processing the raw x-ray images themselves, we achieved an overall accuracy of 69%.

VI. CONCLUSION

In this paper, preliminary work is carried out to evaluate a CNN (GoogLeNet) performance in the task of classifying an x-ray image. The pre-trained model was fine-tuned using openly available datasets of x-rays with no domain knowledge. The initial evaluation of the model shows that it performs with reasonably good accuracy. With more research, and automated diagnosis system can be built that matches or exceeds diagnosis accuracy rates of practicing radiologists, leading to great improvements in regions of the world where access to trained medical personnel is limited.

VII. FUTURE WORK

A future step would be using larger datasets to attempt to increase the classifications' accuracy. A pre-processing module may be implemented to optimize the raw x-rays before training and testing. The number of labels may be increased to allow the model to diagnose an abnormality more specifically. Furthermore, modifications may be made to the network architecture itself to increase the accuracy experimentally.

REFERENCES

[1] B. van Ginneken, S. Katsuragawa, B. M. ter Haar Romeny, K. Doi, and M. A. Viergever, "*Automatic detection of abnormalities in chest radiographs using local texture analysis*," IEEE transactions on medical imaging, vol. 21, no. 2, pp. 139–149, 2002.

[2] L. Hogeweg, C. Mol, P. A. de Jong, R. Dawson, H. Ayles, and B. van Ginneken, "*Fusion of local and global detection systems to detect tuberculosis in chest radiographs,*" in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2010, pp. 650–657.

[3] Hooda, R., Sofat, S., Kaur, S., Mittal, A., & Meriaudeau, F. (2017). Deep-learning: "*A potential method for tuberculosis detection using chest radiography*," IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 497–502, 2017.

[4] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ng, A. Y. (2017). "*CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*," in arXiv: 1711.05225v2 [cs.CV], 2017.

[5] Yao, L., Poblenz, E., Dagunts, D., Covington, B., Bernard, D., & Lyman, K. (2017). "*Learning to diagnose from scratch by exploiting dependencies among labels*," in arXiv: 1710.10501 [cs.CV], PP.1–12, 2018

[6] Raoof, S., Feigin, D., Sung, A., Raoof, S., Irugulpati, L., & Rosenow, E. C. (2012). "*Interpretation of plain chest roentgenogram*," IEEE Computer vision Foundation, Pp. 545–558, 2012

[7] Schalekamp S, Van Ginneken B, Karssemeijer N, Schaefer-Prokop C. "*Chest radiography: New technological developments and their applications. Seminars in respiratory and critical care medicine*". Thieme Medical Publishers, 2014

[8] Rajkomar, A., Lingam, S., Taylor, A. G., Blum, M., & Mongan, J. (2017). "*High-Throughput Classification of Radiographs Using Deep Convolutional Neural Networks*," *Journal of Digital Imaging*, pp.95–101, 2017.

[9] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," 2009

[10] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani et al., "*Automatic tuberculosis screening using chest radiographs*," IEEE transactions on medical imaging, vol. 33, no. 2, pp. 233–245, 2014.

[11] Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

[12] T. Xu, I. Cheng, R. Long, and M. Mandal, "*Novel coarse-to-fine dual scale technique for tuberculosis cavity detection in chest radiographs*," EURASIP Journal on Image and Video Processing, vol. 2013, no. 1, p. 1, 2013.

[13] A. Karargyris, J. Siegelman, D. Tzortzis, S. Jaeger, S. Candemir, Z. Xue, K. Santosh, S. Vajda, S. Antani, L. Folio, et al., "*Combination of texture and shape features to detect pulmonary abnormalities in digital chest x- rays*," International Journal of computer-assisted radiology and surgery, vol. 11, no. 1, pp. 99–106, 2016.

[14] Sagnik M, Ramaprasad P, "*Comparative Study of Convolutional Neural Networks*" SSRG International Journal of Electronics and Communication Engineering 6.8 (2019): 18-21.

[15] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "*Automatic coronary calcium scoring in cardiac ct angiography using convolutional neural networks,*" Medical Image Analysis, vol. 9349, pp. 589–596, 2016.

[16] H. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. Summers, "*Improving computer-aided detection using convolutional neural networks and random view aggregation.*" IEEE Transactions on Medical Imaging, 2015

[17] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "*Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics, and transfer learning*," IEEE Transactions on Medical Imaging, vol. 35, no. 5, p. 1285, 2016

[18] S. Yang, W. Cai, H. Huang, Z. Yun, W. Yue, and D. D. Feng, "*Locality-constrained subcluster representation ensemble for lung image classification*," Medical Image Analysis, vol. 22, no. 1, pp. 102–113, 2015.

[19] ChangmiaoWang, Ahmed Elazab, JianhuangWu, and Qingmao Hu. Lung Nodule Classification Using Deep Feature Fusion in Chest Radiography. Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society, 57:10– 18, Nov 2017.

[20] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae,

Peter Prettenhofer, Alexandre Gram- fort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Ga¨elVaroquaux. API design for machine learning software: experiences from the scikit-learn project. In ECML PKDD Workshop: Languages for Data Mining and Machine Learning, pages 108–122, 2013.

[21] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. CoRR, abs/1512.00567, 2015.

[22] P. Maduskar, R. H. Philipsen, J. Melendez, E. Scholten, D. Chanda, H. Ayles, C. I. S´anchez, and B. van Ginneken, "*Automatic detection of pleural effusion in chest radiographs*," Medical image analysis, vol. 28, pp. 22–32, 2016.

[23] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Du- mitru Erhan, Vincent Vanhoucke, and Andrew Ra- binovich. Going deeper with convolutions. CoRR, abs/1409.4842, 2014.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. CoRR, abs/1512.03385, 2015.