Multi-View Convolutional Recurrent Neural Network for Lung Cancer Nodule Identification

Mian Muhammad Naeem Abid[a], Tehseen Zia[a], Mubeen Ghafoor[a] and David Windridge[b]

[a]Department of Computer Science, COMSATS University Islamabad (CUI), Islamabad, Pakistan
[b]Department of Computer Science, Middlesex University, London, United Kingdom

ARTICLE INFO

Keywords:
Lung Cancer
Nodule Detection
CT Scan Images
Convolutional Neural Networks
Recurrent Neural Networks
Long Short Term Memory

ABSTRACT

Screening via low-dose Computer Tomography (CT) has been shown to reduce lung cancer mortality rates by at least 20%. However, the assessment of large numbers of CT scans by radiologists is cost intensive, and potentially produces varying and inconsistent results for differing radiologists (and also for temporally-separated assessments by the same radiologist). To overcome these challenges, computer aided diagnosis systems based on deep learning methods have proved an effective in automatic detection and classification of lung cancer.

Latterly, interest has focused on the full utilization of the 3D information in CT scans using 3D-CNNs and related approaches. However, such approaches do not intrinsically correlate size and shape information between slices. In this work, an innovative approach to Multi-view Convolutional Recurrent Neural Networks (MV-CRecNet) is proposed that exploits shape, size and cross-slice variations while learning to identify lung cancer nodules from CT scans. The multiple-views that are passed to the model ensure better generalization and the learning of robust features.

We evaluate the proposed MV-CRecNet model on the reference *Lung Image Database Consortium and Image Database Resource Initiative* and *Early Lung Cancer Action Program* datasets; six evaluation metrics are applied to eleven comparison models for testing. Results demonstrate that proposed methodology outperforms all of the models against all of the evaluation metrics.

1. Introduction

Lung cancer is the predominant source of cancerous deaths worldwide, affecting 12.9% of world's population [1]. Among overall lung cancer cases, about 58% occurred in less developed countries [1, 2]. With a survival rate of 17.8% (5-year), lung cancer is regarded as the most lethal form of cancerous disease, primarily due to the fact that symptoms typically appear only when the disease is at an advanced stage. The National Lung Screening Trail (NLST) recently established that the mortality rate of lung cancer can be reduced by at least 20% using an early screening procedure with low-dose Computer Tomography (CT) [3, 4]. The analysis of CT scans (as 3D volumes) is a clinically intensive process that often leads to erroneous decision making. Moreover, there is considerable inter-observer and intra-observer variability amongst radiologists in identifying nodules. Computer Aided Diagnostics (CAD) systems capable of accurate, quick, and cost-effective Lung Cancer Screening (LCS) thus have a clear clinical role. Recognizing this, LCS competitions have recently been organized to engage scientific community in developing effective CT analysis methodologies.

Automated LCS system essentially involve two main phases: nodule detection and nodule classification. In the nodule detection phase, pulmonary nodules are detected from CT volumes. After detection, the nodules are analyzed to determine their malignancy and to decide on a follow-up strategy for patients. To determine the malignancy state, nodules are required to be classified in terms of *metastatic malignant*, *benign* and *primary malignant* categories. The follow-up strategy is based on the characteristics of nodule growth, type and size as defined in the *PanCan* model and the guidelines of *Lung CT Reporting and Data Systems (Lung-RADS)*. However, identification of lung cancer nodules is a challenging task for two main reasons: 1) pulmonary nodules have different sizes, locations and shapes as illustrated in Fig. 1; 2) high false positive rates due to candidate locations having morphologically similar appearances to actual lung cancer nodules as demonstrated in Fig. 2.

In the most recent challenges, including *Lung Nodule Analysis 2016 (LUNA16)* and *National Data Science Bowl 2017*, deep learning approaches have predominated over other conventional machine learning based approaches. A key characteristic of deep learning is that rather than designing ad-hoc/hand-crafted descriptors for image analysis, deep learning is capable of learning these descriptors automatically [5]. Because ad-hoc descriptors consider specific (generally non-hierarchical) aspects of images while disregarding the other aspects, learned task-specific descriptors have been found to outperform them in typical image analysis and medical imaging tasks. In medical imaging specifically, such learned descriptors have, for example, been shown to outperform medical experts in specific tasks within the fields of dermatology [6] and ophthalmology [7].

To date, deep-learning based lung cancer nodule identification has primarily been carried out via Convolutional Neural Networks (CNNs). Although CNNs are recognized

*This work has been supported by Higher Education Commission (HEC) under Grant # 2(1064), and is carried out at Medical Imaging and Diagnostics Lab at COMSATS University Islamabad (CUI), under the umbrella of National Center of Artificial Intelligence (NCAI), Pakistan.

miannaeem1994@gmail.com (Mian Muhammad Naeem Abid);
tehseen.zia@comsats.edu.pk (Tehseen Zia);
mubeen.ghafoor@comsats.edu.pk (Mubeen Ghafoor); d.windridge@mdx.ac.uk (David Windridge)

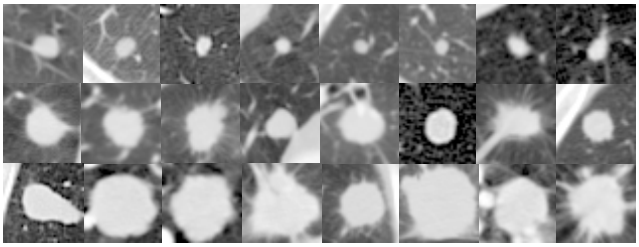


Figure 1: Differently-shaped and sized (small, medium and large) nodules.

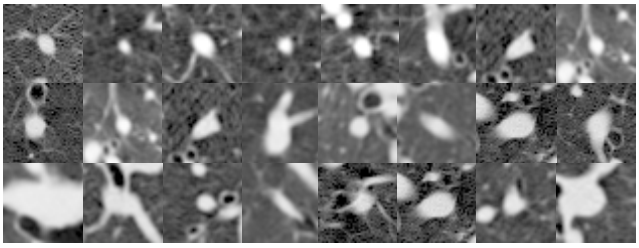


Figure 2: Non-Nodules with a similar morphological appearance to nodules.

for their ability to detect shapes of objects (e.g. cancer nodules), they are by design invariant to the size and positioning of objects. Lately, it has been proposed to capture and exploit the shapes/sizes of nodules using Multi-view CNNs (MV-CNNs). Unlike conventional CNN which consists of single input, MV-CNN integrates multiple inputs. The term multi-views means that different views of the same input are fed to the model. For instance, for lung cancer nodule identification problem, traditional CNN will get a single image as input; on contrary, MV-CNN will get multiple images as input. One case for MV-CNN, which is used in this study, can be to pass different sized cropped images/patches (views) of the nodule to MV-CNN. Advantages of this would be that views with small receptive fields will give more detailed information of the nodules, whereas, views with large receptive fields will also include the surrounding tumor tissues giving broader contextual information about nodules. A key shortcoming of these approaches, however, is that they process 3D volumetric CT scans as individual slices, potentially disregarding useful 3D information. To address this issue, 3D-CNNs have been employed as a means of processing volumetric patterns of cancerous nodules. However, 3D-CNNs involve a potentially very significant number of parameters (typically a multiple of the depth of the scan), requiring a large annotated dataset to learn generalized representations. The lack of availability of large annotated lung cancer CT scan datasets thus restricts the efficacy of 3D-CNNs [8, 9, 10]. To resolve this issue, researchers might instead adopt a Recurrent Neural Network (RNN) based model which, being of a sequential nature, would have the potential to capture and exploit cross-slice variations in order to incorporate volumetric patterns of nodules. Though such a RNN-based model would use significantly fewer parameters relative to 3D-CNNs without modification, it would cater only for cross-slice variations of the nodules and not the shape and size of the nodules. To this end, the objective of this work is to build on this approach and propose a model architecture that can exploit shape, size and cross-slice variations while learning to identify lung cancer nodules from CT scans. Specifically, MV-CNNs are used to exploit the shape and size of nodules and RNNs (in particular Long Short-Term Memory (LSTMs)) are used to encompass cross-slice variation. The proposed Multi-view Convolutional Recurrent Neural Network (MV-CRecNet) model is trained and tested on the *Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI)* and *Early Lung Cancer Action Program (ELCAP)* datasets. Six evaluation metrics are taken into account to measure the performance of proposed technique.

Our principle novelties and contributions can be summarized as:

- We propose a novel model architecture (MV-CRecNet) for learning to identify pulmonary nodules from CT scan images.
- We carry out extensive experimentation on benchmark LIDC-IDRI and ELCAP datasets with this and various other model architectures. The comparative performance of these models is reported.

The rest of the paper is organized as follows: Section 2 sets out Related Work. The MV-CNN, RNN (LSTM) models and the Proposed Methodology (MV-CRectNet) are illustrated in Section 3. Materials and Model-specification are discussed in Section 4. Simulation Results and Discussion are demonstrated in Section 5. Finally, Conclusion and Future Work are presented in Section 6.

2. Related Work

Early detection and classification of cancer nodules has been demonstrated to be beneficial in reducing mortality rates and planning follow up strategies for patients. Various studies can be found in the literature that have the aim of detecting and classifying lung cancer nodules using both conventional machine learning and deep learning approaches.

In early work, Armato et al. [11] used Linear Discriminant Analysis (LDA) for classification of nodules with 187 nodules (juxtapleural and solitary nodules) for validation; the Fisher Linear Discriminant (FLD) classifier was adopted by Messay et al. [12] to detect pulmonary nodules; Teramoto and Fujita [13] proposed a technique for nodule classification incorporating cylindrical filters, with classification via a Support Vector Machine (SVM). Han et al. [14] integrated SVMs with a Hierarchical Vector Quantization approach in their work. Erdal and Aybars [15] proposed a technique for detection of nodules incorporating bespoke features of which five were shape based and the other two were shape and texture based. Surface properties, intensity values and morphological characteristics were incorporated by Way et al. [16] for the classification of malignant & benign nodules (for classification they used a Linear Discrim-

inant Classifier). A supervised learning algorithm based on LDA and Genetic Algorithms (GAs) was developed by Lee et al. [17] for analyzing solitary nodules (benign and malignant). Wavelet transforms were used to characterize images by Orozco et al. [18], with a SVM used for identification of pulmonary nodules. Macedo Firmino et al. [19] proposed using Histogram of Oriented Gradient (HOG) features for nodule characterization and the watershed method for segmentation of lung nodules; SVMs and rule-based classifiers were used for classification.

However, a drawback of conventional machine learning algorithms is that they rely on handcrafted features which have limited scope and capacity to achieve human-level performance. By contrast, deep learning algorithms learn feature extractors automatically from data and realize human-level performance in various domains including medical image analysis. Various deep learning based methods have consequently been proposed for lung cancer nodule identification in recent years.

Stojan Trajanovski et al. [20] proposed a deep learning framework for assessing lung cancer risk using low-dose chest CT scans. To train the learning model they considered a multi-instance weakly labeled approach in which only patient-level annotation is required. Their framework consists in two stages: nodule detection and malignancy risk assessment of the CT scan as a whole (whether the nodules are malignant or benign). A hierarchical SVM was used for nodule detection which gives the information of nodule location, size and shape. This information is provided as the input in the second step of the proposed technique where a wide and deep ResNet-like neural network was employed for cancer risk assessment. NLST data, Kaggle competition data and data from Lahey Hospital and Medical Center (LHMC) are utilized for experimentation with the result that the proposed method achieves 7% better Area Under the Curve (AUC) than the PanCan Risk model on NLST and LHMC data; the model, moreover, performs comparably to a radiologist at patient-level cancer risk assessment.

Kui Liu and Guixia Kang implement a deep learning approach to classifying lung cancer nodules in [21], proposing MV-CNN for lung nodule classification in contrast to traditional CNNs. The authors address binary and ternary classification problems, i.e those having two classes (malignant and benign) and three classes (metastatic malignant, primary malignant and benign), respectively. Experimental data is derived from LIDC-IDRI with results demonstrating that a multi-view approach performs better than single-view strategy, achieving 5.41% error rate for binary classification and 13.91% error rate for ternary classification problem.

In [22], a CNN-based method is proposed by Xinglong Liu et al. for the classification of nodule types using CT scans. Contrary to traditional approaches which deal with pleural-tail, juxta-pleural, vascularized and well-circumscribed nodules types, non-nodules and Ground Glass Optical (GGO) are also taken into account in their study. A multi-view multi-scale CNN is proposed for classification of lung nodules with a spherical surface approximated at the center of nodules from which an estimated radius is achieved for each nodule. After obtaining the radius and (sorted) circular planes, a view independent CNN is pre-trained. Subsequently, the MV-CNN model is trained using maximum pooling. Two datasets: Early Lung Cancer Action Program (ELCAP) and LIDC-IDRI are used for experimentation, with LIDC-IDRI is used for training the CNN model. The proposed model is tested on both ELCAP and LIDC-IDRC datasets which show promising results even for non-nodule and GGO types.

In [23], Francesco Ciompi et al. address the automatic nodule classification problem using six nodule types: part-solid, solid, non-solid, calcified, spiculated and perifissural, according to assessment categories of Lung-RADS and PanCan malignancy model. For classification, multi-stream and multi-scale Convolution Networks (Conv-Nets) are considered. 2D multiple triplets of nodules at multiple scales (10, 20 and 40 mm) are processed simultaneously and a liklihood is computed for each class. Estimation of the nodule size or nodule segmentation is not required. The deep learning model is trained with Multi-centric Italian Lung Detection (MILD) trial data whereas the trained model is validated on Danish Lung Cancer Screening Trial (DLCST) data. For a performance comparison of their proposed deep learning approach, linear SVM classifiers are trained using features from nodules' raw intensity along with learned features from an unsupervised approach (K-means) using raw data. An observer study with four observers including radiologists was conducted in order to compare proposed methodology against human performance. Experimental results demonstrated that the proposed architecture performed better than classical methods (which require hand-crafted features), being comparable to the observers' variability.

Wei Shen et al. [24] propose a multi-crop deep learning method to inspect lung nodule malignancy. Specifically, they use a Multi-crop CNN (MC-CNN) which does not depend on nodule segmentation. Multi-crop pooling crops different sections of feature maps from the convolution layer and then applies max-pooling to substitute for the conventional max pooling layer. Instead of using multiple networks for multi-scale features, the suggested technique uses a single network while reducing computational complexity. Experiments on the prediction of nodule semantics and estimation of nodule diameter were conducted with the LIDC-IDRI dataset. Comparisons carried out against both segmentation independent methods (HOG and Local Binary Patterns (LBP)) and segmentation dependent methods (Auto-encoder and Massive-feet) with results demonstrating the robustness of the proposed technique for classification of lung nodule malignancy suspicion.

2D-CNNs have widely been used for the pulmonary nodule detection task; however, solutions based on 2D-CNNs cannot take full benefit of volumetric information given that lung cancer nodule detection from CT scans is inherently a 3D object detection problem. Recently, researchers have begun to explore the use of 3D-CNNs reflecting the nature of the data in order to take advantage of volumetric information.

Guixia Kang et al. in [25] implemented a 3D MV-CNN for the classification of lung cancer nodules. MV-CNN chain architecture and MV-CNN with Directed Acyclic Graph (DAG) comprising 3D Inception ResNet and 3D Inception are used. Binary and ternary classification using lung CT scan images are conducted. Binary classification contains the two classes *malignant* and *benign*, whereas ternary classification contains three classes: *metastatic malignant*, *primary malignant* and *benign*. Data used for classification is taken from LIDC-IDRI dataset with 10-fold cross validation. Results demonstrate that 3D MV-CNN achieves better performance than 2D MV-CNN in regard to the MV-CNN chain architecture. On the other hand, 3D Inception attained 4.59% binary classification error rate and ternary classification error rate of 7.70%. Comparison of one-view-one-network and multi-view-one-network strategies in this work reveal that the latter achieves lower error rate.

Multi-level Contextual 3D-CNNs are applied by Qi Dou et al. in [26] for nodule identification in CT scans with the objective of reducing false positive rates, accurately discriminating true lung nodules from candidates. The 3D-CNN is compared with 2D-CNN, establishing that that former learns more representative features while encoding a greater degree of spatial information. Multi-level contextual information is additionally used to make the model more robust. The LUNA16 challenge dataset is used for experimentation, the model achieving best performance for false positive rate reduction in the LUNA16 challenge. Results show that the 3D-CNN integrated with multi-level contextual information achieves very promising performance, and which can furthermore be used for other detection problems in medical imaging.

Nonetheless, a disadvantage of 3D-CNN methods is that their parameter-freedom is significantly increased in comparison to 2D-CNNs, potentially leading the model to overfit the training data; it also takes significantly more time and consumes more memory as the size of model increases. Moreover, the availability of annotated 3D data is also an issue; the model of [8] is trained on much less data but cannot learn generalized features for detecting pulmonary nodules.

Besides the 3D-CNN, another commonly used approach for incorporating volumetric/cross-slice information is the RNN. To this end, Petros-Pavlos Ypsilantis and Giovanni Montana in [27] proposed a technique called ReCTnet for the detection of lung cancer nodules in CT scan images. To highlight areas of interest, ReCTnet generates probability maps of three dimensions at pixel level in order to discriminate between normal (non-nodule) and nodule structures. Recurrent layers are merged with convolutional layers to learn spatial dependencies from the slices. First, the architecture is pre-trained without the RNN by using a softmax layer at the end. After training this model, LSTM layers are combined before the fully-connected layers. LIDC-IDRI data is for experimentation with results indicating that on average, 4.5 false positives per scan with a detection sensitivity of 90.5% can be achieved. Comparison with multi-channel CNN and previous studies is carried-out showing that ReCTnet achieves reliable results.

While the above approach uses a single-view CNN for learning feature extractors, we here adopt a MV-CNN for more robust characterization of heterogeneity of nodules. ReCTnet uses unidirectional LSTM, while we use a bidirectional LSTM to capture cross-slice variations in both directions. The previous study integrated two LSTM layers after the CNN layer (pre-trained CNN), whereas we will use one bidirectional LSTM layer after the CNN layers, allowing simplification of the architecture such that the model learns generalized and robust features. In our work, we will thus propose a methodology which incorporates multi-view contextual information while integrating shape, size and cross-slice information of nodules. The benefit of multi-view contextual information is that — with respect to the identification of pulmonary nodules — the model can learn generalized features for small, medium and large nodules as appropriate.

3. Methods

In the following section we discuss MV-CNN, RNN (LSTM) and our proposed technique (MV-CRecNet).

3.1. MV-CNN

In recent years, the research community has begun to use MV-CNNs by virtue of their promising performance with respect to their counterpart Single-view CNNs (SV-CNNs). The main difference between a SV-CNN and a MV-CNN is their input; in MV-CNN, different views of the same input are passed to the model to learn generalized features from the data. Typically, the CNN module is comprised of convolution and pooling layers. To extract features from images, convolutional layers are used [28] with the convolution operation specified as:

$$f_k(x,y) = ReLU\left(b_k + \sum_k \sum_{u,v} I_k(x-u, y-v) * W_k(u,v)\right) \tag{1}$$

Here, the image is denoted by I and the filter for the k'th filter map is denoted by W. The '$*$' sign denotes to the inner product of I and W. After the inner product, a bias term is added denoted by b. '$ReLU$' is a non-linear activation function [29], which is used in this study. These activation functions are used to extract non- linear features.

Convolution is followed by maxpooling which selects the maximum value within a receptive field [30, 31]. Maxpooling is conducted via:

$$M_{kij} = \max_{(s,t)\in \mathbf{R}_{i,j}} (x_{kst}) \tag{2}$$

where M_{kij} is the output of the maxpooling operation for the kth feature map. (s,t) denotes the location of element x_{kst} which is in the pooling region $R_{i,j}$ that covers the receptive field around the position (i,j) [29].

Features obtained after applying the convolutional and maxpooling layers are combined via fully connected layers, which carry-out high level reasoning through interpretation of the features in context [32, 33, 34]. Denser connections

are enabled in fully connected layers, where each neuron is connected with all the neurons in the subsequent layer, as follows:

$$h^t = ReLU(b^t + h^{t-1}W^t) \tag{3}$$

where h^t is the output and h^{t-1} is the input feature vector. W^t denotes the weight matrix and b^t is the bias term. '$ReLU$' is the non-linear activation function, which is used in our study [35].

Fig. 3 shows different views of nodules with (20 x 20), (30 x 30) and (40 x 40) receptive fields. There can be different architectures for MV-CNNs; e.g. a separate network for each view or it can be a single network for multi-views, as illustrated in Fig. 4(a) and 4(b) . In this work, we adopt a separate network for each view for both the 2D MV-CNN and 3D MV-CNN.

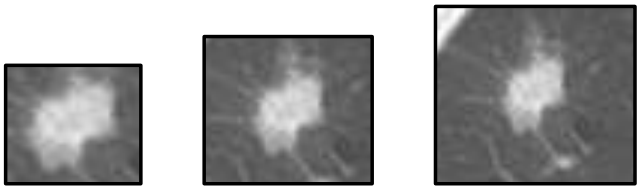
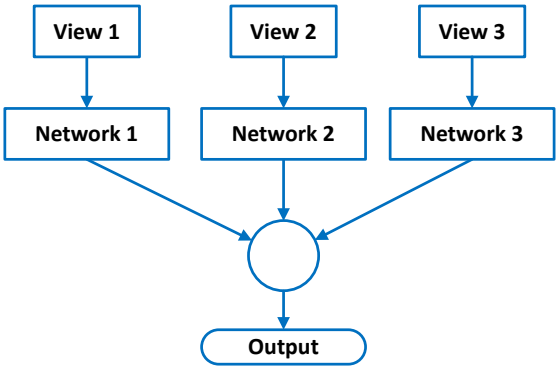
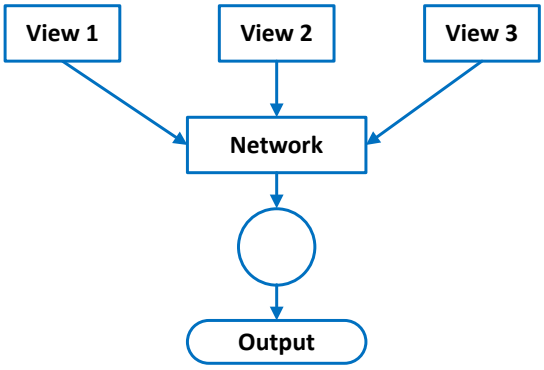


Figure 3: Nodules views with different receptive fields.



(a) Separate network for each view



(b) Single network for multiple views

Figure 4: Variations of MV-CNN architecture.

3.2. RNN / LSTM

The LSTM is a RNN-variant for processing sequential or temporal data. It has shown outstanding results for various problems such as image captioning, language translation/modeling and speech recognition. The specific architecture of a LSTM is demonstrated in Fig. 5. An essential component of LSTM is the memory cell that holds information pertaining to associations between elements of sequential data. The cell is regulated via sigmoid gates including an input gate (I_t), a 'forget' gate (F_t) and an output gate (O_t) as defined below:

$$I_t = \sigma(b_i + [\hat{X}_t, \hat{H}_{t-1}] \cdot W_i) \tag{4}$$

$$F_t = \sigma(b_f + [\hat{X}_t, \hat{H}_{t-1}] \cdot W_f) \tag{5}$$

$$O_t = \sigma(b_o + [\hat{X}_t, \hat{H}_{t-1}] \cdot W_o) \tag{6}$$

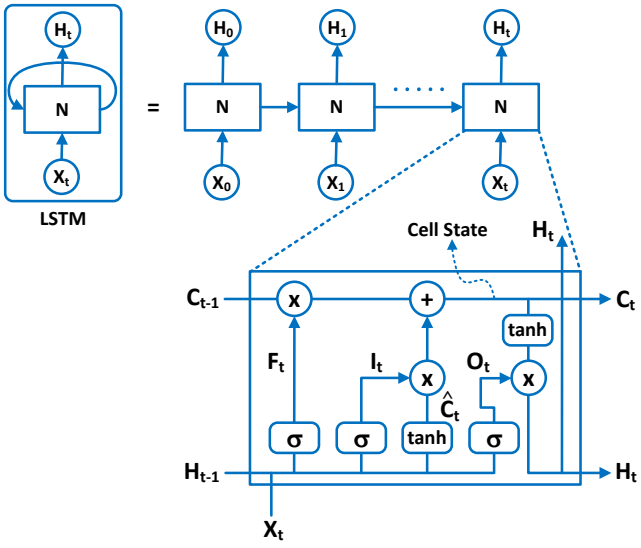


Figure 5: Architecture of LSTM [36].

Here, W_i, W_f and W_o are weight matrices, $\hat{X}_t$ is the input, $\hat{H}_{t-1}$ is the output of the previous layer and b denotes the bias term. The '[]' brackets show the concatenation of $\hat{X}_t$ and $\hat{H}_{t-1}$ as a single vector. The subscript 't' indicates that values of gates are not static, rather that they are capable of change at every time interval t. The cell state is updated as as follows:

$$C_t = (C_{t-1} * F_t) + (\hat{C}_t * I_t) \tag{7}$$

where W_c is a weight matrix and $\hat{C}_t$ is computed as below:

$$\hat{C}_t = tanh(b_c + [\hat{X}_t, \hat{H}_{t-1}] \cdot W_c) \tag{8}$$

Finally, the representation of the sequential data is extracted as follows:

$$H_t = tanh(C_t) * O_t \tag{9}$$

3.3. Proposed Methodology (MV-CRecNet)

This section sets out the architecture of the proposed model (MV-CRecNet), which is structured such that shape, size and volumetric (3D) patterns of nodules can be learned and used for subsequent detection of lung cancer nodules. Suppose that we have a CT dataset $(\hat{I}, L)$ consisting of n samples $\hat{I} = \hat{I}_1, \hat{I}_2, ..., \hat{I}_n$ with corresponding labels $L_i \in \{0, 1\}$, s.t. $L_i = 1$ indicates a nodule and $L_i = 0$ a non-nodule. Each sample $\hat{I}_i$ is a volume of shape (Z, X, Y). The objective is to learn a model function $\hat{y}_i = M(\hat{I}_i)$ where $\hat{y}_i \in [0, 1]$ is the output of the function M that predicts the probability that the volume contains a cancer nodule (i.e. $L_i = 1$). The sample $\hat{I}_i$ consists of volumetric views V_k and each view $V_k^{(i)}$ comprises m slices:

$$V_k^{(i)} = \langle S_{k_1}^{(i)}, S_{k_2}^{(i)}, ..., S_{k_m}^{(i)} \rangle \tag{10}$$

Each slice $S_k^{(i)}$ is processed via a CNN model consisting of convolution, batch normalization, max pooling, and rectified linear unit (ReLU) based activation function with dropout. Despite the single-view processing of individual slices, we allow the model to process slices at multiple views using differently-parameterized CNNs. This enables the model to characterize size-oriented heterogeneity of nodules as different receptive fields can be used to deal with different sizes of nodules. To this end, we use three CNNs to facilitate small, medium and large size nodules. The receptive fields of these CNN architectures in Z, X and Y dimensions are taken as (6 x 20 x 20), (6 x 30 x 30) and (6 x 40 x 40) which covers 58%, 85% and 99% of pulmonary nodules respectively [26]. Another reason to consider 6 as the depth of receptive fields was because if a bigger number is considered as the depth, the size exceeds the size of some CT scans, especially when nodules are in the beginning or last slices of CT scans. The extracted feature maps of all three CNNs are then flattened and concatenated. Mathematically:

$$F_{k_j} = CNN_1\left(S_{k_j}^{(i)}\right) + CNN_2\left(S_{k_j}^{(i)}\right) + CNN_3\left(S_{k_j}^{(i)}\right) \tag{11}$$

These concatenated features are then passed to the LSTM. The slice level features are processed through the LSTM sequentially to capture cross-slice anatomical dependencies of nodules; mathematically as:

$$\psi_{k_j}^{(i)} = \underset{t=j}{LSTM}\left(F_{k_j}\right) \tag{12}$$

Where ψ denotes the output of LSTM. We use the output of the final LSTM layer as an anatomical feature vector for subsequent processing. We utilize the bidirectional LSTM in our model which was discovered to give better results than the unidirectional LSTM. Final classification is performed by using fully-connected networks with sigmoid activation function as follows:

$$h^{(i)} = \rho\left(\psi_{k_j}^{(i)}\right) \tag{13}$$

$$y_i = \sigma\left(h^{(i)}\right) \tag{14}$$

Where ρ denotes fully connected layer and σ denotes Sigmoid activation function. The complete architecture of MV-CRecNet is illustrated in Fig. 6.

4. Materials and Models

In this section we give a description of the test datasets, evaluation metrics and specification of models.

4.1. Dataset and Candidate Generation

Two datasets were considered in this study for the evaluation of proposed technique. Details about the datasets and candidate generation are given below.

4.1.1. LIDC-IDRI

The LIDC-IDRI dataset is considered in this study for the evaluation of proposed technique. The dataset contains 1018 cases deriving from both commercial medical imaging companies and academic centers. It contains thoracic CT scan images with an Extensible Markup Language (XML) files of annotated data conducted by four radiologists with two phases of annotation. During the annotation process, radiologists marked the candidate nodules as nodules $\geq$ 3mm, nodules $<$ 3mm and non-nodules [37]. Candidate positions were extracted using 3 existing detection algorithms [38], [39] and [40]. 1185 nodules with size greater than or equal to 3mm were selected from across the dataset for final panel acceptance by at least 3 out of 4 radiologists. Nodules $<$ 3mm, non-nodules and nodules which were annotated by fewer than 3 radiologists are not considered. In sum, we consider 2370 candidate positions out of which 1185 are nodules and 1185 are non-nodules. In total, 14220 images, 7110 (2D images) and 7110 (3D images), were extracted for presentation to the different models implemented in this study. 80% of the data was used for training and 20% for testing.

4.1.2. ELCAP

ELCAP and Vision and Image Analysis Group (VIA) collaborated to make this dataset available with the purpose of providing a common dataset that can be used to evaluate the performance of different CAD systems [41]. Dataset contains 50 low-dose CT scans with 1.25mm slice thickness and single breath hold for lung cancer detection. Annotations by radiologists of only nodule locations (no non-nodule locations) are also provided. In this study, we considered 902 positions from which 451 are nodules and 451 are non-nodules. As said earlier that ELCAP dataset only contains annotations of nodule locations, therefore, we took non-nodule locations from LIDC-IDRI dataset. The reasons behind considering non-nodules are: firstly, this study consists of binary classification problem and, secondly, to make the models learn the boundaries between nodule and non-nodule data to generalize better. In total, 5412 images are considered out of which 2706 are 2D images and 2706 are 3D images to feed to different models considered in this study. Moreover, 60% data was used for training and 40% for testing.

4.2. Evaluation Metrics

For evaluating the performance of proposed approach, six evaluation metrics are utilized: Accuracy, Sensitivity,

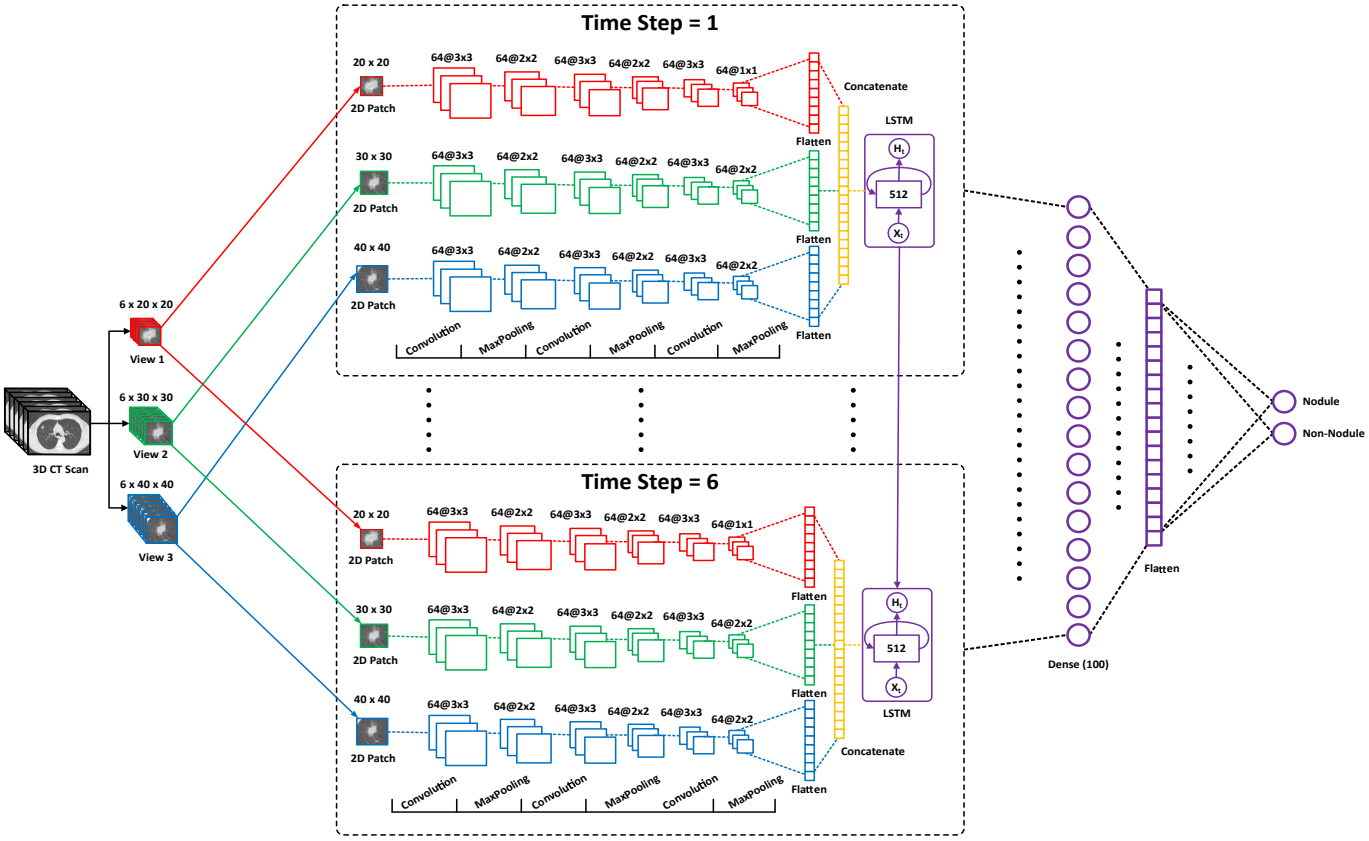


Figure 6: Diagram of the proposed model (MV-CRecNet) in which each layer is shown with different color for better comprehension.

Specificity, Precision, F1-Score and Receiver Operating Characteristic (ROC) curve (via the AUC value).

- True Positive (TP): Number of correct positive (cancerous) examples detected.
- True Negative (TN): Number of correct negative (non-cancerous) examples detected.
- False Positive (FP): Number of incorrect positive (cancerous) examples detected.
- False Negative (FN): Number of incorrect negative (non-cancerous) examples detected.

4.2.1. Accuracy

Accuracy can be described as ratio of number of correctly predicted samples and total number of samples. It is a great measure but only when FP and FN values are almost same. Accuracy is computed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

4.2.2. Sensitivity

Sensitivity is also called Recall or True Positive Rate. When FN has high cost then recall is a good measure to use for evaluating the model. Sensitivity is computed as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{16}$$

4.2.3. Specificity

Specificity is also known as True Negative Rate. Specificity is computed as:

$$Specificity = \frac{TN}{TN + FP} \tag{17}$$

4.2.4. Precision

When cost of FP is high, precision is a good measure for evaluation. Precision is computed as follows:

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

4.2.5. F1-Score

When class distribution is uneven then it is better to use F1-Score rather than accuracy. It is computed as:

$$F1 - Score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{19}$$

4.2.6. ROC Curve and AUC Value

ROC curve demonstrates that how good the model is performing to differentiate between two things, for instance, whether a patient has disease or not. To draw ROC curve, sensitivity and specificity, more precisely (1 − specificity), are used. AUC is a good metric for evaluating model's performance; the higher the AUC score is the better the model is performing.

Table 1
Architectures of 2D-CNNs and 2D MV-CNN

2D-CNNs								
A1			A2			A3		
Lay.	Ker.	Cha.	Lay.	Ker.	Cha.	Lay.	Ker.	Cha.
Input	20x20	1	Input	30x30	1	Input	40x40	1
C1	3x3	64	C1	3x3	64	C1	3x3	64
M1	2x2	64	M1	2x2	64	M1	2x2	64
C2	3x3	64	C2	3x3	64	C2	3x3	64
M2	2x2	64	M2	2x2	64	M2	2x2	64
FC1	-	20	FC1	-	20	FC1	-	20
FC2	-	1	FC2	-	1	FC2	-	1
Sigmoid	-	1	Sigmoid	-	1	Sigmoid	-	1
2D MV-CNN								
$Macro-averaging = \frac{(P1+P2+P3)}{3}$								
Output								

A: Architecture, Lay: Layer, Ker: Kernel, Cha: Channel, C: Convolution, M: Max Pooling, FC: Fully Connected, P: Probability

Table 2
Architectures of 3D-CNNs and 3D MV-CNN

3D-CNNs								
A1			A2			A3		
Lay.	Ker.	Cha.	Lay.	Ker.	Cha.	Lay.	Ker.	Cha.
Input	6x20x20	1	Input	6x30x30	1	Input	6x40x40	1
C1	3x3x3	64	C1	3x3x3	64	C1	3x3x3	64
M1	2x2x2	64	M1	2x2x2	64	M1	2x2x2	64
C2	1x3x3	64	C2	1x3x3	64	C2	1x3x3	64
M2	2x2x2	64	M2	2x2x2	64	M2	2x2x2	64
FC1	-	20	FC1	-	20	FC1	-	20
FC2	-	1	FC2	-	1	FC2	-	1
Sigmoid	-	1	Sigmoid	-	1	Sigmoid	-	1
3D MV-CNN								
$Macro-averaging = \frac{(P1+P2+P3)}{3}$								
Output								

A: Architecture, Lay: Layer, Ker: Kernel, Cha: Channel, C: Convolution, M: Max Pooling, FC: Fully Connected, P: Probability

4.3. Specification of Models

Apart from the proposed methodology, eleven models were implemented for comparison. Description of those models is given in the following subsections.

4.3.1. 2D-CNNs and 2D MV-CNN

The evaluated 2D-CNNs and 2D MV-CNN models have been discussed in Section 3; the specific architectural details of the 2D-CNNs and 2D MV-CNNs are given in Table 1.

4.3.2. 3D-CNNs and 3D MV-CNN

Volumetric 3D-CNNs and 3D MV-CNN were also implemented in this study for comparison to the proposed methodology. The 3D-CNN models consist of 3D convolutional and maxpooling layers; the 3D convolutional layer, in comparison with a 2D convolutional layer, has one more dimension, requiring an additional parameter z and also an additional parameter w for the image I and the kernel W, respectively. The 3D convolution operation is defined as:

$$f_k(z,x,y) = ReLU\left(b_k + \sum_k \sum_{w,u,v} I_k(z-w, x-u, y-v) * W_k(w,u,v)\right) \quad (20)$$

Likewise, the 3D maxpooling layer has an additional dimension in comparison with a 2D maxpooling layer. The size of maxpooling window is thus not only specified for the X/Y dimensions but also for the Z dimension. The equation of the 3D maxpooling operation is hence the same as 2D maxpooling but with an explicit 3rd dimension:

$$M_{khij} = \max_{(r,s,t)\in \mathbb{R}_{h,i,j}} (x_{krst}) \quad (21)$$

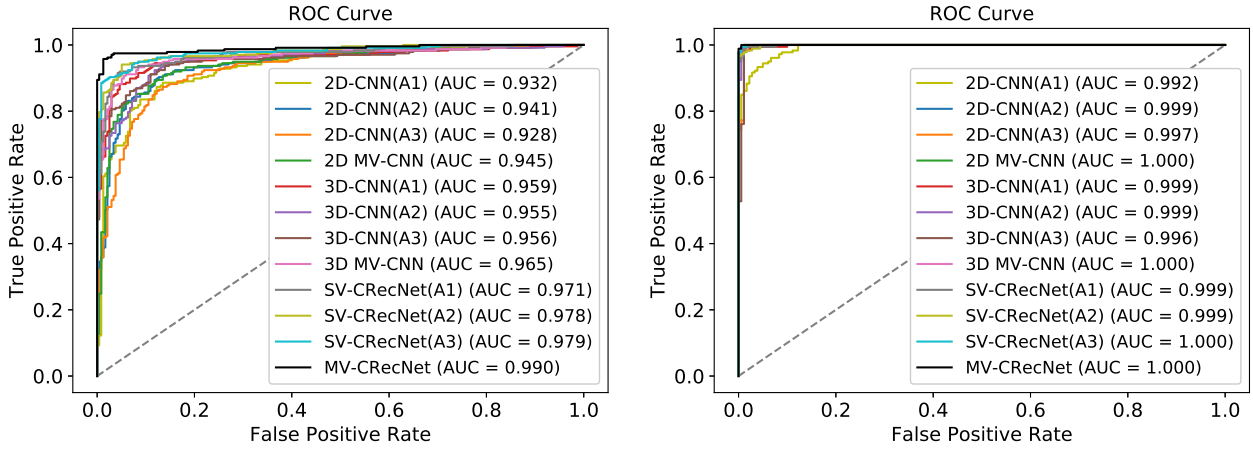
Table 2 presents the architectures of these 3D models.

4.3.3. Single-view CRecNet (SV-CRecNet)

To capture the anatomical dependencies of nodules across slices, we seek to utilize the sequential properties of RNNs. To this end, we merge the CNN model with the LSTM model, which in this work we collectively call SV-CRecNet. In SV-CRecNet, a CNN is used to learn slice level features and an LSTM is then employed to learn cross-slice dependencies of nodules from CT scan images. At each time step, a 2D slice of the CT scan is passed to the model; after obtaining features from the CNN, these are flattened and passed to the LSTM in order to incorporate the anatomical dependencies of nodules across slices. The number of steps depend on the number of slices considered from a CT scan image. Afterwards, an integrated fully-connected layer interprets the features and performs high level reasoning. The features are then once more flattened and the final prediction is produced. The ReLU activation function is used throughout the models except for the last layer which uses a sigmoid activation function. Three different architectures of SV-CRecNet are here evaluated, with model specifications as given in Table 3.

5. Experimental Results and Discussion

In this study, a size and shape aware model, MV-CRecNet, is proposed for lung cancer nodule identification. Its efficacy is determined with respect to six evaluation metrics in order to demonstrate the robustness of the model. Aside from the proposed methodology, eleven further models are implemented for comparison, consisting of 2D-CNN, 2D MV-CNN, 3D-CNN, 3D MV-CNN and SV-CRecNet variants. Table 4 shows the results of all the techniques for all evaluation metrics on LIDC-IDRI and ELCAP datasets. ROC



(a) LIDC-IDRI dataset

(b) ELCAP dataset

Figure 7: ROC curves and AUC scores of all models on different datasets.

Table 3
Architectures of SV-CRecNet Variants

SV-CRecNets								
A1			A2			A3		
Lay.	Ker.	Cha.	Lay.	Ker.	Cha.	Lay.	Ker.	Cha.
Input	6x20x20	1	Input	6x30x30	1	Input	6x40x40	1
C1	3x3	64	C1	3x3	64	C1	3x3	64
M1	2x2	64	M1	2x2	64	M1	2x2	64
C2	3x3	64	C2	3x3	64	C2	3x3	64
M2	2x2	64	M2	2x2	64	M2	2x2	64
C3	3x3	64	C3	3x3	64	C3	3x3	64
M3	1x1	64	M3	2x2	64	M3	2x2	64
Flatten	-	-	Flatten	-	-	Flatten	-	-
LSTM	512	-	LSTM	512	-	LSTM	512	-
FC	-	100	FC	-	100	FC	-	100
Flatten	-	-	Flatten	-	-	Flatten	-	-
Sigmoid	-	1	Sigmoid	-	1	Sigmoid	-	1

A: Architecture, Lay: Layer, Ker: Kernel, Cha: Channel, C: Convolution, M: MaxPooling, FC: Fully Connected

curves and AUC scores are illustrated in Fig. 7(a) and 7(b) for LIDC-IDRI and ELCAP datasets, respectively. Results show that MV-CRecNet learns more generalized and robust features in comparison to other models. Experimental setup, results and discussion are given below.

5.1. Experimental Setup

Experiments were carried out on Intel(R) Core i5-2540M, 2.60GHz CPU, 4 GB RAM, with Windows 10 operating system with additional GPU-based experimentation carried out on Google's Colab cloud service. Models were implemented using the Python programming language and open source Keras library. Results for individual models are given below with discussion in-group.

5.2. LIDC-IDRI

Results and discussion of all implemented models on LIDC-IDRI dataset are given below.

5.2.1. Results of 2D CNN Models

The objective of the experiments is to obtain a model structure useful for detecting and recognizing lung nodules. To this end, we empirically analyze existing model structures and their variations as a baseline. We start with the simplest setting, involving slice-wise processing of CT scans using 2D-CNNs. In this setting, an important question is to determine the size of receptive field suitable for input layer. Three receptive fields, (20 x 20), (30 x 30) and (40 x 40) are hence analyzed in relation to the respective 2D-CNN architectures A1, A2 and A3. Results on LIDC-IDRI dataset demonstrate that the 2D-CNN(A1) architecture achieves better accuracy, specificity and precision than other two models. The reason for this could be that receptive field used for 2D-CNN(A1) is able to successfully identify non-nodules because it provides appropriate contextual information (at least for non-nodules selected in this study). However, the sensitivity of the 2D-CNN(A1) is lower when compared with the other two models because the receptive field only covers 58% pulmonary nodules in the dataset and is thus good for small sized nodules but not the others. The 2D-CNN(A2) has better sensitivity and f1-score than other two models. This is perhaps due to the receptive field used for the 2D-CNN(A2) containing rich contextual information for small sized nodules and apposite contextual information for medium sized nodules, causing the model to classify true positives correctly with high confidence. The reason for the degraded performance of 2D-CNN(A3) in accuracy and f1-score compared with the other two models could be that the receptive field used includes noise in the context of small sized nodules. However, the 2D-CNN(A3) model achieved better sensitivity than the 2D-CNN(A1) model, and also better specificity and precision than the 2D-CNN(A2) model,

Table 4
Results of all models implemented in this study on LIDC-IDRI and ELCAP datasets and their comparison with proposed technique

Models	LIDC-IDRI Dataset					ELCAP Dataset				
	Accuracy	Sensitivity	Specificity	Precision	F1-Score	Accuracy	Sensitivity	Specificity	Precision	F1-Score
2D-CNN(A1)	0.869	0.861	0.878	0.876	0.868	0.925	0.861	0.989	0.987	0.920
2D-CNN(A2)	0.865	0.924	0.806	0.826	0.873	0.942	0.911	0.972	0.970	0.940
2D-CNN(A3)	0.857	0.899	0.814	0.829	0.862	0.964	1.000	0.928	0.934	0.965
2D MV-CNN	0.882	0.911	0.852	0.861	0.885	0.986	0.978	0.995	0.994	0.986
3D-CNN(A1)	0.914	0.873	0.954	0.950	0.910	0.972	0.972	0.972	0.972	0.972
3D-CNN(A2)	0.892	0.941	0.844	0.858	0.897	0.981	1.000	0.961	0.963	0.981
3D-CNN(A3)	0.897	0.911	0.882	0.885	0.898	0.947	1.000	0.895	0.905	0.950
3D MV-CNN	0.918	0.916	0.920	0.920	0.918	0.983	0.967	1.000	1.000	0.983
SV-CRecNet(A1)	0.935	0.911	0.958	0.956	0.933	0.975	0.956	0.995	0.994	0.975
SV-CRecNet(A2)	0.941	0.932	0.949	0.948	0.940	0.967	1.000	0.9333	0.938	0.968
SV-CRecNet(A3)	0.930	0.933	0.928	0.929	0.931	0.961	1.000	0.9222	0.928	0.963
MV-CRecNet	**0.971**	**0.975**	**0.966**	**0.967**	**0.971**	**0.994**	**1.000**	**0.989**	**0.989**	**0.995**

which indicates that it is a good choice for classifying for medium and large sized nodules. To maximize the benefits of the different receptive fields, we analyzed the efficacy of multiple receptive fields using a 2D MV-CNN architecture. A single network for each view was integrated for the 2D MV-CNN; macro-averages of the probabilities deriving from each architecture were then calculated. The 2D MV-CNN model showed improvement in accuracy and f1-score in this context, revealing that passing multi-level contextual information makes the model generalize better; this applies across all evaluation metrics.

5.2.2. Results of 3D CNN Models

2D models cannot take full advantage of volumetric information in CT scans; 3D-CNNs were thus employed to utilize this information. Three 3D-CNNs were implemented: 3D-CNN(A1), 3D-CNN(A2) and 3D-CNN(A3); each CNN takes a different sized volumetric image of magnitude (6 x 20 x 20), (6 x 30 x 30) and (6 x 40 x 40), respectively. Results show that 3D-CNN(A1) obtains better accuracy, specificity, precision and f1-score than the other two 3D-CNNs. This is perhaps because the selected non-nodules in this work are mostly of a small size and a (6 x 20 x 20) receptive field thus provides rich contextual information for these non-nodules. 3D-CNN(A2) achieves better sensitivity than other two models perhaps because the receptive field used for this model allows for rich and apposite contextual information for small and medium sized nodules, respectively, enabling the model to correctly classify true nodules. The 3D-CNN(A3) model achieves moderate results in comparison to other two models, likely due to the receptive field necessarily including noise in the case of small sized nodules, but an appropriate amount of contextual information in relation to medium and large sized nodules. In order to analyze the integration of multiple views with different receptive fields, 3D MV-CNN was applied. Results for this model demonstrate that 3D MV-CNN outperforms all three 3D-CNNs for all evaluation metrics (except sensitivity). The improved results are indicative that the integration of multi-views lets the model learn generalized features. In general, 3D-CNN models performed better than 2D-CNN models, demonstrating that applying 3D models on volumetric information improves results in line with expectation. However, the parameter complexity of the 3D-CNN significantly increases the likelihood of the model overfitting the training data; the unavailability of 3D annotated data is also an issue with this model class.

5.2.3. Results of SV-CRectNets including the Proposed Methodology (MV-CRecNet)

We incorporated LSTMs with 2D-CNNs (which we call SV-CRecNets) in order to force the model to learn shape, size and anatomical dependencies of nodules across slices. For this purpose, we employed three SV-CRecNet architectures: SV-CRecNet(A1), SV-CRecNet(A2) and SV-CRecNet(A3). All three models outclass all of the previously implemented models in this study, even 3D-CNNs. This is perhaps because the 3D-CNN considers nodules purely as volumetric information for which emerging patterns of nodules across progressive slices may not readily be learned. Integration of 2D-CNNs with LSTMs appears to enable learning of anatomical dependencies of nodules across slices such that the model is able to learn how a nodule first appeared, grew to a maximal cross-section and then diminished across slices; that is, it captures a (slice-dependent) encoding of surface gradient changes. Results indicate that, not only with respect to nodule locations, but also with respect to non-nodule locations, SV-CRecNets achieve best-quality results. The receptive field used for SV-CRecNet(A3) also includes noise in case of small sized nodules; however, it achieves the highest sensitivity among the other SV-CRecNets. Reasons for this could be that the learning of cross-slice dependencies enables the model to identify true nodules correctly even in presence of noise. SV-CRecNet(A1) achieves better results than other two models for specificity and precision evaluation metrics, demonstrating that SV-CRecNet(A1)'s receptive field includes rich contextual information for the non-nodule locations considered in this study. SV-CRecNet(A2) obtained the highest accuracy and f1-score, and achieved

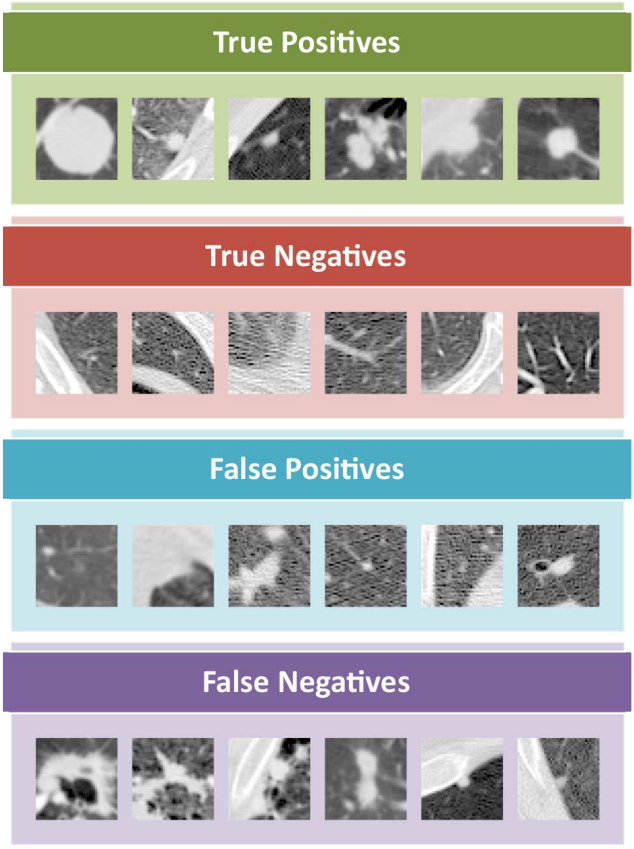


Figure 8: Predictions of proposed model (MV-CRecNet) on various examples.

moderate values for the other evaluation metrics as compared to other two models. This is presumably due to the receptive field of SV-CRecNet(A2) aiding the model in classifying small and medium sized nodules by providing relevant contextual information. In general, SV-CRecNets achieve promising results; however, only single-views are integrated. The question then arises as to what happen if we were to provide multiple views to the model (CRecNet)? The multi-view models implemented previously achieved more promising results than their counterpart single view approaches. Fig. 7(a) shows the ROC curves and corresponding AUC values for all of the models on LIDC-IDRI dataset. It can be seen that each multi-view variant achieves the highest AUC value with respect to their comparative single-view approaches. From this we derive our proposed methodology (MV-CRecNet) in which different views are incorporated. Results show that MV-CRectNet outperforms all of the models in all of the evaluation metrics, demonstrating its robustness. This is because the multiple views passed to the model allow it to learn generalized and robust features for small, medium and large sized nodules; not only nodules but also non-nodules were correctly identified by MV-CRectNet with high confidence. Predictions of MV-CRecNet on various examples are illustrated in Fig. 8.

5.3. ELCAP

Results and discussion of all implemented models on ELCAP dataset are given below.

5.3.1. Results of 2D CNN Models

As mentioned previously that while training models on ELCAP dataset, nodules were considered from ELCAP dataset whereas non-nodules were taken from LIDC-IDRI dataset. On ELCAP dataset, even 2D-CNNs achieved very high values for all evaluation metrics. It can be due to the fact that ELCAP dataset consists of low-dose CT scans, however, LIDC-IDRI dataset contains normal/high-dose CT scans. Therefore, while training the models on different dose CT scans, models may have learned the biases of both datasets and correctly classify majority of examples. So, the model could be learning that if a normal/high dose CT scan comes then it is a negative example and classifies it as non-nodule, whereas, if a low-dose CT scan comes then it is a positive example and classifies it as nodule. Among simple 2D-CNN models, 2D-CNN(A3) has achieved better accuracy, sensitivity and f1-score. Reason could be that the receptive field of this model has covered good amount of contextual information of nodules and correctly classifies them. 2D-CNN(A1) has performed superior to identify non-nodules locations and gained better specificity and precision scores. Nonetheless, it has low sensitivity as compared to other 2D-CNN models. It is because its receptive field was not able to cover appropriate amount of contextual information for nodules; some nodules were bigger in size that is why they did not fit into the size of model's receptive field, as shown in Fig. 9. 2D-CNN(A2) has achieved moderate results in comparison with 2D-CNN(A1) and 2D-CNN(A3). Reason could be its receptive field which not only covered non-nodules correctly but also nodule locations. To combine the benefits of different sized receptive fields, we implemented 2D MV-CNN. Results reveal that 2D MV-CNN obtained highest values, among 2D-CNN models, of accuracy, specificity, precision and f1-score. However, sensitivity score of 2D MV-CNN is less than of 2D-CNN(A3). This is perhaps due to several very large sized nodules in the dataset compared to the receptive field considered. 2D MV-CNN results show that passing multi-views to the model allows it to generalize better.

5.3.2. Results of 3D CNN Models

On ELCAP dataset, after 2D models, 3D models were tested. Reason to apply 3D models was that CT scans are volumetric data (3D data). Therefore, to consider volumetric information of CT scans, 3D-CNNs were implemented. On ELCAP dataset, 3D-CNNs performed superior to 2D-CNN models. 3D-CNN(A2) has achieved better accuracy and f1-score than other simple 3D-CNN models. It could be because receptive field of this model has considered appropriate amount of contextual information of nodules. However, it did not perform better than 3D-CNN(A1) for specificity and precision evaluation metrics. It may be due to the non-nodule locations, for which 3D-CNN(A1)'s receptive field

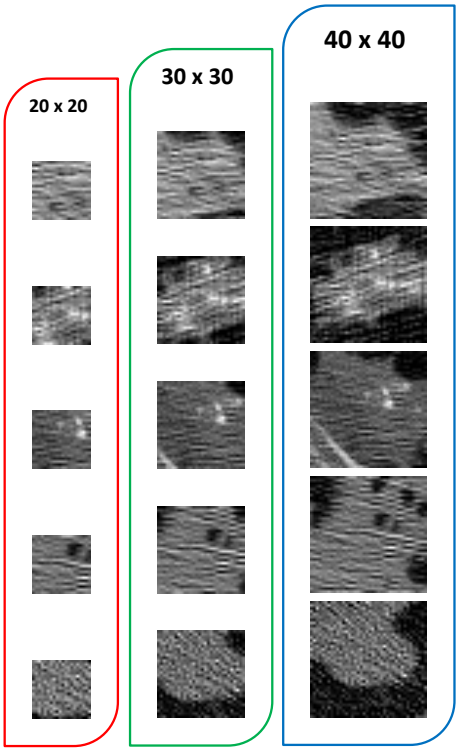


Figure 9: Large sized nodules from ELCAP dataset shown in (20 x 20), (30 x 30) and (40 x 40) receptive fields.

taken into account appropriate contextual information. 3D-CNN(A1) model has shown moderate results as compared to other 3D-CNN models. 3D-CNN(A3) achieved lowest accuracy, specificity, precision and f1-score than other 3D-CNNs. It is perhaps due to the inclusion of noise in case of bigger receptive field integrated for this model. It should be noted that 3D-CNN(A1) and 3D-CNN(A2) has achieved 100% sensitivity, it is because both models were able to get rich contextual information of nodules. To examine the advantages of multi-level contextual information, 3D MV-CNN was implemented. Three views with different receptive fields were provided as input to the 3D MV-CNN model. Results show that 3D MV-CNN outperformed all 3D-CNNs for metrics of accuracy, specificity, precision and f1-score. This is due to the fact that integration of multiple views in a model enables it to differentiate and classify nodules and non-nodules with high confidence. From results it can be analyzed that this model performed better for identifying non-nodules, perhaps because receptive field of this model covered apposite contextual information for non-nodule locations of dataset considered. However, its sensitivity score is less than of 3D-CNN(A2) and 3D-CNN(A3). The volumetric nature of CT scans give 3D-CNNs an upper hand to identify patterns in 3D data. Furthermore, passing multiple views to models let them to generalize better. However, parameters of 3D-CNNs drastically increase which increases the possibility of overfitting; moreover, there is also an issue of unavailability of 3D annotated data.

5.3.3. Results of SV-CRectNets including the Proposed Methodology (MV-CRecNet)

In order to make models to learn shape, size and anatomical dependencies of nodules across slices, 2D-CNNs with LSTMs were combined which we call CRecNets. Three such models which incorporate single views were implemented namely SV-CRecNet(A1), SV-CRecNet(A2) and SV-CRecNet(A3). SV-CRecNet(A1) has achieved highest accuracy, specificity, precision and f1-score as compared to other two models; probably because its receptive field incorporated pertinent contextual information for non-nodules. However, for sensitivity metric this model performed inferior to other SV-CRecNet models, and 3D-CNN(A2) and 3D-CNN(A3), due to the large sized nodules in the dataset whose contextual information was not fully grasped by the receptive field considered, as illustrated in Fig. 9. SV-CRecNet(A2) demonstrated moderate and SV-CRecNet(A3) lowest scores for accuracy, specificity, precision and f1-score evaluation metrics among three SV-CRecNets. Reason for SV-CRecNet(A3) performing lowest could be that receptive field of this model also included noise for candidate positions considered which eventually affected the performance. Nevertheless, it should be noted that SV-CRecNet(A2) and SV-CRecNet(A3) achieved 100% sensitivity. This illustrates that even in presence of noise for SV-CRecNet(A3) model, the consideration of adjacent slices allowed it to identify nodules correctly. SV-CRecNets have achieved good results, nonetheless, only single views are incorporated. From the previous experimentation, it can be observed that multi-view models have performed considerably superior to their counter part single-view models. Thus, what if we provide multiple views to the CRecNet architecture? Fig. 7(b) shows the ROC curves and corresponding AUC values for all the models on ELCAP dataset. Multi-view models have attained highest AUC values as compared to their comparative single-view models. Therefore, we incorporated multiple views in our proposed technique (MV-CRecNet). Results exhibit the robustness of MV-CRecNet which has outperformed all the models by classifying both nodule and non-nodule locations with high confidence. This reveals that passing multi-views to a model which incorporates shape, size and anatomical dependencies, such as MV-CRecNet, allow it to be capable of learning robust and generalized features.

Table 5 shows a comparison of our proposed technique (MV-CRecNet) with other comparable work in the literature. It should be noted that due to the differing data samples, experimental setups and evaluation metrics in the related work, it is difficult to perform a strict comparison. However, most authors have utilized the same dataset (the LUNA16 dataset is a filtered version of LIDC-IDRI dataset, which is why comparison with the papers which have used LUNA16 dataset is also included). Table 5 indicates that our proposed model has achieved very promising results relative to the state of the art.

Table 5
Comparison of proposed technique with existing work

Related work	Dataset	Acc.	Sen.	Spe.	AUC
W. Shen et al. [24]	LIDC-IDRI	0.87	0.77	0.93	0.93
K. Liu et al. [21]	LIDC-IDRI	-	-	-	0.98
X. Liu et al. [22]	LIDC-IDRI	0.92	-	-	-
G. Kang et al. [25]	LIDC-IDRI	-	-	-	0.99
F. Ciompi et al. [23]	DLCST	0.80	-	-	-
Y. Gu et al. [42]	LUNA16	-	0.93	-	-
R. Gruetzemacher et al. [43]	LUNA16	-	0.89	-	0.93
H. Xie et al. [44]	LUNA16	-	0.86	-	0.95
N. Nasrullah et al. [45][46]	LIDC-IDRI	0.89	0.94	0.90	0.99
Z. Xiao et al. [47]	LUNA16	-	0.92	-	-
MV-CRecNet (Ours)	LIDC-IDRI	**0.97**	**0.98**	**0.97**	**0.99**

Acc: Accuracy, Sen: Sensitivity, Spe: Specificity

6. Conclusion and Future Work

Lung cancer is a severe medical threat with symptoms only appearing when the disease is in its advance stage. In this work, we propose a novel deep learning approach for pulmonary nodules identification from CT scan images. The proposed method (MV-CRecNet) exploits shape, size and cross-slice variations in order to learn to identify lung cancer nodules from CT scan images. Multiple views are passed to the model enabling it to generalize better by learning robust features. Besides the proposed methodology, eleven separate models were implemented for comparative evaluation using the LIDC-IDRI and ELCAP datasets. Six evaluation metrics were used to measure the performance of proposed technique with results demonstrating that MV-CRecNet exhibits state-of-the-art performance for all evaluation metrics.

This work has the potential to be extended to various other object identification tasks with multi-view or volumetric data. In future, the proposed technique will hence be tested across a range of medical domains; and with the help of radiologists it can be incorporated in clinical practice.

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012, International Journal of Cancer 136 (5) (2015) E359–E386. doi:10.1002/ijc.29210.

[2] M. C. S. Wong, X. Q. Lao, K.-F. Ho, W. B. Goggins, S. L. A. Tse, Incidence and mortality of lung cancer: global trends and association with socioeconomic status, Scientific Reports 7 (1) (2017) 14300. doi:10.1038/s41598-017-14513-7.

[3] M. R. Sarwar, A. Saqib, Cancer prevalence, incidence and mortality rates in pakistan in 2012, Cogent Medicine 4 (1) (2017) 1288773. doi:10.1080/2331205X.2017.1288773.

[4] R. L. Siegel, K. D. Miller, A. Jemal, Cancer statistics, 2017, CA: A Cancer Journal for Clinicians 67 (1) (2017) 7–30. doi:10.3322/caac.21387.

[5] G. J. S. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, C. I. Sanchez, A survey on deep learning in medical image analysis, Medical Image Analysis 42 (2017) 60–88.

[6] S. S. Han, G. H. Park, W. Lim, M. S. Kim, J. I. Na, I. Park, S. E. Chang, Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network, PLOS ONE 13 (1) (2018) 1–14. doi:10.1371/journal.pone.0191493.

[7] M. Haloi, Towards ophthalmologist level accurate deep learning system for oct screening and diagnosis, CoRR abs/1812.07105. arXiv:1812.07105.

[8] F. Liao, M. Liang, Z. Li, X. Hu, S. Song, Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network, IEEE Transactions on Neural Networks and Learning Systems (2019) 1–12doi:10.1109/tnnls.2019.2892409.

[9] H. Xie, D. Yang, N. Sun, Z. Chen, Y. Zhang, Automated pulmonary nodule detection in ct images using deep convolutional neural networks, Pattern Recognition 85 (2019) 109–119. doi:https://doi.org/10.1016/j.patcog.2018.07.031.

[10] H. Jiang, H. Ma, W. Qian, M. Gao, Y. Li, An automatic detection system of lung nodule based on multigroup patch-based deep learning network, IEEE Journal of Biomedical and Health Informatics 22 (4) (2018) 1227–1237. doi:10.1109/JBHI.2017.2725903.

[11] S. G. Armato, M. L. Giger, C. J. Moran, J. T. Blackburn, K. Doi, H. MacMahon, Computerized detection of pulmonary nodules on ct scans, RadioGraphics 19 (5) (1999) 1303–1311. doi:10.1148/radiographics.19.5.g99se181303.

[12] T. Messay, R. C. Hardie, S. K. Rogers, A new computationally efficient cad system for pulmonary nodule detection in ct imagery, Medical Image Analysis 14 (3) (2010) 390–406. doi:https://doi.org/10.1016/j.media.2010.02.004.

[13] A. Teramoto, H. Fujita, Fast lung nodule detection in chest ct images using cylindrical nodule-enhancement filter, International Journal of Computer Assisted Radiology and Surgery 8 (2) (2013) 193–205. doi:10.1007/s11548-012-0767-5.

[14] H. Han, L. Li, F. Han, B. Song, W. Moore, Z. Liang, Fast and adaptive detection of pulmonary nodules in

thoracic ct images using a hierarchical vector quantization scheme, IEEE Journal of Biomedical and Health Informatics 19 (2) (2015) 648–659. doi:10.1109/JBHI.2014.2328870.

[15] E. Tasci, A. Ugur, Shape and texture based novel features for automated juxtapleural nodule detection in lung cts, Journal of Medical System. 39 (5) (2015) 1–13. doi:10.1007/s10916-015-0231-5.

[16] T. W. Way, B. Sahiner, H.-P. Chan, L. Hadjiiski, P. N. Cascade, A. Chughtai, N. Bogot, E. Kazerooni, Computer-aided diagnosis of pulmonary nodules on ct scans: Improvement of classification performance with nodule surface features, Medical Physics 36 (7) (2009) 3086–3098. doi:10.1118/1.3140589.

[17] M. C. Lee, L. Boroczky, K. Sungur-Stasik, A. D. Cann, A. C. Borczuk, S. M. Kawut, C. A. Powell, Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction, Artificial Intelligence in Medicine 50 (1) (2010) 43–53. doi:https://doi.org/10.1016/j.artmed.2010.04.011.

[18] H. Madero Orozco, O. O. Vergara Villegas, V. G. Cruz Sanchez, H. Ochoa Dominguez, M. Nandayapa Alfaro, Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine, Biomedical Engineering Online 14 (9) (2015) 20. doi:10.1186/s12938-015-0003-y.

[19] M. Firmino, G. Angelo, H. Morais, M. da Camara Ribeiro-Dantas, R. Valentim, Computer-aided detection (cade) and diagnosis (cadx) system for lung cancer with likelihood of malignancy, Biomedical Engineering Online 15 (1) (2016) 17. doi:10.1186/s12938-015-0120-7.

[20] S. Trajanovski, D. Mavroeidis, C. L. Swisher, B. G. Gebre, B. Veeling, R. Wiemker, T. Klinder, A. Tahmasebi, S. M. Regis, C. Wald, B. J. McKee, H. MacMahon, H. Pien, Towards radiologist-level cancer risk assessment in CT lung screening using deep learning, CoRR abs/1804.01901. arXiv:1804.01901.

[21] K. Liu, G. Kang, Multiview convolutional neural networks for lung nodule classification, International Journal of Imaging Systems and Technology 27 (1) (2017) 12–22. doi:10.1002/ima.22206.

[22] X. Liu, F. Hou, H. Qin, A. Hao, Multi-view multi-scale cnns for lung nodule type classification from ct images, Pattern Recognition 77 (2018) 262–275. doi:https://doi.org/10.1016/j.patcog.2017.12.022.

[23] F. Ciompi, K. Chung, S. J. van Riel, A. A. A. Setio, P. K. Gerke, C. Jacobs, E. T. Scholten, C. M. Schaefer-Prokop, M. M. W. Wille, A. Marchiano, U. Pastorino, M. Prokop, B. van Ginneken, Towards automatic pulmonary nodule management in lung cancer screening with deep learning, Scientific Reports 7 (2017) 46479. doi:10.1038/srep46479.

[24] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, J. Tian, Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification, Pattern Recognition 61 (2017) 663–673. doi:https://doi.org/10.1016/j.patcog.2016.05.029.

[25] G. Kang, K. Liu, B. Hou, N. Zhang, 3d multi-view convolutional neural networks for lung nodule classification, PLOS ONE 12 (11) (2017) 1–21. doi:10.1371/journal.pone.0188290.

[26] Q. Dou, H. Chen, L. Yu, J. Qin, P. Heng, Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection, IEEE Transactions on Biomedical Engineering 64 (7) (2017) 1558–1567. doi:10.1109/TBME.2016.2613502.

[27] P. P. Ypsilantis, G. Montana, Recurrent Convolutional Networks for Pulmonary Nodule Detection in CT Imaging, arXiv e-prints abs/1609.09143. arXiv:1609.09143.

[28] C. Y. Lee, S. Xie, P. Gallagher, Z. Zhang, Z. Tu, Deeply-supervised nets, in: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, Vol. 38 of Proceedings of Machine Learning Research, PMLR, San Diego, California, USA, 2015, pp. 562–570.

[29] D. Yu, H. Wang, P. Chen, Z. Wei, Mixed pooling for convolutional neural networks, in: Rough Sets and Knowledge Technology, Springer International Publishing, Cham, 2014, pp. 364–375. doi:10.1007/978-3-319-11740-9_34.

[30] M. Ranzato, F. J. Huang, Y. Boureau, Y. LeCun, Unsupervised learning of invariant feature hierarchies with applications to object recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8. doi:10.1109/CVPR.2007.383157.

[31] W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review, Neural Computation 29 (9) (2017) 2352–2449. doi:10.1162/neco_a_00990.

[32] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, CoRR abs/1207.0580. arXiv:1207.0580.

[33] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: Y. Bengio, Y. LeCun (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.

[34] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: Computer Vision – ECCV 2014, Springer International Publishing, Cham, 2014.

[35] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Vol. 15 of Proceedings of Machine Learning Research, PMLR, Fort Lauderdale, FL, USA, 2011, pp. 315–323.

[36] C. Olah, Understanding lstm networks (Aug 2015, Date Accessed: (Mar 20, 2019)).
URL http://colah.github.io/posts/2015-08-Understanding-LSTMs

[37] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. R. van Beek, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, et. al., 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) (2011) 915–931. doi:10.1118/1.3528204.

[38] K. Murphy, B. van Ginneken, A. Schilham, B. de Hoop, H. Gietema, M. Prokop, A large-scale evaluation of automatic pulmonary nodule detection in chest ct using local image features and k-nearest-neighbour classification, Medical Image Analysis 13 (5) (2009) 757–770. doi:https://doi.org/10.1016/j.media.2009.07.001.

[39] C. Jacobs, E. M. van Rikxoort, T. Twellmann, E. T. Scholten, P. A. de Jong, J.-M. Kuhnigk, M. Oudkerk, H. J. de Koning, M. Prokop, C. Schaefer-Prokop, B. van Ginneken, Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images, Medical Image Analysis 18 (2) (2014) 374–384. doi:https://doi.org/10.1016/j.media.2013.12.001.

[40] A. A. A. Setio, C. Jacobs, J. Gelderblom, B. van Ginneken, Automatic detection of large pulmonary solid nodules in thoracic ct images, Medical Physics 42 (10) (2015) 5642–5653. doi:10.1118/1.4929562.

[41] Elcap public lung image database (Date Accessed: (Mar 22, 2020)).
URL www.via.cornell.edu/lungdb.html

[42] Y. Gu, X. Lu, L. Yang, B. Zhang, D. Yu, Y. Zhao, L. Gao, L. Wu, T. Zhou, Automatic lung nodule detection using a 3d deep convolutional neural network combined with a multi-scale prediction strategy in chest cts, Computers in Biology and Medicine 103 (2018) 220 – 231. doi:https://doi.org/10.1016/j.compbiomed.2018.10.011.

[43] R. Gruetzemacher, A. Gupta, D. Paradice, 3D deep learning for detecting pulmonary nodules in CT scans, Journal of the American Medical Informatics Association 25 (10) (2018) 1301–1310. doi:10.1093/jamia/ocy098.

[44] H. Xie, D. Yang, N. Sun, Z. Chen, Y. Zhang, Automated pulmonary nodule detection in ct images using deep convolutional neural networks, Pattern Recognition 85 (2019) 109–119. doi:https://doi.org/10.1016/j.patcog.2018.07.031.

[45] Nasrullah, J. Sang, M. S. Alam, H. Xiang, Automated detection and classification for early stage lung cancer on CT images using deep learning, in: M. S. Alam (Ed.), Pattern Recognition and Tracking XXX, Vol. 10995, International Society for Optics and Photonics, SPIE, 2019, pp. 200–207. doi:10.1117/12.2520333.

[46] D. Riquelme, M. A. Akhloufi, Deep learning for lung cancer nodules detection and classification in ct scans, AI 1 (1) (2020) 28–67. doi:10.3390/ai1010003.

[47] Z. Xiao, N. Du, L. Geng, F. Zhang, J. Wu, Y. Liu, Multi-scale heterogeneous 3d cnn for false-positive reduction in pulmonary nodule detection, based on chest ct images, Applied Sciences 9 (16) (2019) 3261. doi:10.3390/app9163261.

Mian Muhammad Naeem Abid received the M.S. degree in Computer Science from COMSATS University Islamabad, Pakistan. He is currently working as a Researcher at Medical Imaging and Diagnostics laboratory, COMSATS University Islamabad. His research interest includes Machine Learning (Deep Learning), Medical Imaging and Diagnosis, Computer Vision, Natural Language Processing, so on.

Tehseen Zia is working as Assistant Professor in Department of Computer Science, COMSATS University Islamabad. He holds a Ph.D. in Computer Science (2010) from Vienna University of Technology, Austria, and Master in Computer Science from University of Engineering and Technology, Texila (2003). He received Ph.D. scholarship from Higher Education Commission of Pakistan. His research interest includes Machine Learning and Neuro-Symbolic Computing.

Mubeen Ghafoor is currently serving as Assistant Professor, Department of Computer Science, COMSATS University Islamabad (CUI). He did his PhD from Mohammad Ali Jinnah University (M.A.J.U), Karachi in 2014 in Electronics Engineering with specialization in Image Processing and Computer Vision. Mubeen is currently working as postdoc fellow at University of the West England, UK.

David Windridge is Associate Professor (Reader) in Computer Science and Head of Data Science at Middlesex University, London. His research interests center on the fields of Machine-Learning (A.I.), Cognitive Systems, and Computer Vision (he also has a former research interest in Astrophysics, having obtained his Ph.D. in Cosmology at the University of Bristol, UK). He has played a leading role on a number of large-scale machine-learning projects (including the EU DREAMS4CARS, EPSRC ACASVA and EU DIPLECS projects), and is interested in cross-over areas of data science having won interdisciplinary research grants in areas as diverse as psychology and proteomics. He is Visiting Professor at Trento University, Italy, and the University of Surrey, UK. He has authored more than 100 peer-reviewed publications.