Automatic Detection of Lung Cancer in Computed Tomography Images using Convolutional Neural Networks

Nilushika Udayangani   
Department of Electronic and Telecommunication Engineering  
University of Moratuwa130605M@uom.lk

Harshani Prasangika   
Department of Electronic and Telecommunication Engineering  
University of Moratuwa130458M@uom.lk

Nimeshika Udayangani   
Department of Electronic and Telecommunication Engineering  
University of Moratuwa130604J@uom.lk

Ranga Rodrigo   
Department of Electronic and Telecommunication Engineering  
University of Moratuwaranga@uom.lk

Geethika Abeysinghe   
Department of Electronic and Telecommunication Engineering  
University of Moratuwa130009N@uom.lk

Nuwan Dayananda   
Department of Electronic and Telecommunication Engineering  
University of Moratuwanuwan@uom.lk

***Abstract*— Lung cancer is identified as the most common cancer in the world that causes death. Early detection has the ability to reduce deaths by 20%. In the current clinical process, radiologists use Computed Tomography (CT) scans to identify lung cancer in early stages. Radiologists do so by searching for regions called ‘nodules’, which correspond to abnormal cell growths. But identifying process is time consuming, laborious and depends on the experience of the radiologist. Hence an intelligent system to automatically assess whether a patient is prone to have a lung cancer is a need.**

**This paper presents a novel method which use deep learning, namely convolutional neural networks (CNNs) to identify whether a given CT scan shows evidence of lung cancer or not. The implementation uses a combination of classical feature-based candidate detection with modern deep-learning architectures to generate excellent results better than either of the methods. The overall implementation consists of two stages. Nodule Regions-of-Interest (ROI) extraction and cancer classification. In nodule ROI extraction stage, we select top most candidate regions as nodules. A combination of rule based image processing method and a 2D CNN was used for this stage. In the cancer classification stage, we estimate the malignancy of each nodule regions and hence label the whole CT scan as cancerous or non-cancerous. A combination of feature based eXtreme Gradient Boosting (XGBoost) classifier and 3D CNN was used for this stage. The LUNA dataset and LIDC dataset were used for both training and testing. The results were clearly demonstrated promising classification performance. The sensitivity, accuracy and specificity values obtained for the nodule ROI segmentation and cancer classification showed to be improved in the combined approach of deep learning with classical feature based classifiers compared to the deep-learning only techniques.**

***Keywords—Computed Tomography, Lung Nodules, Convolutional Neural Networks, Deep Learning, Segmentation, Classification***

# Introduction

In the world, cancer will strike two in every five people in their lifetimes. Lung cancer is the most common and number one cause of cancer deaths in both men and women [1]. Lung cancer related deaths accounts for 27% of all cancer related deaths. Unfortunately, most people who develop lung cancer do not develop symptoms until it has become more advanced. The result is late diagnosis, where treatment can be effective, but rarely curative. Early detection is critical for lung cancer since it opens a range of treatment that is not available at later stages. It has been estimated that early detection has the potential to reduce the lung cancer related deaths by 20%. Typically lung cancers are characterized by pulmonary nodules. Pulmonary nodules are regions corresponding to uncontrolled cell growths. The pulmonary nodules can be attached to the lung wall, bronchus, bronchioles and blood vessels or may be isolated within lung region. Computed Tomography (CT) is breakthrough technology for pulmonary nodule detection that heavily used in current clinical process.

CT imaging has the ability to form three dimensional images of the chest with greater resolution of nodule and tumor pathology. A typical lung CT scan is a 3D image of approximate size 30cm ×30cm×40cm. But the nodules we are interested in malignancy estimation for lung cancer detection are confined to very small regions with diameter ranges from 3mm-40mm. Therefore examining the presence of pulmonary nodule in a lung CT scan is literally similar with finding a needle in haystack problem. Figure 1 depicts a pulmonary nodule identified in a CT scan slice. Those nodule regions looks very similar to tiny bronchioles and blood vessels with their intensities and shapes when viewed in 2D slices.

In current clinical process chest radiologists have to manually inspect through the whole CT scan slice by slice to spot out the presence of a pulmonary nodule. The radiologist have to work with approximately 1:1000000 signal to noise ratio to extract out nodules. And for the malignancy estimation of extracted nodules they have to manually extract the nodule contours and calculate shape and texture parameters like area, growth rate, and shape etc. Therefore lung cancer examining for one CT scan roughly takes 30-50 minutes .This makes manual lung cancer diagnosis process into tedious time consuming task which require the expertise domain knowledge in lung anatomy. Yet manual annotations results in lots of false positives and may miss potential nodules.

Therefore computer processing to assist lung cancer detection is one of most timely need of the world. Here we proposed an intelligent system for binary classification problem to detect the presence of lung cancer in a given patient-CT-lung scan. The proposed system use combination of deep learning, feature based machine learning and rule based image processing to assess whether a given CT scan is of cancerous patient or not. It will output top candidate regions for nodules, common shape and texture parameters for each nodule region-of-interest (ROI), malignancy estimation for each nodule ROI and overall malignancy of the whole CT scan. Such a system will dramatically reduce the false positive rate that plagues the current detection technology. And help to find nodules missed by human error.

Regardless of the malignancy outcome, automatic nodule detection can be a big help for radiologists since the nodules can easily be overlooked

And it will help to get patients to access life-saving interventions earlier. It will give radiologists more time to spend with their patients, and the system avoids additional follow-up imaging and interventional treatment. It will be advancing the state of the art in future screening, care and prevention.

The proposed system highlighting its novelty and originality since it use combination of classical feature based classifiers and deep learning classifiers namely Convolutional Neural Networks (CNN) to produce results. Convolutional Neural Network is very popular architecture of deep neural networks which has become the state-of-the-art technology of current computer vision domain. Yet applying it in the domain of medical image processing is very recently being grabbed the attention by research community. Possibility of applying deep learning for cancer classification of CT volumes is very recently examined in 2017 Kaggle Data Science Bowl Competition. Other than that there is no significant research efforts which combining both nodule detection and cancer classification outputs with significant improvement in sensitivity number of false positives and level of automation.

For the training purposes we used two online lung CT scan databases provided by Lung Nodule Analysis 2016 (abbreviated as LUNA 16) competition and the Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) (abbreviated as LIDC-IDRI database).

The proposed method consisting of two stages, namely nodule detection and cancer classification. Since the amount of signal vs noise in a CT scan in the viewpoint of lung cancer detection is almost 1:1000000 the neural nets are not able to learn something from the raw image data. Therefore we have to spoon-feed a neural network with examples with a better signal-to-noise ratio and a more direct relation between inputs and labels. That is why the proposed method having two stages. In nodule detection stage we extract candidate regions for nodules which become the input to the cancer classification stage for malignancy estimation. Each of the stages consists of ensemble models of CNN and feature based classifiers to generate more accurate results.

The rest of the paper is organized as follow…….

# Related Work

The research efforts for lung nodule detection and cancer classification of CT volumes can be divided into two categories. Those are feature based machine learning approaches and deep learning based approaches. Both approaches use image processing techniques like thresholding, rescaling, morphological operations, filtering and segmentation methods to perform some preprocessing tasks. Most of classical feature based approaches first segment the lung, extract features from training data and train a classifier to detect nodule regions. Then shape and texture features are extracted from nodule regions to train a classifier for malignancy classification. Support vector machines (SVM), random forest classifier, XGBoost classifiers k-nearest neighbor (kNN), logistic regression like classifiers are engaged as classifiers. But nodule detection classifiers learned using hand-engineered features often give poor generalization to novel test data. This is because lung nodule detection is inherently more challenging due to the high variability of nodule shape, size and texture. More recently advances in deep learning have enabled extraction of high level features without expertise domain knowledge in lieu of hand engineered features.

Deep learning is the branch of machine learning based on deep neural networks which are neural networks composed of more than one hidden layer. Among the most popular architectures of deep neural networks, CNNs are of particular interest which are widely used in computer vision. E.g.: image classification, super-resolution, semantic segmentation. Recent publications report their usage in medical image segmentation and classification with promising results. For example U-net: Convolutional networks for biomedical image segmentation by O. Ronneberger, P. Fischer, and T. Brox outperforms the prior best methods for segmentation of neuronal structures in electron microscopic stacks in ISBI challenge 2015. And K. Komnitsas et al. use 3D CNN for Brain lesion segmentation in multi-channel MRI patient data with top ranking performance on the public benchmarks BRATS 2015 and ISLES 2015[18].

Deep learning for lung cancer classification very recently grabbed the attention in research community after 2017 Kaggle Data Science Bowl Competition. The winning team of the competition use 3D CNN classifiers for both nodule segmentation and cancer classification. A similar work done by Julio Cesar Mendoza use CNN for lung nodule classification which show that their method is outperforming a base feature-engineering method using the same techniques for other stages. Rushil Anirudh et al proposed 3D CNN for lung nodule detection which works with the availability of weakly labelled data. Their system works with point labels, which specify a single voxel location that indicates the presence of a nodule, and its largest cross sectional area.

We proposed a method combining both classical feature based classifiers and CNNs to produce more accurate results. To the best of our knowledge we are the first to explore lung nodule detection and cancer classification by combined methods which exploit the advantages of each method to generate more accurate results.

# Methodology

Our task is a binary classification problem to detect the presence of lung cancer in patient lung CT scans. Here we exploit the use of deep learning, particularly 2D and 3D convolutional neural networks. But at each stage of the procedure, we followed classical feature based machine learning methods and rule based image processing approaches beside the main deep learning architecture and combined the outputs together to generate more accurate results. Below describes the technological approach we have taken.

Usually a CT scan of lung area is a 3D image of size 30cm × 30cm ×40cm which contains about 100-400 slices in transverse view. But the nodules, interested in malignancy estimation is restricted to regions of 3mm- 30mm diameter resulting a very low signal to noise ratio near 1: 1000 000 leading to a literally finding a needle in haystack problem. Therefore any deep learning architecture will not able to learn something from the raw image data. Given the lung CT scan and the cancer label of that scan, it will unable to find a direct relation between the label and the input scan. Hence our solution consists of two major parts, first extracting the nodule regions of interest thus increasing signal to noise ratio to a great extent, secondly classifying the overall malignancy based on the regions extracted above.

A detailed block diagram explaining the methodology of the system is shown in Figure 3.1. It can be viewed as a combination of several major steps namely image acquisition, pre-processing, nodule ROI extraction, false positive reduction, feature visualization and cancer classification.

For the implementation we used Python 3.5 language, NumPy [1] for N-dimensional array handling, Skimage python library [2] for image processing, SimpleITK [3] for medical image analysis and H5Py [4] for efficiently storing and manipulating huge amount of data.

## *Image Aquisition*

Among various imaging modalities for lung scanning we used low-dose CT scans as the input to our system. In CT scanning, it uses X-rays to get 2D cross sectional images of the body. Therefore a CT scan is a 3D input which consists a stack of 2D detailed images.

For the training and testing stages, we used two public lung images datasets; LUng Nodule Analysis 2016 (LUNA16) and Lung Image Database Consortium image collection (LIDC-IDRI) [27].

### *LUNA16 Dataset*

LUNA16 is a public dataset which provides 888 patient high resolution lung CT scans, made available by the Lung Nodule Analysis 2016 challenge community [23]. This has nodule locations in each scan annotated by radiologists thus very useful for training a nodule extraction methodology. Scans are provided in MetaImage (mhd) format and each CT scan consists of around 100-400 2D slices of 512×512 pixels. Annotations of nodules include position data (x,y,z coordinates) and diameter of each both cancerous and noncancerous nodules. Altogether 1186 nodule locations are provided. Among 1186 nodules, we preserved scans contain last 286 nodules for the test set.

### *LIDC Dataset*

Even though the LUNA16 dataset has the nodule locations annotated, it does not contain the malignancy label for each scan. But later we found that it was drawn from another public dataset called LIDC-IDRI. Here they have provided an assessment on the malignancy and other properties of the nodules. These malignancy labels are ranging from 1 to 5 and provided by 4 radiologists. We inserted the average of these 4 labels as the cancer label for each scan only for those who contained among above 888 scans provided by the LUNA16 dataset.

## *Pre-processing*

The next step of our cancer classification pipeline is the pre-processing. The scans provided have been taken from different scanners having different properties. Before feeding into any kind of classifier it is very important to make these scans as homogeneous as possible. Therefore we carried out several pre-processing steps.

First the mhd formatted scan was read using SimpleITK and at the same time voxel spacing and origin coordinates of the scanner in mm were saved. Then it is required to convert these raw data into Hounsfield Units (HU), which is a measure of radio density. For this we first multiply each slice from its slope and then add the intercept value usually contained in the header file. These HU values have semantic meaning where each substance in the body related to a specific HU value as shown in Table 3.1 [222].

Then the most important step in pre-processing stage is the rescaling all scans to an isotropic resolution so that every voxel represented a volume of 1mm×1mm×1mm. This is needed because we can then feed these data for CNNs without worrying about learning zoom/slice thickness invariance. For this we used SciPy python library [999] methods including interpolation. Since spacing towards the z direction or slice thickness is usually greater than 1mm, after rescaling we ended up with the scans having more slices (more than 400 slices) than the original no of slices.

Then resulted image contains values between -1024HU to 2000HU. But anything higher than 400HU not interesting to us, as they are bones with different radio-density. Therefore we clipped off values outside -1000HU and 400HU and scaled between 0 and 1. This process is called the normalization.

Then other important step in pre-processing is the zero-centering these images in order to reduce the effect of scale differences, by subtracting mean and divided by the standard deviation of the dataset.

## *Nodule ROI Extraction*

As described here the intention is to separate the regions of interest or possible cancer regions from input images by outputting the location of lung nodules. Furthermore it can be viewed as a segmentation problem. Here we followed two separate approaches. One is the deep learning based nodule segmentation while the other is the applying classical feature based image processing techniques.

### *Deep Learning Based Lung Nodule Segmentation*

Given the CT scan, to output a binary mask indicating the boundary and location of each nodule presented in the lung interior, we build and trained a 2D convolutional neural network.

#### *The Network Architecture*

As reviewed in the literature deep learning is a very cutting edge technology promising excellent results in segmentation problems in medical imaging. One very famous convolutional neural network architecture is U-Net[25] which is originally developed for segmenting neuronal structures and up to now it has outperformed the prior best method (a sliding-window convolutional network) [26]. Here we adopted this network to segment lung nodules. It takes a 2D slice as the input and returns a 2D slice of the same size as the output which is the segmentation map of nodules. The architecture is shown in Figure 3.8. To reduce the computational power we down sized the number of neurons in each layer to half size of the original architecture. Unlike the conventional CNNs used in classification problems it is a fully convolutional network, which has no fully connected or dense layers. This is because for segmentation problems the desired output should include the localization i.e., a class label is supposed to be assigned to each pixel. Therefore altogether it has 23 convolutional layers and has up-convolutional layers beside the convolutional and max-pooling layers to upsize the dimensionality reduced after each max-pooling layer.

To build the CNN we explored available deep learning frameworks those provide basic building blocks and application programming interfaces (API) to code in Python. Among them we found Keras and Tensorflow more beneficial than other libraries including Theano, Caffe, CNTK, Pytorch etc. in terms of faster compile times, framework growing speed and development support [888].

#### *Processing Training Dataset*

To train the segmentation neural network we need input images and targets associated with it. As the target we have to generate a binary mask indicating the nodule location using the diameter given in each nodule annotation in LUNA16 dataset. But before fed to the neural network we have to reduce the search space by segmenting only lungs and removing low intensity regions, which greatly reduces the computational complexity. Thus we need to further pre-process each of these images.

Among the 888 patient scans we only selected scans containing at least one nodule and take only the 2D slice which contains nodule centre as well as 2 adjacent slices of it. Thus 3 slices per one nodule generates total 1186×3 input target pairs to train the network. After reading these slices we first segmented the lungs. For this we adopted the methodology proposed by Sasidhar et al. [24] where we used simple image processing and morphological operations to segment the lungs. In a CT scan lungs are visible as darker regions and bright regions inside the lungs are the blood vessels or air. To keep only lung interior separated with background, according to table 3.1, a threshold of -400 HU was used. Then the borders were cleared in the resulted binary image from thresholding to remove the unwanted regions connected to the border of the image. Then the image was labelled and kept only the 2 largest regions. To keep the nodules attached to the lung walls morphological closure operation with a disk of radius 10 was applied yielding the final lung mask as indicated in Figure 3.3. It also shows final segmented lungs which can be obtained through multiplying the generated mask with the original image slice

After segmenting the lungs we have to crop the original image slices into this region of interest. To understand which extent to we have to crop these images, bounding boxes of the labelled lung mask image is calculated. Then we generated a 2D nodule mask indicating a circular region having the given diameter for each image slice and cropped this mask too.

The original size of all scans is 512×512 and varying number of slices in z direction. But after rescaling described in section 3.2, different scans resulted different x, y resolution values. However all slices within a single scan have the same resolution. But above cropping to lung ROI step, would result in having different sized slices in the same scan. On the other hand the deep learning architecture, we are going to train, we would like to have a fixed size input tensor. Therefore what we did was, first finding the largest slice size after both above operations by iterating through all dataset. We got largest slice size as 396×396. Therefore we decided to have the input size fixed at 400×400. So, both mask and image were again zero padded to the size of 400×400.

#### *Training the Model*

To create the training dataset we took slices containing first 900 nodules among 1186 nodules and kept the rest as the test set. As stated in 3.3.1.2 for each nodule we took 3 center slices, thus resulting a total number of 2700 training images and 858 test images. But storing or loading again these images into or from a single array is quite impossible emerging memory errors. Therefore the data was saved to an H5Py dataset. At the training stage, these data is retrieving chunk by chunk from the created HDF5 matrix in Keras.

To train the model, Adam and Stochastic Gradient Descent was used as optimizer algorithms. Adam (short for Adaptive Moment Estimation) is an update to the RMSProp optimizer [2] in which the learning rate is adapted for each of the parameters. Initially the learning rate was set to 10-6 and it was reduced after certain no of epochs to obtain a fine learning process. As the Loss/objective function, the Sorensen-Dice coefficient loss was used. It is the 2 times intersection between the true nodule mask and predicted nodule mask divided by their union (Figure 3.6). Equation (1) explains this metric further.

 (1)

Like in many kind of medical image segmentation problems the interested positive class samples are quite small compared to the large background (negative class). Therefore it is required to apply more importance on correctly predicting the positive class. The reason behind selecting the Dice coefficient loss as the metric is, it indirectly applies more weight to correctly predicting the positive class while considering true negatives as uninteresting defaults. For the training process a NVIDIA Tesla K40c graphical processing unit was used, which has a memory size of 12GB and 2880 CUDA cores. We trained the network more than 300 epochs (no of iterations on training dataset) with a batch size (no of samples to forward pass per one weight update) of 2. Where it took 25min per each epoch consuming 6-7 days for the whole training procedure.

The results obtained at each intermediate step of the learning procedure is indicated in Figure 3.7. It can be seen that after certain no of learning iterations, the output mask where it indicates the possible nodule locations of the input image slice becomes more similar to the true nodule mask, which is shown in second column in figure. Actually output mask obtained from U-Net is a kind of probability map, where highest values indicating the high probable nodule regions. Therefore to obtain a binary mask we should find a proper threshold value. First we set this value as the mean of each predicted slice. But it lead to a huge no of false positives since the mean is a very small number and lots of regions having values higher than that number. We calculated sensitivity and specificity values in terms of area having this threshold set at different values (How we calculated these measures are explained in results section). Increasing the threshold lead to high values of sensitivity but low specificity values. Therefore 0.5 was finally chosen as the threshold to compromise between both these parameters. A 0.5 threshold U-Net output overlaid on original image slice is indicated in Figure 3.8.

### *Feature Based Lung Nodule Segmentation*

As the second approach to get initial nodule candidate regions we applied rule based image processing techniques. Thresholding, morphological operations, connected component properties, edge detection, hole filling, clear border, 3D average like image processing techniques were combined to get the results. A critical step here is the lung segmentation where we followed very similar technique described in section 3.2.1.1. The major steps are; thresholding at -400HU, removing the blobs connected to border, keep only the two largest connected components in each slice, erosion operation with a disk of radius 2 followed by closing operation with a disk of radius 20 to keep the lung wall attached nodules inside the lung mask, taking edge image after Robert filter, filling holes slice wise, and finally superimposing the binary mask with each corresponding slice.

Figure 3.13 showing the 3D view of lung CT scan threshold at -400HU with corresponding lung mask. Selecting the radius of the disk used for the closing operation is critical. For low values of the disk radius, the wall attached nodules kept as part of the lung wall and removed from the lung region. For high values of disk radius lung wall attached nodules were also kept inside the lung region. But when increasing the disk radius the part of bronchus also included as lung regions. Since nodules can also be attached to the bronchus this is desirable for some extent. But since the bronchus has similar intensities as nodules, unnecessarily inclusion of bronchus results in problems in later steps. Figure 3.14 showing the results for two different disk radius for closing operation, 10 and 30.

After lung segmentation we have to generate the mask indicating nodule ROIs. For this we initially suppressed the non-nodule regions by keeping only the regions which having intensity greater than -400HU. The extracted region included whole bunch of blood vessels, bronchioles and part of bronchus with truly nodule regions. Figure 3.15 showing the resulted 3D image after this initial thresholding. Differentiating the nodule regions with bronchioles and blood vessels is very difficult in 2D view taken slice by slice. Figure 3.16 showing non-nodule regions that similarly appear as nodule regions in 2D view. So we performed three operations that take 3D structure in to account when performing non-nodule region suppression. First we applied opening operation with a sphere of radius 1mm in 3D image. Then we removed regions having eccentricity greater than 0.8 in all three planes of the scan, namely sagittal, coronal and transverse plane. Thirdly we removed regions that fail to cover the existence over minimum diameter length to be considered as a nodule in all three planes.

The 3D opening operation with a small sphere removed the nodules attached to the blood vessels and tiny bronchioles. Before filter out regions having low sphericity this step helps to keep the vessel attached nodules in nodule ROI. The next step was to remove the nodule regions that having low sphericity values. The blood vessels and bronchioles have non-spherical elongated nature compared to the nodule regions. But in 2D view taken as slice by slice, these thin elongated structures also appear as circular shapes. So removing areas having high eccentricity values in transverse view is not sufficient. So we filter out areas that having eccentricity greater than 0.8 in transverse view, sagittal view and coronal view. This resulted in three of 3D binary images that contains areas satisfying the eccentricity constraint for each three plane. Binary masks corresponding to sagittal and coronal view are converted into transverse view again. Then we take the intersection of three binary masks and this resulted in the areas satisfying eccentricity condition in all three planes.

Since we only interested in nodules having diameter greater than 3mm, in order to be a nodule, a region should at least exists in four slices. Since the slice thickness was resized to 1mm, 3mm distance is covered by four slices. Usually blood vessels and bronchioles appears in any view disappear over slices rapidly. Therefore we can identify the nodule regions by checking their existence over a neighborhood of slices. For the produced binary 3D mask containing the nodule ROIs from previous steps we iterate over each slice and check whether each region satisfying minimum expansion over neighboring slices. For that we choose previous three slices, current slice and next three slices and add them up. Pixels having summation greater than four (exists in at least four slices) were kept as nodule ROIs while rest is suppressed. This operation performed in all three planes and took the intersection of transverse view mask with sagittal and coronal view masks converted to transverse view.

Then the resulted 3D binary mask was given as the nodule ROI output mask from the image processing technique. Figure 3.17 showing output binary mask which contains nodule ROIs for the selected scan. Figure 3.18 showing the true nodule regions annotated by radiologists for the same scan.

### *Ensemble of Both Methodologies*

The U-Net based segmentation described above, takes only a 2D slice of the scan as the input at one time. To obtain the whole stack of predictions we have to input slice by slice for the net. But as stated, before fed into the net each slice is cropped to its lung ROI by comparing with the lung mask’s bounding boxes. After they are zero padded to size of 400×400. But this is problematic when all these predicted slices are stacked together to get a 3D nodule mask stack, since spatial information with reference to the original coordinate system has been lost. Therefore what we did was before zero padding the images, first remember the top-left corner point’s coordinates with reference to the original 3D image. Then in the U-Net output, we stripped the borders corresponding to these padded zeros and overlay the mask so that its top-left corner point and stored coordinates are coincide.

But being considering only 1 slice at one time, U-Net cannot exploit the inherently 3D structure of the nodules. As a solution we can extend this network to a 3D version of the same architecture. But 3D architectures are quite slow and inflexible, require more resources, and unable to train on the GPU we were used. Therefore, as the solution we got the ensemble of extracted 3D features through feature based method with the U-Net segmented results. Hence the union of two masks was obtained to get the final lung nodule mask which is indicated as in Figure 3.19.

## *False Positive Reduction*

As in Figure 3.19, it can be seen that U-Net tends to output nodules as well as large number of false positives indicating blood vessels as nodules. At the same time feature based method outputs parts of bronchus as the nodule candidates frequently. Thus the binary mask indicating the possible nodule locations obtained through the union of them gives a lot of false positives besides the true positive nodule regions. Therefore next essential step in our classification pipeline is the false positive reduction.

For this purpose we trained a 3D convolutional neural network which takes voxel cubes around nodule candidates and output the probability of that region being a nodule. The 3D CNN adopted here, has a very general architecture as indicated in Figure 3.21. The input to the net is a 36×36×36 voxel cube. Since the largest nodule contained in the LUNA16 dataset has the diameter of 32.3mm and the pixel spacing now is 1mm in each direction this cube size is sufficient to represent the full nodule region around its center. Then the input is followed by 3 convolutional layers to extract fine feature maps. Every convolutional filters have the size of 3×3×3 and max-pooling layers have a stride of 2×2×2. To reduce the over fitting dropout layers were inserted as shown. Then the softmax layer at the end of the network outputs two probabilities for the nodule and non-nodule classes.

To train the network we again used the LUNA16 dataset. Among 1186 nodules, we preserved last 286 nodules for the test set as same as in the U-Net training stage. For each nodule center, 48 positive samples were cut by shifting the cubic center slightly. This caused the data augmentation, where we can increase the training dataset by applying transformations to the input dataset. Then 52 negative samples were cut around randomly selected centers that are outside the true nodule locations from the same scan. All data were saved in a HDF5 dataset as previously thus resulted near 90000 training samples and 28600 test samples each having a binary label of being a nodule or non-nodule. As the optimization function, the categorical-cross entropy between true labels and predicted labels was used. Adam optimizer with learning rate 10-6 was used as the optimization algorithm.

When using this trained network for the prediction purpose of our pipeline, first we labeled the binary mask obtained in nodule ROI extraction stage. Then the center coordinates of all connected components were obtained and 36 sized voxel cubes were cut around these centers. Then each of these cubes were sent through this network and the probabilities of being each of these cubes as nodules were taken to create a 3D stack of a probability mask. Figure 3.22 (right) indicates the 3D structure of this probability mask threshold at 0.9.

## *Cancer Classification*

The important next stage of our pipeline is the cancer (malignancy) classification stage. Given the nodule location, we have to predict the probability of this nodule being cancerous or not. For this purpose we again followed two parallel approaches. One is based on deep learning and the other is based on manually calculated features, usually employed by radiologists to estimate the malignancy of nodules.

### *Deep Learning Based Cancer Classification*

Here we trained a 3D convolutional neural network which takes voxel cubes around nodules and output the probability of that nodule being cancerous or not. Figure 3.23 indicates the network architecture we adopted for the 3D convolutional neural network. It is based on the architecture proposed by Julian Wit [444] with minor modifications. At each nodule center we input a 32×32×32 voxel sized cube to the network. At the very beginning it down sample the z axis resolution of this 3D cube, since for many scans the z-axis was at a more coarse scale than the x and y axes. This makes the net much lighter and do not affect the accuracy. Finally the softmax layer at the very end, outputs the probabilities of this cube being associated with malignant and non-malignant classes.

To train the network we used LIDC dataset, where we extracted the malignancy labels for 1186 nodules. Similarly we preserved last 286 nodules for the test set. For each nodule center, 48 samples were cut by shifting the cubic center slightly in order to increase the no of training samples. This resulted near 43200 training samples and 13728 test samples each having a binary label of being cancerous or non-cancerous. As the optimization function, the categorical-cross entropy between true labels and predicted labels was used, while the Adam optimizer with learning rate 10-5 was used as the optimization algorithm.

### *Feature Based Cancer Classification*

In this approach we developed a feature based classifier for the purpose of extracting the features like nodule equivalent diameter, volume and more complex features like spiculation, lobulation, sphericity, etc those which are used by radiologists to predict the malignancy of a nodule. Figure 3.24 shows the block diagram of this proposed classification scheme.

Even though the results from nodule segmentation stage is very accurate enough to predicting the locations of the nodulous regions, sometimes it is unable to predict nodule contours more precisely since we are applying some morphological operations yielding eroded contours into nodules. Therefore the first step of this approach is the generation of nodule contours in order to accurately calculate above features. So all contours existing in each slice of the input cube is regenerated as shown in the Figure 3.26, where we used Python Skimage functions. Figure 3.26(a) shows the binary mask generated from nodule ROI extraction step, (b) shows the regenerated nodule contour and its binary mask as in (c).

The next important step is to calculate the features. Here we adopted two kind of features, a list of manual features proposed by the LIDC/IDRI dataset and features extracted from the last convolutional layer of the cancer classification CNN discussed above.

As the first set of features, we used several 2D features computed from the biggest axial slice (*S*biggest) of each nodule [30], namely area (the number of pixels in *S*biggest multiplied by pixel resolution), perimeter (the perimeter of *S*biggest multiplied by pixel resolution), convex area (the area of convex hull of *S*biggest ), compactness described by 3.2 and circularity described by 3.3.

 (2)

 (3)

Again we calculated several 3D features namely, volume, surface area, equivalent 3D diameter, sphericity, speculation and lobulation. The volume of a nodule is computed by multiplying the total number of voxels with voxel resolution as per the Dhara et al. [29]. We computed the surface area of a nodule by summing the area of all triangular faces of that nodule mesh. The equivalent 3D diameter, D of a nodule is defined as in 3.4 [28].

 (4)

The sphericity represents the irregularity of the shape of a nodule which is ranging from 0 to 1[28], and defined as 3.5.

 (5)

Then to calculate the spiculation the height and base area of each spicule are determined [29]. The net spiculation of a nodule is computed as 3.6, where ωi is the solid angle subtended at peak point of ith spicule, hi is the height of that spicule, and N is the total number of spicules of a nodule under consideration.

 (6)

 (7)

Then the lobulation is represented in terms of the ratio of total concave surface area, *S*concave and total convex surface area, *S*convex of the nodule mesh as in 3.7. A lobulated surface of a nodule is associated with the uneven growth rate of that nodule [29]. Also malignant nodules have uneven surface due to the presence of spiculation and lobulation, whereas benign nodules have smooth surface (Figure 3.27). We used region growing to get the concave and convex part of the nodule mesh.

Once these features are calculated, another 64 features obtained from the last convolutional layer of the cancer classification CNN were also combined in order to improve the performance of the classifier. As the classification model, an XGBoost classifier [31] was chosen, because it is robust to over fitting and quick to train. The common python implementation of XGBoost also allows for easy modification of the objective function and reweighting the class importance, so that we could reweight the optimization problem to give more weight to cancer cases.

### *Ensemble of Both Methodologies*

The purpose of following two approaches for cancer classification described above are twofold. One trivial fact is to improve the accuracy of the prediction since it is same as the getting ensemble of results from two clinical procedures. Other is to calculate and visualize useful features of detected nodules by our analysis tool. This causes our system to output more human interpretable results, where we will be unable if only the deep learning method is employed.

In our pipeline, after getting the probabilities of each region being a nodule from the false positive reduction stage, we extracted top 8 most probable nodule candidate regions out of them. We choose 8 patches centered on them since when taking less number of patches it gives high specificity but low sensitivity. Increasing this number than 8 will result high sensitivity but low specificity.

Figure 3.28 shows this 8 selected patches in 3D view and a 2D slice containing in 4 patches out of 8. Then these each patch is send through both cancer classification methods and the average of two results are taken as the final probability of being cancerous. For each patch the output probabilities and calculated important features are shown to the user graphically from the analysis tool we developed as indicated in Figure 3.29. Then to predict the malignancy of the overall scan we take the binary OR operation of each of these results. That is if single nodule is turned out to be cancerous we conclude that the overall scan as malignant.

# Results and Analysis

## Nodule ROI Extraction

To measure the performance of the nodule extraction stage of our pipeline we calculated several area measures based on number of pixels of interested region. Four parameters; true positive (TP), false positive (FP), true negative (TN) and false negative (FN) are calculated by the logical AND between ground truth mask and predicted binary mask. Then the sensitivity, specificity, precision and F1-score values were calculated based on these parameters as below.

 (8)

 (9)

 (10)

 (11)

For the performance analysis of this stage 29 test scans was used. First we took the results only from the deep learning segmentation approach and followed through the false positive reduction step to calculate the area measures. Then we recalculated these measures for the results obtained from the ensemble of both deep learning and feature based segmentation methods. The mean values obtained are summarized in Table 4.1.

It can be clearly seen that combination of the two approaches results in higher performance in all measures. To test whether this improvement is significant statistically, we followed a t-test. By comparing P-value 3 measures among 4 have that below 0.05. Thus indicating a significant difference.

## B) Cancer Classification Stage

For the 3D CNN developed for cancer classification, after iterating 55 times on the training dataset we finally achieved a training accuracy of 77.8% and validation accuracy of 79.3%. The training accuracy and the test accuracy variation at various stages of training procedure is shown in Figure 4.3. Note that reasons behind the validation accuracy leading the test accuracy are applying dropout layers and the accuracy calculation procedure in Keras’ APIs.

## C) Overall Results

The accuracy values stated above are for per nodule cancer classification. Then we calculated performance measures for per scan cancer classification. For this purpose we calculated cancer labels for each scan in LUNA16 as one if at least one nodule contained in the scan is cancerous and zero otherwise. The results for 222 test scans including 67 cancerous scans are indicated in table 4.2.

# Conclusions

We have presented a work for lung cancer classification from lung CT volumes using combined method of classical featured based classifiers and CNNs. While the initial results are promising there are areas to further improve the system.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

