B. Tech Project

Pranali Pawar - B14SS011
Shreyansh Chhajer - B14SS017
Shubham Talbar - B14BS015

Automatic Lung Nodule Detection System

Under supervision of Dr. Rajlaxmi Chauhan

Problem Statement

The aim here is to take the sample data, consisting of low-dose CT scan information, and predict what the likelihood of a patient having lung cancer is. The following problem is an actual competition previously hosted by Kaggle as “Data Science Bowl 2017”. The whole problem can be divided into two parts:

1. Detecting pulmonary lung nodules in CT scans
2. Predicting the occurrence of cancer based on the nodules identified

We have analysed and implemented two approaches to the problem. The first approach is based on 3D Convolutional Neural Network and the second one on feature extraction and Support Vector Machines.

3D Convolutional Neural Network based approach

The idea is to use deep learning to segment cancer nodules in CT images. In machine learning, a convolutional neural network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery.

CNNs use a variation of multilayer perceptrons designed to require minimal preprocessing. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics. This means that the network learns the filters that in traditional algorithms were hand-engineered. This independence from prior knowledge and human effort in feature design is a major advantage.

Results Obtained

Due to unavailability of sufficient computational power, as of now we have used a sample dataset of just 20 patients ~1GB. Out of 20, 18 CT scans were used for training and 2 for testing, we did 10-fold cross validation over the dataset. The accuracy achieved was 67% with 10 epochs. With sufficient computational power, we speculate that increasing the number of iterations and training our CNN model over a larger dataset can further improve results.

Feature extraction and SVM based approach

Following approach is based on the conference paper published by Mehre et al in SPIE. The paper utilizes a two-stage approach comprising nodule candidate detection followed by reduction of false positives. The nodule candidate detection involves thresholding, followed by morphological opening. The geometrical features at this stage are selected from properties of nodule size and compactness, and lead to reduced number of false positives. An SVM classifier is used with a radial basis function kernel. The data imbalance, due to uneven distribution of nodules and non-nodules as a result of the candidate detection stage, is proposed to be addressed by oversampling of minority class using Synthetic Minority Over-sampling Technique (SMOTE), and over-imposition of its misclassification penalty.

Results Obtained

So far we have applied 3D connected component labelling algorithm to extract lung fields from input CT scans. Further we have applied multiple thresholding and morphological opening on the extracted lung fields to detect candidate nodules. Further we would extract geometrical features of the detected candidate lung nodules and apply SVM to classify nodules from non-nodules. In order to handle the uneven distribution of above classes we would oversample the minority class (SMOTE) and penalize the misclassification of minority class sample.