On the effect of decomposition granularity on DeTraC for COVID-19 detection using chest X-ray images

Nicole Panashe Mugova[1], Mohammed M. Abdelsamea[1,2,*], Mohamed Medhat Gaber[1,3]
[1]School of Computing and Digital Technology, Birmingham City University, Birmingham, B4 7XG, UK
[2]Faculty of Computers and Information, Assiut University, Assiut, Egypt
[3]Faculty of Computer Science and Engineering, Galala University, Egypt
* Correspondence: mohammed.abdelsamea@bcu.ac.uk

Abstract

COVID-19 is a growing issue in society and there is a need for resources to manage the disease. This paper looks at studying the effect of class decomposition in our previously proposed deep convolutional neural network, called DeTraC (Decompose, Transfer and Compose). DeTraC can robustly detect and predict COVID-19 from chest X-ray images. The experimental results showed that changing the number of clusters (decomposition granularity) affected the performance of DeTraC and influenced the accuracy of the model. As the number of clusters increased, the accuracy decreased for the shallow tuning mode but increased for the deep tuning mode. This shows the importance of using suitable hyperparameter settings to get the best results from the DeTraC deep learning model. The highest accuracy obtained, in this study, was 98.33% from the deep tuning model.

INTRODUCTION

In late December 2019, an outbreak of a virus emerged from Wuhan, China (Wu, Chen, and Chan, 2020). The disease Sars-Cov-2, better known as COVID-19, spread at an alarming rate, affecting countries worldwide. COVID-19 primarily affects the respiratory system (similar to pneumonia) and chest X-rays play a crucial role in assessing the presence, severity, and progression of COVID-19. The virus has significantly affected the healthcare sector, resulting in shortages of staff and necessary protective equipment (Organization, 2020).

X-rays, also known as plain radiography (Johns et al. 1983) are a popular medical imaging technique. X-rays are significantly cheaper than CT scans and are painless, fast, and non-invasive. Understanding chest X-rays requires expert knowledge and experience but can be time-consuming. Using both chest X-rays and artificial intelligence will mean the extent of the disease can be determined. Long-term, diagnostic tools will benefit radiographers and other healthcare professionals involved in the diagnosis process.

Deep learning offers a reliable way to identify abnormalities in medical images, allowing for preventive screening and personalised patient data benefiting both the patient and doctor. Deep learning models have been applied to reduce pressure on healthcare professionals. It has been evident that pretraining deep neural networks on a large generic data set is an effective and efficient way to retrain such models on other tasks. The three types of Neural Networks that form the basis for most pre-trained models in deep learning are – artificial neural networks as multilayer perceptron, convolutional neural network (CNN) and recurrent neural networks (Wang et al., 2019).

CNN is a popular deep learning approach that has shown superior achievements in the medical imaging domain. The primary success of CNN is due to its ability to learn local features automatically from images. One of the most popular strategies for training CNN architecture is to transfer knowledge from a pre-trained network that fulfilled one generic task (e.g., large-scale image recognition) into a new domain-specific task (e.g., COVID-19 detection). Transfer learning (Abbas et al. 2018) is faster and easy to apply, especially without the need for a sufficient annotated dataset for training. Consequently, many researchers tend to utilise and adapt this strategy especially with medical imaging and COVID-19 detection (Abbas et al. 2020b). Transfer learning can be generally accomplished with two major scenarios: a) “shallow tuning”, which adapts the last few classification layers to deal with the new specific task and freezes the parameters of the remaining layers without training; and b) “deep tuning” which aims to retrain all the weights of the pre-trained network from end-to-end manner and requires a huge amount of data to overcome overfitting problem.

Using deep learning techniques to recognise the prominent features in chest X-rays allows for better diagnosis and in some cases is better than a radiologist (Hosny et al., 2018). Amongst the various deep learning techniques, DeTraC (decompose, transfer, and compose) has shown a high accuracy of 93.1% in diagnosing COVID-19. DeTraC is a deep CNN that can be trained using a small number of images (Abbas et al. 2020a), which is beneficial given the limited number of images currently available.

Communications of the ECMS, Volume 35, Issue 1,
Proceedings, ©ECMS Khalid Al-Begain, Mauro Iacono,
Lelio Campanile, Andrzej Bargiela (Editors)
ISBN: 978-3-937436-72-2 / 78-3-937436-73-9(CD) ISSN 2522-2414

One problem that can arise from using deep learning for medical diagnosis is the overfitting problem. This is where the algorithm performs well on the training and validation datasets but when presented with new testing data it does not perform well. This affects the reproducibility of the algorithm, so it is imperative to test different hyperparameters in the hope of finding the optimal hyperparameter settings.

In this paper, the behaviour of DeTraC is investigated when changing the hyperparameters – specifically the number of clusters in the class decomposition component using k-means clustering. A k-means clustering method is an example of an unsupervised learning technique. The main step involved in k-means clustering is to group different instances (examples) based on their similarities. One of the disadvantages of using the k-means clustering algorithm is that the number of clusters needs to be specified, if the number is unsuitable, the performance of DeTraC will be affected. So, changing the value of k will allow the effectiveness of the algorithm to be observed and to quantify its usefulness. Typically, the value of k is important as it determines the number of centroids going around the data (Pham et al. 2005). Although the value of k affects the performance of clustering directly, it affects the performance of DeTraC indirectly. Thus, using cluster performance metrics related to cohesion and separation may not be the best way to investigate the effect of k on DeTraC. In this paper, an experimental exploration on the effect of the number of clusters (i.e., decomposition granularity) on the performance of DeTraC is provided and discussed.

To achieve the main objective, the rest of the paper is structured as follows: The following section provides a brief overview of existing work on Deep Learning, details about the dataset, the workflow of DeTraC, and the experimental study. This is followed by the discussion of the results. Finally, the paper is concluded with a summary and pointers to possible future work.

RELATED WORK

Deep learning has "already left its mark" in healthcare and continues to provide new solutions to current problems in society (Esteva et al. 2021). Developing new advancements such as diagnostic tools benefits healthcare professionals and patients by providing quicker and more accurate diagnoses of certain diseases. The review of the literature has highlighted a gap in research and development. It shows the importance of why scientists and healthcare professionals need to work together, to minimise the effects of new diseases that arise by sharing data and research.

(Stephen et al., 2019) developed a CNN to detect Pneumonia from chest x-ray images. They noted how a CNN was the better choice for image segmentation because of its ability to extract abstract 2D features through learning. This is beneficial for illnesses such as Pneumonia, SARS, and COVID-19 – which all affect the lungs. (Abbas et al. 2020a) proposed a deep CNN called DeTraC (Decompose, Transfer and Compose) to detect COVID-19 from chest X-rays. One of the barriers researchers face when dealing with image classification is data irregularities and (Abbas et al. 2020a) suggests DeTraC can cope with this issue. (Rahaman et al., 2020; Ismael and Şengür, 2021) also use deep learning techniques to compare different pre-trained CNN models. Showing how CNNs are effective tools at image classification, however (Rahaman et al., 2020) note there are essential factors that influence the performance of the individual models, such as, image quantity, model complexity, and the distribution of the dataset. (Rahaman et al., 2020) used 860 images (300 healthy, 300 pneumonia, and 260 COVID-19). Data augmentation techniques were applied to the 1764 images and DeTraC achieved an accuracy of 97.35% (Abbas et al. 2020a). Even though (Rahaman et al., 2020) has a lower number of images, the CNN model performs well with an accuracy of 89.3% using a VGG19 model.

While the other sources use a pre-trained model and show the benefits of using a VGG19 model for image classification, (Stephen et al., 2019) does not rely on a pre-trained model but designs a CNN from scratch. The CNN designed can detect pneumonia from chest X-ray images with high accuracy (93.7%). Despite not relying on a pre-trained model the accuracy achieved is still higher than (Rahaman et al., 2020) but lower than (Abbas et al. 2021). (Stephen et al., 2019) suggests that the model they developed could alleviate some of the problems that researchers face such as reliability and interpretability. However, the method is not as detailed enough for it to be reproduced by someone else without seeing the source code which is not provided.

DeTraC

DeTraC consists of three phases – class decomposition, transfer learning, and class composition. Before the class decomposition phase, a pre-trained CNN model acts as the feature extractor, extracting deep features from the images inputted to build a deep feature space (Abbas et al. 2021). The deep feature space is very important and (Abbas et al. 2021) applied PCA to reduce the high dimensionality of the feature space. Applying PCA to the feature space results in a lower dimensionality and ignores any highly correlated features. Then the reduced feature space decomposes the original classes into decomposed classes (Abbas et al. 2021). Once the representation of each image is constructed based on the associated pre-trained model. In the second phase, DeTraC adds a novel class-decomposition layer to several pre-trained CNN models for the decomposition of classes, in an unsupervised way, and accomplish the training using sophisticated gradient descent optimisation method. Finally, we distinguish between normal and abnormal cases using an error-correction criterion. Class decomposition allows for easing the local

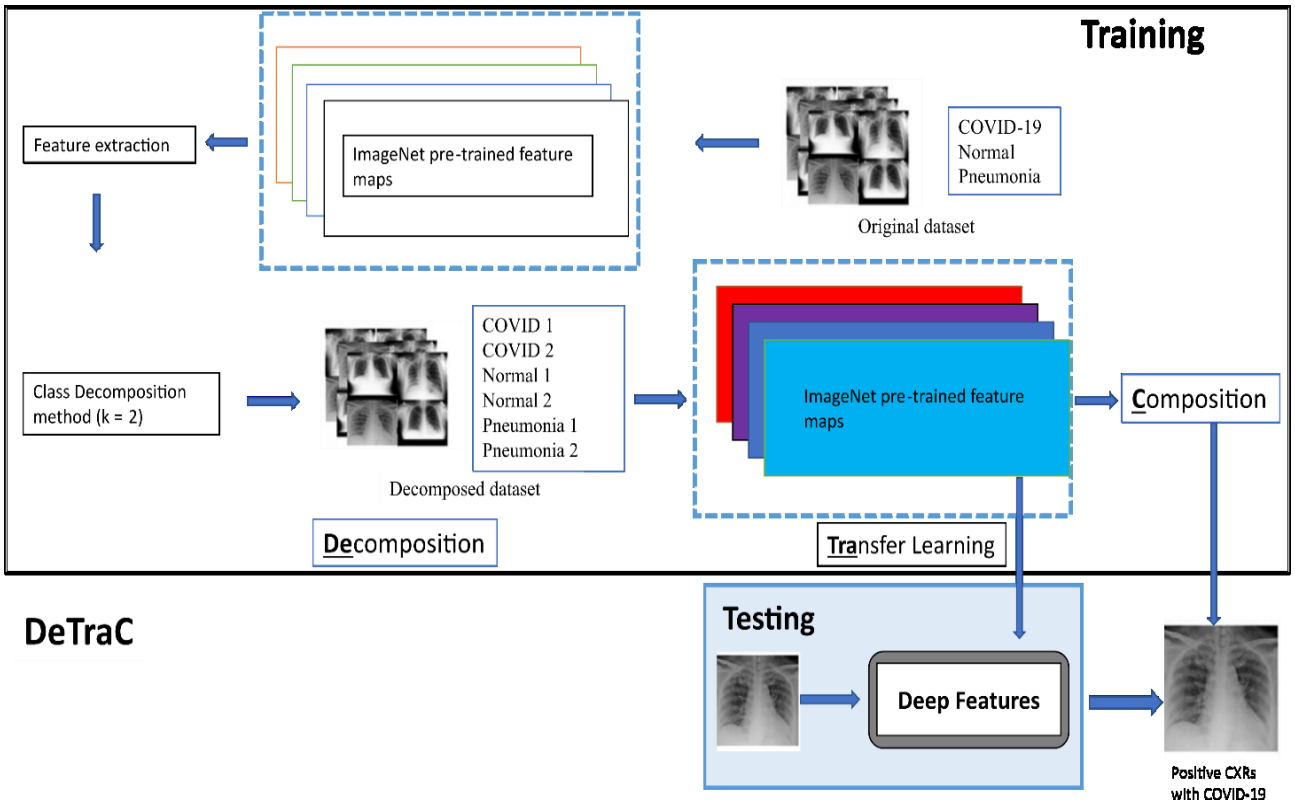

Figure 1: Flow chart outline of DeTraC method used to generate results.

structure of the dataset, and consequently enhancing the effectiveness of the model to deal with any irregularities (Abbas et al. 2020b) in the data distribution. This is achieved by making the complex problem easier to learn.

To illustrate the idea behind class decomposition, assume that the original dataset is denoted as $\boldsymbol{A}$, where $\boldsymbol{A}$ can be represented as:

$$\boldsymbol{A} = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots \dots, \boldsymbol{a}_{\mathrm{n}}\},$$

where n number of images, and each image can be represented as a set of features as:

$$\boldsymbol{a}_{\mathbf{i}} = (\mathrm{a}_{\mathrm{i}1}, \mathrm{a}_{\mathrm{i}2}, \dots, \mathrm{a}_{\mathrm{in}}).$$

Moreover, assume that $\boldsymbol{L}$ is a class category of dataset $\boldsymbol{A}$, then ($\boldsymbol{A}$, $\boldsymbol{L}$) can be rewritten as:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \boldsymbol{L} = \{l_1, l_2, \dots, l_k\},$$

where k is the number of classes while m is the number of features used for each image.

Class decomposition aims to divide each class in a dataset independently into subclasses. For example, if dataset $\boldsymbol{A}$ denoted to CXR images with 2 classes (i.e. normal and abnormal) then each class in $\boldsymbol{L}$ will be divided into two classes, resulting in a new dataset (denoted as dataset $\boldsymbol{B}$) with 4 sub-classes. Therefore, the relationship between dataset $\boldsymbol{A}$ and $\boldsymbol{B}$ can be mathematically described as:

$$\boldsymbol{A}=(\boldsymbol{A}|\boldsymbol{L}) \mapsto \boldsymbol{B}=(\boldsymbol{B}|\boldsymbol{C}),$$

where both $\boldsymbol{A}$ and $\boldsymbol{B}$ have an equal number of instances and $\boldsymbol{C}$ contains labels of the new subclasses, e.g.,

$$\boldsymbol{C} = \sum_{i=1}^{\kappa} \sum_{j=1}^{c} L_{ij} \ , \quad c = 2$$

Accordingly, the feature space for dataset $\boldsymbol{A}$ and $\boldsymbol{B}$ can be illustrated as:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{11} & \dots & a_{1n} & \ell_1 \\ a_{21} & a_{22} & \dots & a_{2n} & \ell_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ell_2 \\ a_{m1} & a_{m2} & \dots & a_{mn} & \ell_2 \end{bmatrix},$$

$$\boldsymbol{B} = \begin{bmatrix} b_{11} & b_{11} & \dots & b_{1n} & \ell_{11} \\ b_{21} & b_{22} & \dots & b_{2n} & \ell_{1c} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ell_{21} \\ b_{m1} & b_{m2} & \dots & b_{mn} & \ell_{2c} \end{bmatrix}$$

Once the training is accomplished those subclasses are recombined back to the original problem, by relabelling patterns of new subclasses (Abbas, Abdelsamea and Gaber, 2020).

In this paper, we study the effect of this decomposition granularity on the performance of DeTraC (with different training modes) in detecting COVID-19 cases using chest X-ray images.

RESULTS

This section includes the dataset that we used and details our hyperparameters setting as well as discussing the experimental results.

The experiments were carried out in Google Colab on a laptop with the following configuration Intel(R) Core (TM) i7-8550U CPU with 8.00 GB RAM.

Dataset

The datasets used were collected from Kaggle (Rahman Tawsifur, 2021), and the owners of the dataset collected the images in collaboration with medical doctors.

The datasets used were composed of:

- 1200 samples of COVID-19 chest X-rays
- 1341 samples of Normal chest X-rays
- 1345 samples of Viral Pneumonia chest X-rays

DeTraC adaptation

In this paper, a shallow-tuning mode was used during the adaptation, weight initialisation, and training of the AlexNet pre-trained model. We used the off-the-shelf CNN local features (extracted from the last fully connected layer) of pre-trained models on the ImageNet dataset. However, due to the high dimensionality of the feature space associated with the images, we applied principal component analysis (PCA) to project the high-dimension feature space into a lower dimension (where the highly correlated features were ignored). This step was important for the class decomposition layer of DeTraC to produce more homogeneous classes and reduce the memory requirements.

In addition to the fact that DeTraC can cope with data irregularities using its class decomposition layer, DeTraC can also provide an efficient solution to overcome the limited availability of training images. This is by transferring knowledge from a generic object recognition task to our specific-domain tasks using ImageNet pre-trained model (e.g., ResNet) as adopted in the transfer learning component of DeTraC.

Hyperparameter settings

The datasets were split into training (80%) and testing (20%) sets. Then, to begin our investigation on the behaviour of the class decomposition component on DeTraC, the elbow method was adopted to change the value of k in the k-means algorithm. To begin with, the value of k was 2 and then was increased by +1 each time until k = 10. The k-means clustering was applied to all the classes (normal, COVID-19, and Pneumonia). Once the method of changing the value of k had been decided, the hyperparameter settings were set. The number of epochs was 5, the batch size was 64, the number of classes was 3, the number of k-folds was 5 and the learning rate for the feature extractor and feature composer was 0.001. They were kept the same for each value of k. Each value of k was run on both the shallow tuning mode and the deep tuning mode to see if there was a difference in the performance. Shallow tuning is where all the pre-trained layers have their weights frozen and the custom classification layer has its weights active. Deep tuning is where all the layers have their weights active (Abbas et al. 2020a). The pre-trained model has 37 layers. The fine-tuning mode allows the user to choose the number of layers to be frozen, but this option was not used for this experiment. The changing accuracy of the model was used as the evaluation metric, see Table 1.

Table 1: Summary table showing the results of changing the value of k on DeTraC.

Value of k	Shallow-Tuning Accuracy (%)	Deep-Tuning Accuracy (%)
2	94.64	97.49
3	94.44	97.66
4	93.98	97.82
5	92.23	97.41
6	91.81	98.01
7	88.91	98.33
8	94.82	98.15
9	92.61	97.78
10	92.24	97.59

DISCUSSION AND CONCLUSION

Class decomposition has been proposed to enhance low variance classifiers facilitating more flexibility to their decision boundaries. In (Abbas et al. 2021), we previously validated DeTraC for the detection of COVID-19 using chest X-ray images when data irregularities challenging problem is presented. This is by adding a class decomposition layer to the pre-trained models, which aims at partitioning each class within the image dataset into sub-classes and then assign new machine labels to the new set, where each subset is treated as an independent class, then those subsets are assembled back to produce the final predictions. Our previous experimental results showed the robustness of DeTraC in the detection of COVID-19 cases from a comprehensive image dataset. High accuracy of 95.12% with a sensitivity of 97.91%, a specificity of 91.87%, and a precision of 93.36%, was achieved in the detection of COVID-19 images from normal, and severe acute respiratory syndrome (SARS) cases.

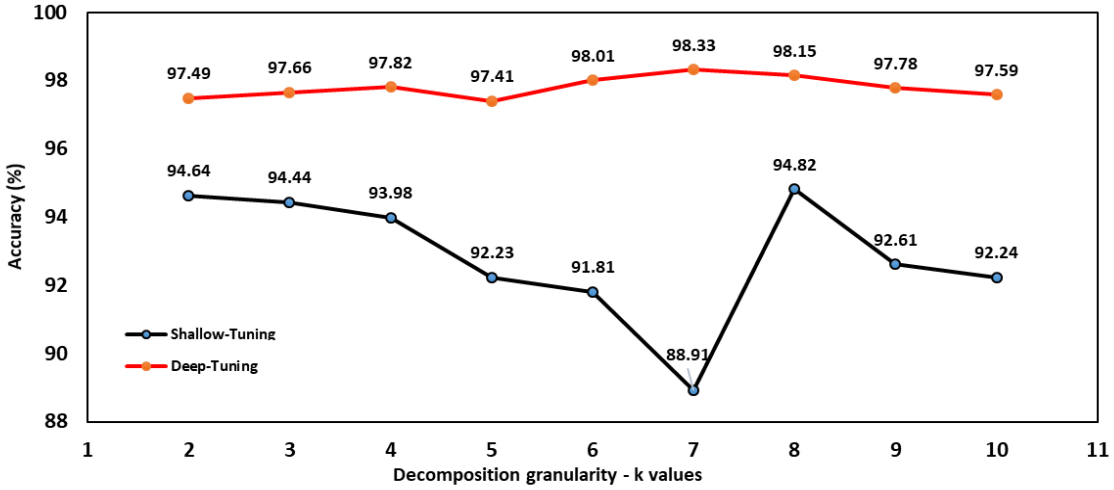

Figure 2: The effect of decomposition granularity on DeTraC.

In this paper, we study the effect of decomposition granularity on our previously proposed convolutional neural network architecture (DeTraC). From the results demonstrated in the previous section, there is a clear link between the value of k and the accuracy of DeTraC. Both the shallow and tuning modes behaved differently.

For the shallow tuning mode, when the value of k increased, the accuracy decreased. However, when k was equal to 8, the accuracy rose significantly (94.82%) before dropping again for k =9 (92.61%) and k = 10 (92.24%). This suggests for the shallow tuning mode, k = 8 is the optimal value to get the highest accuracy.

For the deep tuning mode, when the value of k increased, the accuracy increased also. When k was equal to 7, the accuracy was 98.33% – similar to shallow tuning and then dropped until k =10.

It is interesting, however, to observe that the decomposition granularity has an oppositive behaviour between the shallow and the deep tuning modes. In shallow tuning, the small values of k seem to have a good effect on the model's accuracy, and then higher values have a negative effect, before lifting up again at k=8, see Figure 2. This suggests that it may be the case that with shallow tuning, the learning of a higher number of classes can be more challenging. However, a more steading behaviour has been shown with deep tuning with a peak at k=7. This suggests that with the effectiveness of deep tuning, the decomposition granularity can be investigated carefully to gain the best possible outcome.

This study has shown that changing the parameters affects the behaviour of DeTraC. The value of k influences the overall accuracy of the algorithm in both the shallow and deep tuning modes. This is a positive development and allows for a better understanding of DeTraC and its use as a diagnostic tool.

In the future, it would be interesting to see how different clustering algorithms other than k-means affect DeTraC and if the results obtained would be similar. DeTraC has proved to be an effective tool in predicting COVID-19 and produces reliable results. The highest accuracy for the deep tuning mode was 98.33% which is similar to the 98.23% obtained by (Abbas et al. 2020a) which is a testament to the algorithm's reproducibility even with different images being used.

REFERENCES

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

A. Abbas, M. M. Abdelsamea and M. M. Gaber (2020a) "DeTraC: Transfer Learning of Class Decomposed Medical Images in Convolutional Neural Networks," in IEEE Access, vol. 8, pp. 74901-74913.

Johns, Harold Elford, and John Robert Cunningham. "The physics of radiology." (1983).

Esteva, Andre, et al. "Deep learning-enabled medical computer vision." npj Digital Medicine 4.1 (2021): 1-9.

Hosny, A. et al. (2018) 'Artificial intelligence in radiology', Nature Reviews Cancer. Nature Publishing Group, pp. 500–510. doi: 10.1038/s41568-018-0016-5.

Ismael, A. M. and Şengür, A. (2021) 'Deep learning approaches for COVID-19 detection based on chest X-ray images', Expert Systems with Applications, 164, p. 114054. doi: 10.1016/j.eswa.2020.114054.

Pham, Duc Truong, Stefan S. Dimov, and Chi D. Nguyen. "Selection of K in K-means clustering." Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science 219.1 (2005): 103-119. Organization, W. H. (2020) Coronavirus disease (COVID-19) (Accessed: 23 January 2021).

Parvathy, Velmurugan Subbiah, Sivakumar Pothiraj, and Jenyfal Sampson. "Hyperparameter Optimization of Deep Neural Network in Multimodality Fused Medical Image Classification for Medical and Industrial IoT." Smart Sensors for Industrial Internet of Things. Springer, Cham, 2021. 127-146.

Rahaman, M. et al. (2020) 'Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches the Creative Commons Attribution Non-Commercial License (CC BY-NC 4.0)', Journal of X-Ray Science and Technology, 28, pp. 821–839. doi: 10.3233/XST-200715.

Rahman Tawsifur (2021) COVID-19 Radiography Database Kaggle, Available at: https://www.kaggle.com/tawsifurrahman/COVID19-radiography-database (Accessed: 10 February 2021).

Stephen, O. et al. (2019) 'An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare', Journal of Healthcare Engineering, 2019, pp. 4180949–4180949. doi: 10.1155/2019/4180949.

Wang, B. et al. (2019) 'Deep convolutional neural network with segmentation techniques for chest X-ray analysis', in Proceedings of the 14th IEEE Conference on Industrial Electronics and Applications, ICIEA 2019. Institute of Electrical and Electronics Engineers Inc., pp. 1212–1216. doi: 10.1109/ICIEA.2019.8834117.

Abbas, A., & Abdelsamea, M. M. (2018, December). Learning transformations for automated classification of manifestation of tuberculosis using convolutional neural network. In 2018 13th International Conference on Computer Engineering and Systems (ICCES) (pp. 122-126). IEEE.

Abbas, A., Abdelsamea, M. M., & Gaber, M. (2020b). 4S-DT: Self Supervised Super Sample Decomposition for Transfer learning with application to COVID-19 detection. arXiv preprint arXiv:2007.11450.

Wu, Y.-C., Chen, C.-S. and Chan, Y.-J. (2020) 'The outbreak of COVID-19', Journal of the Chinese Medical Association, 83(3), pp. 217–220. doi: 10.1097/JCMA.0000000000000270.