Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi
International Journal of Engineering Research and Development

UMAGD, (2022) 14(3), s343-s350.

10.29137/umagd.1203617

Cilt/Volume:14 Sayı/Issue:3 Aralık/December 2022 Özel Sayı / Special Issue

Araştırma Makalesi / Research Article

Threshold Based Image Enhancement Method for Low Contrast X-Ray Images Using CLAHE

Buğra Hatipoğlu*[1], İrfan Karagöz [2], Mikail İnal [3]

[1]*Gazi University, Graduate School of Natural and Applied Sciences, Department of Electrical-Electronic Engineering, 06500 Ankara, TURKEY*
[2]*Gazi University, Faculty of Engineering, Department of Electrical-Electronic Engineering, 06560 Ankara, TURKEY*
[3]*Kırıkkale University, Faculty of Medicine, Department of Radiology, 71450 Kırıkkale, TURKEY*

***Başvuru/Received**: 13/11/2022* ***Kabul / Accepted**: 12/12/2022* ***Çevrimiçi Basım / Published Online:** 31/12/2022*

***Son Versiyon/Final Version**: 31/12/2022*

Abstract

While a large dataset in deep learning may seem like a positive factor, it may not always produce good results. Image quality is one of the factors that directly affects model performance, which in turn affects the quality of training. In this study, the effect of low contrast X-ray images on the detection of pneumonia and COVID-19 was investigated. Because the details are extremely important in the detection of these diseases. If the images are low contrast, it can cause some details to be missed in the detection of the disease. This issue can be solved applying adaptive histogram methods such as CLAHE. The CLAHE method can apply various filters to low contrast images to bring them to the desired levels. The data set contains 8849 human chest X-ray images. The Vgg16 model was used for training. Vgg16 is a well-known model architecture in deep learning. The image dimensions are 150x150. Classification performed before low-contrast images were filtered achieved 95.22% accuracy. After filtering based on the threshold value, accuracy increased to 97.35%. In the next stage, by searching for the best values for the parameters, accuracy was increased to 97.86%.

Key Words

"Low Contrast, İmage Processing, X-Ray İmages, Deep Learning, CLAHE"

***Corresponding Author:** bugra.hatipoglu@gazi.edu.tr*

1. Introduction

Deep learning is a subfield of ML (Machine Learning) that involves the use of multi-layered neural networks. Weight and threshold values updated based on the characteristics of the data learned during learning are used (LeCun et al., 2015). It enables learning in a highly automated manner using what is learned from the data sets. In particular, deep learning methods are used to create high-performance models that produce significantly good results for a large number of input data. After training these models, classification, object detection, and segmentation can be performed. However, there are various factors that affect the success of this learning process. The size and quality of the data are extremely important. Therefore, the primary goal is to try to increase and improve the data as much as possible. One way to improve the data is to use image processing methods before training.

The effect of low contrast on classification in deep learning is that it can make it difficult for the model to classify the input data correctly(Dodge & Karam, 2016). This is because low contrast reduces the difference between pixels in an image, which makes it difficult for the model to define the features that are important for correct classification. It causes the visibility of details to decrease. As a result, the model may make more errors or have lower accuracy when classifying low contrast images compared to high contrast images. Therefore, it will be helpful to balance the amount of contrast in the images. Contrast balancing is a technique used in image processing to improve the visual appearance of an image by adjusting its contrast. There are various methods for contrast balancing, including histogram equalization, adaptive histogram equalization, and gamma correction.

In their study, Rahman et al. (2021) examined the effect of various contrast enhancement methods on the detection of COVID-19. They used a data set of 18479 images. The U-Net architecture was used for training the data set and they achieved 96.29% accuracy. In their study, Maurya et al. (2022) developed a fusion algorithm based on Cuckoo Search to adjust the brightness and contrast balance of images. Cuckoo Search is an optimization algorithm. The proposed method offers a new way to improve conflicting features in a balanced manner. In the study, two sets of parameters were produced, one with high sharpness and contrast and the other with high brightness and detail level. These parameter sets are combined to achieve a balanced output. In their study to improve clinical studies, Qiu et al. (2017) used the CLAHE optimization function. CLAHE is an image processing technique that aims to improve contrast in images. In their study, Survarachakan et al. (2021) combined gamma correction with various enhancement filters to perform segmentation on different vessel types. As a result of this study, better results have been obtained using the image processing methods currently used in clinics compared to the results of previous studies. Munadi et al. (2020) used the CLAHE method to improve the images in the dataset for tuberculosis disease. Successful results have been obtained in this study.

In previous contrast improvement studies, various filters and methods were applied to data sets using image processing before classification. It is generally seen that filtering operations are applied to the entire data set. However, not every image in the data set may need the filtering operation. Therefore, in the study, filters were applied to images below a certain threshold value. The most suitable threshold value was selected based on the classification results. This improvement process was tested on X-ray images. The general contrast of the images was normalized with a maximum value of 1 and a minimum value of 0. This process was carried out using the CLAHE method. The best threshold value was found based on the model performance. After the threshold value was determined, tests were conducted to find the appropriate parameter values and successful results were obtained.

2. Materials and Methods

2.1. Dataset

The study includes COVID-19, lobar pneumonia, and healthy chest X-ray images obtained from two different data sets(Chowdhury et al., 2020; Daniel Kermany et al., 2018). There are a total of 8849 chest X-ray images in the data set. Of these, 3616 are COVID-19, pneumonia, 3884 are lobar pneumonia, and 1349 are healthy human chest X-rays. The data set was selected in a way that would not create classification imbalance. Examples of classification imbalances include oversampling and undersampling(Buda et al., 2018). In this data set, the test set is 75% and the training set is 25%.

2.2 Convolutional Neural Networks (CNN)

CNN is a widely preferred deep learning method because of its successes in pattern recognition and classification. CNNs use a process called convolution to separate different parts of images. Through this process, the location of features in the images is determined and classification is performed using these features. These types of artificial neural networks are used in image recognition, speech recognition, and other similar applications.

This model consists of multiple layers, such as the convolutional layer, activation function, pooling layer, and fully connected layer. There are parameters that affect these layers, and by changing these parameters, the performance of the model can be influenced. Changes in these parameters directly affect the efficiency of the CNN. Figure 1 shows an example of a CNN architecture (Albawi et al., 2017).

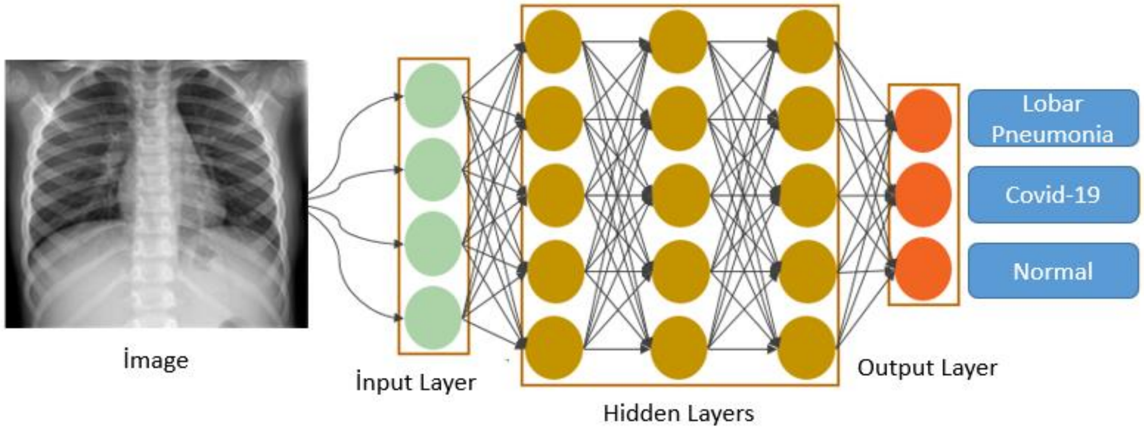

Figure 1. Architecture of the CNN

2.2. Histogram equalization

Histogram equalization is a technique used in image processing and aims to regularize the tonal characteristics of the image. Histogram equalization orders the distribution of brightness values in the image, reducing the differences between the brightness values of the image. This reduces the contrast in the image because as the differences between the brightness values decrease, the contrast of the image also decreases. Firstly, the histogram of the brightness values of the image is plotted. This histogram shows the number of all pixel values in the image. The ratio of brightness values to the image is calculated and redistributed, reducing the differences between the brightness values(Yu Wang et al., 1999).

2.3. Adaptive histogram equalization

Adaptive histogram equalization is an image processing technique that aims to improve the tonality of an image. This technique analyzes the distribution of tones in the image, adjusting the bright and dark areas of the image to increase contrast. Adaptive histogram equalization, unlike standard histogram equalization technique, examines the image in small blocks and equalizes the histograms of these blocks. This allows separate contrast enhancement processes to be performed in each region of the image. This provides a more natural appearance compared to standard histogram equalization technique and increases contrast without degrading the details of the image (Pizer et al., 1987).

2.4. Gamma correction

Gamma correction is a method used to adjust the brilliance and contrast of images by using a "gamma" value as a parameter. The gamma value is a measure used to adjust the brilliance and contrast of images and generally ranges from 0.1 to 10. As the gamma value changes, contrast and brightness are inversely proportional. For example, when gamma is reduced, the brightness of the images decreases but the contrast increases. In general, the gamma correction method is used to adjust the brightness and contrast of images and reduce the tonal difference between the light and dark areas of the images. This makes the images appear more balanced and clear (Rahman et al., 2021).

2.5. Vgg16

Vgg16 is one of the leading convolutional neural network architectures in deep learning. This architecture gained attention by achieving successful results in classification and localization in ILSVRC 2014 VGG models have achieved the best scores in terms of classification accuracy on the ImageNet dataset. Deep models were produced using traditional CNN architecture. VGG16 has 13 convolutional layers and 3 fully connected layers. There are 5 pooling layers between the convolutional layers. The type of pooling layers is maximum pooling method. ReLU is used as the activation function. The softmax function is used in the classification layer. The architecture is shown in Figure 2 (Wang et al., 2020).

Figure 2. Architecture of the Vgg16 Model

2.6. Contrast Limited Adaptive Histogram Equalization Method (CLAHE)

CLAHE is a technique used to increase the contrast in images (Reza, 2004). This technique divides the image into certain regions and performs histogram equalization in each region, thus reducing lighting changes in the image. It produces a clearer image by preserving details in the same way as the adaptive histogram equalization method. There are two important parameters in this method. It is important to choose these parameters correctly.

The clipLimit parameter specifies the limit value used when performing histogram equalization. This parameter limits the distribution of brightness values in the image and thus increases the contrast in the image. The "tileGridSize" parameter determines how many equal parts we will divide the image into. This parameter enables an adaptive histogram equalization process by dividing the image into small regions. The values of these parameters vary depending on the data sets.

2.7. Evaluation metrics

To test the performance of the model, some parameters and evaluation metrics are used. For this purpose, the confusion matrix is used. The confusion matrix includes True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) parameters. TP is the case where the model predicts an example as positive and it is actually positive. For example, correctly predicting a COVID-19 chest X-Ray images. TN is the case where the model predicts an example as negative and it is actually negative. For example, correctly predicting a non-COVID-19 chest X-Ray image. FP is the case where the model predicts an example as negative but it is actually positive. For example, incorrectly predicting a COVID-19 chest X-Ray image as lobar pneumonia. FN is the case where the model predicts an example as positive but it is actually negative. For example, incorrectly predicting a COVID-19 chest X-Ray images as COVID-19 pneumonia. The confusion matrix containing these parameters is shown in Figure 3. The real class in Figure 3 is the actual labels of our images in the class. The predicted class contains the model's predictions for the images.

There are evaluation metrics calculated with the TP, TN, FP, FN parameters to evaluate the performance of the model. These metrics are precision, recall, F1-value, and accuracy metrics. These metrics provide us with information about whether there is a problem of class imbalance in our data set (Luque et al., 2019). Four different combinations based on the positive or negative labels of these classes are given in Figure 3.

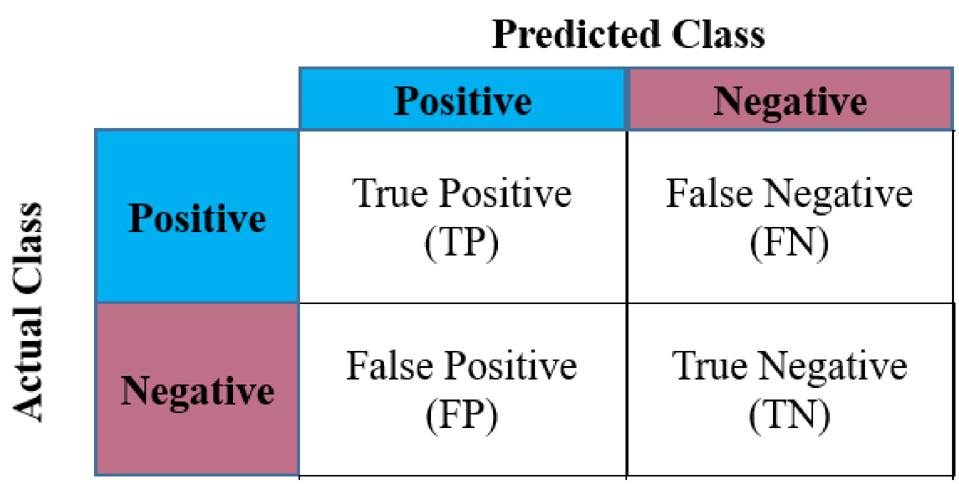

Figure 3. Confusion Matrix

Precision, one of our evaluation metrics, gives the ratio of images that are actually COVID-19 to those we predict as COVID-19. The value obtained should be close to 1. It is defined mathematically as in Equation 1.

$$precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

Recall shows how many images we need to predict as COVID-19 we are able to predict. The value obtained should be close to 1. It is defined mathematically as in Equation 2.

$$recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

The F1-score gives the harmonic mean of recall and precision. If the model is perfect, the F score is equal to 1. It is defined mathematically as in Equation 3.

$$F1 - Score = \frac{\text{2xPrecision x Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

Accuracy is the percentage of images that are correctly classified. The closer the value obtained to 1, the more successful our model is. However, a value of 100% indicates overfitting. It is defined mathematically as in Equation 4.

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

3. Results

In this study, chest X-ray images of individuals with pneumonia, COVID-19, and healthy individuals were classified. In this section, the action of the number of layers and some parameters of the model on the classification accuracy time was examined. First, Vgg16 and Vgg19 models were used to classify 64x64 images. The reason for selecting these sizes is that the training process would take too long with higher pixels. Then, a model with fewer layers was designed in order to decrease the number of parameters. This model has 4 convolutional layers, 4 maximum pooling layers, and 1 fully connected layer. The ReLU function was used as the activation function. The softmax function was preferred in the last layer before the classification stage.

In this study, image improvement was performed for low contrast images at different thresholds. These operations were applied on a dataset containing X-ray images of COVID-19, pneumonia, and healthy individuals. First, the overall contrast of the images must be found. Then, these contrast values are normalized between 0 and 1. Images with an overall contrast value lower than the defined threshold are considered low contrast images. For example, if the overall contrast of the image is 0.35 and the threshold value is 0.40, the image will be filtered. However, if the threshold is 0.30, the filter will not be applied.

After low contrast images were found, the Contrast Limited Adaptive Histogram Equalization (CLAHE) method was used. There are two different parameters that we need to decide when using this method. The value of the "clipLimit" parameter was initially selected as 2. The results for different values were also observed in later stages. The "tileGridSize" parameter determines how many equal regions we will divide the image into. This parameter was initially selected as 8x8 in size. The improvement results in a low contrast image based on the parameter values are given in Figure 4 below.

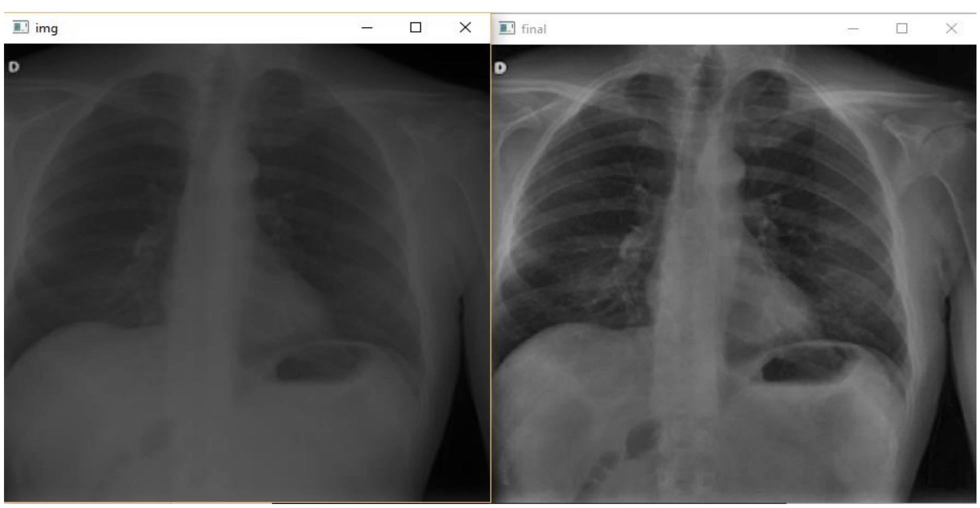

Figure 4. Comparison of Original Image and Filtered Image (cLipLimit=2.0 tileGridSize=8x8)

In the training process, images with 150x150 pixel dimensions were used. The Vgg16 model was used as the training model. This model is a multi-layered convolutional neural network model. During the training process, images with 150x150 pixel dimensions were used. COVID-19 classification was performed with the "clipLimit" parameter set to 2. The results obtained are given in the Table 1.

Table 1. Classification Results for Filtered Images Based on Different Threshold Values

Threshold Value	Precision	Recall	F1-Score	Accuracy	Number of Filtered İmages
0.75-1	0,96	0,96	0,96	%96,47	8849
0.70	0,96	0,96	0,96	%96,34	660
0.65	0,97	0,97	0,97	%96,88	538
0.60	0,97	0,97	0,97	%97,35	389
0.55	0,97	0,97	0,97	**%97,15**	87
0.50	0,96	0,96	0,96	%96,56	55
0.45	0,96	0,96	0,96	%96,43	47
0.40	0,95	0,95	0,95	%95,39	10
0.35	0.96	0.96	0.96	%96,11	10
0-0.30	0.94	0.95	0.95	%95,22	0

Upon examining the results in Table 1, we see that the best results are obtained when the threshold value is 0.60. When the threshold value is 0.60, the accuracy is 97.35%. There are 389 images that pass through the filter at this threshold value, which is 4.39% of the dataset. When classification is performed without filtering, 95.22% accuracy is obtained.

Table 2. Classification Results According to Changes in cLipLimit Parameters

cLipLimit	Precision	Recall	F1-Score	Accuracy
2	0,97	0,97	0,97	%97,35
3	0,97	0,97	0,97	%97,11
4	0,97	0,97	0,97	%97,15
5	0,96	0,96	0,96	%96,11
6	0,95	0,95	0,95	%95,25
7	**0,98**	**0,98**	**0,98**	**%97,86**
8	0,95	0,95	0,95	%95,23
9	0,97	0,97	0,97	%96,84
10	0.97	0.97	0.97	%96,84

In Table 2, the optimal clipLimit parameter value for the model is searched. Accuracy, F1-score, Recall, Precision, etc. are taken into consideration for this parameter. According to the results obtained here, accuracy has reached up to 97.86%.

Table 3. Classification Results According to Changes in tileGridSize Parameters

tileGridSize	Precision	Recall	F1-Score	Accuracy
2x2	0,96	0,96	0,96	%96,16
3x3	0,96	0,96	0,96	%96,25
4x4	0,95	0,95	0,96	%95,75
5x5	0,97	0,97	0,97	%96,88
6x6	0,96	0,96	0,96	%96,02
7x7	0,97	0,97	0,97	%97,24
8x8	**0,98**	**0,98**	**0,98**	**%97,86**
9x9	0,97	0,97	0,97	%96,56
10x10	0.96	0.96	0.96	%95,98

In Table 3, the optimal tileGridSize parameter value for the model is searched. The results of the changes in this parameter are examined according to the Accuracy, F1-score, Recall, and Precision metrics. When the tileGridSize parameter is selected at 8x8 dimensions, the best result of 97.86% is obtained. The visual result of the parameter values that give the best results in the contrast improvement process is given in Figure 5.

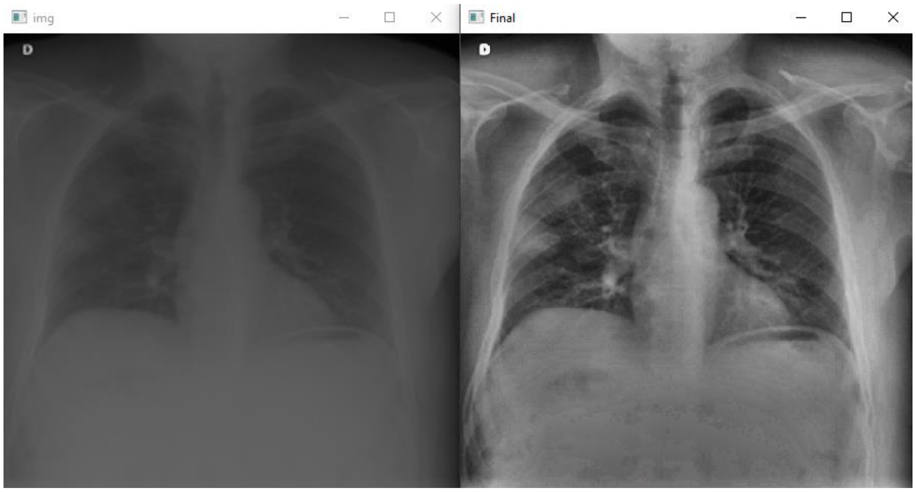

Figure 5. Comparison of Original Image and Filtered Image (cLipLimit=7.0 tileGridSize=8x8)

4. Discussions

In this study, the purpose was to detect low contrast images in the preprocessing stage and apply improvement to these images. Generally, these operations are applied using filters in the preprocessing stage. However, previous studies have applied these filters to the entire dataset. In this study, image improvement was performed based on a threshold. The reason for this is that not all images are low contrast. Therefore, it was thought that contrast balance should be performed adaptively. When we examine Table 1, we can see that the performance of the model is compared by training it with different threshold values. It has been proven that the filtering process is unsuccessful when applied to all images. This result indicates that the threshold-based system has been successful. The performance of the model has been compared by training it with different threshold values. The best threshold value of 0.60 has been selected. The accuracy value here is 97.35%. The number of filtered images is 4.39% of the dataset. However, there has been a significant improvement in the performance of the model. From this, we can conclude that low contrast images reduce the performance of the model. Afterwards, the optimal values of the parameters have been searched for and the accuracy has been increased to 97.86%. After these processes have been performed, the model architecture that we have created has been tested. In addition, successful results have been obtained visually in the study. In Figure 4 and Figure 5, it can be said that the image before and after applying the filter is more detailed visually. This can be helpful during the examination by the experts.

Acknowledgment

This study has been evaluated within the scope of the special issue of engineering sciences research and development applications.

References

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*, *2018-Janua*, 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011

Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. Bin, Islam, K. R., Khan, M. S., Iqbal, A., Emadi, N. Al, Reaz, M. B. I., & Islam, M. T. (2020). Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access*, *8*, 132665–132676. https://doi.org/10.1109/ACCESS.2020.3010287

Daniel Kermany, Kang Zhang, & Michael Goldbaum. (2018). *Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification - Mendeley Data*. https://data.mendeley.com/datasets/rscbjbr9sj/2

Dodge, S., & Karam, L. (2016). Understanding how image quality affects deep neural networks. *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, 1–6. https://doi.org/10.1109/QoMEX.2016.7498955

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231. https://doi.org/10.1016/j.patcog.2019.02.023

Maurya, L., Lohchab, V., Kumar Mahapatra, P., & Abonyi, J. (2022). Contrast and brightness balance in image enhancement using Cuckoo Search-optimized image fusion. *Journal of King Saud University - Computer and Information Sciences*, *34*(9), 7247–7258. https://doi.org/10.1016/J.JKSUCI.2021.07.008

Munadi, K., Muchtar, K., Maulina, N., & Pradhan, B. (2020). Image Enhancement for Tuberculosis Detection Using Deep Learning. *IEEE Access*, *8*, 217897–217907. https://doi.org/10.1109/ACCESS.2020.3041867

Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J. B., & Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, *39*(3), 355–368. https://doi.org/10.1016/S0734-189X(87)80186-X

Qiu, J., Harold Li, H., Zhang, T., Ma, F., & Yang, D. (2017). Automatic x-ray image contrast enhancement based on parameter auto-optimization. *Journal of Applied Clinical Medical Physics*, *18*(6), 218–223. https://doi.org/10.1002/ACM2.12172

Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Abul Kashem, S. Bin, Islam, M. T., Al Maadeed, S., Zughaier, S. M., Khan, M. S., & Chowdhury, M. E. H. (2021). Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Computers in Biology and Medicine*, *132*, 104319. https://doi.org/10.1016/j.compbiomed.2021.104319

Reza, A. M. (2004). Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. *Journal of VLSI Signal Processing*, *38*, 35–44.

Survarachakan, S., Pelanis, E., Khan, Z. A., Kumar, R. P., Edwin, B., & Lindseth, F. (2021). Effects of Enhancement on Deep Learning Based Hepatic Vessel Segmentation. *Electronics*, *10*(10), 1165. https://doi.org/10.3390/electronics10101165

Wang, W., Tian, J., Zhang, C., Luo, Y., Wang, X., & Li, J. (2020). An improved deep learning approach and its applications on colonic polyp images detection. *BMC Medical Imaging*, *20*(1), 83. https://doi.org/10.1186/s12880-020-00482-3

Yu Wang, Qian Chen, & Baeomin Zhang. (1999). Image enhancement based on equal area dualistic sub-image histogram equalization method. *IEEE Transactions on Consumer Electronics*, *45*(1), 68–75. https://doi.org/10.1109/30.754419