
Lung Nodule Classification Using Convolutional Neural Network

Why was this study ?

As cancer, one of the leading cause of death worldwide, with lung cancer being the second most significant diagnosed cancer in both men and women in US [ref] and the dismal five year survival rate of 16% is in part due to lack of symptoms during early stages and lack of effective screening test until recently [ref]. Hence detection at the earliest may decrease the mortality rate. Tumor staging based on coarse and discrete stratification will determine the patient's prognosis in lung cancer. Chest X-rays and sputum cytology, the potential screening tests for lung cancer conclusively proven to be of no value. Subsequently number of studies compared computed tomography (CT) with the chest X-ray helps in identifying lung cancer at the earliest. Radiographic medical images, as shows how patient's inside looks, offers specific information about the changes caused and growth of tumor, helps radiologists in evaluation of prognosis in lung cancer. Later trials have focused on low-dose CT (LDCT) as screening tool. Even-though role of LDCT has established, issues of high false positive rates, radiation risk and cost effectiveness still need to be addressed.

However it seems that the incidence of lung cancer and resulting mortality, fortunately decreasing in both men and women. In the United States in 2011, 115060 and 106070 new cases of lung cancer were seen in men and women respectively. The number of deaths in 2011 was estimated 156940: 84600 in men and 71340 in women. However in men, this represents a continuing decline in incidence and mortality. It occurs by 1980s. Decline is primarily due to the decrease in cigarette consumption, as 80% of lung cancer deaths are caused due to smoking. After long 10 decades by 1990s it seems to be decline in women also. Small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) with the latter category comprising several histological subtypes, including squamous cell cancer, adenocarcinoma and large cell cancer are the major cell types of cancer [ref].

In 1968, Wilson and Junger established the principles of screening for the World Health Organization (WHO) [ref]. The ideal screening test should pose little risk to the patient, should be sensitive for detecting disease at its earliest with least false positives, but acceptable to the patient, and relatively inexpensive to the health system [ref]. The search for the lung cancer screening test were started in 1960s. Early results were promising, but all the tests used had inherent biases. Recent advances in radiomics through applications of artificial intelligence, computer vision and deep learning techniques allows the extraction of numerous quantitative features with minimum pre-processing from radiographic images and solves most the issues and work as ideal screening tool.

The most significant of these biases were lead time, length time, and over diagnosis bias [ref]. The time between early diagnosis with screening and the time in which diagnosis have made without screening is called lead time. The intention of screening is to diagnose a disease at the earliest even before diagnosing the disease without screening. But the early diagnosis cannot guarantee the prolonged life of a person but can affect interpretation of the five-year survival rate [ref]. Length time bias helps in the apparent improvement of survival rate when that

improvement is actually due to selective detection of cancers with a less progressive course while missing cancers that have the most rapidly progressive course. It gives the impression that detecting cancers by screening reduces the dangerous effect of cancer, thus reduces the mortality rate [ref]. Over diagnosis is the diagnosis of disease that will never shows any symptoms or death during ordinarily expected lifetime of patient [ref]. It is actually a side effect of screening for early forms of disease. Even-though screening saves lives, sometimes there is a chance to cause harm to the patient's life due to unnecessary treatments.

Why Computed tomography screening?

The interest in CT as a screening tool developed when CT technology evolved and made it possible to get good images in single breath hold time with less radiation exposure [ref]. CT scan images are combination of a series of X-ray images taken from different angles around the body and uses computer processing to create cross-sectional images of bones, blood vessels and soft tissues inside the body. Hence provide more detailed information than plain X-rays do. It is able to detect very small nodules present in the lungs. Conventional CT was not ideal for screening as radiation exposure was 7 milliSieverts (mSv) and scan time was long [21]. Low-dose CT (LDCT) reduced the radiation exposure to 1.6 mSv in the NLST trial [22]. Low-dose CT delivered images with excellent resolution to detect nodules of 0.5 cm to 1 cm size. Low-dose CT is comparable in sensitivity and specificity of lung nodule detection with the conventional CT mode. The first report was from Kaneko et al., who screened 1369 high-risk participants with both LDCT and chest radiography [23]. CT detected 15 cases of peripheral lung cancer, while 11 of these were missed on chest radiography. Of the non-small cell carcinomas identified, 93% were stage I [23]. Sone et al. authored the second report in the literature with 3958 participants screened with both LDCT and X-ray [24]. Only 4 lung cancers were detected by X-ray, whereas 19 were seen on CT, of which 84% were stage I at resection [24].

More malignant and benign nodules were detected with the LDCT scan when compared with X-ray. LDCT detected about 4 times more lung cancers than X-rays do. CT screening for lung cancer detects more cancers and early disease.

Why deep learning technique?

Artificial intelligence (AI) with its recent development in digitized data acquisition, machine learning and computing infrastructure gradually changes the medical practice. Applications of AI is expanding to the areas of human expert's province. The latest advancements in AI is overwhelming, but it leads to two popular concepts machine learning and deep learning. Due to the supremacy both in terms of accuracy and feature extraction, when trained with huge amount of data deep learning technique gains more popularity.

As a way of making machines intelligent, in every sector machine learning has become necessary. Machine learning is a set of algorithms which parse data and learn from the parsed data, then make decisions from the learned data. Deep learning, subset of machine learning, due to its hierarchical nature achieves great power and flexibility in learning data. It is able to represent the world as nested hierarchy of concepts, with each concepts defined it will relate it to more simpler concepts and helps in more abstract representations in terms of less abstract ones [ref].

Using hidden layer architecture, deep learning technique learn categories incrementally, low level features at the lower layer and high level features at higher layer. Deep learning technique requires high-end machines as compared with traditional machine algorithms. GPU has become an integral part to execute deep learning algorithm.

In order to reduce the complexity of the data and convert the data to a form suitable to accept by the algorithms, machine learning techniques needs to identify the domain expert. But the deep learning algorithm eliminates the need of domain expertise and hard core feature extraction, as it learns high-level features from the data in an incremental manner. Deep learning techniques has the ability to solve the problem end to end, but in the case of machine learning, needs the problem statements to break to different parts to solve and their results should combine at the final stage.

Deep learning algorithm takes long time to train due to large number of parameters, but traditional machine learning requires only few seconds to few hours to train, but the scenario gets reversed in the case of testing phase. At test time deep learning algorithm takes very less time to run, but the test time increases on increasing the size of the data. Even-though deep learning algorithm takes time, the accuracy is ideal and it works similar to human brain with excellent performance. Hence deep learning technique have permeated the entire field of medical image analysis.

Dataset

To conduct the study and to develop the system to classify the images as nodule (cancerous) and non-nodule (non-cancerous) images, sample set of nodule and non-nodule images are required. Sample images of nodules and non-nodules are from the Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) image collection. This database consists of diagnostic and lung cancer screening thoracic CT scans with annotated lesions [ref]. It contains scan of high risk patients with independent annotations of four experienced radiologists, where a final annotation is made when three of four radiologists independently agree on that lesion. In LIDC-IDRI database lesions are classified in to three categories: nodules $> 3\text{mm}$, nodules $= 3\text{mm}$. In this study nodules $> 3\text{mm}$ and non-nodules $\geq 3\text{mm}$ are considered to classify the image as cancerous or non-cancerous. The total dataset contains scans of 1012 patients, from which 1200 scan images are used for this study. The LIDC-IDRI dataset is publicly available which makes the study reproducible.

All the CT scans available in the LIDC-IDRI dataset are in the MetaImage (mhd/raw) format, which is a test-based tagged file format for medical images [ref]. Each .mhd file is stored with separate .raw file that contains all the voxel data. Each CT scan consists of a cross-sectional slices of the chest. Every cross-sectional slice is a two dimension image of 512 by 512 pixels and are called x and y dimensions respectively. Every slice on the pixel contains Hounsefield Unit (HU) value [ref]. HU values are a measure of radio density and are commonly used in CT scans to express the values in standardized and convenient form. The HU value ranges from [-3024; 3071]. Different substances in a human body produces different HU values, hence helps in the classification of images.

Every CT scan images in the LIDC-IDRI dataset are inspected in a two-phase annotation process by the experienced radiologists [ref]. In the initial phase each radiologists independently

marks the lesions to one of the following three categories: nodules > 3mm, nodules = 3mm. in the second phase each radiologists compares their own marks with the anonymized marks of other radiologists. A final annotation is marked when three of the four radiologists agree on a lesion [ref].

Study Analysis

In-order to reduce the mortality rate and increase the survival rate among infected patients of lung cancer, detection of nodules at the earliest is very important. In this study a framework using convolutional neural network (CNN) is developed for the classification of images as cancerous or non-cancerous. The main objective of this study is to reduce the burden on the radiologists in the earlier detection of lung cancer and reduce the processing time of using CNN in the classification of images without the reducing the high accuracy.

Convolutional neural network, class of deep neural network used for the analysis of visual imagery [ref]. As compared with other algorithms of image classification CNNs requires less pre-processing stages. This independence from prior knowledge and human effort in feature design is the major advantage of the use of CNNs. CNNs are commonly used for image classification, the learning process was surprisingly fast and highly accurate [ref]. They are good enough in classifying objects in to fine-grained categories, similar to the human behavior.

CNN, deep learning algorithm, which takes in an input image, then assign importance such as learnable weights and biases to various objects in the image, hence able to differentiate each object in the image one from the other. In CNN architecture the connectivity patterns of neurons is similar to that of the connectivity patterns in human brain [ref]. But the response of individual neurons is restricted only to certain region of the visual field, called as receptive field. Collection of such fields overlap to cover the entire visual area [ref]. The main aim of CNN is to convert image to a form suitable to process, without losing features critical for obtaining good prediction.

The kernels used in the convolutional layers plays very important role in feature extraction and the classification of images. In order to identify the role of kernels in determining the accuracy, pre-trained networks VGG16 and Alexnet is compared. VGG16 with five group of convolutional layers, where each convolutional layer has 3-by-3 kernels is able to give an accuracy of 95.25%, in a processing time of 1703 minutes. It is because of the deep 41 layers of VGG16.

For the analysis 1200 random samples of images from LIDC-IDRI dataset is used, where 695 images were used for training the network and 505 images for validation. Figure 1 shows the accuracy and loss plot of VGG16 network. The confusion matrix shows the true positive, true negative, false positive and false negative which is easy for the analysis. From the analysis true positive gives the correctly classified lung cancer images and false positive gives the misclassification of images, means that the lung cancer is wrongly predicted as non-cancerous image. From the four categories we are able to calculate the sensitivity and specificity of the networks. Confusion matrix is used to identify how many images are depicted correctly and incorrectly as nodule image or non-nodule image i.e. for easy analysis. Figure 2 shows the confusion matrix of VGG16 network.

From the 268 nodule image set VGG16 classifies 250 images as nodule image itself and 18 images as non-nodule images and out of 237 non-nodule image it classifies 234 as non-nodule

and 3 as nodule image.

Alexnet with three group of convolutional layers, where each convolutional layer has different sized kernels, 11-by-11 for the first, 5-by-5 for the second and 3-by-3 for the third group of convolution is able to give an accuracy of 86.93%, in a processing time of 44 minutes. It is because of the lesser 25 layers of Alexnet. Figure 3 shows the accuracy and loss plot of Alexnet.

Figure 4 shows the confusion matrix of Alexnet, where from the 268 nodule image set Alexnet classifies 228 images as nodule image itself and 40 images as non-nodule images and out of 237 non-nodule image it classifies 211 as non-nodule and 26 as nodule image.

Based on the inference obtained new image-net is designed with 5 group of convolutions in each group 5-by-5 sized Laplacian of Gaussian (LoG) kernels are used for convolutions. Usually 3-by-3 or 5-by-5 sized kernels are used for convolution as its performance is more accurate than large sized kernels, where there exist chance of losing features. LoG kernels are used, as laplace operator may detect edges as well as noise, desirable to smooth the image first itself by convolution and then suppresses the noise before detecting the edges. Hence as compared with other kernels helps in accurate feature extraction and classification of images. Designed image-net is a simple serial network with minimal processing time than Alexnet and accurate as VGG16. Figure 5 shows the newly designed network architecture.

Convolutional Layer, Pooling Layer, and fully connected layers are used to build the new CNN. Also ReLU layer is used for the activation functions. Raw input image (CT images) is used as the input to the CNN and the hidden CONV layer computes the output of neurons that are connected to the local regions in the input. ReLU layer applies an element wise activation function and leaves the size of volume unchanged. Pooling layers perform max-pooling function to convert different sized features to unique sized features for easy performance. The class scores will be computed by fully connected layers. Hence as shown in figure 6, designed network is able to give an accuracy of 96.44%, in a processing time of 23 minutes.

Figure 7 shows the confusion metrics of designed network, from the 268 nodule image set the designed network classifies 257 images as nodule image itself and 11 images as non-nodule images and out of 237 non-nodule image it classifies 230 as non-nodule and 7 as nodule image.

Table 1 describes the comparison of three networks, Alexnet, VGG16 and the designed network based on their performance in terms of Accuracy, Sensitivity, Specificity and Processing Time. The designed networks performs well compared with pre-trained CNN models both in terms of accuracy and processing time for the classification of image as nodule and non-nodule images.

Conclusion and Future Work

A convolutional neural network based system for the classification of images as cancerous or non-cancerous image using lung CT image is developed. Lung image with different shape and size of cancerous tissue has fed at the input for training the system. The proposed system was able to detect the presence and absence of cancerous cells with 96.44% accuracy, 95.89% sensitivity and 97.04% specificity in 23 minutes in single CPU workstation.

In near future, the system will be trained with large dataset of multi-resolution images to diagnose the type of cancer with its size and shape. The overall accuracy of the system can be improved using false positive reduction and improving the number of hidden neurons with deep network.

edubirdie.com