



Classification of Lung Nodule Using Hybridized Deep Feature Technique

Malin Bruntha* 

*Corresponding Author, Assistant Prof., Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore – 641114, Tamil Nadu, India. E-mail: malin.bruntha@gmail.com

Immanuel Alex Pandian 

Assistant Prof., Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore – 641114, Tamil Nadu, India. E-mail: immans@karunya.edu

Siril Sam Abraham 

Computer Vision Intern, Vasundharaa Geo Technologies, Pune, Maharashtra, India. E-mail: abrahamcyril77@gmail.com

Abstract

Deep learning techniques have become very popular among Artificial Intelligence (AI) techniques in many areas of life. Among many types of deep learning techniques, Convolutional Neural Networks (CNN) can be useful in image classification applications. In this work, a hybridized approach has been followed to classify lung nodule as benign or malignant. This will help in early detection of lung cancer and help in the life expectancy of lung cancer patients thereby reducing the mortality rate by this deadly disease scourging the world. The hybridization has been carried out between handcrafted features and deep features. The machine learning algorithms such as SVM and Logistic Regression have been used to classify the nodules based on the features. The dimensionality reduction technique, Principle Component Analysis (PCA) has been introduced to improve the performance of hybridized features with SVM. The experiments have been carried out with 14 different methods. It has been found that GLCM + VGG19 + PCA + SVM outperformed all other models with an accuracy of 94.93%, sensitivity of 90.9%, specificity of 97.36% and precision of 95.44%. The F1 score was found to be 0.93 and the AUC was 0.9843. The False Positive Rate was found to be 2.637% and False Negative Rate was 9.09%.

Keywords: CNN, Transfer Learning, GLCM, SVM, PCA.

Introduction

Cancer is second to only the pandemic, Acquired Immuno Deficiency Syndrome (AIDS) in the number of mortalities around the globe. The International Agency for Research on Cancer (IARC) functioning with World Health Organization (WHO) has estimated that there could be 18,078,957 new cancer cases worldwide including people from all ages and both genders in the year 2018. Out of these world-wide cancer occurrences, lung cancer has the highest occurrence with 2,093,876 cases which forms 11.58% of the total.

The cancer occurrence is significant in the continent of Asia (48.4% of the world total) which has the two most populous countries in the world. In Asia, lung cancer has the greatest number of deaths with 1 225 029 (14%) cases in the year 2018. In India, the expected number of cases was estimated to be 1,157,294 which is 13.2% of Asia's total and 6.4% of world's total cancer occurrences in the year 2018. Breast cancer has the highest occurrence with a case load of 14% of 1,157,294 and lung cancer closely follows behind with the case load of 5.9% of the total cancer cases in India. Most of the cancer cases in India go unreported due to poor screening facilities available to the masses. It can be understood from the above statistics that to reduce the cancer mortality rate, screening of a vast number of populations is very vital to ascertain whether or not they are affected with cancer.

Computer Tomography is an efficient way to detect lung pulmonary nodules. As the number of cases increase every year, the task of segmenting and classifying malignant pulmonary modules becomes a Herculean task for expert radiologists. To reduce their workload and increase the detection efficiency, Computer Aided Diagnosis (CADx) has stepped in and is playing a vital role in the early detection of pulmonary lung nodules. Right from the beginning of CADx, machine learning algorithms have played a huge role in detection and classification. Nowadays, deep learning networks like Recurrent Neural Networks (RNN), Auto Encoders, Deep Belief Networks (DBN), and Convolution Neural Networks (CNN) have made a giant stride in medical image analysis. Due to the use of deep learning networks, the accuracy to detect and to classify pulmonary nodules has improved greatly.

Convolutional Neural Network, one of the deep learning networks has gained popularity after 2000. But the foundation for CNN was made in 1980 in the name of neocognitron (Fukushima, K., 1980). The convolutional and pooling (down sampling) layers were introduced in neocognitron neural network. Cresceptron, a modified version of neocognitron (Weng et al., 1993) introduced max-pooling concept. CNN was first proposed in the year 1995 for medical image pattern recognition (Lo et al., 1995). LeNeT (LeCun et al., 1998), a CNN architecture was introduced by LeCun and his team for document recognition application in the year 1998. It was a simple and straightforward architecture. For successful

implementation of any CNN, there is a requirement of computational power and excess data to train the network.

The efficient implementation of CNN was made feasible when the Graphical Processing Unit (GPU) came into the market. The necessity for large number of data for training the CNN was satisfied after the introduction of ImageNet project (Deng et al., 2009). This project was meant for software research in the field of object recognition. It launched a large database that contains almost 14 million images and at least 1 million images were annotated. This ImageNet project has been conducting annual software contest known as ImageNet Large Scale Visual Recognition Challenge (ILSVRC) since 2010 (Russakovsky et al., 2015). Many people participated in this competition with their software program to classify and detect the objects in the ImageNet. AlexNet, a convolutional neural network won ImageNet2012 Challenge with 15.3% of top-5 error rate (Krizhevsky et al., 2012). The successful achievement of AlexNet was due to the usage of GPU and ImageNet dataset. Later, in ILSVRC 2014, a very deep convolutional network called VGG network won the challenge (Simonyan et al., 2014). These networks can be used to detect and classify lung nodules by a method called as transfer learning.

Transfer learning is an approach in which a model that is trained for some other job can be reused for another related job. Instead of building a CNN from scratch, one can reuse the existing models trained on a different dataset to solve a certain job. The aim of transfer learning is to increase learning in the target job by transferring knowledge from the source job (Olivas et al., 2009).

CNNs like AlexNet, VGG16, VGG19 are usually employed for classification tasks. When the top layers are removed, the CNN can be a good feature extractor. Low level features are extracted by initial convolution layers. The middle convolution layers extract mid-level features. The final convolution layers will extract objects. When there are a greater number of convolution layers, we can extract a greater number of features. In this work, VGG19 has been used for feature extraction. The extracted features were concatenated with a popular textural feature known as haralick features. Thus, the hybridized feature set was created. These feature sets were dimensionally reduced using dimensionality reduction technique which is known as Principle Component Analysis (PCA), to lessen the computation time and visualize the features. The reduced set of features were used to train the machine learning classifier SVM and the performance of SVM was analyzed in differentiating malignant nodules from benign nodules.

The classifier performance hinges on the features of the input images. Several features and the classification schemes for lung nodule classification have been analyzed in this section. From the literature survey, many works in pulmonary classification were found to be

based on handcrafted feature extraction techniques. Recently few researchers have been focusing on deep features for training the machine learning algorithms namely Logistic Regression, SVM and Random Forest.

In a seminal paper, Han et al. reported the importance of textural features to classify lung nodules. LIDC database was used. They compared 2D, 3D haralick features, Gabor features and Local binary pattern features for differentiating the lung nodules. Classification was done by using SVM. The performance metrics for 3D haralick features based SVM was found to be 93.86% of AUC (Han et al., 2015).

Dhara et al. (2016) reported a methodology for classifying pulmonary nodules in lung CT images. The classification method was validated using LIDC database with 891 nodules. The sensitivity and specificity of the system were 89.73% and 86.36% respectively.

Hua et al. (2015) reported deep learning framework for pulmonary nodule classification. They compared the performance of various classifiers like deep belief network (DBN), convolutional neural network (CNN), Scale invariant feature transform (SIFT) and Fractal. They found that sensitivity of DBN is 73.4%, specificity of DBN is 82.2%. For CNN, sensitivity was 73.3% and specificity was 78.7%. These deep learning-based networks performed well when compared to feature based SIFT and Fractal algorithms.

Setio et al. (2016) proposed a CAD using multi-view CNN for detecting lung nodules in CT images. 2D patches were extracted for each candidate nodules by the network itself. The sensitivity of the method ranged between 85.4% and 90.1% with 1 and 4 FP/scan.

Tajbaksh et al. (2017) compared Massive Training Artificial Neural Network (MTANN) and CNN for detecting and classifying lung nodules. They considered 50 lung nodules. They compared MTANN with different variants of CNNs namely shallow CNN, LeNet, Relatively deep CNN, AlexNet and fine-tuned AlexNet. They evaluated MTANN and CNNs under two different scenarios of training and testing datasets. One was division protocol and another one was 5-fold protocol. For evaluating lung nodule detection, performance metric is FROC analysis. They proved that MTANN performance is better than CNNs with 2.7 false positives per patient at 100% sensitivity whereas fine-tuned AlexNet generated 22.7 false positives per patient at 100% sensitivity. For classifying nodules as benign and malignant, they conducted experiment with different dataset and evaluated MTANN and CNNs. They found that MTANN performed better than CNNs with AUC of 0.8806.

Materials and Methods

The general flow chart of the proposed approach is illustrated in Figure 1.

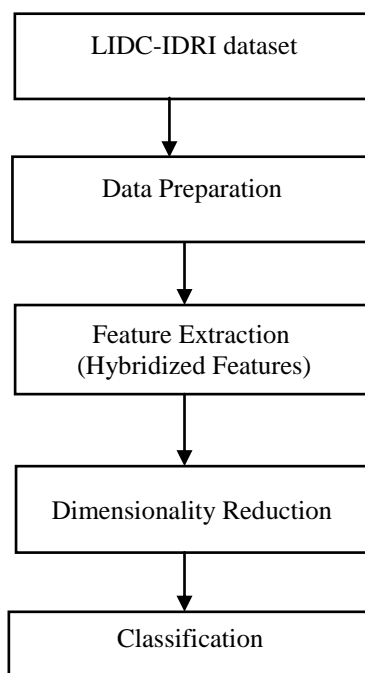


Figure 1. General flow chart of the proposed approach

Data Preparation

The image data set used in this work was taken from the publicly available LIDC-IDRI database. This data set comprises of CT scans of lung cancer patients with annotations from 4 different radiologists. In total, this dataset contains CT scans of 1010 patients. Each scan included an XML file which accounted the details of scan such as slice thickness, resolution and pixel spacing. Size of the nodules, number of nodules, coordinates of each nodule locations, boundary representations of each nodule and characteristics of nodules are available in the XML file (Armato et al., 2011). For this work, more focus was given to categorize the nodules into two categories: malignant or benign.

In LIDC-IDRI database, the malignancy has been assessed in 5 different levels. The range of this assessment was between 1 and 5. Level 1 signifies cancer risk is very less, level 2 represents cancer risk is moderate, level 3 characterizes midway risk for cancer, level 4 embodies moderately suspicious for cancer and level 5 denotes the risk factor for cancer is very high. In this work, levels 1,2 and 3 were combined to find the benign nodules and level 4 and 5 were combined for finding malignant nodules. By carefully analyzing the database, the 2D patches of nodules were extracted and a dataset has been created for the proposed classification scheme. This prepared dataset has benign and malignant nodules of size 64 x 64 with class labels. Class label 0 epitomizes benign nodule and class label 1 denotes malignant

nodule. The dataset derived from LIDC-IDRI contains 4165 benign nodule images and 2526 images are malignant nodule images.

Feature Extraction Techniques

The pixels available in an image carry information. If we process all the information contained in the image, there will be a huge requirement of memory space and computation time. To curb this problem, relevant data will be taken from the input image for further processing. Hence, the process of eliminating irrelevant information from the image without any significant loss in important information about the image is known as feature extraction. These features are a simplified representation of the image and play an important part in machine learning algorithm to differentiate one object from rest of the objects.

a) Handcrafted features

In medical image analysis, the popular imaging modalities are X-ray, CT and MRI. These medical images do not have color information. They are gray scale images. In order to identify the abnormality in such medical images, textural features are very suitable. In this work, the lung nodules were classified into benign and malignant. The benign nodules possess smooth surface whereas the malignant nodules possess uneven surface. Hence, textural features were considered. Gray Level Co-occurrence Matrix (GLCM) method is one of the statistical feature descriptors to take out texture features. These features which are derived from GLCM method are called haralick features (Haralick et al., 1973). The GLCM of an $M \times M$ image has the probabilities $P_{d,\theta}(i,j)$ that two pixels i and j in a given particular direction (θ) which are parted by a pixel distance (d). This method explores the relationship among the neighbourhood pixels in the given image. The intensity of each pixel in the image was quantized into number of times the gray tones i and j are neighbours. The matrices were generated in 4 directions such as 0° , 45° , 90° and 135° . The 14 Haralick features have been calculated from these matrices on each direction separately. This resulted in a feature vector of size 56.

b) Deep Features

Even though GLCM is capable of representing texture features to distinguish malignant from benign nodules, it is not a generalized one and the performance varies for different set of images due to the manual intervention in feature vector calculation. A deep learning model, CNN, contains of sequential layers of convolutional layer, max pooling layer, fully connected layer and softmax classifier (Mastouri et al., 2020). The general architecture of CNN is shown in Figure 2. The architecture of CNN is similar to the neuron connections in the brains of human beings. It follows the visual cortex behaviour of the human brain. Receptive field in the visual cortex plays a major role in the visual system. The neurons present in the receptive

field will give response to the stimulus. The collection of receptive fields is responsible to cover the complete visual area.

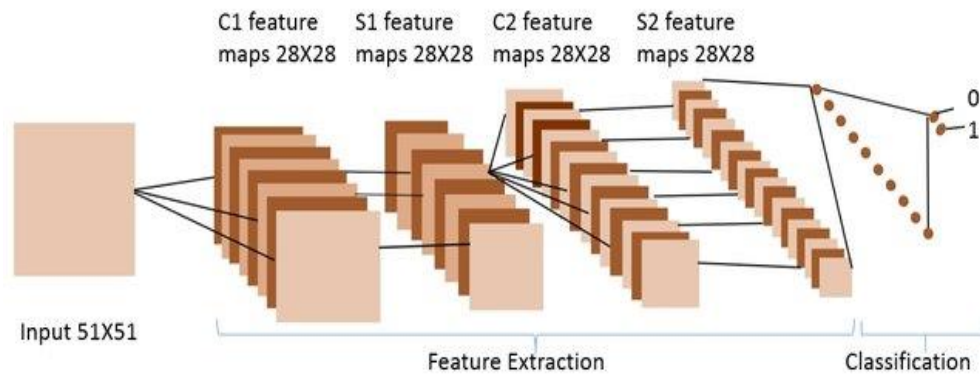


Figure 2. General Architecture of CNN

Convolution Layer

The fundamental building block in CNN is the convolution layer. This convolution layer consists of convolution kernels which carry out convolution operation. The convolutions filters are also called as convolution kernels. For generating feature map, the kernels which are having set of weights have been employed to the input image repeatedly. After generating the feature map, the elements in the feature map are passed through the activation function. The commonly used activation functions in CNN are ReLu or leaky ReLu. The features from the input images are extracted by the help of convolution layers without the help of any manual intervention.

Max-Pooling layer

Convolution layer is followed by max-pooling layer. This max-pooling operation depends on max pooling operator which is a kernel. This filter will select the maximum value from the neighbourhood. This operation is used to downsize the feature set. The max-pooling size and striding can be selected by the designer. The important characteristics of max-pooling operation are location invariant, scale invariant and rotation invariant.

Fully Connected layers

In CNN, fully connected layers are very important entity to classify the images into different groups. When they are passed through convolutional layer and max-pooling layers, they are broken down into features. These features will be flattened by the fully connected input layer and converted into a 1D array. This 1D feature vector will be fed into the first fully connected layer and weights will be applied to predict the class of the corresponding input image. The fully connected output layer i.e., SoftMax layer will assign the probability of each

class with respect to the features. The classified output will be obtained at this final fully connected layer.

Transfer Learning

In this work, transfer learning technique was adopted because of lack of a greater number of lung images. VGG16 and VGG19 models were implemented using Keras Library. VGG16 and VGG19 were pre-trained using ImageNet dataset. The list of all weights variables pertinent to each layer was available. The first part of CNN architecture has feature extraction blocks and the final layers were meant for classification. Here, CNN was used as a feature extraction tool. Hence, the last fully connected layers are removed from the architecture. The feature map which is available at this layer was stored as NumPy array. For visualizing the features, it has been converted into 2D image. Then flattening is performed to convert 2D into 1D array. The following steps are involved for extracting features from the pretrained model.

Step 1: A base model (VGG16/VGG19) is instantiated.

Step 2: Pre-trained weights of VGG16/VGG19 is loaded into the model.

Step 3: The lung nodule dataset is allowed to run on the base model.

Step 4: The features are obtained from the block 5maxpooling layer.

VGG19

One of the variants of VGG model is VGG19. It was developed by Visual Geometry Group, Oxford. The architecture of VGG is shown in Figure 3. This architecture has 16 convolutional layers and 3 fully connected layers. If 16 convolutional layers and 3 fully connected layers are added, 19 layers are obtained (Balaji et al., 2018). Hence, it is called as VGG19. Moreover, it has 5 max-pooling layers and 1 soft-max layer. It is a deep convolutional neural network which is used to classify images. The kernel size of each convolutional layer is 3x3 with striding, $s=1$ and padding, $p=1$. In order to introduce non-linearity, ReLu activation function is applied after every convolution operation. The kernel size of max-pooling layer is 2x2 with striding, $s=2$. In block 1, there are 2 convolutional layers. Each layer has 64 kernels. In block 2, there are 2 convolutional layers and each layer has 128 kernels. In block 3, there are 4 convolutional layers and each layer has 256 kernels. In block 4, there are 4 convolutional layers and each layer has 512 kernels. Block 5 also has 4 convolutional layers and each layer has 512 kernels (Ma et al., 2019). After every block, there is a max-pooling layer. When the image gets transferred from one block to another block, the image size gets reduced and the depth gets increased. Out of 3 fully connected layers, the first two fully connected layers are having 4096 neurons. The last fully connected layer has 1000 neurons. Since there are 1000 classes in ILSCRC, it has been designed like this. The softmax

layer does the classification by giving more probability to respective class neuron (Wani et al., 2020).

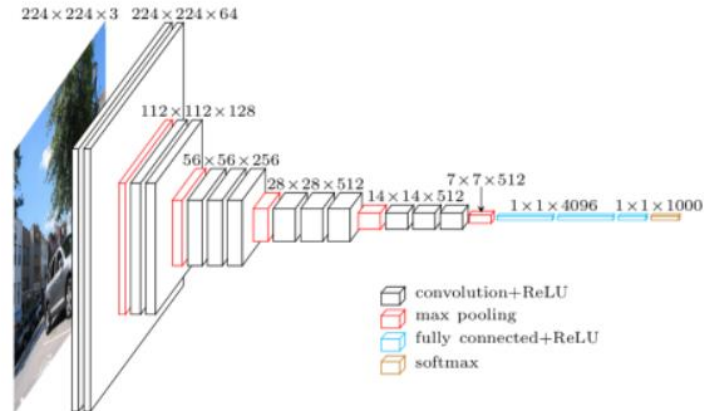


Figure 3. Architecture of VGG

Proposed Model

The block diagram of proposed model for lung nodule classification is illustrated in Figure 4. In the proposed model, the features extracted using GLCM method and the deep features extracted using VGG19 have been concatenated. For extracting features, the architecture of VGG-19 was modified slightly. In the modified architecture, the input layer was followed by block 1 which consisted of 2 convolutional layers and 1 max-pooling layer. Block 1 was followed by block 2 which consisted of 2 convolutional layers and 1 max-pooling layer. It was followed by block 3 which consisted of 4 convolutional layers and 1 max-pooling layer. It was followed by block 4 which comprised of 4 convolutional layers and a max-pooling layer (Ourselin et al., 2016). It was followed by block 5 which comprised of 4 convolutional layers and a max-pooling layer. The features were taken from the max-pooling layer of block 5. The features were stored as NumPy array. The features were visualized. The image size was reduced to $2 \times 2 \times 512$. The depth of the image was 512. Finally, the 2D feature map was converted into 1D array by means of flattening operation. 2048 deep features were obtained through VGG19 and 56 features were obtained through GLCM methodology. The features obtained by these methods were horizontally stacked and 2104 features were finalized.

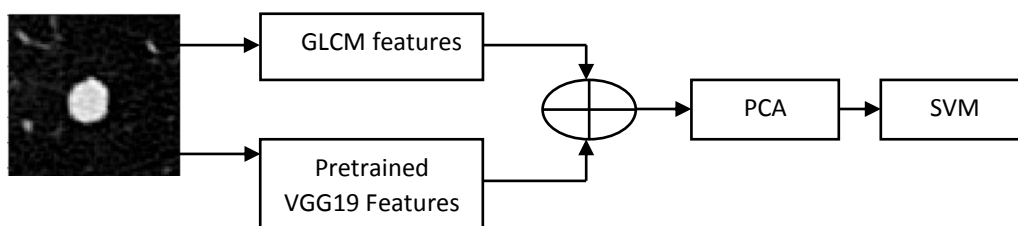


Figure 4. Block diagram of proposed model for lung nodule classification

Impact of PCA

When one wants to reduce the dimensionality of very big data sets, he/she needs to use a dimensionality reduction method. There are many methods available out of which Principal Component Analysis (PCA) is often used method. It will transform a huge set of features into reduced feature set that will preserve much information available. The accuracy is however affected when the number of variables of a data set is reduced. This is done so as to have simplicity. In order to have simplicity, a little trade-off is done that is the accuracy is slightly less. It is easy to explore smaller data sets and the whole process would be much faster. The process also would be much faster for machine learning algorithms. With this idea in mind, PCA was employed to reduce the number of features. At the same time, information was saved as much as possible.

The steps to be followed to obtain reduced feature vector is given below.

1. Standardize the range of all the features(data).

$$f_z = \frac{v - \mu}{\sigma} \quad (1)$$

f_z is standardized value

v is given value

μ is mean

σ is standard deviation

2. Compute the covariance matrix

$$S = v^T v \quad (2)$$

3. Calculate the eigen values(λ) and eigen vectors(V) of the covariance matrix S.

$$\lambda_1, \lambda_2, \dots, \lambda_d \quad (3)$$

$$v_1, v_2, \dots, v_d$$

4. Select the top features from the eigen vectors(V).

Results and Discussions

Fourteen different combinations were trained and tested on malignant and benign nodule dataset which consists of 2526 malignant nodule images (positive class) and 4165 benign nodule images (negative class). 2020 malignant nodule images and 3331 benign nodule images were used for training the classifiers. 506 malignant nodule images and 834 benign

nodule images were used for testing the classifiers. The individual performance of classifiers such as logistic regression and SVM for handcrafted features, deep features and hybridized features were analyzed. 5-fold cross validation was employed for evaluating the performance of the classifiers. The following performance metrics were chosen to evaluate the performance of the classifiers: Accuracy, Sensitivity (Recall), Specificity, Precision, F1 Score, False Positive Rate, False Negative Rate and AUC score.

Accuracy (Acc) tells the number of correct predictions out of all predictions. The other name of Sensitivity (Sen) is recall or True Positive Rate. It is the ratio of number of malignant nodules that are correctly classified as malignant nodules to the total number of malignant nodules in the test dataset. Specificity (Spe) is also called as True Negative Rate. It is the ratio of number of benign nodules that are correctly classified as benign nodules to the total number of benign nodules in the test dataset (Liang et al., 2020). Precision (Pre) is otherwise called as Positive Predictive Value which measures the percentage of them are actually malignant nodules (positive class). It does not consider the benign nodules (negative class).

The harmonic mean of precision and sensitivity is known as F1 score. It gives equal weightage to precision and sensitivity. F1 score lies between 0 and 1. False Positive Rate (Fall-out) is the measure which represents how many benign nodules are misclassified as malignant nodules. False Negative rate represents how many malignant nodules are misclassified as benign nodules. Both FPR and FNR are very dangerous in medical image analysis. These values are expected to be very low. For binary classification, Area Under the receiver operating Curve (AUC) is one of the popular performance metrics. If the train AUC and test AUC value difference is large, the model is overfitting. If that value is lesser than 0.5, the model is under-fitting.

In logistic regression, best alpha (α) and penalty (p) were computed using GridsearchCV method. Logistic loss was considered for logistic regression. L_2 regularization was employed to avoid overfitting. The hyperparameter λ has to be tuned very carefully. If $\lambda=0$, there will be no regularization. This leads to overfitting to the training data. If λ is very large, loss term value will be reduced and this leads to underfitting. For determining best λ , GridsearchCV method was employed. In SVM, radial basis function kernel with $\gamma = 0.01$ was used. Here, c is the hyperparameter. GridsearchCV method was used to tune c value. The threshold was evaluated based on the following equation to generate the confusion matrix.

$$\text{Threshold} = TPR * (1 - FPR) \quad (4)$$

where TPR is True Positive Rate and FPR is False Positive Rate

Analysis of Handcrafted Features

For this experiment, haralick features were taken from GLCM of benign nodule images and malignant nodule images. In all the four directions, pixel distance 1 was considered. Totally, 56 features were extracted. These features were used to train logistic regression and SVM. In logistic regression, the best alpha $a=1e^{-06}$ and penalty $p=12$ were obtained. In SVM, $c=1e^{-10}$ was obtained. For logistic regression, the threshold was 0.277. For SVM, it was 0.282. Table 1 shows the performance measures obtained when logistic regression and SVM was tested using haralick features. Logistic regression was slightly better than SVM with high accuracy, sensitivity, precision and AUC values. Both models did not overfit to the training data. FPR in both models were very large. But FNR was nominal when compared to FPR. The ROC curve of SVM with GLCM features is shown in Figure 5.

Table 1. Performance measures of classifiers with handcrafted features

Method	Parameters		
GLCM+Logistic Regression	Acc: 57.31% Sen: 85.18% Spe: 40.00% Pre: 46.44%	F1 Score: 0.60 FPR: 60.00% FNR: 14.80%	Train AUC: 0.8535 Test AUC: 0.759
GLCM+SVM	Acc: 56.86% Sen: 83.79% Spe: 40.53% Pre: 46.00%	F1 Score: 0.59 FPR: 59.00% FNR: 16.00%	Train AUC: 0.8113 Test AUC: 0.7243

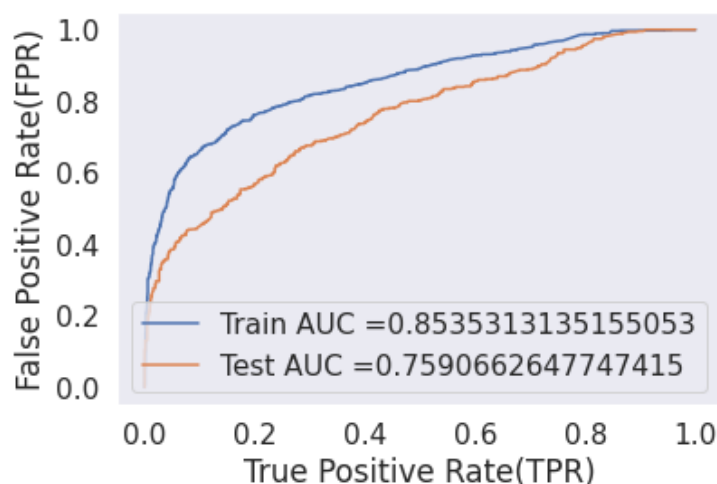


Figure 5. ROC of SVM with handcrafted features (GLCM)

Analysis of Deep Features

The deep features were extracted from VGG16 and VGG19. Table 2 shows the performance metrics of Logistic Regression and SVM when they were tested with deep features. VGG features with SVM gave good results when compared to VGG features with logistic regression. In all the combinations, there was neither overfitting nor underfitting. AUC value was higher than 0.5. The ROC of SVM with VGG19 features is depicted in Figure 6.

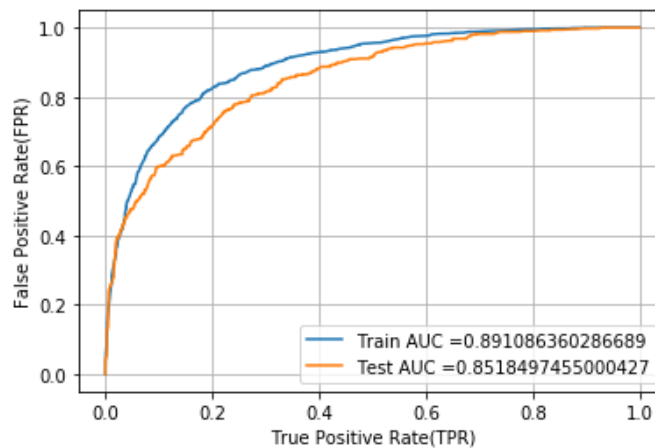


Figure 6. ROC of SVM with deep features (VGG19)

Table 2. Performance measures of classifiers with deep features

Method	Parameters		
VGG16+Logistic Regression	Acc: 58.36% Sen: 62.34% Spe: 55.40% Pre: 46.24%	F1 Score: 0.53 FPR: 44.60% FNR: 36.76%	Train AUC: 0.6633 Test AUC: 0.6426
VGG16+SVM	Acc: 77.84% Sen: 82.21% Spe: 75.18% Pre: 66.77%	F1 Score: 0.74 FPR: 24.80% FNR : 17.78%	Train AUC: 0.8975 Test AUC: 0.8702
VGG19+Logistic Regression	Acc: 74.22% Sen: 80.04% Spe: 70.69% Pre: 62.31%	F1 Score: 0.70 FPR:29.30% FNR:19.96%	Train AUC: 0.8813 Test AUC: 0.8474
VGG19+SVM	Acc: 76.42% Sen: 76.68% Spe:76.26% Pre: 66.21%	F1 Score: 0.71 FPR:23.74% FNR: 23.32%	Train AUC:0.8910 Test AUC:0.8518

Analysis of hybrid features without PCA

In the next model, both deep features and handcrafted features were concatenated by horizontal stacking method. In total, 2104 features were obtained. These features were used to train Logistic regression and SVM. While testing, it was found that the hybridized features with SVM gave very good results. The performance metrics are shown in table 3. Accuracy of 93.28%, Sensitivity of 89.50%, Specificity of 95.6%, Precision of 92.43 were obtained. High F1 score and AUC score were obtained. FPR and FNR were quite low. The ROC of SVM with hybridized features are shown in Figure 7.

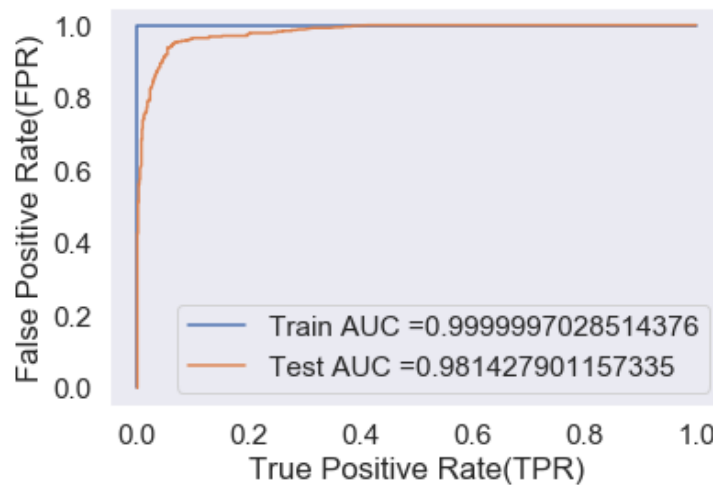


Figure 7. ROC of SVM with hybridized features (GLCM and VGG19)

Table 3. Performance measures of hybrid features without feature reduction technique

Method	Parameters		
GLCM+VGG16+ Logistic Regression	Acc: 58.43% Sen: 70.75% Spe: 50.96% Pre: 46.67%	F1 Score: 0.56 FPR: 49.00% FNR:29.25%	Train AUC:0.6851 Test AUC:0.6743
GLCM+VGG16+ SVM	Acc: 62.38% Sen: 58.89% Spe:64.50% Pre: 50.17%	F1 Score: 0.54 FPR: 35.40% FNR: 41.11%	Train AUC: 0.6671 Test AUC: 0.6735
GLCM+VGG19+ Logistic Regression	Acc: 77.23% Sen: 88.93% Spe:70.14% Pre: 64.37%	F1 Score: 0.75 FPR:29.85% FNR:11.07%	Train AUC: 0.9330 Test AUC: 0.8966
GLCM+VGG19+ SVM	Acc: 93.28% Sen: 89.50% Spe:95.60% Pre:92.43%	F1 Score: 0.91 FPR :4.44% FNR: 10.50%	Train AUC: 0.9999 Test AUC: 0.9841

Analysis of hybrid features with PCA

The hybridized features were dimensionally reduced using PCA in order to minimize the computation time. 2104 features were reduced to 133 features. It is found that the dimensionality reduction improved the results in all aspects. Time complexity and space complexity were reduced due to PCA. After applying PCA, 99% of variance was preserved. The hybridized features GLCM and VGG19 along with PCA outperformed all other combinations with 94.93% of accuracy, 90.90% of sensitivity, 97.36% of specificity, 95.44% of precision, 2.64% of FPR and 9.09% of FNR. The F1 score was 0.93 and test AUC was 0.9843. The ROC of SVM with dimensionally reduced hybridized features is shown in Figure 8. Table 4 depicts the performance measures of hybridized features with PCA.

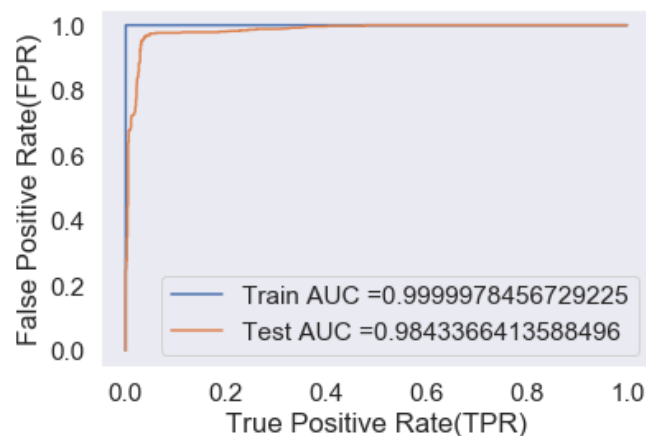


Figure 8. ROC of SVM with dimensionally reduced hybridized features (GLCM + VGG19 + PCA)

Table 4. Performance measures of hybrid features with feature reduction technique

Method	Parameters		
GLCM+VGG16+ PCA+ Logistic Regression	Acc: 78.66% Sen: 80.83% Spe: 77.34% Pre: 68.39%	F1 Score: 0.74 FPR: 22.66% FNR: 19.16%	Train AUC: 0.9065 Test AUC: 0.8774
GLCM+VGG16+PCA+ SVM	Acc: 78.43% Sen: 82.00% Spe:76.25% Pre: 67.69%	F1 Score: 0.74 FPR: 23.74% FNR: 17.98%	Train AUC: 0.9008 Test AUC: 0.8730
GLCM+VGG19+PCA+ Logistic Regression	Acc: 82.80% Sen: 87.55% Spe:79.97% Pre: 72.6%	F1 Score: 0.79 FPR: 20.02% FNR: 12.45%	Train AUC: 0.9637 Test AUC: 0.9264
GLCM+VGG19+PCA+ SVM	Acc: 94.93% Sen: 90.90% Spe:97.36% Pre:95.44%	F1 Score: 0.93 FPR: 2.64% FNR: 9.09%	Train AUC: 0.9999 Test AUC: 0.9843

The performance of all the 14 combinations is summarized in Figure 9, Figure 10 and Figure 11. The proposed models such as VGG19+GLCM features without PCA and with PCA outperform all other combinations.

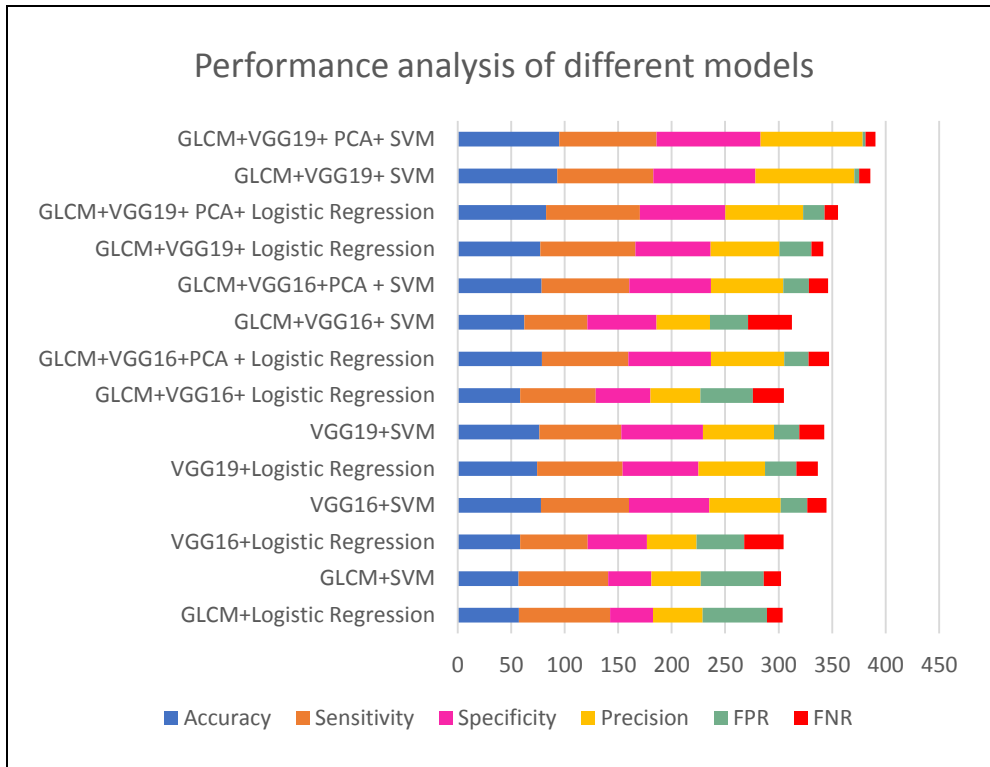


Figure 9. Performance analysis of different models

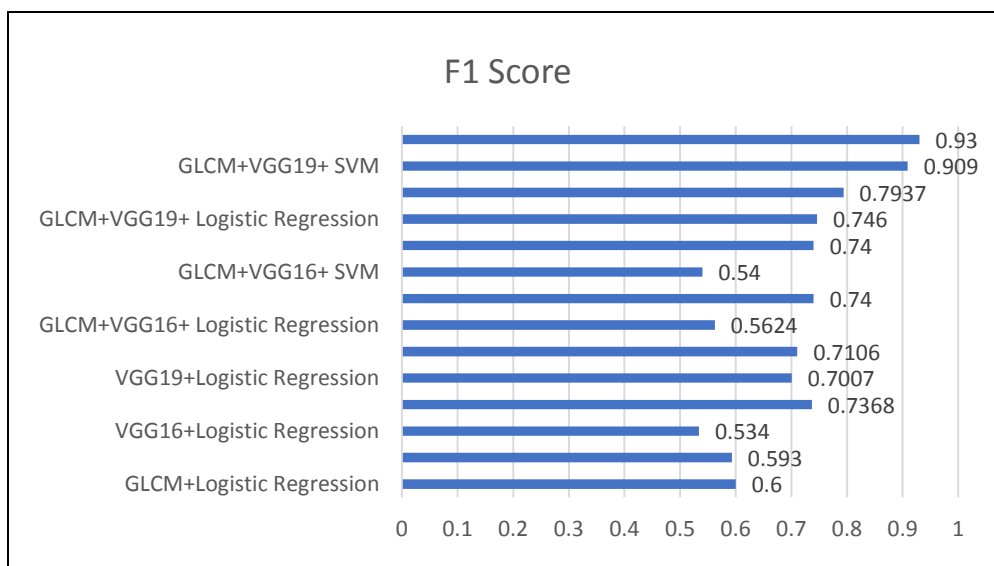


Figure 10. Performance analysis based on F1 score

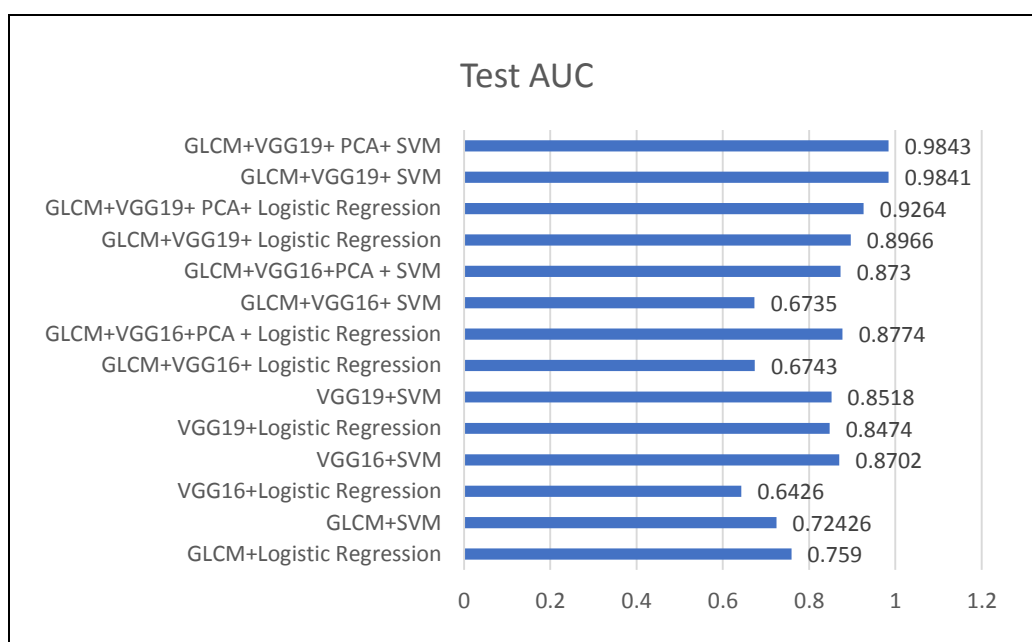


Figure 11. Performance analysis based on AUC

Table 5 gives the comparison of present work with previous reported works. It can be noted that the proposed hybridized methodology gives better output when compared to other methodologies.

Table 5. Comparison with previous works

Other papers	Acc (%)	Sen (%)	Spe (%)	Pre (%)	F1 Score	FPR (%)	FNR (%)	AUC
Proposed approach	94.925	90.9	97.36	95.435	0.93	2.64	9.09	0.9843
Wang et al. (2018)	91.75	-	-	-	-	-	-	0.9702
Raul Victor et al. (2018)	88.41	85.38	-	73.48	0.79	-	-	0.9319
Shen et al. (2015)	86.80	-	-	-	-	-	-	-
Shen et al. (2017)	93	87.14	93	-	-	-	-	0.93
Antonio et al. (2018)	92.63	90.7	93.47	-	-	-	-	0.934
Kumar et al. (2015)	75.01	83.35	-	-	-	-	-	-
Han et al. (2015)	-	-	-	-	-	-	-	0.9270
Dhara et al. (2016)	-	82.89	80.73	-	-	-	-	0.8822
Sarfaraz et al. (2017)	91.26	-	-	-	-	-	-	-
Zhu et al. (2018)	90.44	-	-	-	-	-	-	-
Wei et al. (2018)	85.2	85.80	-	85.8%	0.8580	-	-	0.9863
Wei et al. (2018)	91	92.10	-	92.3	0.9180			0.9840

Conclusion

A hybridized approach was followed in the present work. Handcrafted features and deep features were hybridized using appropriate methods. Support Vector Machine (SVM) and Logistic Regression (LR) which are machine learning algorithms were used to classify the given pulmonary nodules into benign and malignant based on the features. PCA has been introduced to enhance the performance of the hybridized features. The work has been carried out with 14 different methods. It was found that GLCM + VGG19 + PCA + SVM outperformed all other models with an accuracy of 94.93%, sensitivity of 90.9%, specificity of 97.36% and precision of 95.44%. The F1 score was found to be 0.93 and the AUC was 0.9843. The False Positive Rate was found to be 2.637% and False Negative Rate was 9.09%.

References

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4), 433-459.
- Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., ... & Kazerooni, E. A. (2011). The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical physics*, 38(2), 915-931.
- Balaji, K., & Lavanya, K. (2018). Recent Trends in Deep Learning with Applications. In *Cognitive Computing for Big Data Systems Over IoT* (pp. 201-222). Springer, Cham
- da Nóbrega, R. V. M., Peixoto, S. A., da Silva, S. P. P., & Rebouças Filho, P. P. (2018, June). Lung nodule classification via deep transfer learning in CT lung images. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 244-249). IEEE
- de Carvalho Filho, A. O., Silva, A. C., de Paiva, A. C., Nunes, R. A., & Gattass, M. (2018). Classification of patterns of benignity and malignancy based on CT using topology-based phylogenetic diversity index and convolutional neural network. *Pattern Recognition*, 81, 200-212.
- Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE.
- Dhara, A. K., Mukhopadhyay, S., Dutta, A., Garg, M., & Khandelwal, N. (2016). A combination of shape and texture features for classification of pulmonary nodules in lung CT images. *Journal of digital imaging*, 29(4), 466-475.
- Fukushima, K. (1980). Biological cybernetics neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern*, 36, 193-202.
- Han, F., Wang, H., Zhang, G., Han, H., Song, B., Li, L., ... & Liang, Z. (2015). Texture feature analysis for computer-aided diagnosis on pulmonary nodules. *Journal of digital imaging*, 28(1), 99-115.

- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621.
- Hua, K. L., Hsu, C. H., Hidayati, S. C., Cheng, W. H., & Chen, Y. J. (2015). Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoTargets and therapy*, 8, 2015-2022.
- Hussein, S., Cao, K., Song, Q., & Bagci, U. (2017, June). Risk stratification of lung nodules using 3D CNN-based multi-task learning. In *International conference on information processing in medical imaging* (pp. 249-260). Springer, Cham.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- Kumar, D., Wong, A., & Clausi, D. A. (2015, June). Lung nodule classification using deep features in CT images. In *2015 12th Conference on Computer and Robot Vision* (pp. 133-138). IEEE.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324
- Liang, Q., Mu, J., Jia, M., Wang, W., Feng, X., Zhang, B. (Eds.), (2020) *Communications, Signal Processing, and Systems: Proceedings of the 2017 International Conference on Communications, Signal Processing, and Systems*, Springer.
- Lo, S. C. B., Chan, H. P., Lin, J. S., Li, H., Freedman, M. T., & Mun, S. K. (1995). Artificial convolution neural network for medical image pattern recognition. *Neural networks*, 8(7-8), 1201-1214.
- Ma, Y., Xie, Q., Liu, Y., & Xiong, S. (2019). A weighted KNN-based automatic image annotation method. *Neural Computing and Applications*, 1-12
- Mastouri, R., Khelifa, N., Neji, H., & Hantous-Zannad, S. (2020). Deep learning-based CAD schemes for the detection and classification of lung nodules from CT images: A survey. *Journal of X-Ray Science and Technology*, (Preprint), 1-27
- Olivas, E. S., Guerrero, J. D. M., Martinez-Sober, M., Magdalena-Benedito, J. R., & Serrano, L. (Eds.). (2009). *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques*. IGI Global.
- Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G., & Wells, W. (Eds.). (2016). *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II* (Vol. 9901). Springer.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3), 211-252.
- Setio, A. A. A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., Van Riel, S. J., ... & van Ginneken, B. (2016). Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE transactions on medical imaging*, 35(5), 1160-1169.
- Shen, W., Zhou, M., Yang, F., Yang, C., & Tian, J. (2015, June). Multi-scale convolutional neural networks for lung nodule classification. In *International Conference on Information Processing in Medical Imaging* (pp. 588-599). Springer, Cham.

- Shen, W., Zhou, M., Yang, F., Yu, D., Dong, D., Yang, C., ... & Tian, J. (2017). Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognition*, *61*, 663-673.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- Tajbakhsh, N., & Suzuki, K. (2017). Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs. *Pattern recognition*, *63*, 476-486.
- Wang, H., Zhao, T., Li, L. C., Pan, H., Liu, W., Gao, H., ... & Liang, Z. (2018). A hybrid CNN feature model for pulmonary nodule malignancy risk differentiation. *Journal of X-ray Science and Technology*, *26*(2), 171-187.
- Wani, M. A., Bhat, F. A., Afzal, S., & Khan, A. I. (2020). *Advances in deep learning* (Vol. 57). Berlin: Springer.
- Wei, G., Cao, H., Ma, H., Qi, S., Qian, W., & Ma, Z. (2018). Content-based image retrieval for lung nodule classification using texture features and learned distance metric. *Journal of medical systems*, *42*(1), 13.
- Wei, G., Ma, H., Qian, W., Han, F., Jiang, H., Qi, S., & Qiu, M. (2018). Lung nodule classification using local kernel regression models with out-of-sample extension. *Biomedical Signal Processing and Control*, *40*, 1-9.
- Weng, J. J., Ahuja, N., & Huang, T. S. (1993, May). Learning recognition and segmentation of 3-D objects from 2-D images. In 1993 (4th) International Conference on Computer Vision (pp. 121-128). IEEE.
- Zhu, W., Liu, C., Fan, W., & Xie, X. (2018, March). Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 673-681). IEEE.

Bibliographic information of this paper for citing:

Bruntha, M., Pandian, I.A., & Abraham, S.S. (2020). Classification of Lung Nodule Using Hybridized Deep Feature Technique. *Journal of Information Technology Management*, Special Issue, 109-128.