A federated approach for detecting the chest diseases using DenseNet for multi-label classification

K. V. Priya^{1,2} · J. Dinesh Peter¹

Received: 27 February 2021 / Accepted: 14 July 2021 / Published online: 28 July 2021
© The Author(s) 2021

Abstract

Multi-label disease classification algorithms help to predict various chronic diseases at an early stage. Diverse deep neural networks are applied for multi-label classification problems to foresee multiple mutually non-exclusive classes or diseases. We propose a federated approach for detecting the chest diseases using DenseNets for better accuracy in prediction of various diseases. Images of chest X-ray from the Kaggle repository is used as the dataset in the proposed model. This new model is tested with both sample and full dataset of chest X-ray, and it outperforms existing models in terms of various evaluation metrics. We adopted transfer learning approach along with the pre-trained network from scratch to improve performance. For this, we have integrated DenseNet121 to our framework. DenseNets have a few focal points as they help to overcome vanishing gradient issues, boost up the feature propagation and reuse and also to reduce the number of parameters. Furthermore, gradCAMS are used as visualization methods to visualize the affected parts on chest X-ray. Henceforth, the proposed architecture will help the prediction of various diseases from a single chest X-ray and furthermore direct the doctors and specialists for taking timely decisions.

Keywords Deep learning · Transfer learning · Multi-label classification

Introduction

Early diagnosis of chronic diseases aims at detecting the presence of diseases as early as possible so that the patients will get better treatments in time. When the treatments get delayed, then the chances of survival are less, and it may lead to worse situations in future. Early detection of diseases helps the patients and doctors to make important decisions on their treatments and the expenses. It also helps the patients and their dependents to get advice and proper guidance to face the challenges in the future. The World Health Organization (WHO) suggests diverse screening programs, which is an unexpected methodology in comparison to early finding. Screening is characterized as the identification of unrecognized disease in an evidently sound, assessments or different

strategies that can be applied quickly and effectively to the objective population. A screening program should remember all the components for the screening cycle from welcoming the objective population to getting to viable treatment for people determined to have the diseases. To support this screening programme, we propose an optimized deep neural network framework.

Nowadays, deep learning is being explored in healthcare applications for disease prediction. However, the significant issue experienced in the clinical space is the absence of enormous dataset with legitimate class names. Because of the unavailability of large data sets from hospitals or test centers, most of the researchers depend on online datasets. Many researchers have already made a comparison of various machine learning algorithms for disease predictions using X-ray image dataset [1].

Deep learning models already shown the capacity of deriving more information about the examples of real evaluation data than customary characterization systems. The high level deep learning designs can be utilized to distinguish the meaning of various highlights from patient records just as to become familiar with the contributions of each component that detect a patient's danger for various diseases. In any

✉ K. V. Priya
priyakv06@gmail.com

¹ Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

² Department of Computer Science and Engineering, Sahrdya College of Engineering and Technology, Thrissur, Kerala, India

Fig. 1 Sample X-ray images from chest X-ray 14

case, exact forecast of chronic disease chances stays a difficult issue that warrants further examinations. Therefore, it is unavoidable to think of a useful asset for doctors to recognize patterns in patient information that demonstrate chances related with particular sorts of chronic diseases. The proposed model can be the better solution for classifying various diseases from chest X-ray images.

System of transfer learning is accepted to speed up the performance of deep neural networks. Transfer learning is an approach of predictive modeling on different but the same kinds of problems. The principle thought behind transfer learning is to another move the knowledge secured in one process to another [2].

The two ways to deal with transfer learning are: (1) calibrating the parameters in the pre-trained network as demonstrated by the given problem [2] and (2) utilizing the pre-trained network in place of feature extractor. At that point, these extracted features will be utilized to train the new classifier. In the deep learning area, transfer learning infers, reusing the weights in at any rate one layer from a pre-trained network model. As per the requirements, it is possible to keep the weights as fixed, fine-tuning the weights or accommodating the weights exactly the same as the pre-trained model for training the new classifier.

Chronic diseases are one of the significant reasons for death among adults in practically all nations and the pace of influenced individuals will increment by 17% in the following 10 years. It is accounted for that six of every ten adults in the US have an chronic disease and four out of ten adults have at least two. In addition, it has been accounted for that of the 58 million passings in 2005, around 35 million will be because of chronic diseases. Early examination of chronic diseases targets recognizing the presence of diseases as exactly on time as possible with the objective that the patients will improve treatments on time. When the treatments get delayed, then the chances of survival are less, and it may lead to worse situations in future. Multi-label disease classification algorithms help to predict various chronic diseases at an early stage. Binary classification in this domain help only to predict any one chronic disease. In the proposed work, it is possible to predict various chronic diseases from a single chest X-ray and it will help the doctors and specialists for taking accurate decisions.

For the execution of this model, chest X-ray images provided by the NIH (National Institutes of Health) Clinical Center is drawn out, and the same dataset is open from Kag-

gle. It is accessible on an open-source platform [3]. The chest X-ray dataset be composed of 112,120 front facing CXR pictures from 30,805 one of a kind patients. All the pictures are named with 14 unique diseases, for example, atelectasis, consolidation, infiltration, pneumothorax, edema, emphysema, fibrosis and so on. The samples from this chest X-ray 14 dataset are exhibited in Fig. 1.

Related works

The greater part of the current works have embraced convolutional networks for the chest X-ray classification [4,5]. X-ray images have been utilized to recognize lung cancer and other lung illnesses utilizing distinctive deep learning models [6,7]. The ideal analysis of different lung infections such as pneumothorax or atelectasis, pneumonia, pulmonary edema, COPD, asthma and so on, additionally should be dealt with. Several deep neural network frameworks were already developed for the early detection of these diseases. Lung knobs from chest CT scans were recognized by implementing a deep CNN with multi-label prediction [7]. In this work, they compared with 2D CNN [7]. They concluded that 3D CNN helps to utilize spatial 3D contextual information and thereby leads to the generation of discriminative features by the training with 3D images. To identify the tiny nodules, they additionally think of a variational nodule forecast procedure, contains cube expectation and clustering [7]. But this framework cannot be used for the classification of different disease types. To minimize the false-positive rate in classification of the lung nodules is proposed in [8,20] with a convolutional neural network. By investigating the attributes of CT scan images, they could lessen the false-positive rate in the classification [8].

The hybrid framework combines spatial transfer network (STN), data augmentation and VGG along with convolution neural network has given better accuracy for the revelation of lung diseases from X-rays [9]. This work was named as hybrid CNN VGG Data STN (VDSNet). Based on input parameters such as age, X-rays, gender and view position, a binary classification was done [9], and disease was predicted. This model had resized the input images as 64*64 image size for the classifier. The architecture contains three main stages as spatial transformer layers, extraction of features layers and classification layers [9].

Spatial transformer layer incorporates λ (lambda) to move the normal routing $[-0.5:0.5]$, batch normalization and a spatial transformer to eliminate the most significant features for lung illness classification [9]. To separate key features, a location network also used in the spatial transformation layer. The different metrics used are recall, precision, and F-score [9]. In the extraction stage of feature layers, pre-trained model VGG16 has used. VGG16 [10] gives 92.7% accuracy

rate in ImageNet which contains million of image dataset with thousand classes. The architecture of VGG16 contains 13 convolutional layers, 5 max pooling and 3 dense layers for better performance. VGG 16 is one of the best models accessed by ILSVRC in 2014. AlexNet gets improved with the more kernel-sized filters such as 11 in the primary layer, 5 in the second layer and that too one after another [10].

Beomhee Park et al. [11] built up a curriculum learning strategy to improve the classification precision of diverse injuries in X-ray images evaluating for aspiratory irregularities to see different pneumonic anomalies including nodule[s], pneumothorax, pleural effusion, consolidation etc. with chest-X-rays of two different origins. The model was aligned with two phases: first-the examples of thoracic anomalies were recognized and then, pre-trained dataset called ResNet-50 on the ILSVRC were used to tune the model using the entire images from two different datasets [11].

The convolutional neural network (CNN) architecture ResNet have been utilized commonly because of its introduction in the ILSVRC (ImageNet Large Scale Visual Recognition Competition). They proved that when the layers become deeper, ResNet could perform better. They have used a 50-layer architecture for their learning model and the multi-classification problem with Softmax function [11]. The abnormal patterns are visualized with the help of class activation maps (CAM). The results of CAM and AUC together yields missorted in not many patients by means of consolidation or nodule[s]. The outcomes of CAM were extracted for all the trained classes and those results shown on the independent X-ray images. This highlights of inferred diseases can be used by the experts for accurate decision making [11]. Different estimates such as sensitivity, specificity, and area under the curve (AUC) are measured separately for each of the two selected datasets. In the first dataset, they could achieve 85.4%, 99.8% and 0.947 for the mentioned measures and 97.9%, 100.0%, and 0.983 in the second dataset. This model helped to train the system easily with high-scale CXR images, and this opened the door to detect more diseases from Chest X-rays. Sebastian Gundel et al. [12] come up with a deep neural network-DenseNet architecture [13]. This incorporates five dense blocks and an aggregate of 121 convolutional layers. Each dense block comprises various dense layers that incorporate batch normalization, improved linear units, and convolution [13]. In the proposed network [12], a transition layer is added between each dense block that incorporates batch normalization, convolution, and pooling, to diminish the dimensions. A global average pooling layer (GAP) is likewise included. The former model is initialized through the pre-trained ImageNet model [12], DenseNet 121. They saw that this perform multiple tasks convolutional neural network could effectively arrange an ample scope of irregularities in the X-ray images of chest. While learning anomaly detection, the framework was upheld by extra features, for example spa-

tial data and standardization on an dataset of 297,541 images [13]. This framework could function admirably for classification of 12 unique irregularities, characterization of their area, and segmentation of heart and lung projections. With extra data of lung and heart segmentation and spatial labels, with a versatile normalization technique, they could improve the irregularity classification execution to a normal AUC as 0.883 on 12 diverse irregularities [13].

Another Dense Net model—CheXNet which contains a dense convolutional Network of 121 layers [13] trained on the dataset of chest X-ray. DenseNets improve the progression of data and gradients, building the enhancement of dense networks manageable [14]. The weights of this pre-trained model are adopted for our model to improve the classification accuracy.

Methodology

The overall architecture works in three different stages. In the first stage, the required preprocessing should be done to achieve accurate results. Second, the model is fine tuned with the help of a pre-trained DenseNet-DenseNet 121 to train the model to detect multiple diseases from a single chest X-ray images. At last, grad-CAM were extricated to confine and visualize the irregular examples on the chest X-rays. The proposed model involves the usage of densely connected convolutional networks, especially DenseNet121 [13]. The major problem with conventional CNN appears when CNN gets more intense. This is due to the fact that the mapping of data from input to output layer ends up being tremendous so much that they can vanish before showing up at the output layer. Another problem encountered in the high-layer networks is that many of the layers are redundant. DenseNets improve the network design between layers presented in different architectures. DenseNets require less parameters than a comparable convolutional neural network, there can be no compelling reason to figure out redundant feature maps. Compared to other convolutional neural network, DenseNet requires less number of parameters as it is not using redundant feature maps while learning. The other regularly utilized CNN is ResNets which has just demonstrated that numerous layers are scarcely contributing. The parameters of ResNets are more than DenseNets in light of the fact that each layer has its weights to learn. All things considered, the number of layers in DenseNets are exceptionally limited (for example 12 filters), and they add new feature maps [13]. Another issue in dense networks raise while training, it is only because of the stream of data and gradients. In DenseNet, each layer procure the gradients through the input and loss function. Each and every layer in DenseNet generate k features and those features get concatenated with the featured captured from the previous layers. The result of this concatenation operation out

to be given contributions to the following layer. Instead of summing up the output feature maps with the new feature maps in other dense networks, DenseNet concatenates these output feature maps with the new feature maps. The following equation represents the concatenation of feature maps in DenseNets:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]). \quad (1)$$

DenseNet basically of different dense blocks and the feature maps are in same dimension inside each dense block. Although the filters in its number may change within the blocks. The layers among them known as transition layers. Transition layers are meant for downsampling and includes batch normalization, a 1×1 convolution and a 2×2 pooling layers. Compared to VGG and ResNets, DenseNet perform better because of the dense number of associations.

The new 32 feature maps are added to the previous feature maps in the dense network. As a result of this, there is shift from 64 feature maps to 256 feature maps. This change in feature maps significant at rear of sixth layer. As already mentioned, transition block continues with 1×1 convolution through 128 filters. This is also pursued through a pooling window of 2×2 across a stride of 2. It bringing about the feature maps into equal parts. The weights of our federated model are initialized through the weights obtained from a pre-trained model, DenseNet121 on ImageNet.

As our proposed model focused on transfer learning [15, 16], separate sets of training, testing and validation data samples are chosen from the NIH chest X-ray dataset. During initial stage, it is necessary to check data leakage. Data leakage means the X-rays of the same patient present in the testing, training and validation dataset. For preprocessing, various data generator methods such as normalization using mean and standard deviation on both validation and training dataset have been used. In addition, a separate standardization is required on the testing set.

Prior to contributing the images into the model, input images are resized and normalized dependent on the standard deviation and mean. In each epoch, input is shuffled to get accurate results. The main problem with the chest X-ray dataset is the class imbalance problem. Figure 2 shows the class irregularity issue that the Hernia pathology has the best imbalance with the extent of positive training cases being about 0.2%. Yet, even the Infiltration pathology, which has minimal measure of imbalance, has just 17.5% of the training cases marked positive.

Cross-entropy is an appropriate cost function feasible on classification problems. To express the probabilities or class distributions, it takes advantage of activation function at the output layer. Cross-entropy function is better choice when dealing with an imbalance dataset. The fraction of less dominant class is get multiplied with the part associated with the

Fig. 2 Class imbalance problem in the dataset

Fig. 3 After implementing loss entropy

Fig. 4 After solving class imbalance problem

dominant class. This facilitate the reduction of loss value to a smaller value. If there is only one class in our dataset, then this loss value become zero. With the utilization of ordinary cross-entropy loss function work on this highly unbalanced dataset, the calculation will be boosted to focus on the majority class (for example negative for this situation), since it offers more to the loss. To beat this issue, the average cross-entropy loss function can be altered over the whole training set D of size N as follows:

$$L_{\text{cross-entropy}}(D) = -\frac{1}{N} \left(\sum_{\text{positive}} \log(f(x_i)) + \sum_{\text{negative}} \log(1 - f(x_i)) \right). \quad (2)$$

Utilizing this formulae, it is noticeable that if there is a huge imbalance with not many positive training cases, for instance, at that point the loss will be overwhelmed by the negative class. Equation (3) and (4) can be used to calculate the contribution of each class's probability distributions and it is used to find the cross-entropy.

Adding the contribution over all the training cases for each class (for example pathological condition), at that point the contribution of each class (for example positive or negative) is

$$\text{freq}_p = \frac{\text{number of positives}}{N} \quad (3)$$

$$\text{freq}_n = \frac{\text{number of negatives}}{N} \quad (4)$$

In Fig. 3, it is obviously distinguished that the contribution of positive cases is fundamentally lower than that of the negative ones. Be that as it may, the main focus here is to make the contributions to be equivalent. One method of doing this is by increasing every model from each class by a class-explicit weight factor, w_{pos} and w_{neg} with the overall contribution of each class is the equivalent. To do this, we need

$$w_{\text{pos}} \times \text{freq}_p = w_{\text{neg}} \times \text{freq}_n, \quad (5)$$

which can be simplified as

$$w_{\text{pos}} = \text{freq}_{\text{neg}} \quad (6)$$

$$w_{\text{neg}} = \text{freq}_{\text{pos}}. \quad (7)$$

Using the above equation, it could balance the contribution of positive and negative labels.

After solving the class imbalance problem, train the model with the help of DenseNet 121. The proposed model is trained with the weights obtained from the pre-trained model. Figure 5 consists of the proposed architecture in which the pre-processed image is given to the pre-trained DenseNet 121 and then to the Global Average 2D pooling. The processed image is then fed to a dense net.

Localization using grad-CAM

After training the network for classification, gradient class activation maps (grad-CAM) are generated [17,18]. To visualize what CNNs are actually looking at, grad-CAM can be utilized. The gradients derived from the output layer of

convolutional model get utilized to create heat maps. These heat maps are generally created through grad-CAM (gradient weighted class activation map) that features the significant areas of an input image. To interpret convolutional neural networks, grad-CAM is an excellent technique. The results of CAM and AUC get joined to produce significant results. The localization of the disease patterns is important for the easy analysis of classification methods. The outcome of all trained samples were autonomously drawn out with the grad-CAMs for the predicted results. In addition, these results are highlighted on the X-rays will help the specialists for their analysis.

These representation procedures help to comprehend the network and may likewise be valuable as almost resembles the visual conclusion for introduction to radiologists [19]. The classification results and localization utilizing grad-CAMs for test illnesses are shown in Fig. 6. The localization of the disease patterns is important for the easy analysis of classification methods. The outcome of all trained samples were autonomously drawn out with the grad-CAMs for the predicted results. Steps involve in the visualization using grad-CAM are as follows:

1. Hook into model output and last layer activations.
2. Get gradients of last layer activations for output.
3. Compute the value of the last layer and gradients for an input image.
4. Compute weights from gradients by global average pooling.
5. Compute the dot product between the last layer and weights to get the score for each pixel.
6. Resize, take ReLU, and return cam.
7. Show the labels with the top 4 AUC.

Results

Figure 7 depicts the training loss curve acquired in our model, and portrays the training process and the way in which the network learns. During an epoch, the loss function is determined across each data item, and it is ensured to give the quantitative loss measure at the given epoch.

The performance of this multi-class classification issue can be pictured utilizing the receiver operating characteristics (ROC) curve. The performance of a classification model at all classification thresholds can be shown on a graph called ROC curve. This curve outlines two different parameters such as true-positive rate (TPR) and false-positive rate (FPR). The term true-positive rate is used to represent recall. An ROC curve plots TPR versus FPR at different classification thresholds. In this model, ROC curves are plotted for every disease, as demonstrated in Fig. 8. The ROC curve is plotted with true-

Fig. 5 A federated approach for detecting the chest diseases using DenseNet

Fig. 6 Grad-CAMs for different diseases of the same patient

Fig. 7 Training loss curve

Fig. 8 ROC curve **Table 1** Prediction accuracy of various diseases

Disease	Prediction accuracy	Disease	Prediction accuracy
Atelectasis	0.882	Pneumothorax	0.886
Cardiomegaly	0.809	Consolidation	0.724
Effusion	0.904	Edema	0.985
Infiltration	0.742	Emphysema	0.959
Mass	0.872	Fibrosis	0.918
Nodule	0.874	Pleural thickening	0.583
Pneumonia	0.811	Hernia	0.966

Table 2 Comparison study of prediction accuracies with existing models

Disease	Pathology Wang et al. [6]	Yao et al. [21]	CheXNet	Federated approach
Atelectasis	0.716	0.772	0.8094	0.882
Cardiomegaly	0.807	0.904	0.9248	0.809
Effusion	0.784	0.859	0.8638	0.904
Infiltration	0.609	0.695	0.7345	0.742
Mass	0.706	0.792	0.8676	0.872
Nodule	0.671	0.717	0.7802	0.874
Pneumonia	0.633	0.713	0.7680	0.811
Pneumothorax	0.806	0.841	0.8887	0.886
Consolidation	0.708	0.788	0.7901	0.724
Edema	0.835	0.882	0.8878	0.985
Emphysema	0.815	0.829	0.9371	0.959
Fibrosis	0.769	0.767	0.8047	0.918
Pleural thickening	0.708	0.765	0.8062	0.583
Hernia	0.767	0.914	0.9164	0.966

positive rate (TPR) on the X -axis against the false-positive rate (FPR) on the x -axis.

The accuracy of different diseases obtained in our model is represented in Table 1.

The proposed federated approach is compared with the existing results and found that our model outperforms in the prediction of chest diseases such as mass, nodule, infiltra-

tion, atelectasis, effusion, pneumonia, edema, emphysema, fibrosis, hernia and so on. Table 2 gives a comparison study of various approaches on multi-label classification of chest X-ray images. The table gives comparison of the prediction accuracies of existing models with our proposed model. This gives a conclusion that our model performs better compared to existing models.

Conclusion

This work presents a high level deep learning network architecture upgraded for problems which need multi-label classification. The proposed network model can be prepared without any preparation and moreover can be tweaked with the weights embraced from the pre-trained model on the ImageNet. The proposed architecture demonstrated the performance improvement using weighted loss entropy to balance class imbalance problems. The loss entropy gives better execution particularly to those classes with relative lesser instances, for example Hernia. The output of this model visualized using grad-CAMs and obtained a better visual explanation of the proposed classification architecture. The results obtained by the grad-CAMs will be useful to the radiologists because of the highlighted areas of multiple disease affected areas on the X-ray images. This will assist radiologists with auditing and interpreting the different diseases in a single X-ray image. The future work will focus on the better techniques for pre-training on various datasets especially for multiple disease prediction. The proposed model can be fine tuned with the use of appropriate optimization techniques such as stochastic gradient descent with momentum, AdaGrad, RMSProp, and Adam Optimizer. Optimization algorithms are responsible for reducing losses and providing the most accurate results possible.

Declarations

Conflict of interest The authors declare that they 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

- Baltruschat IM, Nickisch H, Grass M, Knopp T, Saalbach A (2019) Comparison of deep learning approaches for multi-label chest X-ray classification. *Sci Rep* 9:1–10
- Ribani R, Marengoni M (2019) A survey of transfer learning for convolutional neural networks. In: *Proceedings of SIBGRAPI-T*. IEEE, IV-B, pp 47–57
- NIH sample Chest X-rays dataset (2020) <https://www.kaggle.com/nih-chest-xrays/sample>. Accessed 5 June 2020
- Sun W, Zheng B, Qian W (2017) Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. *Comput Biol Med* 89:530–539
- Sun W, Zheng B, Qian W (2016) Computer aided lung cancer diagnosis with deep learning algorithms. In: *Proceedings of SPIE, medical imaging*, vol 9785
- Wang X, 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: *2017 IEEE conference on computer vision and pattern recognition (CVPR)*. pp 3462–3471
- Gu Y, Lu X, Yang L, Zhang B, Yu D, Zhao Y, Gao L, Wu L, Zhou T (2018) Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs. *Comput Biol Med* 103:220–31
- Setio AAA, Traverso A, de Bel T, Berens MSN, van den Bogaard C, Cerello P, Chen H, Dou Q, Fantacci ME, Geurts B et al (2017) Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal* 42:1–13
- Bharati S, Podder P, Mondal MRH (2020) Hybrid deep learning for detecting lung diseases from X-ray images. *Inform Med Unlocked* 20:100391
- Simonyan K, Zisserman A (2015) VERY Deep Convolutional Networks for large-scale image recognition. Published as a conference paper at ICLR
- Park B, Cho Y, Lee G, Lee SM, Cho Y-H, Lee ES, Lee KH, Seo JB, Kim N (2019) A curriculum learning strategy to enhance the accuracy of classification of various lesions in chest-PA X-ray screening for pulmonary abnormalities. *Nature Sci Rep* 9:15352. <https://doi.org/10.1038/s41598-019-51832-3>
- Gundel S, Ghesu FC, Grbic S, Gibson E, Georgescu B, Maier A, Comaniciu D (2019) IEEE, multi-task learning for chest X-ray abnormality classification on noisy labels. *IEEE* (2019)
- Huang G, Liu Z, Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, pp 2261–2269
- Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Ball RL, Langlotz C, Shpanskaya K, Lungren MP (2017) Ng AY CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning
- Zhi W, Yueng H, Chen Z et al (2017) Using transfer learning with convolutional neural networks to diagnose breast cancer from histopathological images. In: *Proceedings of ICONIP 2017, III-A0a, 3, IV-B*. pp 669–676
- Mehra R et al (2018) Breast cancer histology images classification: training from scratch or transfer learning? *ICT Express* 4(4):247–254 (III-A0a, 3, IV-B)
- Selvaraju RR et al (2016) Grad-CAM: why did you say that? Visual explanations from deep networks via gradient-based Localization. pp 1–5. [arXiv:1610.02391v2](https://arxiv.org/abs/1610.02391v2)
- Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp 2921–2929
- Pasa F, Golkov V, Pfeifer F, Cremers D, Pfeifer D (2019) Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization. *Sci Rep* 9:1–9
- Song Q, Zhao L, Luo X, Dou X (2017) Using deep learning for classification of lung nodules on computed tomography images. *J Healthc Eng* 2017:8314740. <https://doi.org/10.1155/2017/8314740>. PMID: PMC5569872

21. Yao, Li, Poblenz, Eric, Dagunts, Dmitry, Covington, Ben, Bernard, Devon, and Lyman, Kevin. Learning to diagnose from scratch by exploiting dependencies among labels. arXiv preprint [arXiv:1710.10501](https://arxiv.org/abs/1710.10501), 2017
22. Yao, Li, Poblenz, Eric, Dagunts, Dmitry, Covington, Ben, Bernard, Devon, and Lyman, Kevin (2017) Learning to diagnose from scratch by exploiting dependencies among labels. [arXiv:1710.10501](https://arxiv.org/abs/1710.10501)

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