Segmentation of Chest Radiography Scans for COVID-19

Navya Agarwal[1,1*†] and Ananya Srivastava[1,1†]

[1*]Department of Artificial Intelligence and Data Science, Indira Gandhi Delhi Technical University for Women, Madrasa Road Opposite St. James Church, Kashmere Gate, 110006, Delhi, India.

*Corresponding author(s). E-mail(s): navya.agarwal.in@gmail.com;
Contributing authors: amit.ritu.ananya@gmail.com;
†These authors contributed equally to this work.

Abstract

For more than two years, the world has seen the COVID-19 pandemic exact its toll on humanity. While the worst of times have subsided, people must not forget that there were weeks when the hospitals were full to capacity, testing kits were not available for everyone which created a panic situation. Thus, any technological advancement that allows faster diagnosis of the COVID-19 infection with high accuracy can be very useful for the healthcare industry. Since COVID-19 attacks the epithelial cells that line the respiratory tract, X-rays can be used to analyze the health of a patient's lungs. However, the images of various viral cases of pneumonia are similar to those of COVID-19. Therefore, it is difficult to distinguish COVID-19 from other viral cases of pneumonia. This research paper aims to investigate the utility of Deep Learning models in the rapid and accurate detection of COVID-19 from chest X-ray images. CNN classifiers are trained to classify COVID-19 lung scans from normal and Viral Pneumonia infected lung scans. The goal is to provide better methods for estimations that can help the healthcare system to prepare for epidemic outbreaks.

Keywords: CNN, COVID-19, Deep learning, Classification

1 Introduction

COVID-19 tested the limits of the entire healthcare industry. The testing and diagnosis of the infection, could not be done at the pace the infection was spreading due to the lack of testing kits. This created the need for a much faster method of diagnosis that could be carried out out quickly for a large number of people. Chest X-ray is one such option.

However it comes with its own set of challenges such as X-Ray scans of COVID-19 being similar to those of Viral Pneumonia and lack of availability of large enough data sets for training of complex deep learning models.

This paper first surveys the existing work done in the field and then talks about our own deep learning models that investigate the best methods to go about classifying COVID-19 X-ray scans.

Before implementing our models, we carry out an elementary data analysis, which reveals the significant class imbalance within the data set.

Next we implement Convolutional Neural Network models which are best for working with images. We implement a CNN without data augmentation and another one with data augmentation. We also implement a VGG-16 pre-trained model for classification.

We plot both the accuracy and loss graohs for training v/s validation for all three models. We also plot the confusion matrix for the first two models to show how data augmentation helps to reduce mis-classification for the minority class.

2 Literature Survey

Objectives	Parameters	Data set	Result	Conclusion
[1]Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network	Accuracy	80 samples of normal CXR images (with 4020 $\times$ 4892 pixels) from the Japanese Society of Radiological Technology (JSRT). CXR images, which contains 105 and 11 samples of COVID-19 and SARS (with 4248 $\times$ 3480 pixels)	DeTraC outperformed all pre-trained models with a large margin in most cases	To increase the efficiency and allow deployment on hand-held devices, model pruning, and quantisation could be utilised.

[2]DeTrac: Transfer Learning of Class Decomposed Medical Images in Convolutional Neural Networks	Multi-classes confusion matrix	Histological images of human colorectal cancer of 5000 histological images (with 150×150 pixels), divided into three classes: tumour epithelium, stroma and mucosal glands	DeTraC with ResNet achieved the highest accuracy of 99.1 % while DeTraC with VGG16 and GoogleNet was behaving almost the same (with 99.8 % and 99.7 % accuracies, respectively), in case of the colorectal cancer dataset	The application of class decomposition to other deep learning architectures like Recurrent Neural Networks (RNNs) and Long Short Term Memory (LSTM) for sequence models can be explored.
[3]Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Network	Accuracy	315 chest X-ray images of COVID-19 patients obtained from the open source repository shared by Dr. Joseph Cohen containing chest X-ray/CT images of patients with ARDS, COVID-19, MERS, pneumonia, SARS	Obtained the best performance as a classification accuracy of more than 98 %	Inception V3 model exhibits an excellent performance in classifying COVID-19 pneumonia by effectively training itself from a comparatively lower collection of images.

[4]Design of Accurate Classification of COVID-19 Disease in X-Ray Images Using Deep Learning Approach	Sensitivity, specificity, precision, accuracy	standard datasets for evaluation named Cohen's dataset, which consists of 60k around images with 400 positive COVID-19 x-ray images	The proposed CNN + Histogram Oriented Gradient(HOG) method shows good detection accuracy in a fast and effective manner	In the future, more lung disease detection should be incorporated along with the algorithm. A dataset from several sources allows the development of robust models.
[5]Self-Super Vision v/s Transfer Learning: Robust Biomedical Image Analysis Against Adversarial Attacks	Accuracy	Chest X-rays of patients who have and don't have pneumonia. The right ventricle segmentation data set is a set of cardiac MRI images of 16 patients.	Self-super vision trained models learn more robust features than the ImageNet based transfer learning models. It is established that self-super vision learning with adversarial training as default approach for better performance with small amount of labeled data by exploiting a huge volume of unlabelled data	This approach can be used as a base for deciding on active selection of data to be labelled in the active learning framework

[6]Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data	Accuracy	Review paper	While it is not possible to suggest an optimal architecture for a specific task, it is observed that ensembles of networks typically perform better than individual models	A final issue for deep learning researchers to consider is frequently referred to as 'explainable AI'. Systems which produce classification labels without any indication of reasoning raise concerns of trustworthiness for radiologists.
[7]Unsupervised Deep Anomaly Detection in Chest Radiographs	Receiver operating characteristic (ROC)	The Radiological Society of North America (RSNA) Pneumonia Detection Challenge database	The system could correctly localize various lesions or anomalies, namely, a lung mass, cardiomegaly, pleural efusion, bilateral hilar lymphadenopathy, and even dextrocardia	Future work will focus on the improvement of performance in anomaly detection and visualization, with the aim to clinically apply an all-purpose initial screening tool for any type of anomaly and even for any modality including 3D images

[8]Viral Pneumonia Screening on Chest X-Rays Using Confidence-Aware Anomaly Detection	area under the receiver operator curve (AUC), sensitivity, specificity, and, accuracy	Two in-house X-ray image datasets, X-VIRAL and X-COVID, were used for this study. The X-VIRAL dataset contains 5,977 viral pneumonia cases, 18,619 non-viral pneumonia cases, and 18,774 healthy controls	Binary classification using ResNet achieves the an accuracy of 78.52 %, a sensitivity of 78.28 %, a specificity of 78.56 %, and an AUC of 86.24 %. An anomaly detection model always outperforms (particularly in terms of sensitivity) the corresponding binary classification model	Anomaly detection is superior to binary classification methods, and learning model confidence is useful to predict failures, greatly reducing the false negatives. CAAD model achieves an AUC of 83.61 % and sensitivity of 71.70 % on the unseen X-COVID data set.

3 Methodology

Our research lies in the field of classification. The metric used to judge our results is the accuracy of our classification model.

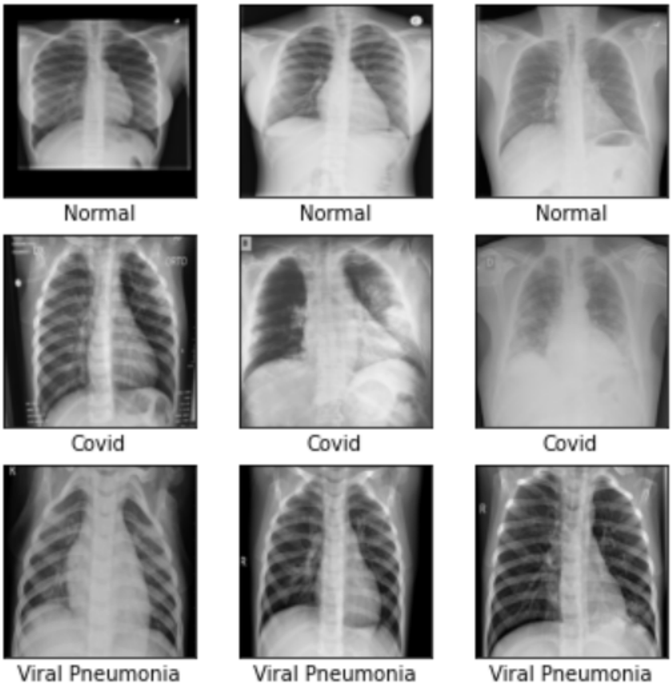
3.1 Data Collection

A team of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh along with their collaborators from Pakistan and Malaysia in collaboration with medical doctors have created a database of chest X-ray images for COVID-19 positive cases along with Normal and Viral Pneumonia images. This dataset was available on Kaggle. It consists of 3616 positive COVID-19 scans, 10216 normal lung scans and 1345 Viral Pneumonia scans.

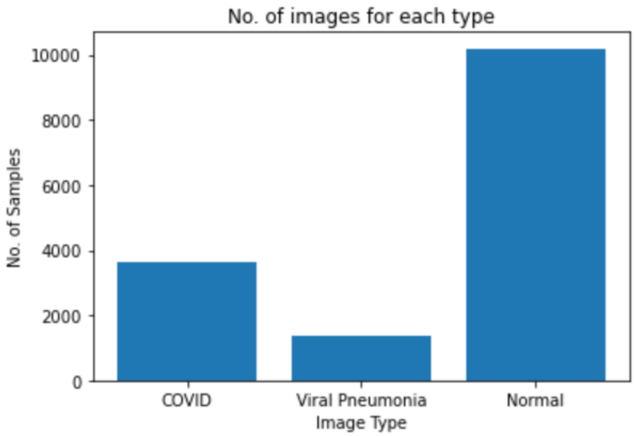
3.2 Data Analysis

Elementary data analysis was performed on the data set to gain some crucial insights. Sample images of the lung scans for the different categories are shown below.

Every image is of data type uint8 and of dimensions 299x299. Even though these are black and white X-ray images, there are 3 colour channels - red, green and blue.



Next we plotted a bar graph of the number of images in each category. From the grah, we can conclude that there is a significant imbalance amongst the three classes that needs to be handled.



3.3 Implementation of Deep Learning Models

3.3.1 CNN Model without data augmentation

- **Data loading and Pre-processing:** We create a function loadImages that reads images from directories and adds them to a list and also creates a list of their target labels. It performs normalisation and resizes the images to 100x100 for ease of processing.

- The labels for the three classes are decided as follows: Normal = 0, COVID-19 = 1, Viral Pneumonia = 2.

- The images for the specific labels are stored in three separate lists for normal, COVID and viral pneumonia labels using the loadImages function. These three lists and their target labels lists are stacked into two lists – data and targets.

- **Data Splitting:** The whole data is split into training, validation and testing data with proportions as 70 %, 10 % and 20 % using the train-test-split from sklearn.

- We build a CNN model using Sequential from sklearn with the following specifications:

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 98, 98, 32)	896
max_pooling2d (MaxPooling2D)	(None, 49, 49, 32)	0
dropout (Dropout)	(None, 49, 49, 32)	0
conv2d_1 (Conv2D)	(None, 47, 47, 16)	4624
max_pooling2d_1 (MaxPooling 2D)	(None, 23, 23, 16)	0
dropout_1 (Dropout)	(None, 23, 23, 16)	0
conv2d_2 (Conv2D)	(None, 21, 21, 16)	2320
max_pooling2d_2 (MaxPooling 2D)	(None, 10, 10, 16)	0
dropout_2 (Dropout)	(None, 10, 10, 16)	0
flatten (Flatten)	(None, 1600)	0
dense (Dense)	(None, 512)	819712
dense_1 (Dense)	(None, 256)	131328
dense_2 (Dense)	(None, 3)	771

- Regularizers are added to the convolutional layers and dropout layers are also added in the model to prevent overfitting.
  Regularizers: Technique that helps the model to generalise better such that it gives better accuracy on the unseen/ test data.
  Dropout layers: Dropout layer helps to prevent over-fitting by ignoring certain randomly selected neurons during the training phase i.e. these neurons are not considered during a particular forward or backward pass. This prevents the network from relying too much on single neurons and forces all neurons to learn to generalize better.

- The model is compiled using "adam" optimizer and "Sparse Categorical Cross Entropy" as the loss function. The metric for training is chosen as "accuracy".
  ADAM (Adaptive Moment Estimation): The results of the Adam optimizer are generally better than every other optimization algorithm, have faster computation time, and require fewer parameters for tuning, because of which, it is recommended as the default optimizer for most of the applications.
  Sparse Categorical Cross Entropy: It is used when there are two or more classes present in our classification task. It also requires our labels to be in the form of integers. However, unlike Categorical Cross Entropy, it does not require us to perform one hot encoding.

- Then the model is trained with batch size 64 and maximum number of epochs 50. A call back is also passed to the model for early stopping of the training of model in case the validation loss does not decrease in 3 consecutive epochs.

- The model stops training after 25 epochs, eventually attaining a training data accuracy of 94.51 % and validation data accuracy of 91.69%

- We save this model in HDF5 format for easy reloading. It contains the model's architecture, weights values, and compile() information. It is a light-weight alternative to SavedModel.

- Accuracy and loss values are plotted for both training and validation data, and we see that the model is well fitting. Both under and overfitting of the model is avoided.

- Then we predict the target values for the testing data and plot the confusion matrix for it and calculate the accuracy. The accuracy for the test data on the trained model is found to be 91.55

3.3.2 CNN Model with data augmentation

- While performing the Exploratory Data Analysis, we see that there is a significant class imbalance between the normal, COVID and viral pneumonia classes.
- Hence to eliminate this imbalance, we perform data augmentation using the ImageDataGenerator() from sklearn. Then we create an iterator object for it that is passed to the model during compilation.
- The rest of the steps are performed same as the previous model.
- On plotting the accuracy and loss values for both training and validation data, we see that the values for both models are comparable. Even the testing accuracy of both models do not have a significant difference.
- However, the CNN model with data augmentation achieves its target of reducing the mis-classification in the minority class, which in this case is the Viral Pneumonia class.

3.3.3 VGG-16 Pre-trained Model

Keras Applications are deep learning models that are made available alongside pre-trained weights. These models can be used for prediction, feature extraction, and fine-tuning. The model we used was VGG16 (Visual Geometry Group) VGG-16 is a convolutional neural network that is 16 layers deep. We load a pre-trained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224.

- The images were present in different directories on the basis of the class they belonged to so we created a list containing the paths of each of the image files in our dataset and another list containing the label of the image which we extracted from the directory name in which it was present.
- Then we created a dataframe using Pandas with 2 columns: a 'File Path' column containing the paths of the images and a 'Category' column which contained the label of each image.
- We split the data into training, validation and testing images using SciKit's train-test-split method, dividing the dataframe in the ratio of 75%, 12.5% and 12.5% for each. This method shuffles our dataset and divides into various categories or strata on the basis of the image label.

- Image Preprocessing: We used the Image-Data-Generator method that generates batches of tensor image data (using a dataframe) with real-time data augmentation. Data augmentation is a technique to increase the diversity of your training set by applying random (but realistic) transformations, such as image rotation. We also performed normalization techniques such as rescaling our images. Here are a few parameters that we changed:

  1. class-mode = set to 'categorical', the y-col parameter contains the target labels/classes
  2. x-col = column in dataframe that contains file paths of images
  3. target-size = size to which all images are resized

  We created a training data generator and a similar validation data generator.

- We instantiated a VGG16 Model (described above) while leaving out the fully connected top layers because we want to connect the model layers to that we define.

- The layers that we defined and then added to the base model:

  1. Pooling Layer: An average pooling layer
  2. Flatten: A layer that flattened the multi-dimensional images to 1D arrays.
  3. Dense: A single hidden layer with 128 neurons and the Rectified Linear Unit activation function (add relu math function here)
  4. Dropout: The Dropout layer randomly sets input units to 0 with a frequency of rate at each step during training time, which helps prevent overfitting. Inputs not set to 0 are scaled up by 1/(1 - rate) such that the sum over all inputs is unchanged.
  5. Dense: Output layer containing 3 neurons and using Softmax activation function (add softmax math function here)

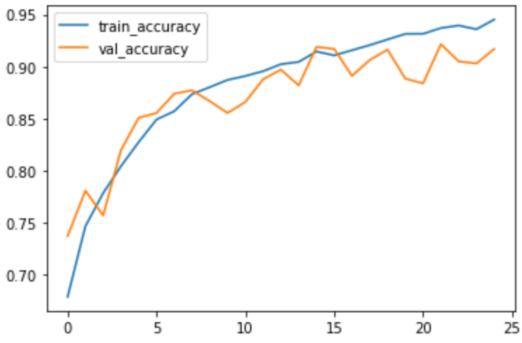
- Then we compiled our model and using the Categorical Cross Entropy Loss (also called as Softmax Loss. It is a Softmax activation plus a Cross-Entropy loss. If we use this loss, we will train our model to output a probability over the n classes for each image. It is used for multi-class classification) The optimizer that we used was Adam (It stands for Adaptive Moment Estimation. It computes adaptive learning rate for each parameter) The metric to gauge performance was chosen as accuracy of the model.

- Finally, we fit the model to the train data generator and executed the model for 10 epochs, eventually attaining a training data accuracy of 91.34% and validation data accuracy of 94.62%

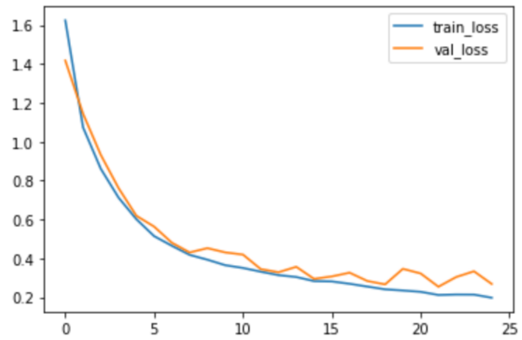
4 Results

4.1 CNN model without data augmentation

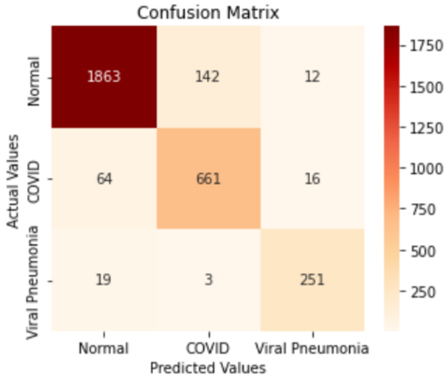
Training v/s Validation Accuracy



Training v/s Validation Loss



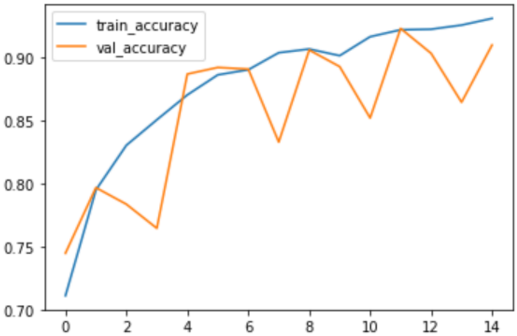
Confusion Matrix for Testing Data



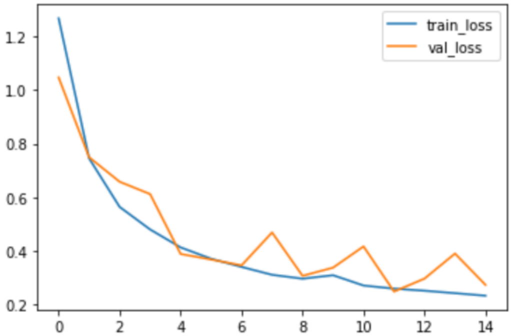
When the model was run on testing data, the accuracy was found to be 91.55%.

4.2 CNN model with data augmentation

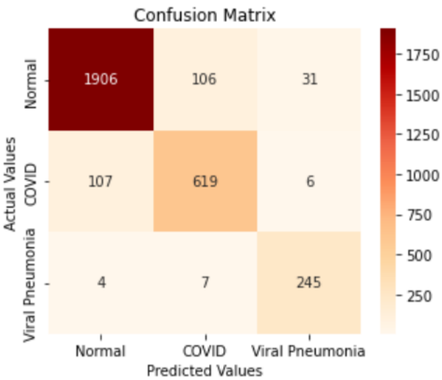
Training v/s Validation Accuracy



Training v/s Validation Loss



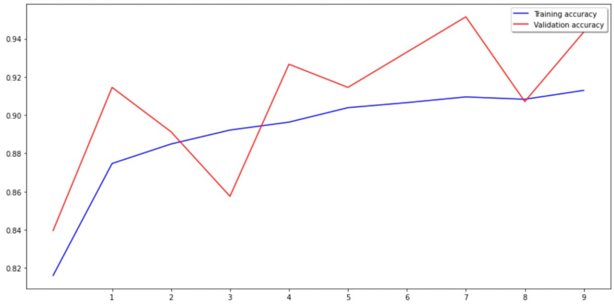
Confusion Matrix for Testing Data



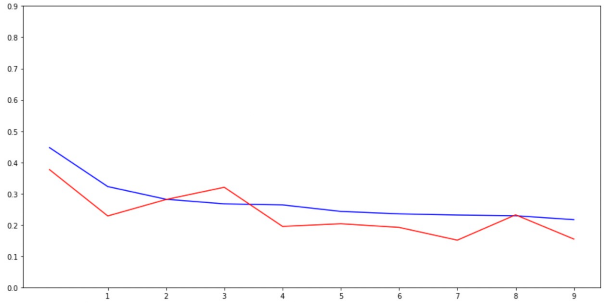
When the model was run on testing data, the accuracy was found to be 91.39%.

4.3 VGG-16 Pre-trained Model

Training v/s Validation Accuracy



Training v/s Validation Loss



After 10 epochs, the model attained a training accuracy of 91.3% and validation accuracy of 94.35%

5 Conclusion

After a thorough analysis of our implemented models and the literature survey, we can conclude that transfer learning combined with methods that handle class imbalance and prevent over-fitting provide a much higher accuracy in classifying COVID-19 chest X-Ray scans from those of normal and Viral Pneumonia X-Ray scans.
Transfer Learning gives us a very good starting point for working on problem statements relating to computer vision and NLP given vast the compute and time resources required to develop neural network models on these problems. Using pre-trained models, we were able to obtain a higher accuracy as they are previously trained on huge amounts of data.
Our models also investigated the use of image data augmentation to improve performance. We conclude that augmenting data instead of random oversampling is a better way to handle class imbalance as it prevents over-fitting.
While there are huge strides being taken in the field of Computer Vision for medical science, there is a lot of room for improvement. Collection and open

source availability of more data sets in the field would encourage more work in this direction.

References

[1] Abbas, A., Abdelsamea, M.M. Gaber, M.M. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Appl Intell 51, 854–864 (2021). https://doi.org/10.1007/s10489-020-01829-7

[2] A. Abbas, M. M. Abdelsamea and M. M. Gaber, 'DeTrac: Transfer Learning of Class Decomposed Medical Images in Convolutional Neural Networks', in IEEE Access, vol. 8, pp. 74901-74913, 2020, doi: 10.1109/ACCESS.2020.2989273

[3] S. Asif, Y. Wenhui, H. Jin and S. Jinhai, 'Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Network', 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 2020, pp. 426-433, doi: 10.1109/ICCC51575.2020.9344870

[4] Design of Accurate Classification of COVID-19 Disease in X-Ray Images Using Deep Learning Approach Joy Iong-Zong Chen DOI: https://doi.org/10.36548/jismac.2021.2.006

[5] D. Anand, D. Tank, H. Tibrewal and A. Sethi, 'Self-Supervision vs. Transfer Learning: Robust Biomedical Image Analysis Against Adversarial Attacks', 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 1159-1163, doi: 10.1109/ISBI45749.2020.9098369

[6] N. Kussul, M. Lavreniuk, S. Skakun and A. Shelestov, 'Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data', in IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 5, pp. 778-782, May 2017, doi: 10.1109/LGRS.2017.2681128

[7] Nakao, T., Hanaoka, S., Nomura, Y. et al. Unsupervised Deep Anomaly Detection in Chest Radiographs. J Digit Imaging 34, 418–427 (2021). https://doi.org/10.1007/s10278-020-00413-2

[8] J. Zhang et al., 'Viral Pneumonia Screening on Chest X-Rays Using Confidence-Aware Anomaly Detection', in IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 879-890, March 2021, doi: 10.1109/TMI.2020.3040950