




# A generalized framework for lung Cancer classification based on deep generative models

Wessam M. Salama<sup>1</sup> · Ahmed Shokry<sup>2</sup> · Moustafa H. Aly<sup>3</sup> 

Received: 5 May 2021 / Revised: 15 February 2022 / Accepted: 28 March 2022 /

Published online: 14 April 2022

© The Author(s) 2022

## Abstract

A new generalized framework for lung cancer detection and classification are introduced in this paper. Specifically, two types of deep models are presented. The first model is a generative model to capture the distribution of the important features in a set of small class-unbalanced collected CXR images. This generative model can be utilized to synthesize any number of CXR images for each class. For example, our generative model can generate images with tumors with different sizes and positions in the lung. Hence, the system can automatically convert the small unbalanced collected dataset to a larger balanced one. The second model is the ResNet50 that is trained using the large balanced dataset for cancer classification into benign and malignant. The proposed framework acquires 98.91% overall detection accuracy, 98.85% area under curve (AUC), 98.46% sensitivity, 97.72% precision, 97.89% F1 score. The classifier takes 1.2334 s on average to classify a single image using a machine with 13GB RAM.

**Keywords** Lung cancer · Computer aided diagnosis · ResNet50 · Convolutional neural network · Variational auto-encoders

---

✉ Wessam M. Salama  
wessam.salama@pua.edu.eg

Ahmed Shokry  
ahmed.shokry@alexu.edu.eg

Moustafa H. Aly  
mosaly@aast.edu

<sup>1</sup> Department of Basic Science, Faculty of Engineering, Pharos University, Alexandria, Egypt

<sup>2</sup> Computer and Systems Engineering, Alexandria University, Alexandria, Egypt

<sup>3</sup> Department of Electronics and Communications Engineering, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt

## 1 Introduction

Lung cancer is usually considered as the largest cause of cancer-related fatalities across the world. Lung cancer has the greatest mortality rate of any kind of cancer. As a result, it is the world's most deadly kind of cancer. According to the World Health Organization, this epidemic claimed the lives of 1.59 million people in 2012 and 0.158 million people in 2015 [23]. The survival rate is governed by the fact that treatment can be started as soon as feasible in order to optimize survival and reduce death. Lung cancer is defined as a severe illness produced by abnormal cells that proliferate and ultimately form a tumor.

Deep learning has becoming increasingly popular in medical image analysis as artificial intelligence has advanced. Deep convolutional neural networks [24] give a feasible approach to extracting deeper and more powerful features. Advanced convolutional neural networks improve prediction accuracy as well.

Pneumonia is a seasonal respiratory lung illness that can be fatal in children under the age of five and elderly adults over the age of 60 [31]. In hospitals, pneumonia is diagnosed using a variety of imaging modalities, including a chest X-ray (CXR) and a computed tomography CT [1]. Though CXR, radiographs are the most cost-effective diagnostic technique for identifying pneumonia; their diagnosis needs highly educated radiologists since these images frequently interact with other clinical diseases of the lungs. Manual pneumonia identification is a time-consuming process that frequently leads in subjective variances, delaying diagnosis and treatment. Furthermore, the severity of the pneumonia infection might be obscured on X-Ray images.

Machine learning techniques are frequently employed in Computer Aided Diagnosis (CAD) systems [34] to identify and classify lung cancer. The feature extraction stage, in which the system learns the relevant characteristics in CXR pictures, is critical for these systems. [8]. The collected characteristics are then used as an input into a classifier model for lung cancer detection and classification [28]. Furthermore, with the assistance of large parallel architecture GPUs, these techniques have made significant progress.

Deep learning solutions can learn to generate a high-level feature directly from raw CXR images. Convolutional Neural Networks (CNN) are commonly used deep model in these systems [15]. It can automatically learn the representation of complex features directly from the data itself. CNN-based lung tumor detection systems have usually two-stage. An offline phase utilizes a deep CNN model that is trained using a set of classified CXR images (training data). An online phase, that takes a lung CXR image and determines whether it contains tumors or not. CNN-based systems, have been successfully applied to the problem of detection and classification of lung tumors. In addition, with the support of parallel GPUs, these technologies have gained tremendous success. On the other hand, CNN and its variants are not able to significantly improve performance. This is because CNN and deep learning models, in general, are data-hungry. In order to achieve the expected good performance, they require large amounts of training data (classified images).

The ResNet50 model is a typical deep model in deep learning-based systems [19]. ResNet50-based algorithms have been effectively used to the identification and categorization of CXR images [32]. Recently, a subclass of deep learning called as transfer learning has demonstrated promise in CXR image identification [35]. Transfer learning allows the employment of a pre-trained model that was previously designed for another comparable application. Several studies have been carried out to extract deep characteristics from CXR pictures using pre-trained networks. These studies illustrate the capacity of transfer learning to deal with

smaller datasets. Examples are AlexNet [5] and GoogLeNet [29] used in research work on the grading of glioma from CXR images [21].

In [7], the limits of the present methods for lung cancer detection and categorization are discussed. According to [7], the authors offer a comprehensive overview and analysis of the most important approaches for computer-aided identification of lung nodules. The authors compared current methods for detecting lung nodules by conducting literature searches with selection criteria based on validation dataset types, nodule sizes, number of cases, types of nodules, extracted features in traditional feature-based classifiers, sensitivity, and false positives (FP)/scans. According to the authors' analysis, existing detection methods are frequently tailored for certain datasets and can detect just one or two types of nodules. They observed that, in addition to high sensitivity and low FP/scans, methods for identifying lung nodules must detect a wide range of nodules with good accuracy in order to enhance radiologists' performance. Furthermore, they worked through deep-learning methods in depth and concluded that features that must be carefully chosen in order to enhance the overall accuracy of the system. Given the medical significance of the identification and classification problem, the performance of cutting-edge methods was inadequate. The automatic solutions developed using ResNet50 and its variants could not achieve an influential improvement in performance. This is because ResNet50 and deep learning models, in general, are data-hungry [11]; i.e. They need a significant volume of training data in order to produce the anticipated successful results (classified images). This comes with extra cost due to the need to perform the tedious CXR images collection and classification process.

Although transfer learning methods can alleviate the requirement for huge amounts of training data, the success of such solutions depends on the pre-trained model. If the required task (i.e., tumors detection or classification) and the task that the pre-trained model is trained for are too dissimilar, the accuracy will be poor using transfer learning [14]. Finally, existing solutions using either ResNet50 or other transfer learning models are trained using class-unbalanced datasets (with respect to number of training samples for each class), lowering their detection and classification accuracy [15, 17].

In [25], the authors offer a pathology simulation in pictures, using X-rays of the chest as a model medical picture. Moreover, a generative adversarial network (GAN) was utilized to generate artificial pictures from a small annotated dataset. The authors' used a mix of actual and artificial pictures to train a deep convolutional neural network (DCNN) to detect disease in five different types of chest X-rays. Furthermore, the authors showed that supplementing the original unbalanced dataset with GAN produced pictures enhanced the performance of the proposed DCNN in chest pathology classification when compared to the same DCNN trained with the original dataset alone.

Although GANs can learn a more general distribution for the input features than VAEs, they suffer from significant drawbacks which make it a bad choice given the medical significance of the classification problem. This can be more significant if the GAN is trained using small unbalanced data, like the collected available one. The drawbacks include the vanishing gradients, where the discriminator is too good to the point that generator training fails to learn the important features of the mode collapse, where GANs fail to produce a wide variety of the generated outputs, and finally GANs frequently fail to converge.

In this paper, we propose a novel framework that can be seamlessly integrated with any of the existing CXR based detection and classification systems by processing their small, unbalanced training dataset to produce a larger balanced dataset which is suitable for training deep learning models. In particular, we employ two types of deep learning models. The first

model is an innovative deep learning model to capture the distribution of the important features in a group of a small lass-unbalanced images dataset. Then, by using the distribution of the important features, the system can collect any number of CXR images for each class. Thus, the system can automatically convert a small unbalanced dataset into a larger balanced one. The second model is the classifier that is trained using the large lass-balanced dataset to detect lung tumors in CXR images. The generator model can provide a dense class-balanced dataset that is required for training the classifier model. This makes it having a large impact on many real-world applications that require dense class-balanced training data. Examples include lung tumor detection, classification, and grading. However, as the classifier model is trained using the newly generated data from the generator model, the accuracy of the classification (i.e. tumor detection) is affected by the accuracy of the generator model. This limitation can be solved if we carefully train the generator model until it reaches certain good accuracy.

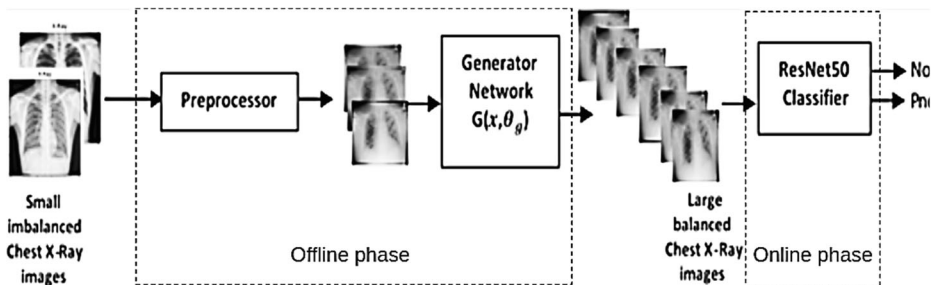
The rest of the paper is structured as follows. Section 2 defines the proposed structure as well as the particulars of the various versions. Section 3 describes the dataset and the proposed framework assessment. Section 4 displays and discusses the collected data. Section 5 is dedicated to the major findings.

## 2 Methodology

Figure 1 shows the proposed framework. As shown, small imbalanced and small size CXR lung images are fed into the preprocessor that resizes and normalizes the input CXR lung images. The framework contains two deep models: the generator model  $G(x, \theta_g)$  and the ResNet50 classifier, where  $x$  is the input to each model and  $\theta_g$  is the generator network parameters. The generator learns the distribution of the important features for each class in the preprocessed images, then, it converts the small unbalanced preprocessed images dataset to a larger balanced one by generating a large number of samples for each class. Finally, the large balanced dataset is used to train another deep model which is ResNet50 classifier.

### 2.1 Input data

The input to the system is a set of CXR lung images [4]. This set is small in size and imbalanced with respect to number of training images per class. The CXR lung images are in the standardized medical image format [3]. These images are stored as a two dimensional (2D)



**Fig. 1** The proposed framework. The framework works in two phases: Offline phase for pre-processing the input images and training the generator network to generate more images for each class. Online phase for training the ResNet50 classifier to classify the images to Normal and Pneumonia classes

grayscale images, which allows us to store each CXR images as a matrix. Each matrix entry stores values ranging from 0 to 255. This set represents a compromise between the simplicity of storing image information (256 values fit correctly in 1 byte) and the flexibility of the human eye (humans distinguish a limited number of shades of the same color). The grayscale images are then passed through the preprocessor.

## 2.2 Preprocessor

This module aims to resize and normalize the input CXR images. Firstly, to make all the input images have a fixed size, we resize the grayscale input images to be images of size  $224 \times 224$  pixels. This allows these images to be fed into deep classifier model with a fixed input size. Secondly, normalizing the input data generally speeds up learning and leads to faster convergence. To this end, we normalize the grayscale input images to be in the range 0 to 1. The images are normalized in intensity values. A minimum-maximum normalization technique is followed to scale the intensity values between 0 and 1. These preprocessed images are then passed through the generator network.

## 2.3 Generator

The goal of the generative model is to synthesize new CXR lung image samples for each class. The generative model is trained using the preprocessed images. Then, it can generate new CXR lung image samples that reflect the typical patterns in the preprocessed images. Our generative model is a modified version of the variational auto-encoder [12]. In general, the auto-encoders are specific types of feed forward neural networks, where the input is the same as the output [6]. They are used to learn key representational information for a dataset in an unsupervised manner within a low-dimensional latent space by compressing the input into a lower-dimensional latent-space representation,  $z$ , (embedding), and then reconstructing the performance from this representation. As a result, the latent space learns to collect the most critical knowledge required for reconstruction. However, the latent space embedding may be sparsely distributed that makes the key information to spread across several clusters in the latent space. Meanwhile, the empty space between clusters provides no valuable details, so sampling from it yields useless data. A new restriction is applied on variational auto-encoders (VAEs). The latent space embeddings need to follow certain predefined distribution  $p(z)$  [16]. This distribution is usually selected as normal distribution [22]. The network is thus forced to completely use the latent space by requiring the latent space embedding to obey the usual distribution. As a result, information is spread in such a way that any point in the latent space can be sampled to produce new images that represent the usual trends in the original small CXR lung images dataset. Therefore, we depend on VAE to generate new CXR lung images. Instead of using regular feed forward layers in our VAE, we apply convolutional layers as they may use sliding filter maps to recognize local patterns (e.g., tumors) regardless of their location in the lung image; convolutional layers, on the other hand, are better suited to recognize local patterns independent of their position than feed forward networks. Table 1 explains the generative model parameters.

We would like to note that the generator and the network parameters are determined experimentally. Tables 1 and 2 discuss the parameters and the hyper-parameters used in the experiment of both generator and classifier models. The learning rate controls how fast the network parameters (i.e. the weights) are updated. The batch size controls the number of

**Table 1** Generator network parameters

Parameter	Default value
Learning rate	0.0001
Batch size	16
Dropout rate	0.1
Number of training epochs	2000
Size of input layer	256 × 256
Size of output layer	256 × 256
Optimization technique	SGD

samples before the network updates its parameters. The dropout rate determines the percentage of parameters that are dropped while training to prevent the over-fitting. The number of epochs controls the number of training iterations.

Figure 2 shows our convolutional variational auto-encoder (CVAE) architecture. The CVAE network has two main components: the encoder (conv) and the decoder (de-conv). The encoder network consists of several convolutional layers followed by a fully connected layer, while the decoder networks comprises of a fully connected layer followed by convolutional layers.

The encoder compresses the CXR image input ( $x \in \mathcal{X}$ ) to obtain the hidden (latent) representation,  $z$ , and network parameters. Since the encoder must acquire an effective encoding of the image data into a lower-dimensional area, the latent space,  $z$ , is commonly referred to as a bottleneck. The encoder is known as  $Q_{\phi}(z/x)$ . This distribution can be sampled in order to obtain noisy values for the representations  $z$ . The input image of size  $224 \times 224$  is passed through the encoder network,  $Q_{\phi}(z/x)$ , which has two convolutional and max pooling layers to get the most important characteristics from the input image,  $x$ . The output from the encoder is transformed to one vector (i.e. flattened), then passed through a fully connected layer (FC) to get the mean,  $\mu$ , and the standard deviation,  $\Sigma$ , of the encoder distribution  $Q_{\phi}(z/x)$ . We then sample  $z$  from the  $Q_{\phi}(z/x)$  and pass it through the decoder network,  $P_{\theta}(x/z)$ , with two de-convolutional and up-sampling (i.e. nearest neighbors) layers to reconstruct the image  $\tilde{x}$ . We

**Table 2** Summary of the CNN layers for the ResNet50 architecture

Layer No.	Type	Size	Kernel/Stride	Activation
	Input	$224 \times 224 \times 3$	–	–
1	Convolution	$64 \times 3 \times 3$	2	ReLU
2	Convolution	$1 \times 1 \times 64$		
	Convolution	$3 \times 3 \times 64$	2	ReLU
	Convolution	$1 \times 1 \times 256$		
3	Convolution	$1 \times 1 \times 128$		
	Convolution	$3 \times 3 \times 128$	2	ReLU
	Convolution	$1 \times 1 \times 512$		
4	Convolution	$1 \times 1 \times 256$		
	Convolution	$3 \times 3 \times 256$	2	ReLU
	Convolution	$1 \times 1 \times 1024$		
5	Convolution	$1 \times 1 \times 512$		
	Convolution	$3 \times 3 \times 512$	2	ReLU
	Convolution	$1 \times 1 \times 2048$		
	Output	2	–	SVM

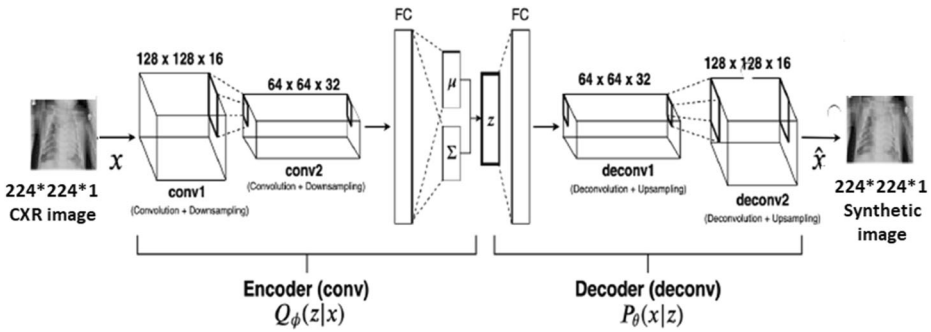


Fig. 2 The generative network structure

force the generative network to reconstruct the input images while making  $z$  follows the normal distribution.

The decoder takes the latent representation,  $z$ , as input and generates parameters for the data’s probability distribution, as well as weights of prejudice  $\theta$ . The decoder is denoted by  $P_{\theta}(x|z)$ . The loss function is the negative log-likelihood function defined as follows [20].

$$L(\varphi, \theta) = L1 + L2 \tag{1}$$

$$L1 = -E_{z \sim Q_{\varphi}(z|x)} [\log P_{\theta}(x|z)] \tag{2}$$

$$L2 = KL(Q_{\varphi}(z|x) \parallel p(z)) \tag{3}$$

The loss is made up of two terms. Next, consider the reconstruction loss, which is the data’s predicted negative log-likelihood. This phrase compels the decoder to learn how to decrypt the data. The regularizer is the second word. The divergence between the encoder’s distribution  $Q_{\varphi}(z|x_i)$  and the predefined distribution  $p$  is the Kullback-Leibler [2] divergence ( $z$ ). Assuming that  $p(z) = N(0, 1)$ , the regularizer makes the latent representations,  $z$ , adhere to the usual normal distribution. To optimise the loss in relation to CVAE network parameters  $\varphi$  and  $\theta$ , we use stochastic gradient descent. We employ the convolutional variational auto-encoder (CVAE) models for the different classes to learn the joint distribution  $P(X_i)$  of input features over the small training images  $X_i$  for each class  $i$ . After finishing the training phase, the network can generate new lung images by sampling the latent variables  $z \sim N(0, 1)$ , then decode  $z$  to get a new lung image sample  $\hat{x}$ . So, one can convert a small imbalanced dataset to a larger balanced one. The large dataset generated by the generator network is then fed into the classifier.

Initially, at iteration 100 from the training of generator network, the generator produces vague lung images. As more one increases training iterations, as more one gets clearer CXR lung images that reflect the typical patterns in the CXR lung images.

### 2.4 Classifier model

The ResNet50 plays an important role in the classification filed. This deep learning model is pre-trained on the ImageNet [10] dataset (1.3 million natural images) with

1000 classes. The patch size for these images is  $224 \times 44$  pixels, having three color channels. The CXR lung images will be concatenated along its third dimension to completely leverage these deep models in this work. Moreover, instead of 1000 classes, the final classification layer is replaced by a new layer for the classification of two classes: normal and pneumonia. Then, certain parameters must be set in order to begin fine-tuning the CXR lung dataset. Firstly, the number of iterations and the learning rate are adjusted to  $10^6$  and  $10^{-5}$ , respectively, in the ResNet50 classifier. The epoch number is 50, and it is found that the momentum should be 0.8, and the weight decay is set to  $5 \times 10^{-4}$ . These settings are used to ensure that the conditions for medical lung cancer diagnosis are fine-tuned.

### 3 System performance

A classifier can be analyzed using a number of machine assessment metrics such as the accuracy, AUC, the precision, and the F1 score.

**Accuracy** The classifier's accuracy is the percentage of accurate predictions it makes and is represented as [27].

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP} \quad (4)$$

where  $TP$ ,  $FN$ ,  $TN$ , and  $FP$  denote True Positive, False Negative, True Negative, and False Positive, respectively.

**Precision** Precision is defined as the proportion of correctly predicted positive observations to the total number of predicted positive observations [27].

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

**F1 score** The F1 score is determined by taking the weighted average of precision and recall. It is used as a statistical tool to evaluate the classifier's accuracy. As a result, this score finds all false positives and false negatives [27].

$$F1 = \frac{2 (Precision \times Sensitivity)}{(Precision + Sensitivity)} \quad (6)$$

**Area under the ROC curve (AUC)** In the medical diagnostic scheme, the AUC is used, which describes a method for analyzing models depending on the average of each point on the ROC curve. A model with a higher AUC value performs best as a classifier.

**Receiver operating characteristic (ROC)** The true positive rate (TPR) as a function of the false positive rate (FPR) is known as ROC curve. The sensitivity (recall) is defined as TPR and is represented as [27].



$$Sensitivity = \frac{TP}{TP + FN} \quad (7)$$

## 4 Results and discussion

The CXR is a critical instrument used to measure lung disorders in order to show the efficacy of the proposed system. With a Softmax classifier, the ResNet50 classifier is used to identify the selected radiograph images as benign or malignant.

The performance of the lung tumor classification framework depends on a combination of generator model parameters and classifier model parameters. There are three distinct framework settings.

### 4.1 Classifier model

The deep convolutional neural network is used as a classifier model. The network is trained using the small training dataset (after pre-processing). The network architecture is described in Sec. 2.4. Table 1 summarizes the classifier parameters.

### 4.2 Transfer learning

A pre-trained modified VGG network followed by a dense layer with 256 neurons and the output layer with two neurons is trained using the small training dataset (after pre-processing). The RMSprop is used as an optimizer with a learning rate of 0.0004. The loss function is categorical cross-entropy. The dropout rate is chosen to be 0.5.

### 4.3 Proposed framework

Our proposed framework includes two networks: the generator network and the classifier. The preprocessed images are used to train the generator network described in Sec. 2.3. With increasing the training time (i.e. number of training iterations), the network generates clearer images that eventually reflect the typical pattern of the lung CXR images. We used the generator network to convert the original small training dataset to a larger one with 1000 samples (500 for each class). The newly generated large dataset is then used for training the classifier network for detection. The classifier network is described in Sec. 2.4. We tried different architectures for the generator and the classifier networks. Tables 1 and 2 contain the generator and the classifier best parameters in terms of the detection accuracy.

**Table 3** Distribution of CXR images

CXR Database: Images (Normal 1585 and Pneumonia 4275)			
Age Range of patients (Years)			
30–40	<b>41–50</b>	<b>51–60</b>	<b>61–70</b>
1120	1460	1830	865
Patients Gender			
Male			<b>Female</b>
3984			1876

**Table 4** The suggested frameworks' average classification outcomes for k-fold cross-validation using the CXR database

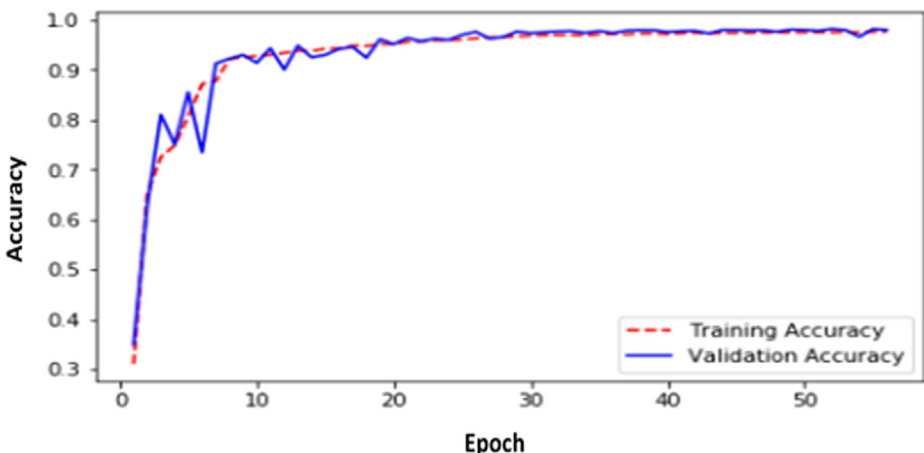
Model	value of k-fold	Accuracy	AUC	Sensitivity	Precision	Specificity	F1 score
Variational auto encoder+ ResNet50	1	98.88	99.73	98.87	97.89	97.35	97.88
	2	98.98	99.67	97.79	97.97	97.68	97.99
	3	98.89	97.86	98.86	96.88	96.78	96.86
	4	98.79	98.48	97.94	97.86	97.88	97.96
	5	98.97	98.54	98.88	97.99	97.65	98.76
	<b>Average Results</b>	<b>98.91% ±0.22</b>	<b>98.85% ±0.29</b>	<b>98.46% ±0.31</b>	<b>97.72% ±0.26</b>	<b>97.46% ±0.29</b>	<b>97.89% ±0.28</b>
ResNet50		<b>95.35% ±0.38</b>	<b>95.85% ±0.43</b>	<b>94.99% ±0.21</b>	<b>95.65% ±0.32</b>	<b>95.88% ±0.22</b>	<b>95.78% ±0.38</b>
CNN	Average Results	92.88% ±0.18	91.98% ±0.25	92.68% ±0.13	92.67% ±0.14	91.99% ±0.23	92.87% ±0.24
Transfer learning	Average Results	94.35% ±0.35	93.99% ±0.32	94.21% ±0.29	94.65% ±0.28	94.85% ±0.18	94.34% ±0.37

**Table 5** A comparison of many classification models based on various CNN architectures and datasets, as well as our proposed model

Reference	Data set	Accuracy (%)	Sensitivity	Specificity
[25]	Chest X-Ray	92.10	–	–
[13]	LIDC-IDRI	94.75	94.7	95.1
[33]	LIDC-IDRI	94.6	94.8	94.3
[9]	LIDC-IDRI	93	89	–
[18]	LIDC-IDRI	84.2	84.3	84
[30]	LIDC-IDRI	95	93.5	90.2
Our proposed framework	<b>Chest X-Ray</b>	<b>98.91</b>	<b>98.46</b>	<b>98.89</b>

Finally, the proposed framework consists of two different models for two different functions. The generator model target is to generate more data samples for each class in a small class-unbalanced data. On the other hand, the classifier model is used only for classification. The generator model needs to be trained first to generate the synthetic samples. The original samples and the synthetic ones are then used for training the classifier. Hence, the two networks cannot be trained in an end-to-end manner. The generator network can be employed dependently from the classifier network. It can be used to generate more data samples for any image-based machine learning problem that needs a dense balanced dataset for training. The classifier, on the other hand, is used for any image-based classification. Hence, our framework is general for different detection and classification problems as one can replace the classifier of the lung tumor detection with another one that handles another problem (e.g. grading the tumors).

Table 3 presents the details of data set distribution. The k-fold cross-validation procedure is used to verify the diagnosis outcomes [26]. In k-fold cross-validation, data is randomly divided into k groups of roughly equal size, and each group is used once for testing and (k-1) times for training. In our work, k-fold values ranging from 1 to 5 are used, and the average classification results for all k-fold value and standard deviation of the ResNet50 and variational auto encoder with ResNet50 using CXR database are shown in Table 4. Table 4 summarizes the performance metrics such as Accuracy, AUC, Sensitivity, Precision, Specificity and F1-score for the different approaches. The results reveal that the different approaches can classify the positive

**Fig. 3** Our proposed framework accuracy

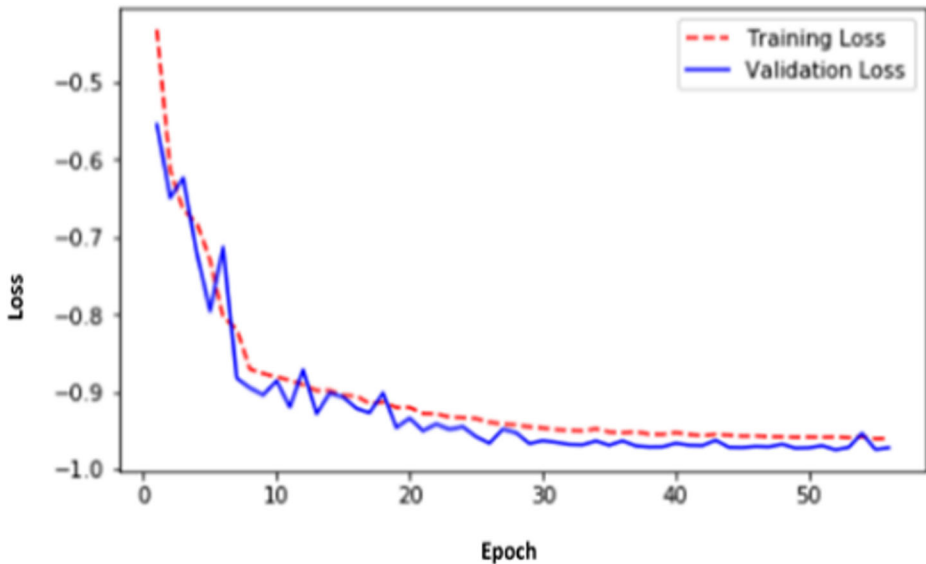


Fig. 4 Our proposed framework loss

samples (i.e. the samples that have a tumor) as good as the negative samples. Finally, the table shows that our proposed framework can detect tumors better than other approaches. This is because of the ability of the generative model to augment the classifier with newly generated samples which highlights the promise of the proposed framework as an accurate low-overhead lung tumor detection system.

For example, in 5-fold cross-validation, data is randomly divided into five classes, with one group serving as a testing set and the other four serving as a training set. Averaging the classification outcomes of the five rounds yields the 5-fold average classification accuracy, sensitivity, precision, and AUC. This approach ensures the validity of the current data. The proposed framework acquired 98.91% overall detection accuracy, 98.85% AUC, 98.46%

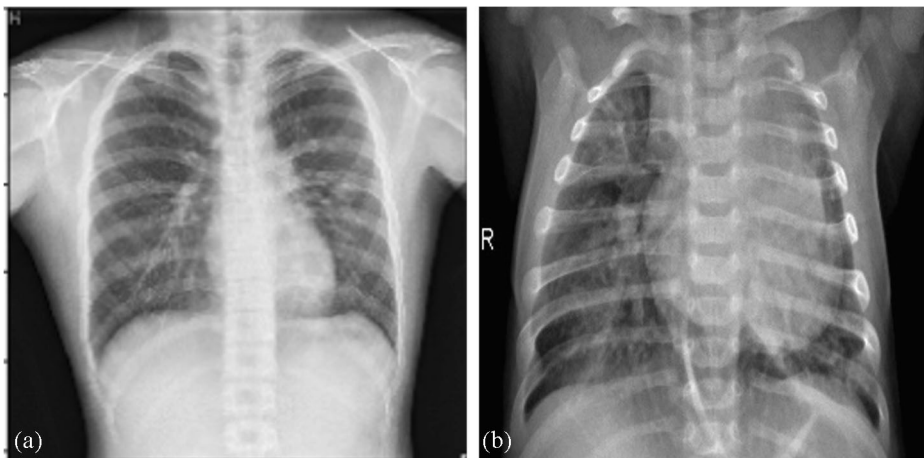
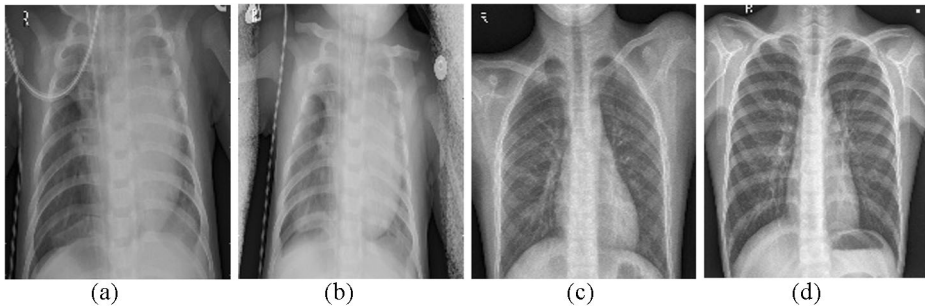


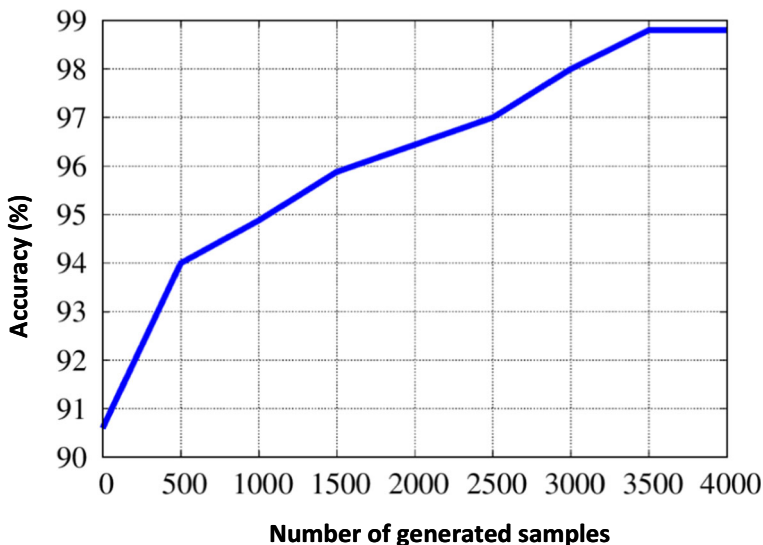
Fig. 5 CXR dataset used for testing: (a) Normal, (b) Pneumonia



**Fig. 6** An example of CXR dataset classification: (a) True Positive, (b) True Negative, (c) False Negative, (d) False Positive

sensitivity, 97.72% precision, 97.89% F1 score and 1.2334 s computational time on a Kaggle Notebook GPU cloud (2 CPU cores and 13 GB RAM) using Python software outperforming state-of-the-art approaches which highlight the promise of the proposed framework. Table 5 explains a comparison between different classification methods based on different Convolutional Neural Networks (CNN) architectures and datasets and our newly proposed models.

When compared to other approaches, the suggested system achieves the best accuracy (98.91%). Furthermore, the suggested model has a higher AUC than the other structures found in the literature. The relation in Table 4 demonstrates that the planned work outperforms other current schemes. Figures 3 and 4 demonstrate the efficiency of the proposed models by contrasting the consistency and loss between experiments and the training CXR lung dataset, indicating that the training and validation datasets are adequately representative. These graphs are known as learning curves, and are widely used as diagnostic methods for our models that learn incrementally from a training dataset. With each modification during testing, the model can be tested on the training dataset and a hold out validation dataset, and graphs of the assessed results can be generated to display learning curves. These curves demonstrate that the



**Fig. 7** Effect of number of generated samples on the accuracy

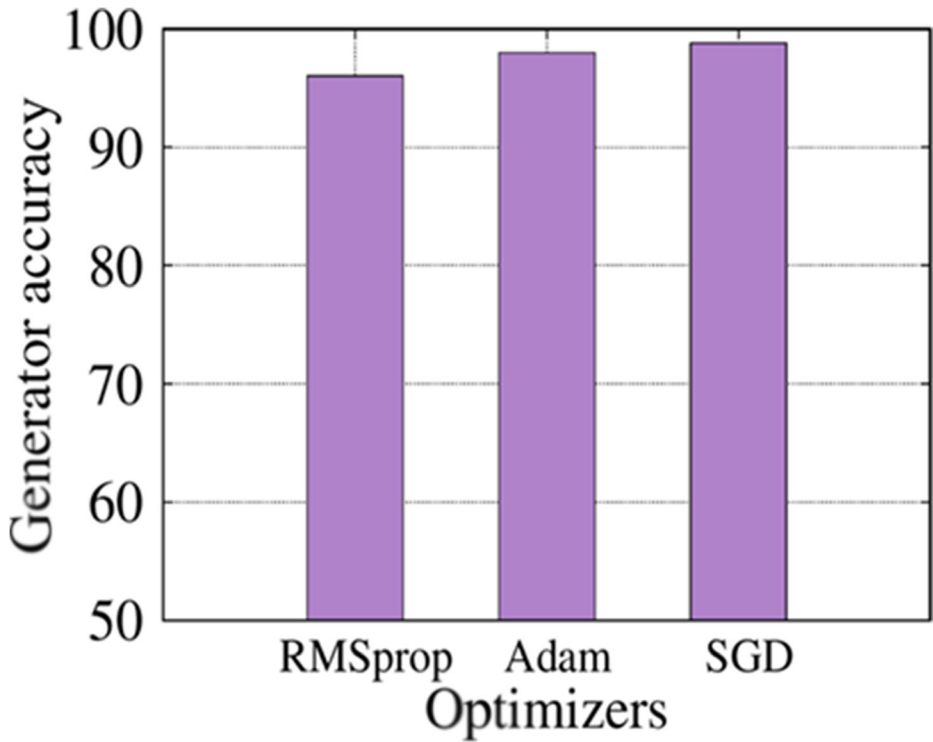


Fig. 8 Different optimizers of our proposed algorithm (RMSprop, Adam and SGD)

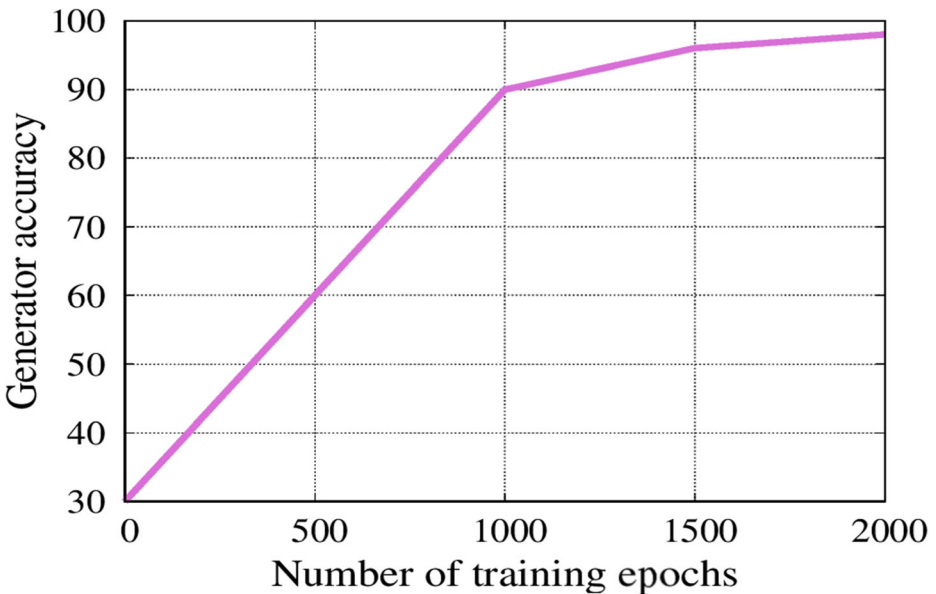
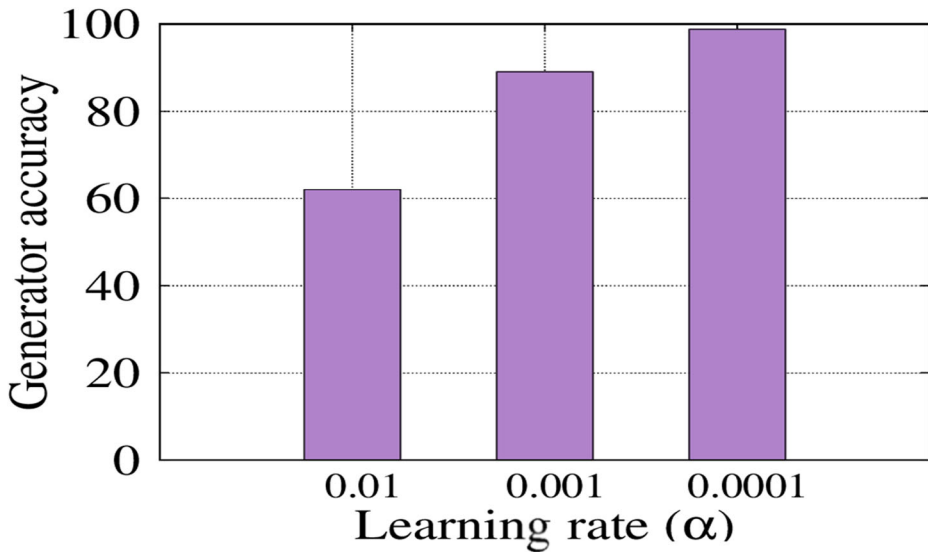


Fig. 9 Effect of number of training epochs (iterations) on the generator training accuracy



**Fig. 10** Effect of choosing the learning rate on the generator accuracy

proposed models greatly improve precision. Figure 5 illustrates image extracts from the research CXR dataset. Furthermore, Fig. 6 displays a classifier visualization on four patients. The classifier properly diagnoses patients A and B. The classifier wrongly diagnoses patients C and D. Both figures demonstrate that the efficiency of our model rapidly increases and stabilizes to greater than 97% accuracy. This paper demonstrates that deep learning models can accurately classify XCR lung pictures. Figure 7 shows the effect of increasing the number of additional generated training samples on the system accuracy. The figure shows that, as expected, increasing number of training samples increases the system accuracy due to the increase of the quality of the learned model. The good news is that the system performance saturates after about 4000 samples.

#### 4.4 Discussion and limitations

Given the heuristic nature of deep learning models, one cannot generalize the model parameters on the different lung cancer input datasets. However, the different lung cancer datasets share the same input features and class labels. These input features and target labels are first turned into some initial input vector space and target vector space. Each layer, in the deep learning model, operates one simple transformation on the data that goes through it. The model layers perform a complex nonlinear transformation, and is broken down into a series of simple ones. This complex transformation attempts to map the input space to the target space. This transformation is parametrized by the weights of the layers, which are iteratively updated based on how well the model is currently performing. The key property that makes the model converges is that it must have a differentiable transformation, which is required in order to be able to learn the model parameters via gradient descent. Initially, the best model parameters are unknown. Then, theoretically, by minimizing the cost function in Eq. (1), you will find the best parameter values. The way to do this is to feed a training dataset into the model and adjust the parameters iteratively to make the cost function as small as possible.

To this end, we experiment with the different optimizers (e.g. RMSprop, Adam and SGD) to minimize the negative log-likelihood loss in Eq. (1) for the first 2000 iterations as shown in Fig. 8. The figure shows that the SGD optimizer performs better than RMSprop and Adam. Figure 9, shows the effect of the number of training epochs (iterations) on the generator training accuracy. The figure shows also that the model accuracy saturates after 2000 iterations. Figure 10, further, displays the effect of choosing the learning rate on the generator accuracy. The generator model archives the best accuracy at 0.0001.

## 5 Conclusion

In this paper, a novel generalized framework for lung cancer detection and classification is introduced. The proposed framework used two different deep models in two different phases. The first was a convolutional variational generative Auto-encoder model. It worked offline to convert a small class-unbalanced dataset to a larger balanced one. The second was the ResNet50 classifier which was trained in the offline phase using the large balanced dataset to classify CXR images into benign and malignant in the online phase. The accuracy, sensitivity, precision, AUC, F1 score and computational time are utilized to judge our system performance. A retraining scenario is introduced, where the training is initialized by weights of a network that has already been trained using another dataset. This work is very helpful and proves that there is no need for human interface with pre or post processing or hand-crafted features. The robustness of the proposed framework is investigated by evaluating the learning curves. Also, the proposed systems are compared with other recent systems, and the obtained results show the superior performance of the proposed systems. The ResNet50 classifier achieves the best performance: 98.91% accuracy, 98.85% AUC, 98.46% sensitivity, 97.72% precision and 97.89% F1 score and computational time 1.2334 s. For future work, we are extending our work to address different detection and classification tasks such as brain tumor detection and classification. Moreover, we are modifying our framework to use other generative models such as generative adversarial network.

**Funding** Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

## Declarations

**Conflict of interest** The authors have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



## References

1. Abhir B, Prabhu GA, Rajinikanth V, Thanaraj KP, Satapathy SC, Robbins DE, Shasky C, Zhang YD, Tavares JM, Raja NS (2020) Deep-learning framework to detect lung abnormality—a study with chest X-ray and lung CT scan images. *Pattern Recognit Lett* 129:271–278
2. Alexander M, Hensman J, Turner R, Ghahraman Z (2016) On sparse variational methods and the Kullback-Leibler divergence between stochastic processes in proceedings of the 19<sup>th</sup> international conference on artificial intelligence and statistics (AISTATS) Cadiz, Spain: 231–239.
3. Ali A, Mofrad FB, Pouladian M (2018) Inter-patient modelling of 2d lung variations from chest x-ray imaging via Fourier descriptors. *J Med Syst* 42(11):233
4. <http://www.via.cornell.edu/lungdb.html>. Accessed April 2020.
5. Bhandary A, Prabhu GA, Rajinikanth V, Thanaraj KP, Satapathy SC, Robbins DE, Shasky C, Zhang YD, Tavares JMR, Raja NSM (2020) Deep-learning framework to detect lung abnormality—a study with chest X-ray and lung CT scan images. *Pattern Recogn Lett* 129:271–278
6. Bilwaj G, Hovda D, Martin N, Macyszyn L (2016) Deep learning in the small sample size setting: cascaded feed forward neural networks for medical image segmentation. In: *Medical Imaging Computer-Aided Diagnosis International Society for Optics and Photonics*: 9785 978521
7. Furqan S, Raja G, Frangi AF (2019) Computer-aided detection of lung nodules: a review. *J Med Imaging* 6(2):020901
8. Geert L, Kooi T, Bejnordi BE, Setio A, Ciompi F, Ghafoorian M, Der J, Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis *42*: 60–88
9. Guanghui H, Liu X, Zheng G, Wang M, Huang S (2018) Automatic recognition of 3D GGO CT imaging signs through the fusion of hybrid resampling and layer-wise fine-tuning CNNs. *Med Biol Eng Comput* 56(12):2201–2212
10. He K, Girshick R, Dollár P (2019) Rethinking image net pre-training. In: *Proceedings of the IEEE international conference on computer vision*: 4918–4927
11. Jae CM, Chung MJ, Lee JH, Lee KS (2019) Performance of deep learning model in detecting operable lung cancer with chest radiographs. *J Thorac Imaging* 34(2):86–91
12. Joo W, Lee W, Park S, Moon IC (2020) Dirichlet variational autoencoder. *Pattern Recogn* 107:107514
13. Jose G, Skaria S, Varun VV (2018) Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans in medical imaging computer-aided diagnosis: 10575 105751I. *International Society for Optics and Photonics*
14. Li X, Yves G, Davoine F (2020) A baseline regularization scheme for transfer learning with convolutional neural networks. *Pattern Recogn* 98:107049
15. Lo S-CB, Chan H-P, Lin J-S, Li H, Freedman MT, Mun SK Artificial convolution neural network for medical image pattern recognition. *Neural Netw* 8(7–8):1201–1214
16. Lopez R, Regier J, Jordan MI, Yosef N (2018) Information constraints on auto-encoding variational bayes. In: *Advances in Neural Information Processing Systems*: 6114–6125
17. Loris N, Ghidoni S, Brahmam S (2017) Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recogn* 71:158–172
18. Marie PL, Nielsen MB, Lauridsen CA (2019) Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the lide-idri database: a systematic review. *Diagnostics* 9(1): 64–69
19. Masafumi Y, Kasagi A, Tabuchi A, Honda T, Miwa M, Fukumoto N, Tabaru T, Ike A, Nakashima K (2019) Yet another accelerated sgd: resnet-50 training on imagenet in 74.7 seconds, arXiv preprint arXiv: 1903.12650
20. Masataro A, Set Cross Entropy (2018) Likelihood-based permutation invariant loss function for probability distributions, arXiv preprint arXiv:1812.01217
21. Paras L, Sundaram B (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 284(2):574–582
22. Ping G, Chen J, Pan T, Pan J (2019) Degradation feature extraction using multi-source monitoring data via logarithmic normal distribution based variational auto-encoder. *Comput Ind* 109:72–82
23. Priyanshu T, Tyagi S, Nath M (2019) A comparative analysis of segmentation techniques for lung cancer detection. *Pattern Recognit Image Anal* 29(1):167–173
24. Rahul P, Hawkins SH, Balagurunathan Y, Schabath MB, Gillies RJ, Hall LO, Goldg of D.B (2016) Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. *Tomography* 2(4):388–395
25. Salehinejad H, Valaee S, Dowdell T, Colak E, Barfett J (2018) Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks In 2018 IEEE international

- conference on acoustics, speech and signal processing (ICASSP), April 2018, Calgary, AB, Canada, IEEE p 990–994
26. Stone M (1974) Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc Ser B Methodol* 36:111–147
  27. Takaya S, Rehmsmeier M (2017) Pre-rec: fast and accurate precision–recall and ROC Curve Calculations in R. *Bioinformatics* 33(1):145–147
  28. Tanzila S, Sameh A, Khan F, Shad SA, Sharif M (2018) Lung nodule detection based on ensemble of hand crafted and deep features. *J Med Syst* 43(12):332
  29. Tiantian F (2018) A novel computer-aided lung cancer detection method based on transfer learning from googlenet and median intensity projections. In: 2018 IEEE international conference on computer and communication engineering technology (CCET): 286–290
  30. Ting Z, Zhao J, Luo J, Qiang Y (2017) Deep belief network for lung nodules diagnosed in CT imaging. *Int J Perform Eng* 13(8):1358–1370
  31. Ulaş B, Bray M, Caban J, Yao J, Mollura DJ (2012) Computer-assisted detection of infectious lung diseases: a review. *Comput Med Imaging Graph* 36(1):72–84
  32. Xiaosong W, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM (2017) Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases In: *Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition*: 2097–2106
  33. Xie Y, Zhang J, Xia Y, Fulham M, Zhang Y (2018) Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. *Inf Fusion* 42: 102–110
  34. Xie H, Yang D, Sun N, Chen Z, Zhang Y (2019) Automated pulmonary nodule detection in CT images using deep convolutional neural networks. *Pattern Recognit* 85(18):109–119
  35. Yiwen X, Hosny A, Zeleznik R, Parmar C, Coroller T, Franco I, Mak RH, Aerts H (2019) Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res* 25(11):3266–3275

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.