IET Image Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Deep3DSCan: deep residual network and morphological descriptor based framework for lung cancer classification and 3D segmentation

ISSN 1751-9659
Received on 30th August 2019
Revised 27th January 2020
Accepted on 2nd March 2020
doi: 10.1049/iet-ipr.2019.1164
www.ietdl.org

Gaurang Bansal[1], Vinay Chamola[2] ✉, Pratik Narang[1], Subham Kumar[1], Sundaresan Raman[1]

[1]Department of Computer Science, BITS Pilani, Pilani, India

[2]EEE Department, BITS Pilani, Pilani, India

✉ *E-mail: vinay.chamola@pilani.bits-pilani.ac.in*

Abstract: With the increasing incidence rate of lung cancer patients, early diagnosis could help in reducing the mortality rate. However, accurate recognition of cancerous lesions is immensely challenging owing to factors such as low contrast variation, heterogeneity and visual similarity between benign and malignant nodules. Deep learning techniques have been very effective in performing natural image segmentation with robustness to previously unseen situations, reasonable scale invariance and the ability to detect even minute differences. However, they usually fail to learn domain-specific features due to the limited amount of available data and domain agnostic nature of these techniques. This work presents an ensemble framework Deep3DSCan for lung cancer segmentation and classification. The deep 3D segmentation network generates the 3D volume of interest from computed tomography scans of patients. The deep features and handcrafted descriptors are extracted using a fine-tuned residual network and morphological techniques, respectively. Finally, the fused features are used for cancer classification. The experiments were conducted on the publicly available LUNA16 dataset. For the segmentation, the authors achieved an accuracy of 0.927, significant improvement over the template matching technique, which had achieved an accuracy of 0.927. For the detection, previous state-of-the-art is 0.866, while ours is 0.883.

1 Introduction

Lung cancer with a standardised incidence rate of 230 cases per million is one of the most widespread cancers in the world [1], constituting about 19% of deaths by cancer [2]. Lung cancer results in more than 1.59 million deaths, more than colon, prostate and breast cancers every year [3]. More than 6.1 million deaths result from direct tobacco use, and about 900,000 deaths occur from exposure to second-hand smoke [4]. The escalating pace of lung cancer incidents combined with the nature of its recurrence, is becoming a significant health care issue. Therefore, the need for a solution is of utmost importance which can help in early diagnosis.

The increase in survival rate has been slow for lung cancer in comparison to most cancers. More than 62% of the diagnosed cases occur at the age of 65 or above [5] leading to the survival rate for lung cancer being only 10%. National Lung Screening Trial (NLST) in their report demonstrate on how helical computed tomography (CT) scans can help in reducing the mortality rate of lung cancer by 20% [6]. Multifarious scanning processes have provided an expanding number of high-resolution CT scans [7]. CT scans in clinical practice can capture subtle granularities of lung nodules. However, the high sensitivity of CT imaging leads to voluminous data with complex ambiguities being generated [8]. Therefore, it is tough for radiologists to differentiate the lesions from healthy tissues. With the advancement of computerisation and data processing, computer aided design (CAD) system shows great potential in diagnostic assistance.

Segmentation is used for highlighting the nodule in the CT scan, while classification is used after this to detect whether the nodule is malignant or benign. An end-to-end deep learning architecture which takes the input image and classifies it as cancerous or not does not make the intermediate steps, the basis of prediction clear. In the medical domain, where we have to deal with such sensitive issues, the prediction basis, why is the network saying what it is saying has to be clear to both the radiologists and the patients. In the proposed pipeline architecture, the segmented nodules are generated, and they can be critically examined by radiologists to check for correctness. Any errors which have crept in can also be identified and corrected. Although there are many existing techniques in literature for lung cancer nodule segmentation and classification, there is scope of performance improvement in both these areas. Accurate recognition of lesions is immensely challenging owing to factors like low contrast variation, visual similarity between benign and malignant nodules, and so on. This paper proposes a novel approach for recognising lesions which focuses on the inner structures of nodule voxels. The objective is to segment the pulmonary region from outside and each of the lesions as well.

In this paper, a deep learning framework consisting of a residual network (ResNet) is employed to produce the segmentation results. Using the segmented nodules, an approach for tumour diagnosis is dilineated which utilises the power of morphological and deep learning descriptors. The overall architecture pipeline is depicted in Fig. 1. The key contributions of this paper can be summarised as follows:

1. The system proposed here achieves an accuracy of 88.3% for the cancer classification phase.
2. A 3D segmentation network uses volumetric segmentation instead of slice-by-slice processing. It achieves a Dice coefficient of 0.958.
3. Visual analysis of nodules is a sophisticated and tiresome task even for an expert radiologist as the difference between cancerous and non-cancerous lesions is very minute. In this paper, we delineate a pipeline for obtaining 18 morphological descriptors.

The rest of the paper is organised as follows. Related work in this domain is highlighted in Section 2. Section 3 describes the dataset. Proposed Deep3DSCan framework is explained in Section 4. Section 5 discusses the cancer classification of patients. Results of the research and conclusion are presented in Sections 6 and 7, respectively.

© The Institution of Engineering and Technology 2020

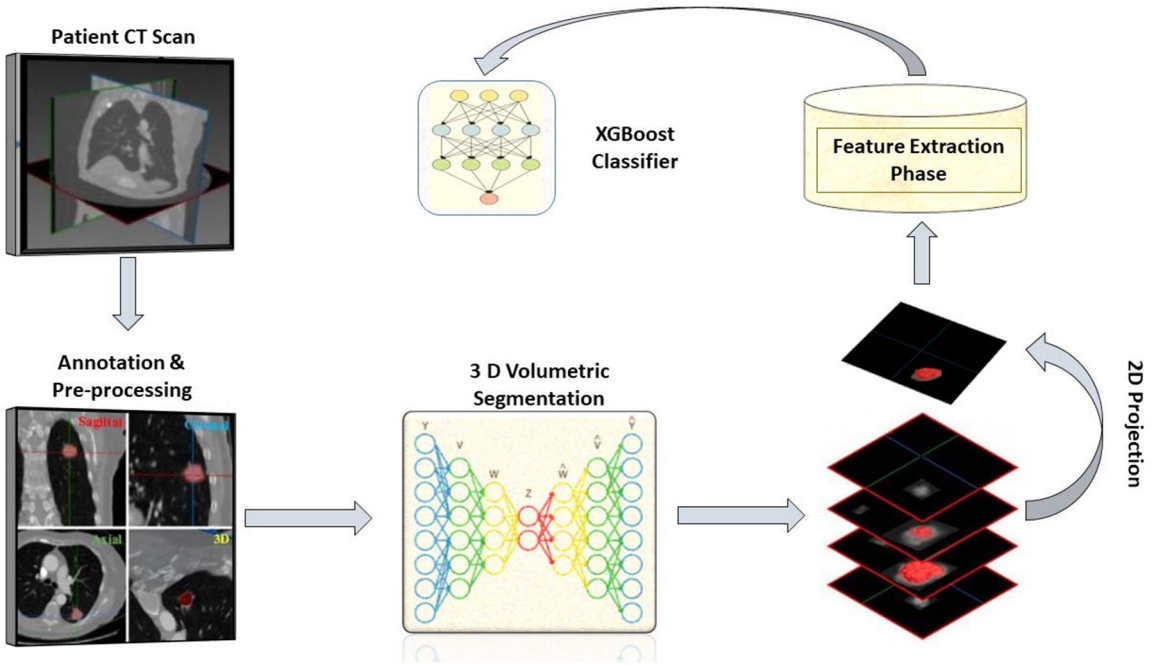

Fig. 1 *System model consisting of the various phases*

2 Literature review

It has been shown in existing literature that advancement in automated designs can help radiologists in early cancer prediction [6, 9]. CAD models aid in the diagnostic determination of pulmonary nodules and in curtailing false positive instances [10]. A diagnostic model is generally composed of four phases. Pre-processing is the first phase and it aims at highlighting the region of interest (ROI). The second phase is segmentation of parenchyma for reducing the search area. Feature extraction is the third phase followed by cancer diagnosis, which is the fourth phase. This section reviews various state-of-art approaches that have been developed for cancer diagnosis.

Most of the models proposed till date focus on enhancing one of the four phases mentioned above. Considering the first phase, image pre-processing involves reduction of noise in images. Research in pre-processing techniques has been extensive. Numerous filtering and statistical techniques for the same have been proposed in the past [11, 12].

The efficiency of CAD model depends on its segmentation phase. Most of the existing segmentation algorithms in image processing domain [13, 14] rely on edge detection and contour curvature extraction. However, issues like inhomogeneities and small density variation in the CT scans limit the application of these techniques. Uzelaltinbulat and Ugur [15] used thresholding for segmenting the lung scans using the difference between greyscale pixels of lesions and the exterior region. Recently, the application of neural networks in the domain of computer vision has attracted many researchers. Variety of neural network architectures such as convolutional networks, capsule networks and deep belief networks (DBNs) have been used for lung image segmentation. The neural network automatically learns the regions of interest using available ground truth. Initially, the research began with 2D convolutional networks [16]. However, in the recent past there has been a paradigm shift to medical-specialised architectures such as U-Net [17] which are specifically designed for extraction of lesions from a medical dataset [18, 19]. Our network differs from U-Net in two major ways: It has 3D filters which perform volumetric convolutions on 3D input and uses a Dice coefficient as the loss function, instead of cross-entropy. Dice coefficient by design performs better when there is a class imbalance, which in this case is between the number of voxels present in the foreground and number of voxels present in the background.

The general trend in radiology classification involves handling an enormous collection of quantitative imaging components. Medical images are converted to mineable data for calculating feature descriptors. The most naive approach is to extract the images and feed it to a standard perceptron model. But with the escalating volume of CT scans, this method is not scalable. However, the scope of feature extraction is not just limited to the perceptron model. Researchers have extended the classical technique of extracting features from images. Abdullah and Shaharum [20] suggested using a feed-forward neural network for lung X-rays. However, the approach employed only three variant features consisting of area, perimeter and shape. Kuruvilla and Gunavathi [21] extended the model by using a set of six discrete parameters. da Silva-Sousa *et al.* [22] utilised geometrical, gradient and spatial properties for improving the model accuracy. After extracting the ROI, authors calculated scale-invariant gradient (SIFT) using ten descriptors.

The application of neural networks is not limited to the segmentation phase; rather their ability to extract relevant features has been utilised in other areas as well. The neural networks family, consisting of convolutional neural networks (CNNs), recurrent neural networks and residual neural networks, have the capability of learning a large number of descriptors which provide enhanced accuracy over traditional handcrafted feature extraction approaches. Kumar *et al.* [23] employed features from neural networks for lung cancer classification on CT images. Hua *et al.* [24] also examined the impact of deep learning on the Lung Image Database Consortium (LIDC) dataset. They accomplished a sensitivity of 70.4 and 73.3% using DBNs and CNNs, respectively. Ciompi *et al.* [25] and van Ginneken *et al.* [26] have used OverFeat [27] (a pre-trained CNN) for classification of pulmonary lesions. They demonstrate the potential of pre-trained CNNs in therapeutic and pharmaceutical cases.

Existing model designs use either geometric features that obtain structure specific shape, texture and morphological facets or determine descriptors from neural network architectures. Geometric features are domain specific but challenging to extract. On the other hand, neural network descriptors are domain agnostic but have advantages like effectiveness in previously unseen situations and the ability to detect even minute differences and providing scalability.

In this paper, an attempt is made to fuse these two diverse categories of feature techniques to create a unified set of attributes which will likely surpass either category of techniques. This paper also discusses 3D residual model architecture for nodule extraction using volumetric segmentation.

© The Institution of Engineering and Technology 2020

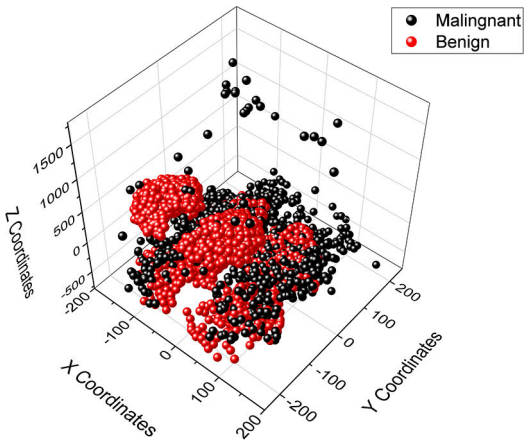

Fig. 2 *Distribution of location of lesions in x, y and z coordinates for 800 patients. Benign lesions (500) are marked as red, while malignant or non-cancerous (300) are marked as black*

3 Dataset

In this work, a publicly available dataset has been used – Lung Nodule Analysis 2016 Challenge [28], commonly referred to as LUNA16. LUNA16 is a subset of the LIDC dataset [29] consisting of CT scans of 888 patients and removing scans with a slice thickness greater than 2.5 mm. The CT scans are stored in MetaImage (mhd/raw) format. Each .mhd file is stored with a separate .raw binary file for the pixeldata. The developers of the challenge provide the ground truth but do not specify the malignancy information about the patient. They also provide information of 551,065 candidates and their label information categorising as 'non-cancerous' or 'cancerous' using detection algorithms.

Cancerous nodules are labelled as 1 while non-cancerous nodules are labelled as 0. However, there is a possibility of multiple candidates existing for single patient. However, there is a possibility of multiple candidates existing for single nodule. For classification phase, the candidate anotation is used to label the patients. Any patient's scan containing even 1 cancerous nodule is labelled as cancerous. The distribution of location of lesions for the patients are shown in Fig. 2.

4 Deep3DSCan framework description

The proposed framework can be subdivided into various stages. The pre-processing phase takes the input CT scans for every patient and applies image processing techniques to decrease the noise and highlight volume of interest (VOI). VOI (nodules) are extracted in the segmentation phase. Then classification of nodule is done on the basis of fused features extracted from 3D ResNet and handcrafted features from morphological techniques. The framework is explained in detail in following sections.

4.1 Pre-processing

The pre-processing step consists of a series of operations for reducing the exploration region. A study of CT scans revealed that there is a substantial radio-density variation between the pulmonary region and surrounding neighbourhood. Thus, thresholding was applied to highlight the pulmonary region from the rest of the search space. Linear metamorphosis of the attenuation coefficient is carried out and the pixel values are transformed into Hounsfield units (HU). HU is employed as the measure of the radio-density. The HU values for the region-of-interest exists between -1000 and -320. Other areas such as air, bone, water, blood were masked out. The mean is calculated for all images to zero-centre the data. In this stage, the patients are given labelled as cancerous and non-cancerous as discussed in the previous section.

4.2 Segmentation architecture

Segmentation is the process of constructing segments of an image by partitioning it using appropriate techniques. Enhancing the resolution of images has produced a multifold increase in diagnostic capabilities [19]. The CT scans are 3D in nature, and dealing with such voluminous information is a challenging task. In medical volumes, the anatomy of interest is a quite small region compared to the whole image. In many of the images, the size of the image is about 36,000 times larger than the lesions present in them. Shallow segmentation networks which are a few layers deep fail to segment such small lesions effectively. A 3D segmentation network is designed which gives an efficient solution for extracting the nodules using the concept of V-Net [30].

The proposed segmentation technique substantially differs from previous state-of-art approaches. Unlike others [31], we refrain from patch-wise image processing. Patch-wise segmentation merely examines the geometric neighbourhood and is thus susceptible to failure. The segmentation architecture is depicted in Fig. 3. At each stage, convolution operations are performed to extract and learn features from the data. This is a fully convolutional network, there are no pooling layers in this architecture, which speeds up backpropogation updates and requires lesser memory, which speeds up training [32]. The segmentation architecture can be subdivided into two halves. As we proceed down the network, the receptive field of the network increases, which is the area of the image which can be percieved by the particular stage. The left branch using 'down convolution' performs compression, while the right section with 'up convolution' decompresses the signal back to its original dimensions. There are skip connections which go from each stage in the left part of the network to the corresponding stage in the right.

Each stage in the proposed segmentation architecture consists of three convolutional layers. In a residual setup, not only is the output of the first convolution layer passed to the second and second to third layer, but the output of the first convolution layer is concatenated to the output of layer 3 (and so on). These connections are known as 'skip connections'. How these layers function is described in Section 4.3. The convolution operations are performed with $2 \times 2 \times 2$ voxels wide kernels. The 'down' convolution operations reduce the resulting feature map at every stage to half, while doubling the number of channels. The right portion extracts features and reconstructs the higher resolution feature map from the lower stages to assimilate the necessary information. The skip/residual connections coming from the left half of the network ensure that the assimilation will contain both the high-level features, coming from the stage below as well as the corresponding low-level features extracted in the corresponding left stage. This is important as otherwise important contextual information would be lost. The skip connections coming from the left half of the network ensure that the assimilation will contain both the high level features, coming from the stage below as well as the corresponding low level features extracted in the corresponding left stage. The feature map is converted into two probabilistic segments as foreground and background by applying a softmax filter. The foreground is the main ROI, while the background contains additional information which is also important for providing the proper context. The probability of a voxel belonging to ROI is calculated using thresholding. All voxels having probability of 0.4 or more form the part of foreground. The expression for Dice coefficient [33] is given as follows:

$$\text{Dice coefficent}(D) = \frac{2\sum_i^N (p_i g_i)}{\sum_i^N (p_i)^2 + \sum_i^N (g_i)^2} \tag{1}$$

where p_i is the predicted binary segmentation, g_i is the ground truth binary volume and the number of voxels are represented by N. The proposed segmentation network is based on the Dice loss function, the gradient to be optimised is given as follows:

© The Institution of Engineering and Technology 2020

$$\frac{\partial D}{\partial p_j} = 2.\left[\frac{g_j\left(\sum_i^N p_i^2 + \sum_i^N g_i^2\right) - 2p_j\left(\sum_i^N p_i g_i\right)}{\left(\sum_i^N p_i^2 + \sum_i^N g_i^2\right)^2}\right] \quad (2)$$

The input image is broken into multiple batches of $128 \times 128 \times 64$ voxels called blocks. The input to the network is a tensor of the form $[16, 128, 128, 64, 1]$ where the image is broken into chunks. This step is required to avoid performance degradation because of redundant computation and to discard the less useful information. The LUNA16 dataset is used to train network. The dataset contains an annotation of nodules which have size > 3 mm. The block size of $128 \times 128 \times 64$ also ensures that there would generally be one nodule per volume. If the same nodule is present in two volumes, the network treats it as if there are two nodules. This assumption does not create any issues of duplicate nodule identification, as features are concatenated in-order resulting in a tensor as if there was a single nodule only. The foreground and background outputs have the same spatial dimensions as that of input, with size of $128 \times 128 \times 64$ voxels. Segmentation results of patient are depicted in Fig. 4.

4.3 Residual network features

With the evolution of deep neural networks (DNNs), experts have employed more sophisticated techniques involving CNNs to address different problems, like AlexNet [34]. One of the revolutionary neural network architectures which was introduced in this domain is the deep ResNet, also known as ResNet [35]. The motivation behind using the ResNet architecture over conventional neural networks is to address two commonly faced issues.

Firstly, as features are learned in the layers and the depth of DNN increases, the performance of model will increase as long as the model is not over-fitting. However, it is generally seen that the accuracy drops as the depth of the network architecture increased beyond a certain threshold. As the gradient progresses along the network, the backpropagation error becomes very small, and the weights of layers hardly change. This is the vanishing gradient problem [36], which results in the layers learning very little information, nothing tangible. The second problem is the degradation problem. As the number of layers increase, performing optimisation on a huge parameter space becomes difficult, leading to higher training error and compounding problems like overfitting. The residual learning framework addresses these issues effectively and eases the training of these networks by using skip-connections. These connections forward the gradient every few layers, so that the gradient is not lost. These connections help the the convolutional layers learn contextual information, which goes a long way in preventing overfitting. These connections also help in decreasing the training time. It also provides a check-and-balance system where it is possible to guage the effectiveness of depth, by comparing the difference between the output of the residual block and its input. If it is not useful then the entire block can be done away with. In this work, a 2D ResNet model is preferred instead of other DNN architectures.

Fig. 5 depicts on how ResNet uses a sequence of layers stacked together. The functionalities of each layer are described in brief here:

• *Convolutional layer*

Convolutional layer, as the name suggests, applies a 'convolution' operation. The convolution operation on the input results into the feature map. The feature map is passed to activation function or ReLU layer.

• *Rectified linear layer*

Rectified linear units layer is commonly known as ReLU. It applies $\max(0,x)$ as an element-wise activation function.

• *Batch normalisation*

ResNet heavily relies on batch normalization (BatchNorm). The basic idea behind batch normalisation is to limit covariate shift by normalising the activations of each layer. It allows each layer to learn on a more stable distribution of inputs.

After the segmentation step, 2D slices of the 3D image are taken and passed to the ResNet model as input. 2D input is used in this case as only the region which contain the nodule is needed in this part of the architecture pipeline. 3D convolution operations are much more computationally intensive compared to 2D operations, hence unnecessary information is discarded. Each channel from the network containing the nodule referred to as a ROI is the input to ResNet. Segmentation results in 2 VOI for each patient of size 128×128. These ROI are used to train a ResNet architecture for cancer classification. The 16 ROI are fed simultaneously to the network. The input to ResNet has dimension $(16, 128, 128)$. The first dimension stands for batch size, while the second and third dimensions are x and y axes, respectively.

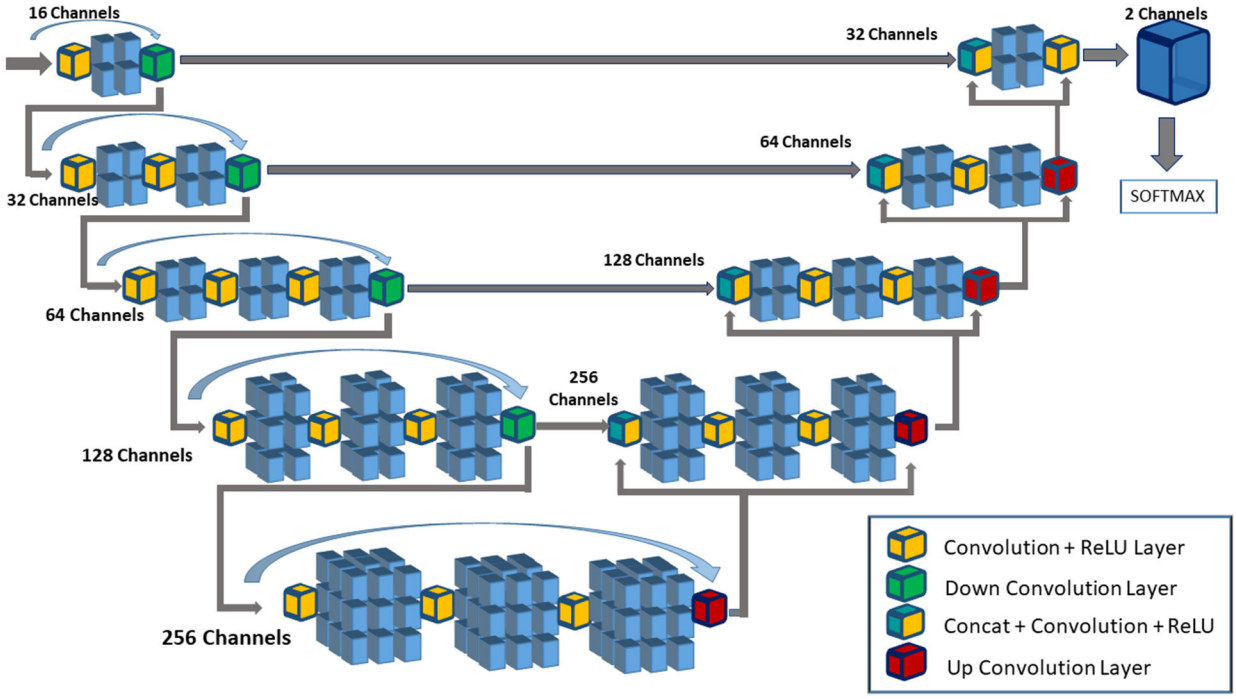

Fig. 3 *Segmentation network architecture. The input to the network has the same spatial dimensions of 16 channels, with each channel of size* $128 \times 128 \times 64$ *voxels*

© The Institution of Engineering and Technology 2020

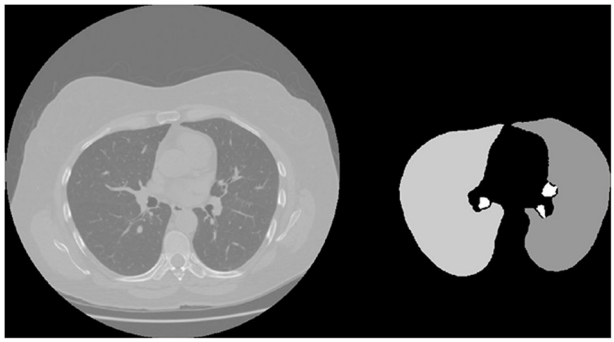

Fig. 4 *Segmentation results. Left image depicts the original scan. Right image depicts the image after segmentation for a patient. White portions are the nodules, grey are the lungs and black is the background*

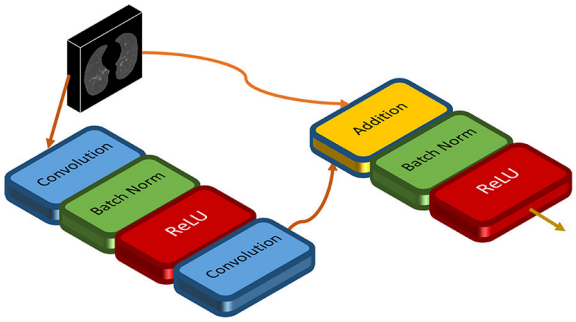

Fig. 5 *Depiction of different stages in ResNet architecture*

The network architecture in a 2D CNN is designed in a way that it reuses its weights in the ROI dimension. ResNet convolves the two spatial dimensions and reduces the input tensor of shape $(16, 128, 128)$ to an output tensor with binary classification 0 or 1. For training the architecture, the malignancy information of each patient has been used which is available in the LUNA16 dataset. This information is mapped to the respective input coming from the segmentation architecture. The penultimate layer is removed and features from morphological descriptors are added. Instead of using a softmax layer, a XGBoost classifier is used. The motivation behind classification with the 'max' function to form binary classification is that the 'cancerous' nodule also determines the cancer likelihood of the whole patient. This final output scalar will be the prediction for a patient. This setup is directly end-to-end trainable from 16 ROI per patient to a single prediction per patient.

4.4 Morphological descriptors

Understanding of morphological properties is essential for proper classification. Although ResNets have emerged as the most widely used techniques for classification yet, they still fail to learn domain-specific features or characteristics that can be very vital in classification. For extracting morphological features, a bounding box is generated around the image. The image is scanned from the beginning along the axis and tries to fit the smallest rectangle that contains the ROI or nodules. Four images are generated for each ROI by applying rotation, transpose, inverse on the original ROI matrix. Fig. 6 gives information about the morphological features that are calculated on each ROI. For the sake of brevity, the mathematical proof of the following descriptors is skipped. For details, the interested reader may refer to 'Practical Conic Sections: The Geometric Properties of Ellipses, Parabolas and Hyperbolas' [37] and 'The Ellipse: A Historical and Mathematical Journey' [38].

All the features from the patient are concatenated in order to form a feature vector consisting of 4×18 features. Finally, the feature vector (along with tuples) is flattened to form a feature vector of size 120 and padding up with 0, in cases where no nodule is found. The feature vector is passed on to XGBoost classifier.

ATTRIBUTES	DESCRIPTION
AREA	Number of pixels of the region.
CONVEX AREA	Number of pixels of convex hull image, which is the smallest convex polygon that encloses the region.
BOUNDING BOX COORDINATES	Coordinates tuple of bounding box (min row, min col, max row, max col)
ECCENTRICITY	Eccentricity is second-moments as the region. The eccentricity is the ratio of the focal distance (distance between focal points) over the major axis length. The value is in the interval [0, 1).
EQUIVALENT DIAMETER	The diameter of a circle with the same area as the region.
EULER NUMBER	Euler characteristic of region. Computed as: number of objects - number of holes.
EXTENT	Ratio of pixels in region to pixels in bounding box. Computed as : area / (rows * cols)
FILLED AREA	Number of pixels of the region will all the holes filled in. It describes the area of the filled image.
LOCAL CENTROID	Centroid coordinate tuple (row, col) relative to region bounding box
MEAN INTENSITY	Value with the mean intensity in the region.
MOMENTS	Spatial moments
MOMENTS HU	Hu moments (translation, scale and rotation invariant).
ORIENTATION	The angle between the x axis and the major axis of the ellipse that has the same second moments as the region, ranging from -π to π in counter clockwise direction.
PERIMETER	Perimeter of object which approximates the contour as a line through the centre of border pixels using a 4-connectivity.
SOLIDITY	Ratio of pixels in the region to pixels of the convex hull image
ASPECT RATIO	Ratio of major Axis to minor axis
INERTIA TENSOR REGION	Inertia tensor of the region for the rotation around its mass.
INERTIA TENSOR EIGEN VALUES	Eigen values of the inertia tensor

Fig. 6 *Description of morphological features used for classification*

4.5 Processing time

The pre-processing stage took 24 h, once the best techniques to be used were identified. The training of the segmentation architecture took six days on a Tesla K-80 GPU on Google Colab. Access was reset every 12 h as per their policy, hence weights were saved and then loaded when the next subset was used for training. The dataset, model and weights were stored on google drive, which was mounted every time a new session was started. Training of ResNet took two days, about 40 h on a GeForce GTX 760, while the application of morphological descriptors took 6–8 h for all the images.

5 Cancer classification

Features from two phases are fused together and passed onto a classifier. XGBoost is chosen as a classifier as it provides automatic handling of sparsity patterns. Sparsity can occur primarily because of two reasons: (i) presence of erroneous values in the data due to segmentation, and (ii) frequent zero entries. Studies [26, 39, 40] show XGBoost dominates classification of structured datasets.

However while training the classifier, it tend to become biased because of skewedness of data. So, synthetic minority over-sampling technique (SMOTE) [41] was used with the random state being set to 2, for handling data imbalance. SMOTE decreases the ratio of the number of samples in the majority class over the number of samples in the minority class. It is also important to have a balanced test dataset, otherwise it is possible that a model always giving deterministic output say '1' may result in high accuracy.

The test data is constituted of 60 patients. 30 patients of test data were cancerous and 30 patient were non-cancerous. The classifier was trained with number of estimators as 1000, and random state being set to 4242. Binary:logistic and logloss were chosen as objective function and evaluation metric, respectively.

6 Numerical results

Different researchers have used different parameters for exhibiting their performance. In this paper, some extensive parameters are

© The Institution of Engineering and Technology 2020

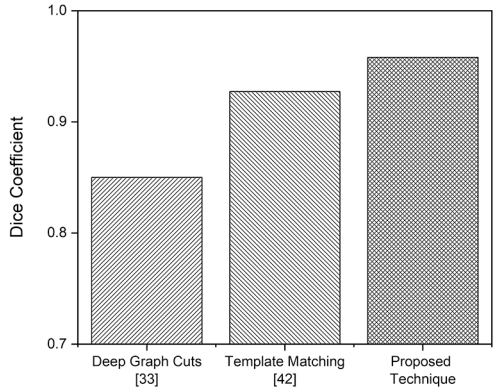

Fig. 7 *Comparion of Dice coefficient for the segmentation task. Ground truth for evaluation of Dice coefficient is provided in dataset itself*

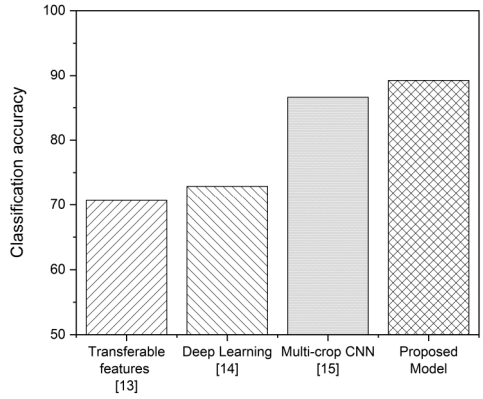

Fig. 8 *Comparion of number of morphological features that have been used in previous lung cancer classification model designs*

shown on which this model achieves a better performance to existing works in literature.

To highlight the performance of segmentation phase, the Dice coefficient metric is used. Dice coefficient measures how similar the segmentation results are to the ground truth. The ground truth is available in the dataset, that was used to find the dice.The segmentation architecture is trained on 800 instances. 80 patients scans were used as test data to measure the efficiency of this model. Fig. 7 depicts the comparison of different CAD models that have used Dice coefficient metric. This model achieves a dice coefficient of 0.958, significantly higher than deep graph cut [33] and template matching techniques [42] which achieve 0.75 and 0.927, respectively. The superior performance of the segmentation architecture is because of the volumetric convolutions it performs, thereby using all the information which is present in the CT scan, in contrast to the typical slice by slice approach which is taken by 2D networks. The latter results in loss of information while processing, as well as a lack of context between slides.

Fig. 8 highlights another major contribution of this paper, delineating the approach for obtaining morphological descriptors. Eighteen morphological features have been considered, significantly more than [20–22].

The confusion matrix (Fig. 9) was generated for the test dataset consisting of 60 patient instances. Test data comprises 30 positive instances and 30 negative instances. This model could identify 27/30 positive and 26/30 negative instances correctly.

Accuracy is used to evaluate the overall classification relevance as it provides a wider scope of comparison with other works. Table 1 depicts comparison of the proposed model with three other lung cancer classification model designs. It outperforms other techniques achieving an accuracy of 88.3%, owing to improved segmentation and feature extraction phases surpassing [43–46] models.

However, accuracy is not a very conclusive metric. In some cases where the dataset is imbalanced, a model which labels every single person as non-cancerous might result in a high accuracy. Out

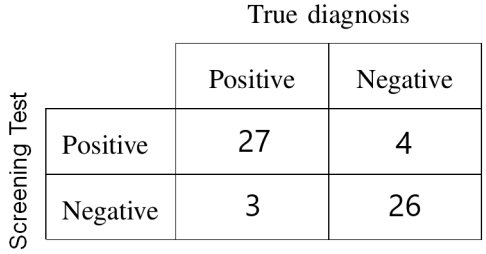

Fig. 9 *Confusion matrix*

Table 1 Accuracy table

Model	Accuracy	Enhanced technique
Alakwaa *et al*. [43]	0.866	segmentation
Qin *et al*. [44]	0.730	pre-processing
Sóñora-Mengana *et al*. [45]	0.760	segmentation
Bhatia *et al*. [46]	0.840	classification
proposed model	0.883	segmentation, classify

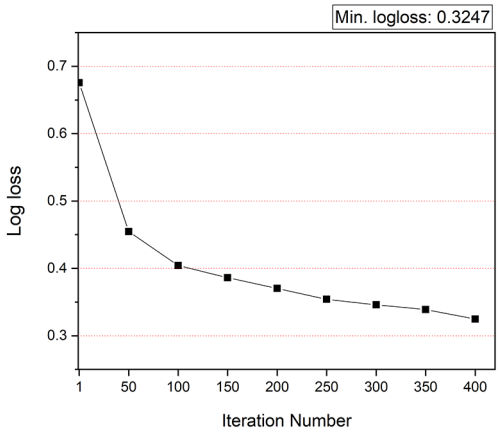

Fig. 10 *Logloss values for the proposed model on test data. XGBoost was trained to optimise the logloss. This curve shows how the logloss varies with number of iterations achieving a minimum value of 0.3247*

of the two classes which need to be identified, one category has an overwhelming majority over the other.

Logloss is a metric which measures the performance of a classification model used by Kaggle [47]. It takes variation from actual label as the degree of uncertainity giving a more nuanced view of model performance. The objective of models is to minimise this value. Fig. 10 gives logloss curve for XGBoost model. The minimum logloss achieved is 0.32.

Receiver operating characteristic (ROC) is one of the most trusted and widely used metric to assess the accuracy of radiological imaging systems. ROC calculated the confidence level that an abnormality is present or not. The area under the ROC curve (AUC) is used as a performance metric. Higher the AUC, better the model. The proposed architecture achieves an AUC of 0.883 excelling over [43, 48–50]. The comparison of AUC is shown in Fig. 11. In order to give a conclusive picture of the model, five different metrics are calculated, as shown in Table 2. Here, TP stands for true positive, TN for true negative, FP for false positive and FN for false negative.

Sensitivity, also commonly known as true positive rate, is the ratio of true positive cases to all the classified positive cases. Specificity, or true negative rate, is the ratio of true negative instances to total negative labelled instances. The proposed framework achieves a sensitivity and specificity of 0.8710 and 0.8966, respectively. While accuracy is related to how close an answer is to the actual value, precision relates the closeness of the two measurements to each other. False negative rate is also used to evaluate the performance. It is the proportion of positives which yield negative test outcomes. Another widely used metric that utilises both precision and recall is F1 score. It is the harmonic mean of precision and recall. The proposed CAD model achieves a

© The Institution of Engineering and Technology 2020

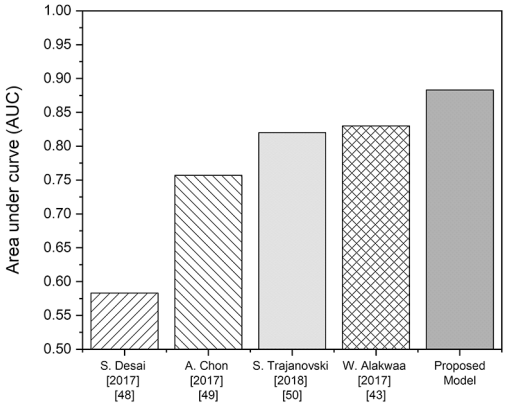

Fig. 11 *Comparion of area under ROC curve for different CAD models*

Table 2 Metric table

Measure	Value	Derivation
ROC area	0.88	$\frac{TP}{TP+FN}$
sensitivity	0.8710	$\frac{TP}{TP+FN}$
specificity	0.8966	$\frac{TN}{FP+TN}$
precision	0.9000	$\frac{TP}{TP+FP}$
accuracy	0.8833	$\frac{TP+TN}{TP+TN+FP+FN}$
F1 score	0.8852	$\frac{2*TP}{2*TP+FP+FN}$
false omission rate	0.103	$\frac{FN}{TN+FN}$
markedness	0.797	$\frac{TP}{TP+FP}-\frac{FN}{TN+FN}$

precision of 0.9000. Calculated F1 score is 0.8852, a significant improvement over 0.740 achieved by Kuan *et al.* [31]. A simple two-layer three-node neural network with threshold activation functions is NP-complete [51]. A general expression for the overall time complexity of all convolutional layers in a CNN was presented by He and Sun [52] and demonstrated by Sinha and Ajmera [53]. CNNs are composed of a lot of shared weights and biases, which increase the computational complexity drastically. But despite this high computational complexity, there are several modifications that enable faster and efficient learning like ReLU activation functions and regularisation. The running time depends heavily on the hardware configuration and the platform used for training the model. The space complexity is dependent on the size of the input image and the batch size, which is used for training for a particular model. All of the experiments in this paper are performed on Google Colab, on a Tesla K-80 GPU with 12 GB of VG-RAM. The code is written in python3 based on PyTorch.

7 Conclusion

This paper presents the Deep3DSCan framework for lung nodule segmentation and cancer classification. Classification of a patient as cancerous or does not involve the segmentation phase. We employ a skip network-based architecture that uses volumetric segmentation. Segmentation achieves a Dice coefficient of 0.958. The proposed framework achieves significantly better results than the existing deep learning model design. Deep learning techniques being domain agnostic and having limited interpretability, hamper the capability in diagnostic assistance. The proposed framework fuses handcrafted features to features extracted from the ResNet model, providing a better enhancement than each class of feature generation designs. Through results we showed the performance gains of the proposed model compared to the traditional schemes used.

8 Acknowledgments

This work was supported by the NVIDIA GPU Research Grant.

9 References

[1] Pakzad, R., Mohammadian-Hafshejani, A., Ghoncheh, M., *et al.*: 'The incidence and mortality of lung cancer and their relationship to development in asia', *Transl. Lung Cancer Res.*, 2015, **4**, (6), pp. 763

[2] Timmerman, R., Paulus, R., Galvin, J., *et al.*: 'Stereotactic body radiation therapy for inoperable early stage lung cancer', *J. Am. Med. Assoc.*, 2010, **303**, (11), pp. 1070–1076

[3] Torre, L.A., Bray, F., Siegel, R.L., *et al.*: 'Global cancer statistics, 2012', *CA: Cancer J. Clin.*, 2015, **65**, (2), pp. 87–108

[4] Rafiemanesh, H., Mehtarpour, M., Khani, F., *et al.*: 'Epidemiology, incidence and mortality of lung cancer and their relationship with the development index in the world', *J. Thoracic Disease*, 2016, **8**, (6), pp. 1094

[5] Malvezzi, M., Bosetti, C., Rosso, T., *et al.*: 'Lung cancer mortality in european men: trends and predictions', *Lung Cancer*, 2013, **80**, (2), pp. 138–145

[6] Vansteenkiste, J., Dooms, C., Mascaux, C., *et al.*: 'Screening and early detection of lung cancer', *Ann. Oncol.*, 2012, **23**, (suppl_10), pp. x320–x327

[7] Hu, H., He, H.D., Foley, W.D., *et al.*: 'Four multidetector-row helical ct: image quality and volume coverage speed', *Radiology*, 2000, **215**, (1), pp. 55–62

[8] El-Baz, A., Beache, G.M., Gimel'farb, G., *et al.*: 'Computer-aided diagnosis systems for lung cancer: challenges and methodologies', *Int. J. Biomed. Imaging*, 2013, **2013**, pp. 53–60

[9] Arimura, H., Katsuragawa, S., Suzuki, K., *et al.*: 'Computerized scheme for automated detection of lung nodules in low-dose computed tomography images for lung cancer screening1', *Acad. Radiol.*, 2004, **11**, (6), pp. 617–629

[10] Lee, Y., Hara, T., Fujita, H., *et al.*: 'Automated detection of pulmonary nodules in helical ct images based on an improved template-matching technique', *IEEE Trans. Med. Imaging*, 2001, **20**, (7), pp. 595–604

[11] Bhattacharyya, S.: 'A brief survey of color image preprocessing and segmentation techniques', *J. Pattern Recognit. Res.*, 2011, **1**, (1), pp. 120–129

[12] Beck, A., Teboulle, M.: 'A fast iterative shrinkage-thresholding algorithm for linear inverse problems', *SIAM J. Imaging Sci.*, 2009, **2**, (1), pp. 183–202

[13] Armato, S.G., Sensakovic, W.F.: 'Automated lung segmentation for thoracic ct: impact on computer-aided diagnosis1', *Acad. Radiol.*, 2004, **11**, (9), pp. 1011–1021

[14] Leader, J.K., Zheng, B., Rogers, R.M., *et al.*: 'Automated lung segmentation in x-ray computed tomography: development and evaluation of a heuristic threshold-based scheme1', *Acad. Radiol.*, 2003, **10**, (11), pp. 1224–1236

[15] Uzelaltinbulat, S., Ugur, B.: 'Lung tumor segmentation algorithm', *Procedia Comput. Sci.*, 2017, **120**, pp. 140–147

[16] Miah, M.B.A., Yousuf, M.A.: 'Detection of lung cancer from ct image using image processing and neural network'. Electrical Engineering and Information Communication Technology (ICEEICT) 2015 Int. Conf. on, Dhaka, Bangladesh, 2015, pp. 1–6

[17] Ronneberger, O., Fischer, P., Brox, T.: 'U-net: convolutional networks for biomedical image segmentation'. Int. Conf. on Medical image computing and computer-assisted intervention, Munich, Germany, 2015, pp. 234–241

[18] Wright, J., Yang, A.Y., Ganesh, A., *et al.*: 'Robust face recognition via sparse representation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, **31**, (2), pp. 210–227

[19] Liu, M., Zhang, D., Shen, D., *et al.*: 'Ensemble sparse classification of alzheimer's disease', *Neuro Image*, 2012, **60**, (2), pp. 1106–1116

[20] Abdullah, A.A., Shaharum, S.M.: 'Lung cancer cell classification method using artificial neural network', *Inf. Eng. Lett.*, 2012, **2**, (1), pp. 49–59

[21] Kuruvilla, J., Gunavathi, K.: 'Lung cancer classification using neural networks for ct images', *Comput. Methods Programs Biomed.*, 2014, **113**, (1), pp. 202–209

[22] da Silva-Sousa, J.R.F., Silva, A.C., de Paiva, A.C., *et al.*: 'Methodology for automatic detection of lung nodules in computerized tomography images', *Comput. Methods Programs Biomed.*, 2010, **98**, (1), pp. 1–14

[23] Kumar, D., Wong, A., Clausi, D.A.: 'Lung nodule classification using deep features in ct images'. Computer and Robot Vision (CRV) 2015 12th Conf. on, Halifax, Canada, 2015, pp. 133–138

[24] Hua, K.L., Hsu, C.H., Hidayati, S.C., *et al.*: 'Computer-aided classification of lung nodules on computed tomography images via deep learning technique', *OncoTargets Ther.*, 2015, **8**, 2015–2022

[25] Ciompi, F., Chung, K., Van-Riel, S.J., *et al.*: 'Towards automatic pulmonary nodule management in lung cancer screening with deep learning', *Sci. Rep.*, 2017, **7**, pp. 1–10

[26] van Ginneken, B., Setio, A.A., Jacobs, C., *et al.*: 'Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans'. Biomedical Imaging (ISBI), 2015 IEEE 12th Int. Symp. on, New York, NY, USA, 2015, pp. 286–289

[27] Sermanet, P., Eigen, D., Zhang, X., *et al.*: 'Overfeat: Integrated recognition, localization and detection using convolutional networks', arXiv preprint arXiv:13126229, 2013

[28] Grand.Challenge: 'Lung nodule analysis 2016', 2016, (accessed August 7, 2018). Available at https://luna16.grand-challenge.org/

[29] Armato III, S.G., McLennan, G., Bidaut, L., *et al.*: The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans', *Med. Phys.*, 2011, **38**, (2), pp. 915–931

[30] Milletari, F., Navab, N., Ahmadi, S.A.: 'V-net: fully convolutional neural networks for volumetric medical image segmentation'. 2016 Fourth Int. Conf. on 3D Vision (3DV), Stanford, CA, USA, 2016, pp. 565–571

[31] Kuan, K., Ravaut, M., Manek, G., *et al.*: 'Deep learning for lung cancer detection: Tackling the kaggle data science bowl 2017 challenge', arXiv preprint arXiv:170509435, 2017

© The Institution of Engineering and Technology 2020

[32] Kumar, S., Raman, S.: 'Lung nodule segmentation using 3-dimensional convolutional neural networks'. Soft Computing for Problem Solving, Vellore, India, 2020, pp. 585–596
[33] Wu, X., Zhong, Z., Buatti, J., *et al.*: 'Multi-scale segmentation using deep graph cuts: robust lung tumor delineation in mvcbct'. Biomedical Imaging (ISBI 2018), 2018 IEEE 15th Int. Symp. on, Washington DC, USA, 2018, pp. 514–518
[34] Krizhevsky, A., Sutskever, I., Hinton, G.E.: 'Imagenet classification with deep convolutional neural networks'. Advances in neural information processing systems, Lake Tahoe, USA, 2012, pp. 1097–1105
[35] He, K., Zhang, X., Ren, S., *et al.*: 'Deep residual learning for image recognition'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 770–778
[36] Pascanu, R., Mikolov, T., Bengio, Y.: 'On the difficulty of training recurrent neural networks'. Int. Conf. on Machine Learning, Atlanta, GA, USA, 2013, pp. 1310–1318
[37] Downs, J.: '*Practical conic sections: the geometric properties of ellipses, parabolas and hyperbolas*' (Courier Corporation, New York, USA, 2003)
[38] Mazer, A.: '*The ellipse: a historical and mathematical journey*' (John Wiley & Sons, New Jersey, USA, 2011)
[39] Farag, A., Ali, A., Graham, J., *et al.*: 'Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose ct scans of the chest'. 2011 IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro, Chicago, IL, USA, 2011, pp. 169–172
[40] Pignon, J.P., Arriagada, R., Ihde, D.C., *et al.*: 'A meta-analysis of thoracic radiotherapy for small-cell lung cancer', *New England J. Med.*, 1992, **327**, (23), pp. 1618–1624
[41] Chawla, N.V., Bowyer, K.W., Hall, L.O., *et al.*: 'Smote: synthetic minority over-sampling technique', *J. Artif. Intell. Res.*, 2002, **16**, pp. 321–357
[42] Yang, B., Xiang, D., Yu, F., *et al.*: 'Lung tumor segmentation based on multi-scale template matching and region growing'. Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging, Int. Society for Optics and Photonics, Houston, TX, USA, 2018, vol. 10578, p. 105782Q
[43] Alakwaa, W., Nassef, M., Badr, A.: 'Lung cancer detection and classification with 3d convolutional neural network (3d-cnn)', *Lung Cancer*, 2017, **8**, (8), pp. 409–417
[44] Qin, P., Chen, J., Zhang, K., *et al.*: 'Convolutional neural networks and hash learning for feature extraction and of fast retrieval of pulmonary nodules.', *Comput. Sci. Inf. Syst.*, 2018, **15**, (3), pp. 517–531
[45] Sóñora-Mengana, A., Papavasileiou, E., García-Naranjo, J., *et al.*: 'Evaluation of data balancing techniques. application to cad of lung nodules using the luna16 framework', *Revista Científica de Ingeniería Electrónica, Automática y Comunicaciones*, 2018, **39**, (3), pp. 57–67, ISSN: 1815-5928
[46] Bhatia, S., Sinha, Y., Goel, L.: 'Lung cancer detection: a deep learning approach'. Soft Computing for Problem Solving, Bhubaneshwar, India, 2019, pp. 699–705
[47] Kaggle.: 'Kaggle data bowl', 2017, (accessed August 7, 2018). Available at https://www.kaggle.com/c/data-science-bowl-2017/
[48] Desai, S.: 'Exploring 3d convolutional neural networks for lung cancer detection in ct volumes'
[49] Chon, A., Balachandra, N., Lu, P.: '*Deep convolutional neural networks for lung cancer detection*' (Standford University, Stanford, USA, 2017)
[50] Trajanovski, S., Mavroeidis, D., Swisher, C.L., *et al.*: 'Towards radiologist-level cancer risk assessment in ct lung screening using deep learning', 2018, arXiv preprint arXiv:180401901
[51] Blum, A., Rivest, R.L.: 'Training a 3-node neural network is np-complete'. Advances in Neural Information Processing Systems, Denver, CO, USA, 1989, pp. 494–501
[52] He, K., Sun, J.: 'Convolutional neural networks at constrained time cost'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015, pp. 5353–5360
[53] Sinha, H., Ajmera, P.K.: 'Upgrading security and protection in ear biometrics', *IET Biometrics*, 2019, **8**, (4), pp. 259–266

© The Institution of Engineering and Technology 2020