

Detection of Lung Nodules on X-ray Using Transfer Learning and Manual Features

Imran Arshad Choudhry* and Adnan N. Qureshi

Faculty of Information Technology, University of Central Punjab, Lahore, Pakistan

*Corresponding Author: Imran Arshad Choudhry. Email: i.arshad@ucp.edu.pk

Received: 16 November 2021; Accepted: 12 January 2022

Abstract: The well-established mortality rates due to lung cancers, scarcity of radiology experts and inter-observer variability underpin the dire need for robust and accurate computer aided diagnostics to provide a second opinion. To this end, we propose a feature grafting approach to classify lung cancer images from publicly available National Institute of Health (NIH) chest X-Ray dataset comprised of 30,805 unique patients. The performance of transfer learning with pre-trained VGG and Inception models is evaluated in comparison against manually extracted radiomics features added to convolutional neural network using custom layer. For classification with both approaches, Support Vectors Machines (SVM) are used. The results from the 5-fold cross validation report Area Under Curve (AUC) of 0.92 and accuracy of 96.87% in detecting lung nodules with the proposed method. This is a plausible improvement against the observed accuracy of transfer learning using Inception (79.87%). The specificity of all methods is >99%, however, the sensitivity of the proposed method (97.24%) surpasses that of transfer learning approaches (<67%). Furthermore, it is observed that the true positive rate with SVM is highest at the same false-positive rate in experiments amongst Random Forests, Decision Trees, and K-Nearest Neighbor classifiers. Hence, the proposed approach can be used in clinical and research environments to provide second opinions very close to the experts' intuition.

Keywords: Lungs cancer; convolutional neural network; hand-crafted feature extraction; deep learning; classification

1 Introduction

Lung cancer is one of the deadliest forms of cancer has a high mortality rate. It is the second most common cancer among men and women in the United States [1]. According to World Health Organization (WHO), 9771 new cases of lung cancer have been reported in both men and women of Pakistan for the year 2018 which is 5.6% of the total number of lung cancer cases reported. It is commonly diagnosed in men and accounts for 14.5% of total cases in men and 8.4% of total cases in women [2].



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lung cancer is further categorized into two categories based on cell size: Small Cell Lung Cancer (SCLC) and Non-small Cell Lung Cancer (NSCLC). The former is considered to be highly malignant with early metastasis and poor prognosis and accounts for 15%–20% of all cases of lung cancer [2]. SCLC is further categorized into two stages, namely, extensive stage and limited stage. The cancer is limited to one side of chest in the limited stage while it spreads to the whole lungs and lymph nodes in the extensive stage. Every two out of three patients are diagnosed with extensive stage and have to undergo regular chemotherapy sessions as the treatment [2].

As mentioned, this form of cancer is the leading cause of mortality around the world and a timely detection mechanism in the early stages is critically imperative for treatment regimen. Computer Aided Diagnosis (CAD) has gained the attention of researchers for its early diagnosis followed by appropriate treatment [3]. Precise assessment of pulmonary nodules can help ascertain the degree of lung cancer [4]. Currently, a fine needle biopsy is the most common method for examining the malignancy status of the pulmonary nodules but it is a severely painful invasive test for the patient. Therefore, computer aided detection mechanisms are a dire need of time to support the clinical diagnostic methods and save time and lives. Thus, non-invasive approaches, such as Computer tomography (CT) scans can be considered as a replacement and this procedure takes lesser time and is completely painless [4].

In image processing and computer vision applications, a feature is a measurable part of the information that has some unique capability to describe the image(s). These features are further classified into several classes (sometimes called a segment). But the classification of grayscale images such as x-ray images is completely spatially blind. Usually, for generation of Region of Interest (ROI) or cluster, it uses pixel intensity. The feature space (especially the location of the pixels and gray level intensity) of Region of Interest (ROI) is limited. The problem of in-homogeneity due to background contribution and quasi-homogeneity due to noise in grayscale images arises. Different approaches are used for medical image classification. Some approaches are presented in this thesis such as Region-based [5] or Clustering based [6,7], Atlas based Hybrid-classification based, partial differential equation, and thresholding based [8].

The main theme of this article is to extract manual features of medical images such as lungs (x-ray images) and concatenation with auto features extracted by Convolutional Neural Network (CNN) to find the precise and pertinent features which are used for classification.

2 Background

Different approaches were adopted for manual feature extraction into different environments. Cuenca et al. [5] and Freixenet et al. [9] have published their work and proposed classification based concept of nodule detection using 3D region growing algorithms. Initially, they used a selective enhancement filter and thresholding approach. They achieved 71.8% accuracy with 0.8 False Positive (FP). They presented that region-growing based algorithms have poor results as compared to thresholding based classification approaches.

For dividing pixel into different groups it used some criteria functions or grouping algorithms (use similarity measures-these algorithms are used to find the similarity between two points ($A = \{a_1, a_2, \dots, a_n\}$, $B = \{b_1, b_2, \dots, b_n\}$) and make groups of similar pixel Eq. (1) can be generalized by defining Minkowski, Manhattan and Euclidean Distance) such as Expectation Maximization (EM), K-MEANS and Fuzzy C-Means (FCM) algorithm. For nodule classification, Filho et al. [10] proposed a concept of Quality thresholding algorithm and Javaid et al. [11] have developed a CAD

based system using K-MEANS algorithm for nodule detection.

$$A = \{a_1, a_2, \dots, a_n\}$$

$$B = \{b_1, b_2, \dots, b_n\}$$

$$\begin{aligned} Minkowski(A, B) &= \left(\sum_{i=1}^n |a_i - b_i|^p \right)^{\frac{1}{p}} \\ \text{Cosine} &= \frac{\sum a_i b_i}{\sqrt{\sum a_i^2} \sqrt{\sum b_i^2}} \\ \text{Euclidean} &= \sqrt{\sum |a_i - b_i|^2} \\ \text{Manhattan} &= \sum |a_i - b_i| \end{aligned} \tag{1}$$

Assefa et al. [12] developed the concept for nodule detection that combines template matching and multiresolution algorithms and reduces false-positive. They achieved sensitivity from range 84% to 91% by using template matching (For object detection and classification, this approach has a brute force algorithm. In this approach, an image is divided into sub-image or templates that contain the region of interest of an image. The sub-image or template slides over the whole image to explore the desired template. For searching or matching templates over the image, the most well-known searching algorithms are used such as Normalized Cross Correlation (NCC), Cross-correlation, Sum of Squared Difference (SSD), and Sum of Absolute Difference (SAD)). The only demerit of the template matching approach is selecting metric for template matching and literature depict that it takes too long time for compute the correlation.

Gong et al. [13] proposed a concept based on dynamic self-adaptive matching and Fisher Linear Discriminant Analysis (FLDA). They have used OTSU and 3D region growing algorithms for classification and for noise smoothing they used Gaussian smoothing operation. The main drawback of this approach, it works well in binary classification and fails in multi-classification problems.

In the last 20 years, computer vision based applications have used Scale-Invariant Feature Transform (SIFT), Haar Cascades, Speeded Up Robust Feature (SURF), Histogram of Oriented Gradients (HOG), and various statistical features such as Difference of Variance, Entropy, Energy, and Sum of Variance. Recently the researchers have shifted the paradigm from hand-crafted feature extraction to automatic feature extraction methods such as deep learning approaches. The pertinent reasons for the transition from hand-crafted to automatic feature extraction methods are:

1. Handcrafted feature extraction methods are time consuming-manually setting and tuning the bounding box on a data set or getting the required portion of images is tedious.
2. Sometimes data set have very low image quality to extract the required portion of pertinent features.
3. Handcrafted feature extraction requires the active participation of medical expertise to get precise information.

Automatic feature extraction such as CNN networks can extract features from data set. It gives them random weights to all the available features and during training, it adjusts these weights to extract the meaningful features. Convolution is the first layer of CNN, which extracts features from the input

image. Convolution learns the image features considering small squares of input data to establish the relationship between pixels. Eq. (2) depicts the mathematical operation of CNN that takes two inputs such as portion of an image and a filter/kernel. The final output of the convolution between image and filter/kernel is called “Feature Map”.

$$\sigma \left(b + \sum_{l=0}^4 \sum_{m=0}^4 w_{l,m} a_{j+l,k+m} \right) \quad (2)$$

A variety of computer aided diagnostic methods are using both image processing and deep learning approaches and Convolutional Neural Network (CNN) is one of the examples [14]. A Convolutional Neural Network (CNN) is a neural network that consists of one or more convolutional layers. These layers can be thought of as filters to the input data or some function that is to be applied to the whole or partial image i.e., input data. The number of layers depends on the number of features to be extracted from the input image or number of functions (conv(), batch Normalization() and max Pooling() etc.) needed to be applied to it. They are sometimes called Deep Convolutional Neural Networks (Deep CNNs) due to the large number of convolutional layers used in them.

The deep neural network requires a large amount of labeled datasets to work efficiently. While the number of annotated datasets that are publicly available are few, Imran et al. [15] have proposed a multi-task learning model for learning a classifier, for chest X-ray images along with a loss function (Tversky depicted in Eq. (3)) for convergence. It extracts the feature from the dataset in a complete black box manner, and it consists of several convolution layers with some additional layers such as max-pooling, dropout, and activation. These layers are learnable and transfer these weights into the next subsequent layer and further pass to the classifier as a vector. The classifier assigns a label to a vector or sometimes assigns multiple labels to a single vector.

$$T(\alpha, \beta) = \frac{\sum_{i=1}^N (p_{0i} g_{0i})}{\sum_{i=1}^N (p_{0i} g_{0i}) + \alpha \sum_{i=1}^N (p_{0i} g_{1i}) + \beta \sum_{i=1}^N (p_{1i} g_{0i})} \quad (3)$$

where α and β are used to control the magnitude and p_{oi} is the probability of pixel and g_{oi} is the loss of gradient.

Qin et al. [16] apply deep learning (which is a subset of machine learning that works on the principle of artificial neural networks) on X-ray images of the chest. The main focus is performing the lung segmentation (method of identifying the boundary of lungs from the surrounding tissues) efficiently. In the modern era of technology, machine learning and pattern recognition techniques have been widely used in computer vision (Image processing) and other AI-based system [16].

Baltruschat et al. [17] focused on chest x-ray radiography which is the most familiar type of examination used in imaging departments. The machine-driven detection tools in the radiology and clinical activity-flow could have enough combat on the level of care. In this research, they analyze and canvas the quality of two advanced and progressive image preprocessing methodologies at first formulated for getting image data points by radiologists, for the achievement and demonstration of deep learning methodologies. Jaiswal et al. [18] propose a deep learning model based on Mask-RCNN to localize pneumonia on Chest X-Ray images. The authors incorporate the local and global features for pixel-wise segmentation and report plausible performance on X-Ray images. It fails to segment pneumonia in low quality images and takes extra computation cost to analyze high quality images.

In study, Hussain et al. [19] explained concept of Reconstruction Independent Component Analysis (RICA) and Sparse filter features for detection of lungs cancer using machine learning algorithms. They have used multiple machine learning algorithms such as Gaussian Radial Base

Function (GRBF), Decision Tree, Support Vector Machine (SVM), and Naive Bayes to classify lungs cancer. Using RICA and Sparse filters, they have achieved plausible results with the jackknife cross validation technique. Kesim et al. [20] proposed a small Convolutional Neural Network (CNN) model for X-Ray image classification using Japanese Society of Radiological Technology (JSRT) dataset. The authors tested their model on JSRT dataset and achieved 86% accuracy. Bhandary et al. [21] customized a pre-trained CNN model AlexNet for detection of abnormalities on lungs X-Ray images. The authors focus on pneumonia detection in this study using customized AlexNet and it introduced a new threshold filter for feature ensemble strategy and achieved classification accuracy 96%. Cao et al. [22] introduce Variational Auto Encoder (VAE) in each layer of the Convolutional Neural Network (CNN) model. Their model was based on U-Net, the most widely used model for segmentation. Variational Auto Encoder (VAE) is used to extract symmetrical semantic information from both right and left thoraxes. The attention methodology used spatial and channel information to segment the region of interest in lungs to improve the segmentation accuracy. Salman et al. [23] explored the concept of deep learning methodologies on X-Ray images collected from Kaggle, Github, and Open-i repository. The author proposed a Convolutional Neural Network and apply on X-Ray images collected from Github.

We have observed that manual annotations task is very time costly and there is a high risk of human errors. Therefore, the aim of this research is to evaluate hybridization of manually extracted and convolutional features for classification and detection of lung cancer nodules which can significantly minimize reporting time and maximize accuracy.

3 Proposed Methodology

3.1 Datasets

The NIH chest X-Ray dataset comprised of 30805 unique patients with disease labeled data. Image size is 120×12 and has 15 classes (14 diseases [Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural-thickening, Cardiomegaly, Nodule, Mass, Hernia] and one for “No Finding”). Fig. 1 shows the sample images of NIH X-Ray dataset. The labels are extracted by Natural Language Processing algorithms based on text-mining disease classification under supervision by medical experts. As we are interested in Cancer detection, hence, we consider Nodule as class of interest. Because of data imbalance, we combine Lung Image Database Consortium image collection (LIDC-containing 244,527 low-dose lung images from 1010 lung unique patients) data into the experiments.

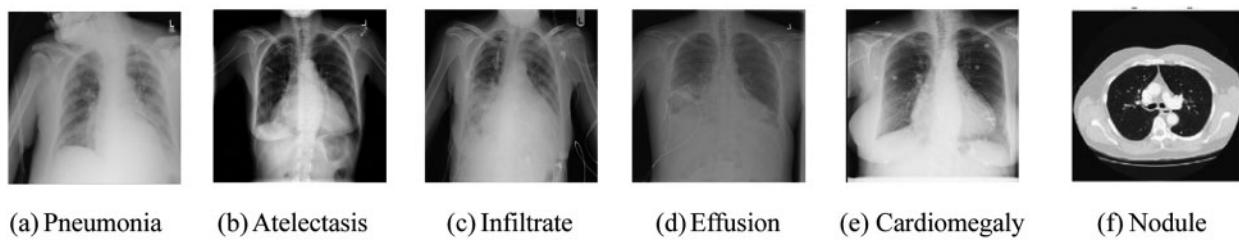


Figure 1: Sample of x-ray National Institutes of Health (NIH) chest X-ray dataset

Evaluating deep learning models can be quite difficult and tricky. Normally, we split the dataset into different ratios of training and testing sets. One of the most widely used statistical techniques to evaluate the performance of deep learning models and to avoid overfitting the model is Cross

Validation or K-Fold Cross Validation. For improving machine learning model prediction, we used k-fold cross validation technique as depicted in Fig. 2. In k-fold cross validation, the dataset is completely shuffled and to make sure that our inputs are not biased. Further, we divide the dataset into k size equal portions with no overlapping. According to the requirement and environment the size of k is set to 10 or 5. We have used 5 to split the dataset into 5 equal sized portions. Apart from k-folding, we improved the robustness of our proposed novel model and pre-trained models using data augmentation technique. Using this technique, we can generate several samples of under-sampled class images which helps in almost balancing the ratio of distribution. In this step, we artificially synthesize x-ray images from original x-ray images with the help of minor alteration in original images such as rotations, horizontal/vertical flipping, and scaling, zooming, padding and random brightness. We have achieved better results using data augmentation techniques, which prevent data scarcity, increases generalization, and resolve class imbalance issues.

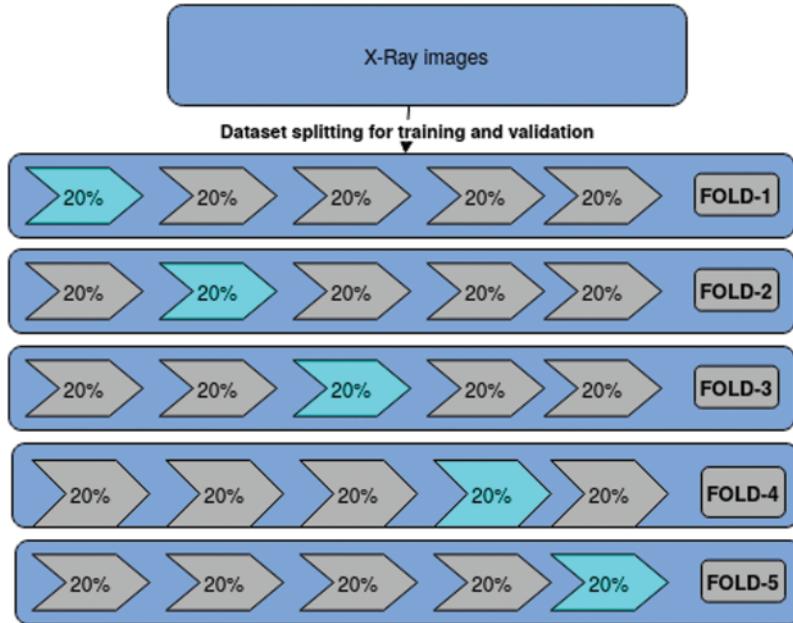


Figure 2: Dataset representation of training and validation procedure employed in the 5-fold cross validation

Mostly, Lungs X-ray are gray-scale dataset with have large amount of noise and have low quality images due to using different protocols during acquisition of images. From low quality dataset, extraction of visual features is quite a challenging task. So here, we have applied some contrast enhancement algorithms on these low quality images and achieve some better performance and efficacy. One of the most widely technique used in image processing for background equalization and feature extraction is morphological operations (especially bottom-hat and top-hat transformation).

$$\begin{aligned} Img_{top} &= In_{img} - (In_{img} \circ SE) \\ Img_{bot} &= In_{img} - (In_{img} \bullet SE) \\ Img_{enh} &= In_{img} + Img_{top} - Img_{bot} \end{aligned} \quad (4)$$

In top-hat morphological operation depicted in Eq. (4), we apply opening operation on an input image with the help of Structural Element (SE) and then subtract from the original image to get the bright features and objects from the original image. In the bottom-hat morphological operation

depicted in Eq. (4), we apply closing operation on an input image with the help of Structural Element (SE) and apply addition with the original image to extract the dark features and objects from the original image. After closing and opening operations on an image, we combine the operations subtracting the result of bottom-hat and adding the result of top-hat morphological operation from an image to get the enhanced image. Before applying the morphological operations, finding an appropriate and suitable Structural Element (SE) for achieving better enhancement is also a challenge. For getting better and appropriate SE, we use Edge Content (EC) [24] for automatically selecting Structural Element (SE) based on contrast matrix. The vector of Edge Content (EC) is the magnitude of the gradient vector of pixel position (x, y) of an input image (In_{img}). The depicted Eq. (5) presents the calculation of Edge Content (EC), where (i, j) is the block size of an input image and Fig. 3 shows the original and processed images. Fig. 4 presents the abstract flow of proposed methodology.

$$\nabla In_{img}(x, y) = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} In_{img}(x, y) \\ \frac{\partial}{\partial y} In_{img}(x, y) \end{bmatrix}$$

$$EC = \frac{1}{(i \times j)} \sum_x \sum_y |\nabla In_{img}(x, y)|$$
(5)

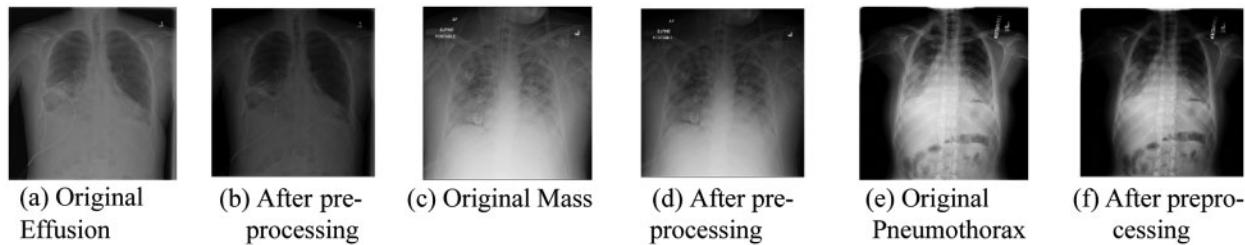


Figure 3: In preprocessing, we have applied some contrast enhancement algorithms on these low quality images and achieve some better performance and efficacy. One of the most widely technique used in image processing for background equalization and feature extraction is morphological operations (especially bottom-hat and top-hat transformation)

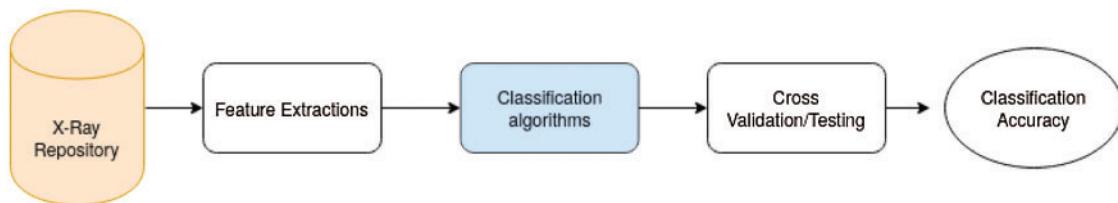


Figure 4: Abstract level of proposed methodology

3.2 Feature Extraction

Manually, we extract statistical features such as Sum of Variance, Entropy, Energy, and Difference of Variance from the NIH X-ray dataset and saved the features in different vectors for future use with pre-trained models (if necessary). Before moving into transfer learning, we train different machine learning algorithms depicted in Eq. (6) such as Random Forest-it creates several individual trees during

training and acquired prediction values from all individual trees to announce the final prediction value, Decision Tree-for finding categorical features that will use the entropy and Gini depicted in Eq. (6) to find the maximum information gain, K-Nearest Neighbor-it saves all the training data and finds the nearest node for features similarity and Support Vector Machine-it handles non-linear and linear data using kernel and properly tackles the outlier values on our extracted features (vectors). The training and validation accuracy is depicted in Fig. 5. We have used the k-fold cross validation concept to train machine learning classification algorithms and achieve plausible constant results on Random Forest (RF) as compared to other classification algorithms. Initially, Decision Tree algorithm achieves optimal results and after ingesting k-folding data it decreases the training accuracy. The K-Nearest Neighbor algorithm gives us a plausible result when K is 3 to 26. When we increase the value of K, it decreases the training accuracy and validation accuracy has huge spikes. In the Support Vector Machine (SVM) algorithm, we use k-fold cross validation to achieve better results as compared to others. It takes a long time to train the SVM with k-fold cross validation but gives us better results.

$$\begin{aligned}
 Gini &= 1 - \sum_{i=1}^c (p_i)^2 \\
 Entropy &= \sum_{i=1}^c -p_i * \lg_2(p_i) \\
 SVM &= mx + b = 0 \rightarrow SVM - h = W^T x + b = 0 \rightarrow SVM - d = \frac{ax_0 + bx_0 + c}{\sqrt{(a^2 + b^2)}}
 \end{aligned} \tag{6}$$

where p_i^2 is the probability of class c and b present the intercept of line on the y-axis, W represents the weight vector.

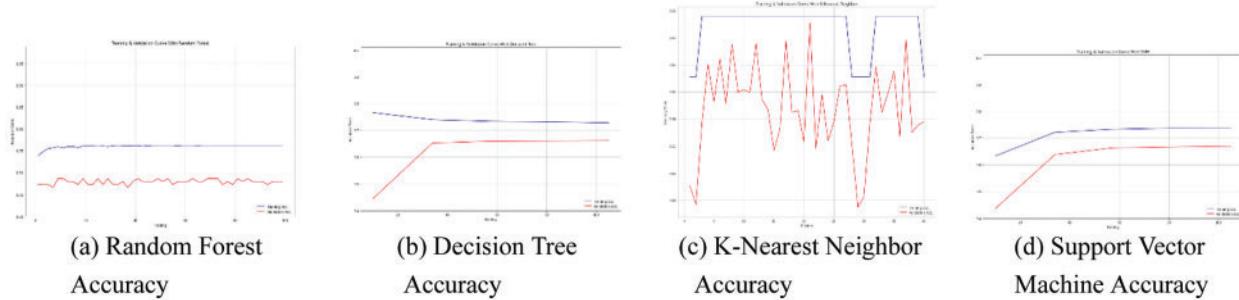


Figure 5: Machine learning classification algorithms (Random forest, decision tree, K-nearest neighbor and support vector machine) training and validation on National Institutes of Health (NIH) chest X-ray dataset

The achieved results from machine learning classification algorithms failed to meet our threshold. So we use pre-trained deep learning models to achieve better accuracy as compared to machine learning classification algorithms. In pre-trained models, we use the concept of transfer learning with fine tuning of custom Fully Connected (FC) layers.

The manually extracted features are used as the input of the machine learning classification algorithms as depicted in Fig. 6. The features are evaluated using ROC and AUC metrics and selection is performed using hand-crafted feature selection algorithms.

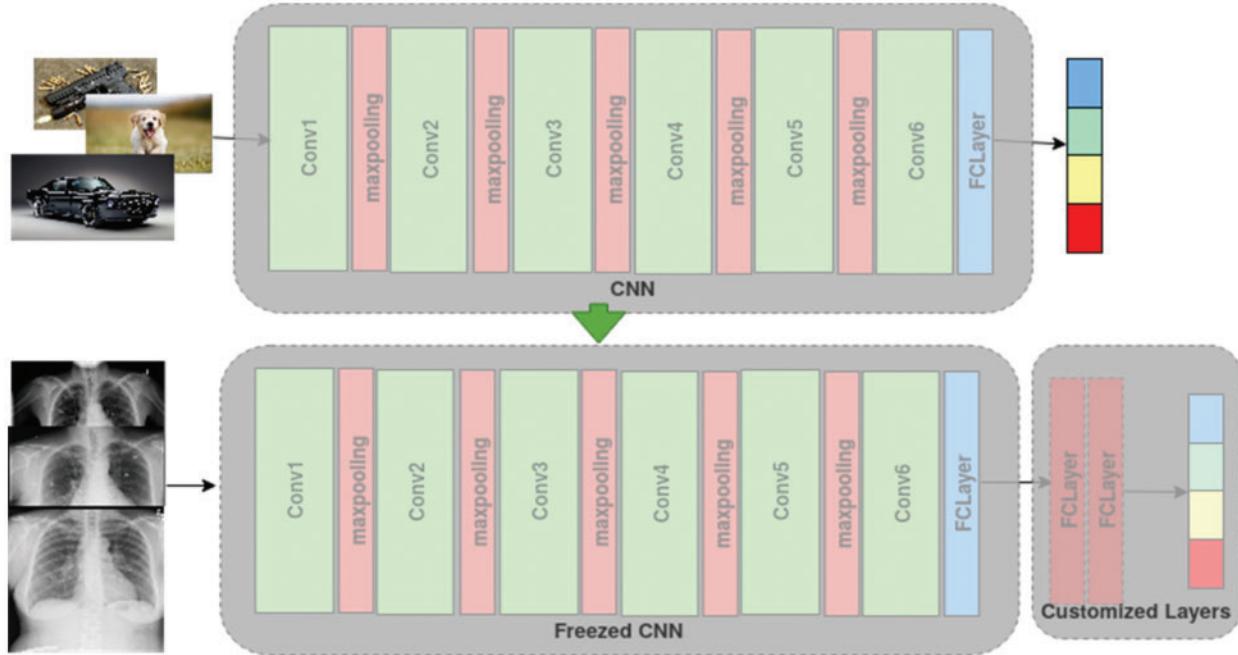


Figure 6: Complete detail and architecture of transfer learning concept with pretrained models: above model is the birdview of the pretrained model and bottom model is freezed and embedded with custom fully connected layers

Further, we extract the radiomics features of the chest x-ray customized dataset. These features are divided into five groups such as shape and size based features (volume, flatness area, surface area and etc.), gray-level co-occurrence matrix (GLCM-Energy, Entropy, Sum of Average, Difference of Variance, Difference Entropy, Difference Entropy and etc.) depicted in Tab. 1, size zone matrix (SZM-small area emphasis, Gray Level Non-uniformity) and run length matrix (RLM-Short Run emphasis, Zone Percentage, Site Zone Non-uniformity and etc.). We extract some features from the radiomics and pass these features to machine learning based classification algorithms (such as Support Vector Machine). Fig. 7 represents the SVM classification ROCs of multiple classes.

Table 1: GLCM features

Features	Mathematical expression	Features	Mathematical expression
Energy	$\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} G(i,j)^2$	Entropy	$-\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} G(i,j) \ln[G(i,j)]$
Sum of average	$\sum_{i=0}^{n-1} iG_{x+y}(i) \dots$	Sum of entropy	$-\sum_{i=2}^{2n} G_{x+y}(i) \lg[G_{x+y}(i)]$
Difference of entropy	$-\sum_{i=0}^{n-1} G_{x-y}(i) \lg[G_{x-y}(i)]$	Difference of variance	$-\sum_{i=0}^{n-1} iG_{x-y}(i)(i - DEnt)^2$

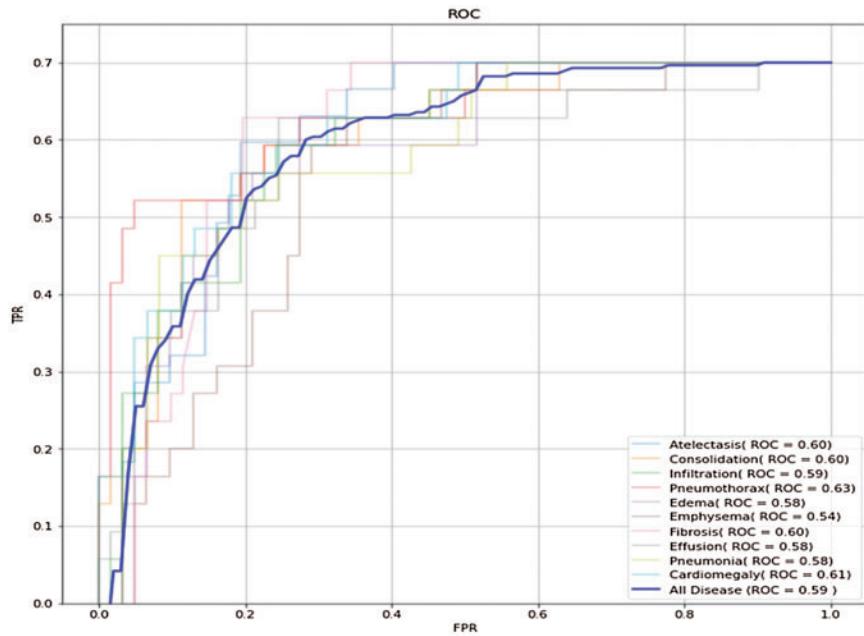


Figure 7: Handcrafted feature extraction with the help of radiomics methodologies from NIH chest X-ray and classify with traditional machine learning classifier (SVM)

It may be observed that the ROC achieved using pre-trained models is not satisfactory and AUC is close to 0.5 in detecting all abnormalities. Generally, radiomics features require expert annotated masks and a manual bounding box. Hence, improved results are expected using the hybrid framework proposed in this research. The depicted results in Fig. 7 are not up to the mark, so we move towards automatic deep feature extraction using deep learning methodologies.

3.3 Hybrid Framework

The proposed model would be intended to perform better than the existing models and a better prediction accuracy would be proved through experimentation and results. Fig. 4 depicts the proposed methodology and our proposed and customized CNN model is depicted in Fig. 8:

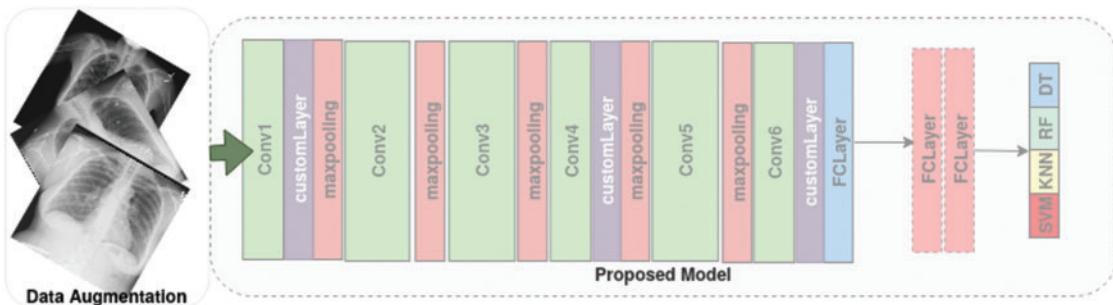


Figure 8: Our proposed customized CNN model have automated deep feature extraction techniques and also embedded handcrafted features in the middle of the proposed CNN model

In the proposed model, the input images would first be preprocessed by applying horizontal and vertical flips, random crops, 45° rotations and color distortion. From the processed images, we manually extract the meaningful feature using multiple handcrafted feature methods and these features will guide the CNN towards classification. The preprocessed set of images would be passed through convolution, batch normalization and leakyRelu model which would further be passed to another set of the aforementioned layers that gets divided into two sets of N similar layers running in parallel. The parallel set of layers would be concatenated at the end and passed again through another set of three layers i.e., convolution layer, batch normalization and leakyRelu, followed by a convolutional and leakyRelu layer before passing through a fully convolutional neural network. The processed image data gets through another leakyRelu layer before going through another fully convolutional neural network which gives out the final output which is to be passed to N number of clusters. As observed, the above model does not get trained on a set of input images, the proposed methodology makes use of unsupervised learning.

Our proposed Convolutional Neural Network (CNN) model achieves robust results and improved performance with handcrafted features. By progressive extraction of the deep features from the input image on each CNN layer, the initial layers learn about edges, boundary analysis and the last layers can identify lungs specification such as Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural_thickening and Cardiomegaly.

Apart from automatic feature extraction by CNN layers, we also embed handcrafted features with our proposed customization layer that also intensify the performance of classification accuracy. Eq. (7) depicts the mathematical operation of the basic CNN model, where In_{img} is an input image F is a filter/kernel size of dimension f_1 and f_2 .

$$(In_{img} * F)_{ij} = \sum_{u=0}^{f_1-1} \sum_{v=0}^{f_2-1} In_{img}(i-u, j-v)F(u, v) \\ = \sum_{u=0}^{f_1-1} \sum_{v=0}^{f_2-1} In_{img}(i+u, j+v)F(-u, -v) \quad (7)$$

The output of the above Eq. (7) is followed by Eq. (8), where the Filter is horizontally or vertically flipped during operation.

$$(In_{img} * F)_{ij} = \sum_{u=0}^{f_1-1} \sum_{v=0}^{f_2-1} F_{u,v} \cdot In_{(img)i+u,j+v} + bias \quad (8)$$

$vec_{i,j}^{layer}$ is the input vector of a layer during convolution and $output_{i,j}^{layer}$ is the output layer, after convolution and $f(\cdot)$ is an activation function of an activation layer

$$vec_{i,j}^{layer} = \sum_u \sum_v weight_{u,v}^{layer} output_{i+u,j+v}^{layer-1} + bias^{layer} \\ output_{i,j}^{layer} = f(x_{i,j}^{layer}) \quad (9)$$

For calculation cost (C) of network, the depicted Eq. (10) express the mathematical formulation to find cost value of a network, where fc presents actual forecasting cost, n f c is network predicted cost.

$$C = \frac{1}{2} \sum_{fc} (a_{fc} - n_{fc})^2 \quad (10)$$

Mostly the above equations are used during CNN training, so for integrating manual features within Convolution layer, we introduce a novel concept of hybrid convolution which comprises automated feature extraction and fusion of handcrafted features to improve the efficacy of classification

of medical images. In the depicted Eq. (11) $MF_{img_{u,v}}^T$ presents the ingestion of handcrafted features in automated features.

$$(In_{img} * F)_{ij} = \sum_{u=0}^{f_1-1} \sum_{v=0}^{f_2-1} F_{u,v} . ((In_{(img)i+u,j+v}) + (MF_{img_{u,v}}^T)) + bias \quad (11)$$

We apply backpropagation using the chain rule with subtracting of our handcrafted feature before finding the ∂ derivative. For finding gradient on individual weights chain rule is applied. The depicted Eq. (12) shows the subtraction of handcrafted features that are placed in a single vector ($MF_{img_{u,v}}^T$).

$$\begin{aligned} \frac{\partial C}{\partial weight_{u',v'}^{layer}} &= \sum_{i=0}^{H-f_1} \sum_{j=0}^{weight-f_2} \frac{\partial C}{\partial vec_{i,j}^{layer} - (MF_{img_{u,v}}^T)} \frac{\partial vec_{i,j}^{layer} - (MF_{img_{u,v}}^T)}{\partial weight_{u',v'}^{layer}} \\ &= \sum_{i=0}^{H-f_1} \sum_{j=0}^{weight-f_2} \delta_{i,j}^{layer} \frac{\partial vec_{i,j}^{layer} - (MF_{img_{u,v}}^T)}{\partial weight_{u',v'}^{layer}} \end{aligned} \quad (12)$$

We apply k-fold cross validation techniques to handle the overfitting issue and achieve generalization during training. In k-fold cross validation technique, the complete dataset is divided into different folds, and combinations of folds are transferred for training. For achieving credible results from the proposed CNN model, we apply data re-sampling techniques to handle the imbalanced classes. We apply different data augmentation techniques such as rotations, horizontal/vertical flipping, and scaling, zooming, padding, and random brightness to synthesize the NIH X-ray dataset to resolve the imbalance issue. After that, we apply morphological operations (combination of opening and closing) to handle the contrast issues in the NIH X-ray dataset. In morphological operations, we apply Edge Content (EC) for optimal selection of Structural Element (SE). Using optimal SE, we get an arguable result to fill and eliminate the smaller cracks and gaps in NIH X-ray dataset.

Fig. 9 represents the activation maps (lines, edges, and texture patterns) of CNN model of the second layer and deep (N) layer.



Figure 9: Middle layer activation map: Visual activation maps of our proposed CNN model. Later layers of CNN model, construct the features by ingesting deep features from the previous layers

4 Result and Discussion

For extraction of deep features, we used pre-trained models (VGG and Inception) and classify those features with traditional machine learning classifiers (Decision Tree, Random Forest, K-Nearest Neighbor and Support Vector Machine). We have removed the top layer from the pre-trained model and embedded dense (FC) layers with by-default optimizer (Adam) with regularization to avoid the overfitting issue and set hyperparameters (learning rate = $1e - 7$, batch size = 32 and epochs = 100) based on cognitive content. We have set an early stopping mechanism and used the k-fold cross validation concept to acquire optimal generalization performance on the pre-trained models.

In the field of medical imaging, a large number of annotated datasets are required for better classification accuracy which is still a big challenge. Training on the smaller dataset can lead to overfitting and mostly deep learning models require a large amount of dataset for optimal classification accuracy. For training on the large medical dataset, we require enough processing power and time and also a large amount of medical annotated dataset. To handle this overcome, initially, we apply the most widely used pre-trained models such as VGG and Inception which are already pre-trained on a large natural dataset for different classification tasks. Therefore, we adopt the concept of transfer learning (pre-trained model) with our custom layers (Fully Connected) Layers. Fig. 6 depicts the concept of transfer learning. In transfer learning layer, we use the pre-trained low level features of those models and embed our custom Fully Connected layer for fine tuning. We have frozen the convolutional layers of those pre-trained models and use our customer Fully Connected layer for training. The training, validation and testing accuracy of those pre-trained models and the complete details and architecture of the pre-trained models that are utilized in this study have listed below:

A well-known and smaller pre-trained model VGG is a convolutional network model that can achieve 92.7% top-5 test accuracy in ImageNet dataset. VGG network is considered a simplified model because it only comprises convolutional layers that are stacked over each other. Max pooling handles reduction in volume size. Fully connected layers are then followed by SoftMax. VGG comprises two versions that are VGG16 and VGG19. VGG16 architecture is made up of 16 layers. The layer composition consists of 13 convolutional layers, 5 pooling layers and 3 fully connected layers. We used with formal machine learning classifier Support Vector Machine (SVM) for optimal results. We freeze the middle and last layers of the VGG model and embed custom Fully Connected (FC) layers to extract features and send those features to machine learning classification algorithms. Using transfer learning with a machine learning classifier, we achieve better results as compared to handcrafted features with machine learning classifiers and other pre-trained models. Furthermore, we embed handcrafted features into the VGG model and acquire optimal results.

Fig. 10 presents the automated extracted features (obtained from pre-trained VGG model) and handcrafted features, Support Vector Machine (SVM) have improved the results as compared to other machine learning classifiers (Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbor (KNN)).

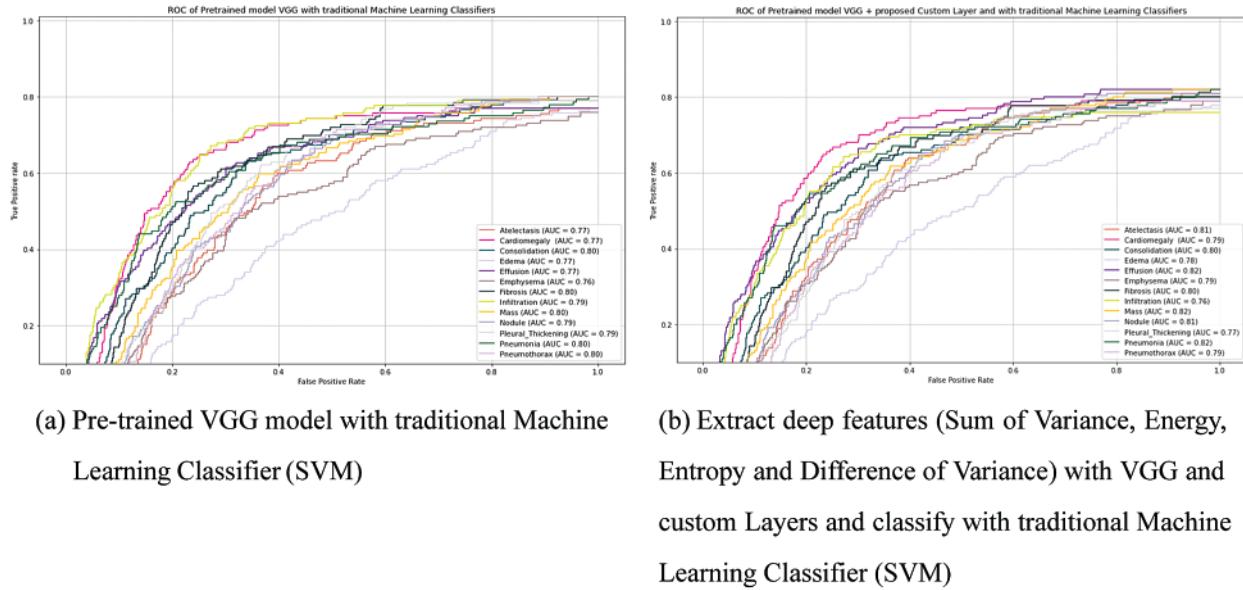


Figure 10: LEFT: Using pre-trained model (VGG) with fine tuning fully connected (FC) layers for feature extraction and then classify those features with traditional machine learning classifier (Support vector machine). RIGHT: Acquired high dimensional features with VGG and custom layers and forward those feature to support vector machine

Inception is a convolutional neural network that is used for object detection and image analysis. It was first used as a module by GoogleNet. It has shown an accuracy of 78.1% on ImageNet dataset. It was developed on the different ideas of different researchers. Inception has different versions such as Inception V1, V2, and V3. It performs convolution on the input with 3 different sizes i.e., $(1 \times 1, 3 \times 3,$ and $5 \times 5)$. It also performed max pooling and the output will be concatenated and sent to the next inception module. To achieve advisable results in a lesser time and with more features, we apply another pre-trained model (Inception) model. Using the Inception block and Hebbian principle, it concatenates $1 \times 1, 3 \times 3,$ and 5×5 convolutional layers into a single output layer. We freeze down the middle and last layers of the Inception model and interject custom Fully Connected (FC) layers to acquire features and transmit those features to machine learning classification algorithms. Using transfer learning with a machine learning classifier, we come through advisable results as compared to handcrafted features with machine learning classifiers and other pre-trained models such as VGG. Fig. 11 shows the results of simple automated feature classification with SVM and integrating handcrafted features with the Inception model to achieve better results.

The depicted Figs. 12 and 13 show the results of our novel proposed hybrid CNN with handcrafted features and classify with traditional machine learning classifiers. Tab. 2 presents the accuracy, specificity, and sensitivity of pre-trained models such as VGG and Inception and the proposed model with a traditional machine learning classifier (SVM). We observe that the proposed model achieved high sensitivity due to preprocessing module and feature grafting of manual with CNN features.

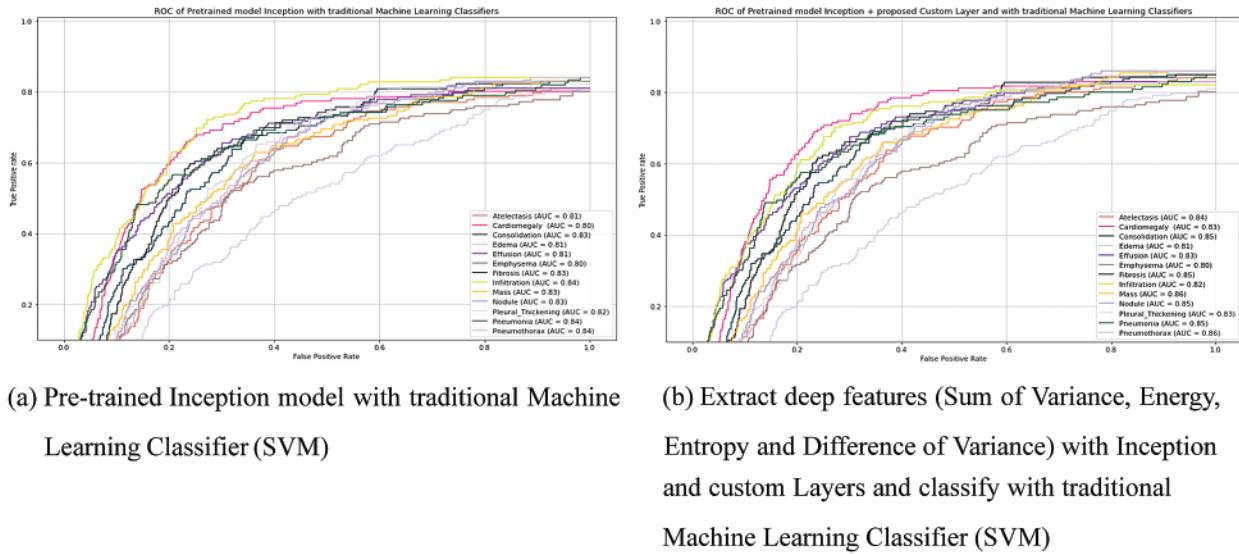


Figure 11: LEFT: Using pre-trained model (Inception) with fine tuning fully connected (FC) layers for feature extraction and then classify those features with traditional machine learning classifier (Support vector machine). RIGHT: Acquired high dimensional features with inception and custom layers and forward those feature to SVM

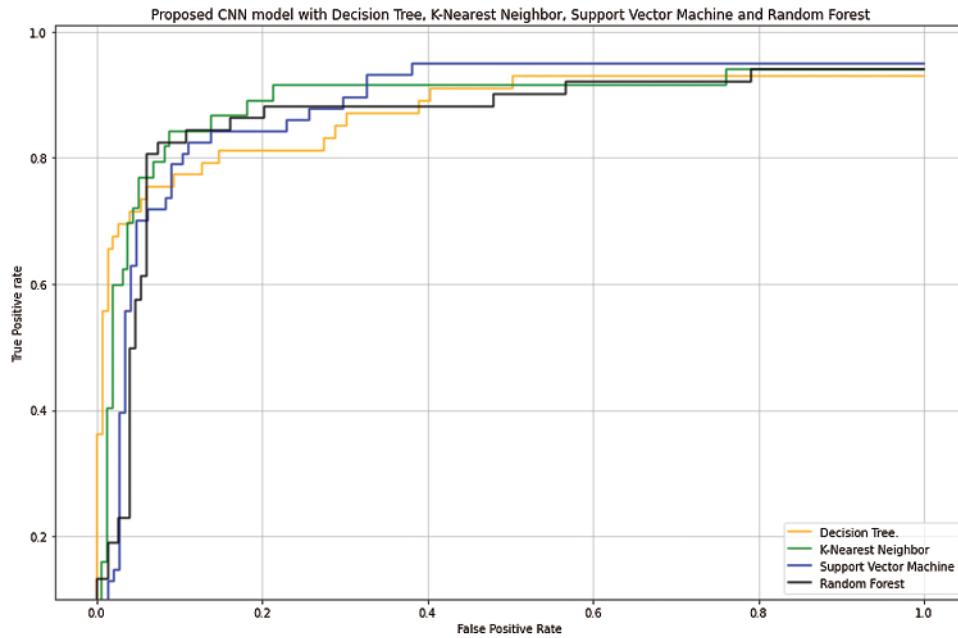


Figure 12: Acquired high dimensional features with our novel proposed and custom layers (for fusion of handcrafted features) and forward those feature to traditional machine learning classifier (Support vector machine)

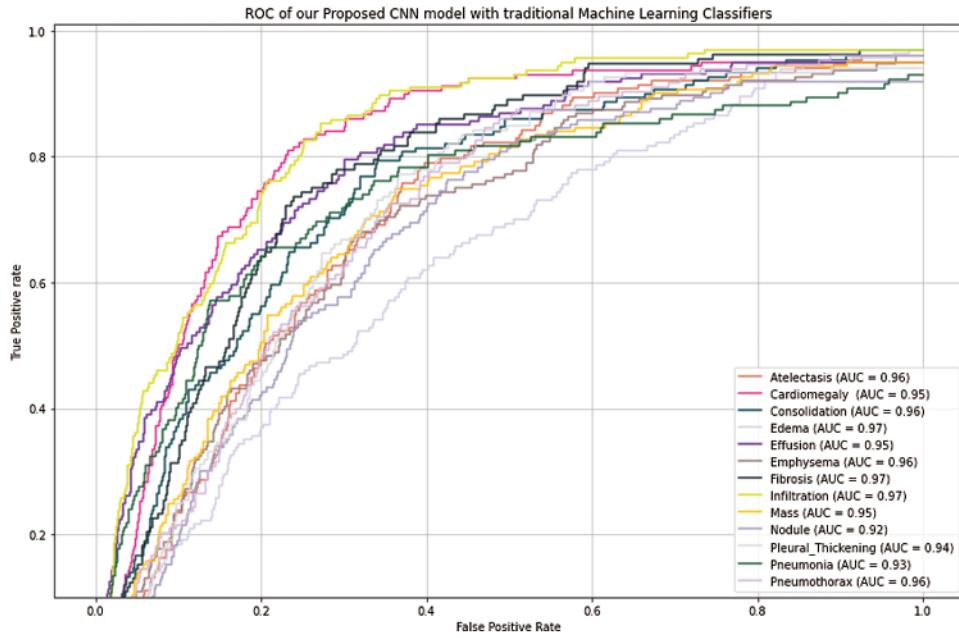


Figure 13: Using our novel proposed CNN model with fine tuning fully connected (FC) layers for feature extraction and then classify those features machine learning classifiers (Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN) and Support Vector Machine (SVM))

Table 2: Assessment results of novel proposed CNN model with pre-trained CNN models and traditional machine learning classifier Support Vector Machine (SVM)

Method	Accuracy	Specificity	Sensitivity
VGG + SVM	78.10	99.73	65.69
Inception + SVM	79.87	99.69	66.46
Proposed model + SVM	96.87	99.77	97.24

The proposed CNN model are further classified with other traditional machine learning classifier such as Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbor (KNN). The result of the traditional machine learning classifier is depicted in Fig. 13. We observe that Support Vector Machine give us plausible results on the fusion of automated feature with hand-crafted features as compared to other classifiers such as K-Nearest Neighbor (KNN), Decision Tree (DT), and Random Forest (RF).

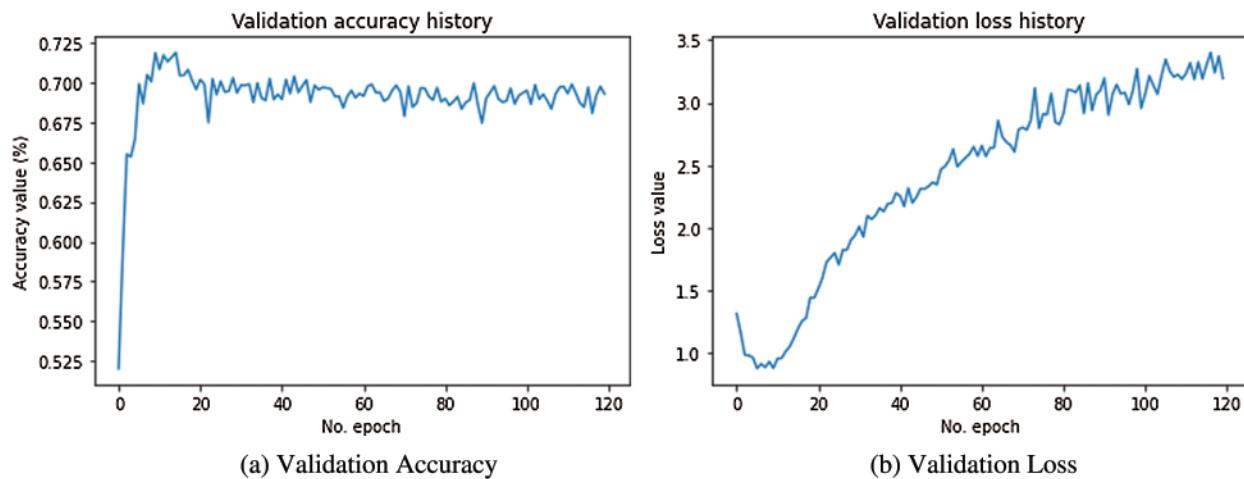


Figure 14: Validation accuracy and loss of LIDC dataset on our proposed model

5 Conclusion

In this paper, we have proposed a hybrid framework utilizing both handcrafted features as well as features extracted from convolution neural networks to categorize the lung cancer images into classes of interest. In this novel technique, the manual features, radiomics GLCM, are inserted into the middle of the proposed sequential CNN model to improve performance and the technique is evaluated on publicly available NIH and LIDC datasets. It is observed that GLCM features significantly improve the performance as compared to pre-trained VGG and Inception models and this is validated using 5-fold cross-validation. The SVM classifier is used in experiments as it gives significantly robust results than the KNN, Decision Trees, and Randon Forest for classification of features. This novel technique gives us plausible results as depicted by sensitivity 97.24% which is a significant improvement compared to 65.69% with VGG + SVM and 66.49% with Inception + SVM approaches. Also, specificity is 99.77% and overall improved accuracy is 96.87%. In future work, we plan to insert the manually extracted features into the initial and/or ending layers of CNN and evaluate performance. For validating our proposed model, we have trained with the LIDC dataset and acquired the validation accuracy and loss as shown in Fig. 14.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] W. Chen, H. Wei, S. Peng, J. Sun, X. Qiao *et al.*, “Hybrid segmentation network for small cell lung cancer segmentation,” *IEEE Access*, vol. 7, pp. 75591–75603, 2019.
 - [2] International Agency for Research on Cancer *et al.*, “Global cancer observatory,” World Health Organization. [Online]. Available: <http://gco.iarc.fr>. Accessed: 8, 2018.
 - [3] G. Wei, H. Ma, W. Qian, H. Jiang and X. Zhao, “Content-based retrieval for lung nodule diagnosis using learned distance metric,” in *IEEE 39th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, JeJu, Korea, pp. 3910–3913, 2017.

- [4] H. Jiang, H. Ma, W. Qian, G. Wei, X. Zhao *et al.*, “A novel pixel value space statistics map of the pulmonary nodule for classification in computerized tomography images,” in *IEEE 39th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, JeJu, Korea, pp. 556–559, 2017.
- [5] J. J. S. Cuenca, W. Guo and Q. Li, “Automated detection of pulmonary nodules in ct: False positive reduction by combining multiple classifiers,” *International Society for Optics and Photonics in Medical Imaging 2011: Computer-Aided Diagnosis*, vol. 7963, pp. 796338, 2011.
- [6] E. E. Nithila and S. S. Kumar, “Automatic detection of solitary pulmonary nodules using swarm intelligence optimized neural networks on ct images,” *An International Journal Engineering Science and Technology*, vol. 20, no. 3, pp. 1192–1202, 2017.
- [7] F. V. Farahani, A. Ahmadi and M. H. F. Zarandi, “Hybrid intelligent approach for diagnosis of the lung nodule from ct images using spatial kernelized fuzzy c-means and ensemble learning,” *Mathematics and Computers in Simulation*, vol. 149, pp. 48–68, 2018.
- [8] G. Aresta, A. Cunha and A. Campilho, “Detection of juxta-pleural lung nodules in computed tomography images,” *International Society for Optics and Photonics in Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, pp. 101343, 2017.
- [9] J. Freixenet, X. Munoz, D. Raba, J. Marti and X. Cufi, “Yet another survey on image segmentation: Region and boundary information integration,” in *Springer European Conf. on Computer Vision*, Berlin, Heidelberg, pp. 408–422, 2002.
- [10] A. O. D. C. Filho, W. B. de S, A. C. Silva, A. C. de Paiva, R. A. Nunes *et al.*, “Automatic detection of solitary lung nodules using quality threshold clustering, genetic algorithm and diversity index,” *Artificial Intelligence in Medicine*, vol. 60, no. 3, pp. 165–177, 2014.
- [11] M. Javaid, M. Javid, M. Z. Rehman and S. I. A. Shah, “A novel approach to cad system for the detection of lung nodules in ct images,” *Computer Methods and Programs in Biomedicine*, vol. 135, pp. 125–139, 2016.
- [12] M. Assefa, I. Faye, A. S. Malik and M. Shoaib, “Lung nodule detection using multi-resolution analysis,” in *IEEE ICME Int. Conf. on Complex Medical Engineering*, Beijing, China, pp. 457–461, 2013.
- [13] J. Gong, J. Liu, L. Wang, B. Zheng and S. Nie, “Computer-aided detection of pulmonary nodules using dynamic self-adaptive template matching and aflda classifier,” *Physica Medica*, vol. 32, no. 12, pp. 1502–1509, 2016.
- [14] H. J. Chen, S. J. Ruan, S. W. Huang and Y. T. Peng, “Lung x-ray segmentation using deep convolutional neural networks on contrast-enhanced binarized images,” *Mathematics*, vol. 8, no. 4, pp. 545, 2020.
- [15] B. A. Z. Imran and D. Terzopoulos, “Semi-supervised multi-task learning with chest x-ray images,” in *10th Int. Workshop Machine Learning in Medical Imaging (MLMI), Held in Conjunction with MICCAI*, Shenzhen, China, vol. 11861, pp. 151, 2019.
- [16] C. Qin, D. Yao, Y. Shi and Z. Song, “Computer-aided detection in chest radiography based on artificial intelligence: A survey,” *Biomedical Engineering*, vol. 17, no. 1, pp. 113, 2018.
- [17] I. M. Baltruschat, L. S. H. Ittrich, G. Adam, H. Nickisch, A. S. Bach *et al.*, “When does bone suppression and lung field segmentation improve chest x-ray disease classification,” in *IEEE 16th Int. Symp. on Biomedical Imaging (ISBI)*, Venice, Italy, pp. 1362–1366, 2019.
- [18] A. K. Jaiswal, P. Tiwari, S. Kumar, D. Gupta, A. Khanna *et al.*, “Identifying pneumonia in chest x-rays: A deep learning approach,” *Measurement*, vol. 145, pp. 511–518, 2019.
- [19] L. Hussain, M. S. Almaraashi, W. Aziz, N. Habib and S. R. S. Abbasi, “Machine learning-based lungs cancer detection using reconstruction independent component analysis and sparse filter features,” *Waves in Random and Complex Media*, vol. 31, pp. 1–26, 2021.
- [20] E. Kesim, Z. Dokur and T. Olmez, “X-ray chest image classification by a small-sized convolutional neural network,” in *IEEE Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, Istanbul, Turkey, pp. 1–5, 2019.
- [21] A. Bhandary, G. A. Prabhu, V. Rajinikanth, K. P. Thanaraj, S. C. Satapathy *et al.*, “Deep-learning framework to detect lung abnormality-a study with chest x-ray and lung ct scan images,” *Pattern Recognition Letters*, vol. 129, pp. 271–278, 2020.

- [22] F. Cao and H. Zhao, "Automatic lung segmentation algorithm on chest x-ray images based on fusion variational auto-encoder and three-terminal attention mechanism," *Symmetry*, vol. 13, no. 5, pp. 814, 2021.
- [23] F. M. Salman, S. S. Abu-Naser, E. Alajrami, B. S. Abu-Nasser and B. A. M. Alashqar, "Covid-19 detection using artificial intelligence," *International Journal of Academic Engineering Research (IJAER)*, vol. 4, no. 3, pp. 18–25, 2020.
- [24] R. Kushol, M. H. Kabir, M. S. Salekin and A. A. Rahman, "Contrast enhancement by top-hat and bottom-hat transform with optimal structuring element: Application to retinal vessel segmentation," in *Springer Int. Conf. Image Analysis and Recognition*, Montreal, QC, Canada, pp. 533–540, 2017.