ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika

ISSN(p): 2338-8323 | ISSN(e): 2459-9638 | Vol. 11 | No. 2 | Halaman 424 - 438
DOI : http://dx.doi.org/10.26760/elkomika.v11i2.424 April 2023

Study on Image Processing Method and Data Augmentation for Chest X-Ray Nodule Detection with YOLOv5 Algorithm

ADINDRA VICKAR EGA, WUWUS ARDIATNA

National Research and Innovation Agency Republic of Indonesia (BRIN), Indonesia
Email : adin009@brin.go.id

Received 15 Februari 2023 **|** *Revised* 14 Maret 2023 **|** *Accepted* 21 Maret 2023

ABSTRAK

Saat ini, deteksi dini kanker paru-paru dapat dilakukan dengan sistem Computer Aided Diagnosis (CAD) berbasis AI. Oleh karena itu, peningkatan perfoma sistem CAD sangat diperlukan. Dalam studi ini, berbagai teknik pengolahan citra dan augmentasi data diterapkan untuk mengevaluasi performa deteksi nodul paru-paru pada citra X-Ray dada dengan algoritma YOLOv5. Dataset publik yang terdiri dari 1500 data latih dan 516 data uji beserta dengan anotasi nodulnya digunakan dalam simulasi. Hasil simulasi menunjukkan bahwa model YOLOv5 menghasilkan presisi, recall, dan nilai mAP yang tinggi dengan nilai masing-masing 0,811, 0,776, dan 0,858, walaupun tidak menggunakan teknik pengolahan citra dan augmentasi data. Hasil validasi silang dengan dengan dataset publik JSRT dengan augmentasi data tiga kali menunjukkan bahwa YOLOv5s memiliki performa yang lebih baik untuk deteksi nodul pada paru-paru dibandingkan variasi model YOLOv5s lainnya, dengan nilai presisi 0,719 dan nilai recall 0,630.

***Kata Kunci**: nodul paru, YOLOv5, X-Ray dada, pengolahan citra, augmentasi data*

ABSTRACT

Recently, early detection of lung cancer can be performed with AI-based Computer Aided Diagnosis (CAD) system. Therefore, the performance improvement of CAD is urgently needed. In this study, various image processing and data augmentation techniques were carried out to evaluate the performance of lung nodule detection on chest X-Ray images using YOLOv5 object detection algorithms. Public dataset consist of 1500 train and 516 test data along with the annotated nodules were used. The simulation results showed that the YOLOv5 model produced high precision, recall, and mAP@0.5 values of 0.811, 0.776, and 0.858, respectively although no data augmentation and image processing techniques were performed on the previous dataset. Cross-validation results with JSRT public dataset with three times the augmentation data sets showed that YOLOv5s has better performance for nodule detection with the other model with precision and recall of 0.719 and 0.630, respectively.

***Keywords**: lung nodule, YOLOv5, X-Ray, image processing, data augmentation*

1. INTRODUCTION

Lung cancer is one of the most dangerous diseases that causes the highest death rate in the world. Based on the Global Cancer Statistics data in the year 2020, lung cancer has the second highest number of new cases after breast cancer. However, lung cancer has the highest dead cases (1.796.144), compared with the dead cases caused by breast cancer (684.996) **(Sung et al., 2021).** Early detection of lung cancer holds an important role in order to increase the percentage of patient survival rate up to 40%. Unfortunately, more than 70% of the patients with lung cancer are diagnosed when the tumor has reached the advanced stage, so that the medical treatment such as therapy or operation becomes difficult to be conducted. This is related with the inaccurate and insensitive lung cancer early detection method, which causes undetected lung nodules at the early stage of the disease **(Knight et al., 2017).**

Recently, the method used for early detection of lung cancer is a deep learning-based Computer Aided Diagnosis (CAD) system, which can assist radiologists in detecting lung nodules on X-Ray images **(Chen et al., 2011)**. Previous research has shown that the accuracy of radiologist readings can be significantly improved if the CAD system has a low false positive (FP) rate with a sensitivity of 80% **(Matsumoto et al., 1993).** However, the detection accuracy is still relatively low. This is due to the general visual appearance of nodules on chest X-ray images, blocked by the shadow of the patient's ribs **(Gang et al., 2018).** The application of bone shadow suppression techniques in previous studies showed an increase in nodule detection rates, in contrast, the detection error rate also increased **(Berg et al., 2015)**.

In the previous study, machine learning models with statistical features were often used to solve the lung nodule detection problem in chest X-Ray images. Orban, et. al. has developed machine learning-based lung nodule detection with Support Vector Machine (SVM) model by using the statistic feature from the texture, geometry, and the location of the nodule. Several Image processing techniques such as nodule enhancement, resizing, and normalizing were used to pre-process the chest X-Ray image dataset. The result shows sensitivity up to 77% with 3 False Positive Rate (FPR) **(Orbán & Horváth, 2011)**. Ogul, et.al. also has developed the machine learning-based lung nodule detection with SVM model and additional image processing techniques such as grey-level transform and weighted convergence index (WCI) filter in the pre-processing stage. The result shows 80% sensitivity with an average of 6.4 False Positive (FP) **(Oğul et al., 2015)**. Thus, further research needs to be conducted in order to evaluate the appropriate artificial intelligence (AI) model and image processing technique to improve the detection accuracy for lung nodule detection problem.

Udeshani et.al. has conducted the comparison of recognition rate performance results from neural networks based on pixel-based intensity values and based on first-order and second-order statistical features, in detecting nodules chest X-ray images. The results showed that his study was better by using an input pixel-based intensity value with a recognition rate of 96%, compared to inputs using feature-based by 88% **(Udeshani et al., 2011)**. Li, et.al. has proposed solitary feature-based pulmonary. The proposed method has better results for the lung nodule detection than previous studies using JSRT datasets. As many as 80% of lung nodules can be detected with an average false positive (FP) of 2 per images, and 93% detection with an average FP of 5 **(Li et al., 2018)**. Ausawalaithong, et.al has proposed automatic lung cancer predictions using transfer learning. From base model, parallel retraining was carried out, ChestX-ray14 dataset for the introduction of lung nodule cases using JSRT dataset with three models (A, B, and C). The results showed that model A produced the highest accuracy with 84.02%. Model B produced the highest specificity with 80.95 ± 20.59%.

Compared to model B, model C has higher accuracy and sensitivity, at 74.43 ± 6.01% and 74.68 ± 15.33% as well as smaller standard deviation values across all parameters **(Ausawalaithong et al., 2018)**.

Improvement of lung nodule detection performance can be done by eliminating rib shadows on chest X-ray images **(Gordienko et al., 2017)**. Pre-processed datasets without ribs, resulting in better accuracy and loss compared to the same dataset after lung segmentation. Li, et.al. proposed a CNN-based multi-resolution. The results showed that for the JSRT database public datasets, more than 99% of lung nodules can be detected if the false positive value per image (FPs/image) is 0.2. The proposed method also yields a FAUC value of 0.982 and an R-CPM value of 0.987 **(Li et al., 2020).** Schultheiss, et.al. used a one-stage detector (RetinaNet)-based CNN model for the training phase using 257 private dataset and 154 additional public dataset (JSRT), both with annotated nodules x-ray images. The training process for pulmonary segmentation was carried out using the U-Net architecture. The CNN performance then compared with two radiologists in detecting lung nodules from 75 cases (test phase). The results showed that from 75 cases, RetinaNet produced 43 True Positive (TP) identifications, 26 False Positive (FP) and 22 False Negative (FN). In comparison, the performance of both radiologists are 42±2 TP, 28±0 FP, and 23±2 FN **(Schultheiss et al., 2020).**

Luo, Y. et.al. used YOLOv5 and ResNet50 for the detection of lung abnormalities. In his research, YOLOv5 was used for localization of abnormality areas or regions of interest (ROI) in chest X-ray images. Meanwhile, ResNet is used to avoid gradient explosion problems in DL for classification. The dataset used was VinBigData, consisting of 18000 image scans, of which 15000 data are used as trained datasets, and 3000 data as test datasets. The results showed that using the same dataset, the YOLOv5 + ResNet50 model produced higher mAP and precision values (0.254 and 0.512) compared to the results from YOLOv5 (0.244 and 0.494), Fast RCNN (0.234 and 0.485), and EfficientDet (0.231 and 0.479) **(Luo et al., 2021)**.

Unlike the previous machine learning technique in which the statistical features were prepared manually, feature extraction in deep learning can be conducted automatically. YOLO stands for You Only Look Once and it is one of the most accomplished and friendly AI architecture for deep learning-based object detection models with state-of-the-art performances **(Maindola, 2021)**. As for Fast R-CNN has reduced the running time of the detection networks, exposing region proposal computation as a bottleneck **(Ren et al., 2015)**. Meanwhile, RetinaNet uses a feature pyramid network to effectively detect objects at multiple scales and introduces a new loss, the Focal loss function, to ease the problem of the extreme foreground-background class imbalance **(Humbarwadi, 2020)**. These deep learning-based object detection techniques have their own advantages. However, these techniques have not been found much as other machine learning techniques. Thus, this paper aims are to analyze the impact of image processing to chest X-Ray image datasets on the detection accuracy and determine the most suitable image processing and object detection technique to be used as well as to improve the lung nodule detection performance and reduce the possibility of nodule detection errors.

2. MATERIALS AND METHODS

Figure 1 describes the flow of steps conducted during the research. Starting with the preparation of dataset in the local drive, pre-processing, image processing and data augmentation, until datasets were ready to be upload to the Google Drive folder to conduct the model training and validation using the Google Colab.

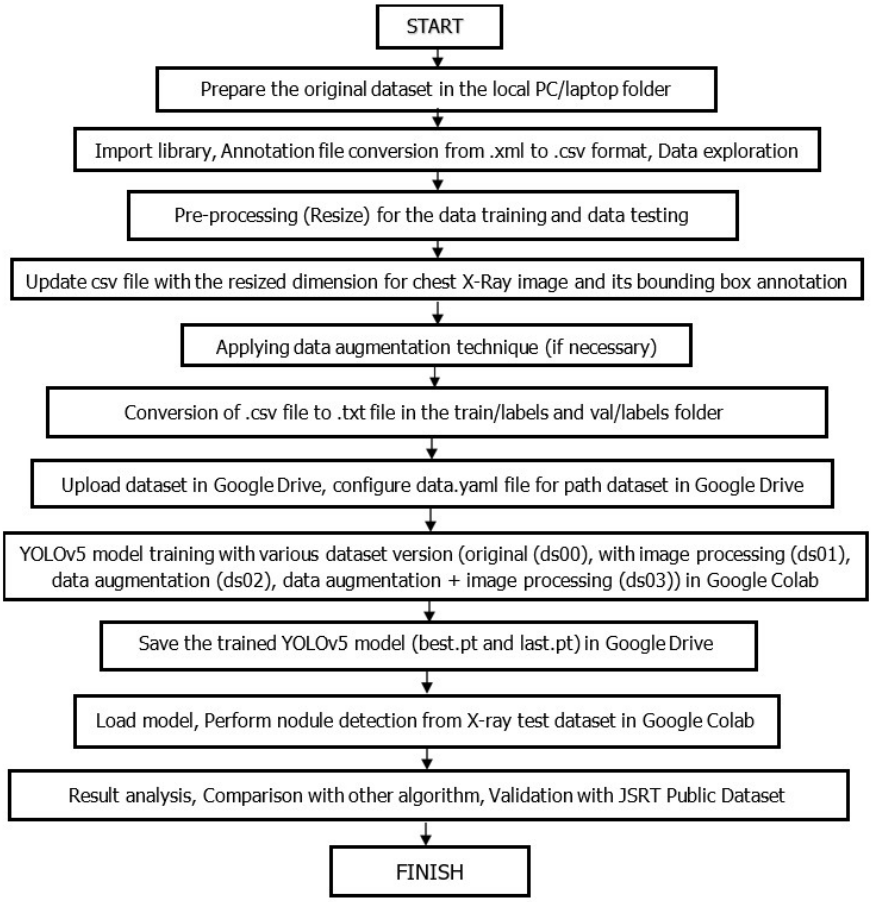
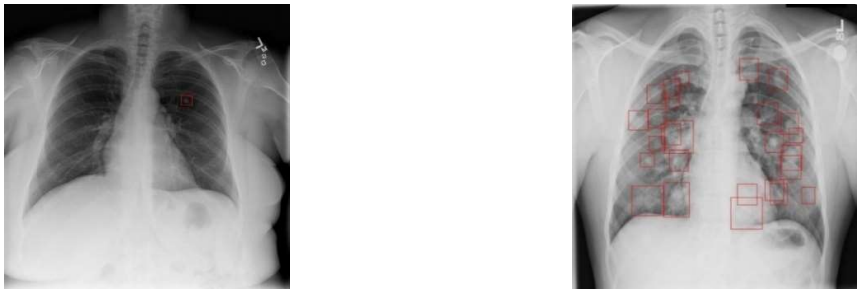


Figure 1. Flowchart of the Method Steps Conducted in the Study

2.1 Dataset

The dataset used in this study was the chest X-Ray (CXR) image which was provided by Zhang Jin **(Jin, 2018)** in the Kaggle website. It consists of 1500 (74,4%) and 516 (25,6%) chest X-ray chest image data for training and testing, respectively with annotated nodule labels for each image, in the .xml format file. In the dataset, there were not exist for the healthy CXR image samples which did not contains any nodule. Figure 2 describes the visual distinction between the considered normal CXR image which contains single real nodules in Figure 2(a), with the abnormal CXR image with multiple real nodules that contains 26 nodules in Figure 2(b). The lung nodule has the typical visual view of irregular small to large round shape that contains in the chest area.



(a) **(b)**

Figure 2. Sample of CXR Image Dataset (a) Normal with Less Nodule (b) Abnormal

2.2 YOLO

To perform the detection, YOLO splits the input image into N grids or cell, with each cell having an equal dimensional region of S x S. Each cell is responsible for detection and location of the object contains with bounding box coordinate, object label, and probability of the object being present in the cell. To provide unnecessary overlapping bounding boxes, YOLO repeatedly suppresses all the bounding boxes with lower probability scores and intersection over union (IOU), by using the non-maximal suppression (NMS), until final bounding boxes were obtained. Figure 3 below describes the architecture model of the YOLOv5. It consists of three main parts, which are backbone, neck, and head. YOLOv5 has a cross stage partial network from Darknet model (CSPDarknet) as the backbone to perform the feature extraction. Then, PANet (path aggregation network) in the neck is used for feature fusion with the feature pyramid network (FPN) structure in the second layer. Lastly, the head part of the YOLOv5 layer is used for multi-scale detection with each of the corresponding feature maps **(Ding et al., 2021)**.

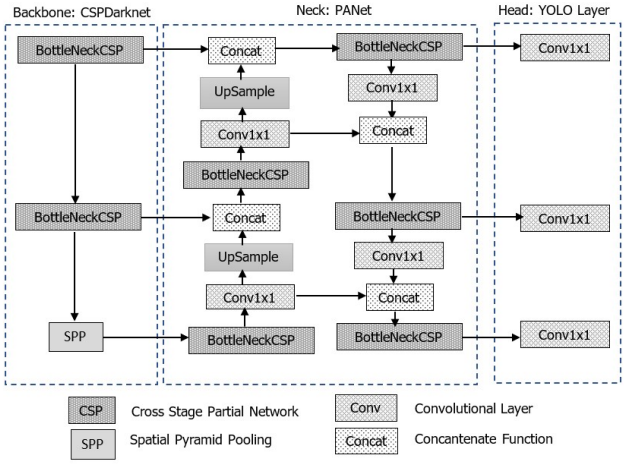


Figure 3. Architecture of the YOLOv5 Model (Ding et al., 2021)

2.3 Data Augmentation

Local Jupyter Notebook was used to pre-process the dataset and apply augmentation techniques to increase the number of the dataset along with the variability. The techniques used were the basic geometry augmentation technique such as scale, rotate and flip. During the implementation, the combined techniques can be conducted sequentially or randomly for each image on the dataset. Figure 4 below describes the sample result of the rotate -45° augmentation technique. The transformed chest X-Ray image during data augmentation process was followed with its corresponding bounding boxes annotation.

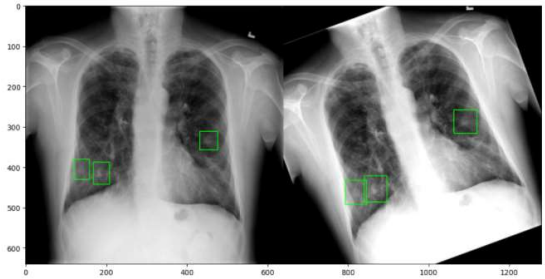


Figure 4. Sample of Data Augmentation Technique with Rotate -45°

The final pre-processed and augmented dataset is then uploaded to Google Drive folder for training and the testing of the model in the Google Colab platform.

2.4 CLAHE

CLAHE (Contrast Limited Adaptive Histogram Equalization) is an image processing technique that enhances the contrast of an image by redistributing its pixel intensity values **(Hayati et al., 2023)**. The traditional histogram equalization techniques can often result in over-amplifying noise and create unwanted artifacts, particularly in images with local variations in contrast. CLAHE solves this problem by dividing the image into smaller regions, while also constraining the maximum amplification of the histogram equalization to avoid over-enhancement. In order to implement CLAHE, a sliding window approach is used where the image is divided into overlapping tiles or patches. For each patch, a histogram equalization is applied, and the intensity values are mapped into new values based on the cumulative distribution function of the histogram. To constrain the amplification of the histogram equalization, a clipping function is applied to limit the range of the cumulative distribution function. The output of the CLAHE algorithm is an image with improved contrast and reduced local variation in brightness.

2.5 Model Evaluation

In this experiment, the model performance was evaluated with precision, recall, and mean average precision (mAP). With default intersection over union (IoU) threshold value of 0.5, hence the detection results with IoU value greater or equal 0.5 was defined as True Positive (TP). If it less than 0.5, the detection was defined as False Positive (FP). If the model could not detect objects while the ground truth exists on the image, the detection was defined as False Negative (FN). Based on the number of TP, FP, and FN, the precision and recall of the model can be calculated by using the Equation (1) and Equation (2) :

$$Precision = (TP)/\,(TP + FP) \tag{1}$$

$$Recall = (TP)/\,(TP + FN) \tag{2}$$

$$Mean\ Average\ Precision\ (mAP)\ = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \tag{3}$$

Mean average precision (mAP) was the average of average precision (AP) as described in Equation (3), where n was the number of class, while APk was the average AP for certain class. Basically, AP is defined as the area under precision-recall curve. In case of multi-class, the AP value for each class were averaged in order to obtain the mAP value. However, since there is only one class exist in this nodule detection case, hence there is difference between the AP and mAP in this case. Two types of mAP was used in this study, which were mAP@.5 and mAP@.5:0.95. The mAP@.5 parameter is the definition of mAP with IoU threshold of 0.5, while the mAP@.5:0.95 parameter has the definition of the average of mAP with varied IoU threshold from 0.5 to 0.95 with step value of 0.05. The higher the mAP value, the more accurate the detection result obtained from the model used **(Jung et al., 2018)**.

3. RESULTS AND DISCUSSION

3.1 YOLOv5 Training Result with Variation of Resized Original Image Dataset (ds00)

Table 1 describes the training results of YOLOv5 model with various resized original image dataset without applying any further image processing technique. The precision, recall, and mAP value tend to decrease as the image dataset getting resized to the smaller resolution down to 256 x 256 pixels. The training with higher resolution of the image dataset could generates higher evaluation metric results. Model with image dataset size of 1280x1280 pixels gives a high precision value of 0.812, and mAP@0.5 of 0.855 from training results with 52

epochs. However, the most optimum results was achieved by using the image size dimension of 800x800 and 640x640 pixels with mAP@0.5 value of 0.865 and 0.858, respectively.

Table 1. YOLOv5 Training Results with Variation of Resized Original Dataset

No	Model	P	R	mAP@0.5	mAP@ 0.5:0.95	Epoch
1	YOLOv5s-ds00_1280x1280	0.812	0.784	0.855	0.499	52/199
2	YOLOv5s-ds00_1024x1024	0.811	0.765	0.851	0.509	106/199
3	YOLOv5m-ds00_1024x1024	0.783	0.788	0.831	0.479	50/199
4	YOLOv5s-ds00_800x800	0.797	0.793	0.865	0.512	49/49
5	YOLOv5s-ds00_640x640	0.811	0.776	0.858	0.504	49/49
6	YOLOv5s-ds00_416x416	0.794	0.747	0.835	0.473	49/49
7	YOLOv5s-ds00_256x256	0.743	0.668	0.746	0.399	199/199
8	YOLOv5m-ds00_256x256	0.783	0.641	0.750	0.413	192/199

The number of epochs in this case was standardized to 50 epochs due to the stabilize value of the metrics after 50 epochs, as it can be seen on Figure 5. The increasing of object losses in Figure 5 also becomes the main reason to set the training up to 50 epochs, in order to prevent the overfitting issue.

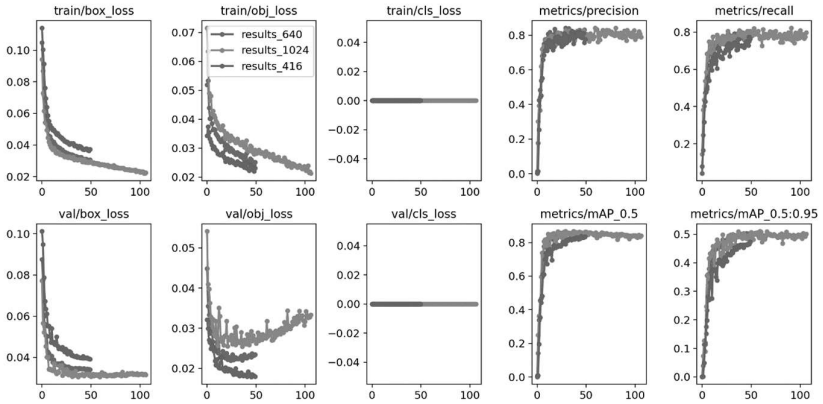


Figure 5. Comparison of YOLOv5 Model Train Result with respect to Image Size Variation

The use of different YOLOv5 model, as for an example, the medium size of the YOLOv5 model (YOLOv5m), could improve the training results for nodule detection. However, the improvement of those parameter was relatively small compared with the higher computation load, as well as the longer training duration of the model. Therefore, the small version of the YOLOv5 (YOLOv5s) was used in the next simulation to prevent training failure in Google Colab due to the high computation load. The YOLOv5s model was used to investigate further the effect of image processing and data augmentation technique to the training results.

3.2 YOLOv5 Training Result with Variation of Image Processing Technique (ds01)

Table 2 describes the training results of YOLOv5s model with variation of image processing technique applied to the original Chest X-Ray (CXR) dataset. While the original dataset was denote as 'ds00', the modified original dataset due to the image processing technique was denoted as 'ds01'. Sample of the modified CXR image dataset due to various applied image processing technique is described in Figure 6. YOLOv5 model training with this technique generally not better than the previous training results by means of using the original dataset, although there is a slightly improvement of the training results with applied combination of DIP to the CXR image size of 1024x1024 pixel. Among the image processing technique used, the DIPN generates the highest mAP@0.5 value with 0.857 for 640x640 pixels. It might be

caused by the lung nodule that generally has the higher intensity of grey color, compared to the intensity of grey color at the bone which overshadow it. Thus, when it is inverted, the nodule can be seen clearly since it has darker intensity compared to its background color intensity. Meanwhile, the combination of CLAHE and DIPCN produce the highest precision among the other technique. It can be caused by the implementation of CLAHE which overcome the drawback of histogram equalization, that prevent the increase of additional noise on the background, which is usually generates from the usual histogram equalization technique.

Table 2. YOLOv5 Training Results with Variation of Resized Original Dataset

No	Model	P	R	mAP@0.5	mAP@ 0.5:0.95	Epoch
1	YOLOv5s-ds01_1024x1024_DIP	0.811	0.789	0.86	0.509	66/199
2	YOLOv5s-ds01_640x640_DIP	0.809	0.743	0.84	0.497	49/49
3	YOLOv5s-ds01_640x640_DIP+	0.8	0.729	0.825	0.485	49/49
4	YOLOv5s-ds01_640x640_DIP++	0.774	0.744	0.824	0.483	49/49
5	YOLOv5s-ds01_640x640_DIPN	0.782	0.804	0.857	0.503	49/49
6	YOLOv5s-ds01_640x640_DIPC	0.8	0.744	0.834	0.481	49/49
7	YOLOv5s-ds01_640x640_DIPCN	0.826	0.743	0.839	0.481	49/49
8	YOLOv5s-ds01_640x640_DIPNC	0.781	0.763	0.831	0.485	49/49
9	YOLOv5s-ds01_256x256_DIP	0.759	0.622	0.726	0.388	199/199
10	YOLOv5s-ds01_256x256_bse_DIP	0.689	0.643	0.709	0.375	199/199

DIP = Grayscale, Histogram Equalization (HE)
DIP+ = Grayscale, Median Filter 5x5, Histogram Equalization
DIP++= Grayscale, HE, Top Hat Transform, Black Hat Transform
DIPNC = Negative + CLAHE
DIPN = Negative Image
DIPC = CLAHE
bse = bone shadow exclusion
DIPCN = CLAHE + Negative

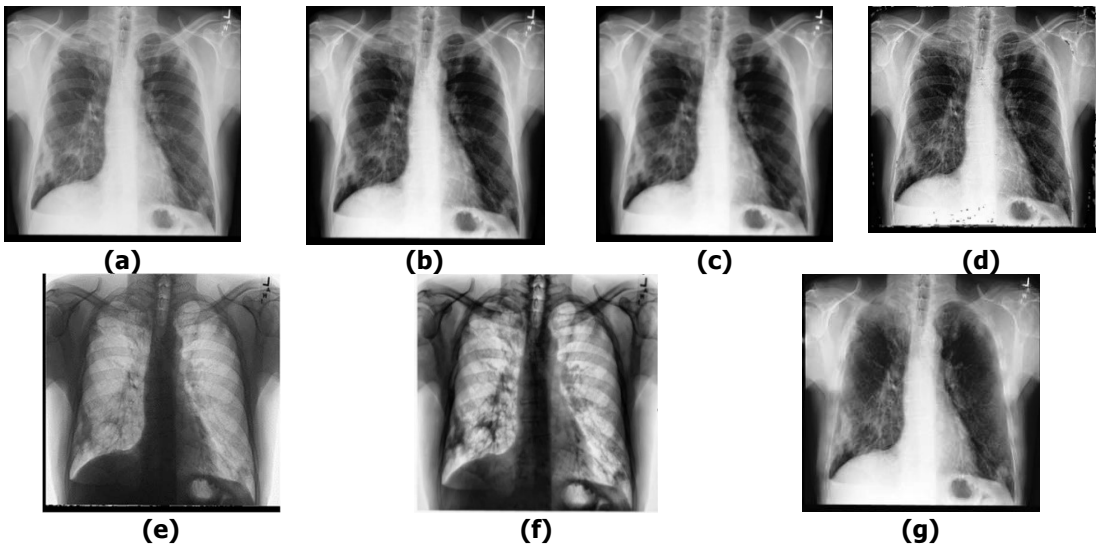


Figure 6. (a) Original Image (b) After DIP (c) After DIP+ (d) After DIP++ (e) After DIPN (f) After DIPCN (g) After Bone Shadow Exclusion (bse)

Another technique implemented in this study was the bse or more known as rib suppression, which has purpose to remove the bone shadow which cover the lung nodule. Initial training phase should be conducted first for the model to recognize the CXR image with bone and without bones. From those training results, CNN model with mask filter were obtained so that it can remove the bone shadow of the CXR input image during the testing phase. In this paper, rib suppression is only conducted with the image size of 256x256. It is because the high

computational load that generate crash or training failure when it is conducted with size of 640x640, even at 416x416. Even though, the trained model at 256x256 could also be used for mask filtering of bone shadow for the input image of 640x640, but the filtering results is not significant so that it seems like does not have any effect at all for the bone shadow removal. Based on Table 2, the implementation of rib suppression generates lower precision and mAP compare with other technique. It can be caused by the lung nodule that also removed along with the bone shadow during the rib suppression. To prevent the missing nodule, nodule enhancement technique should be done at first before the rib suppression technique, so that the number of nodules did not decrease of missing completely from the original image.

3.3 YOLOv5 Training Result with Variation of Data Augmentation Technique (ds02)

In this stage, YOLOv5 model training simulation to detect lung nodule on CXR image was conducted by using a single data augmentation technique, randomly between scale, rotate, flip, multiply, and gaussian blur. The method was denoted as 'aug'. The second method was to apply three data augmentation technique (scale, rotate, and flip) for each image in the training dataset, denoted as 'aug3x'. The first method (aug) generates additional number of train dataset from 1500 to 3000 data. The second method (aug3x) generates additional training dataset from 1500 to 6000 data, each augmentation technique uses 1500 original image data independently for scale, rotate, and flip augmentation technique.

Table 3. YOLOv5 Training Results with Variation of Data Augmentation Technique

No	Model	P	R	mAP@0.5	mAP@ 0.5:0.95	Epoch
1	YOLOv5m-ds02_640x640_aug	0.793	0.728	0.819	0.466	14/14
2	YOLOv5s-ds02_640x640_aug	0.796	0.719	0.814	0.461	49/49
3	YOLOv5s-ds02_640x640_aug3x	0.828	0.769	0.860	0.500	49/49
4	YOLOv5s-ds02_416x416_aug	0.762	0.694	0.785	0.416	19/19

Table 3 shows that the implementation of single data augmentation technique randomly, generates lower precision, recall, and mAP, compared with the results without data augmentation. It might be caused by some inappropriate technique such as multiply (pixel intensity value was multiplied) or gaussian blur (additional blur effect). Besides, since one image was augmented randomly, thus each one of every single image dataset was augmented diversely or uneven so that the model is hardly to recognize the transformation effect on the image due to the insufficient number of data from augmentation results. The training results with three different data augmentation technique, independently on each CXR original image dataset with the total number of training dataset by 6000, yield in higher precision and mAP value compared with the training results with the original 1500 CXR image dataset. The improvement of the precision was from 0.811 to 0.828 and the mAP@0.5 from 0.858 to 0.860.

3.4 YOLOv5 Training Result with Combination of Data Augmentation and Image Processing Technique (ds03)

Table 4, the training result with the combination of data augmentation and image processing. The image processing technique used in this simulation were DIP, DIPN, and also the DIPCN. The results shows that the combination of tripled data augmentation technique ('aug3x') and DIPN produces the highest precision, mAP@0.5, and mAP@0.5:0.95 with the value of 0.83, 0.845, and 0.499, respectively. The precision value obtained in this simulation is the highest, although the recall and mAP value is less than the best results from the previous simulation. Table 5 shows the lung nodule detection results on the CXR test image dataset from several best trained YOLOv5 model in the previous section. For additional comparison purpose, FasterRCNN+ResNet-50 and RetinaNet+ResNet-50, were also used. In the test dataset, there are total of 1975 nodules from 516 test image dataset. The trained YOLOv5s model with using the original dataset (ds00) and image size of 640x640 pixels yields 1735 nodule detection out

of 1975 nodule (87.8%). The trained YOLOv5s model with the DIPN produce 2651 nodules detected (134.2%). This indicates that not all the detected nodules in this case were TP. Therefore, by comparing the number of real nodules in the test dataset with the number of detected nodules from the trained model, the number of FP and FN can be determined.

Table 4. YOLOv5 Training Results with Combination of Data Augmentation and Image Processing Method

No	Model	P	R	mAP@0.5	mAP@ .5:0.95	Epoch
1	YOLOv5s-ds03_800x800_aug_DIP	0.776	0.712	0.801	0.442	49/49
2	YOLOv5s-ds03_416x416_aug_DIP	0.757	0.673	0.771	0.417	29/29
3	YOLOv5s_ds03_640x640_aug_DIP	0.78	0.692	0.782	0.445	49/49
4	YOLOv5s_ds03_640x640_aug3x_DIP	0.8	0.765	0.838	0.492	49/49
5	YOLOv5s_ds03_640x640_aug3x_DIPN	0.83	0.744	0.845	0.499	49/49
6	YOLOv5s_ds03_640x640_aug3x_DIPCN	0.805	0.755	0.845	0.477	49/49

Table 5. Comparison of Nodule Detection Results with Several Trained Model

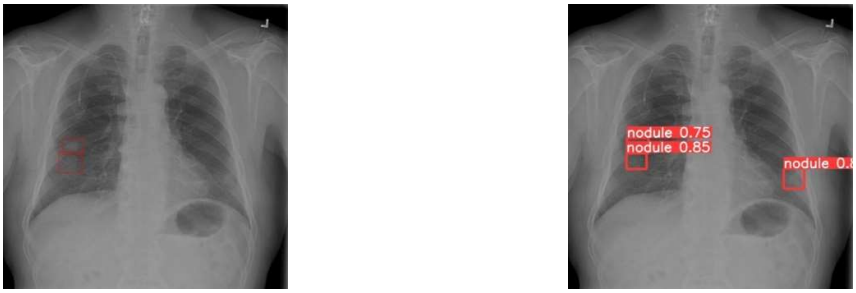
No	Filename (*.jpg)	Σreal nodule	Σnodule detected for each trained model					
			YOLOv5s ds00_640	YOLOv5s ds01_640 _DIPN	YOLOv5s ds02_aug3x_640	YOLO v5s ds03_640 _aug3x_ DIPN	Faster RCNN _ds00 _640	Retina Net-_ds00 _640
1	00000021_000	1	1	1	1	1	2	1
2	00000093_000	4	2	4	3	3	8	1
3	00000105_005	1	2	3	1	1	2	1
4	00000199_002	1	1	3	0	1	4	0
5	00000199_005	3	0	2	0	0	1	0
..	……	….	….	….	….	….	….	….
511	00029909_004	17	15	28	15	12	17	3
512	00030162_029	2	2	2	2	2	2	1
513	00030280_003	1	2	4	1	2	4	0
514	00030536_004	1	2	3	1	1	5	2
515	00030536_006	1	2	2	2	1	2	2
516	00030541_000	2	3	4	3	3	4	0
∑ Total Nodule		1975	1735	2651	1790	1693	2369	651
Detection (%)		-	87.8	134.2	90.6	85.7	119.9	32.9

Table 6. Details of TP, FP, and FN (Results from YOLOv5s_ds00_640 model)

No	Model	Detected	Real	TP	FP	FN
1	00000021_000	1	1	1	0	0
2	00000093_000	2	4	2	0	2
3	00000105_005	2	1	1	1	0
4	00000199_002	1	1	1	0	0
5	00000199_005	0	3	0	0	3
	……	….	….	….	….	….
511	00029909_004	15	17	15	0	2
512	00030162_029	2	2	2	0	0
513	00030280_003	2	1	1	1	0
514	00030536_004	2	1	1	1	0
515	00030536_006	2	1	1	1	0
516	00030541_000	3	2	2	1	0
∑ Total		1735	1975	1624	111	351

Table 6 describes the details of TP, FP, and FN from the nodule detection results, by using the results from the trained YOLOv5s model with original dataset without image processing and data augmentation (YOLOv5s_ds00_640). As for example, the detected results in the sample image "00030541_000.jpg" in Figure 7 were three nodules. However, there are only exist two

real nodules, thus the FP value was one. It can be caused by additional one detected nodule that miscategorized as nodule, instead of the background. Meanwhile in Figure 8, the other detected nodules in the image "00000093_000.jpg" which detected two nodules, whereas the actual number of real nodules was four. Two nodules were not detected and considered as background by the model. Therefore, the two undetected nodules were clarified as FN.



(a) **(b)**

Figure 7. Comparison Between (a) Real Nodule (b) YOLOv5s_ds00_640 Detection Results on "00030541_000.jpg" image.

(a) **(b)**

Figure 8. Comparison Between (a) Real Nodule (b) YOLOv5s_ds00_640 Detection Results on "00000093_000.jpg" image.

Table 7. Summary of TP, FP, FN, and Metric Evaluation of Nodule Detection Results

No	Model	∑Real	∑Detect	TP	FP	FN	P	R
1	YOLOv5s_ds00_640	1975	1735	1624	111	351	0.936	0.822
2	YOLOv5s_ds01_DIPN_640	1975	2651	1907	744	68	0.719	0.966
3	YOLOv5s_ds02_aug3x_640	1975	1790	1654	136	321	0.924	0.837
4	YOLOv5s_ds03_aug3x_DIPN_640	1975	1509	1449	60	526	0.960	0.734
5	FasterRCNN-ResNet-50_ds00_640	1975	2369	1810	559	165	0.764	0.916
6	RetinaNet-ResNet-50_ds00_640	1975	651	647	4	1328	0.994	0.328
7	YOLOv5x_ds00_640	1975	1713	1604	109	371	0.936	0.812

Table 7 describes the summary of TP, FP, and FN of the detection results from the various models which were compared. By obtaining the total of TP, FP, and FN, thus the precision (P) and recall (R) for each model from the test dataset detection results can be calculated by using Equation (1) and Equation (2). The detection result from the trained YOLOv5s_ds00_640 model generates nodule detection with the value of TP, FP, FN, P, and R were 1624, 111, 351, 0.936, and 0.822, respectively. The highest precision 0.960 was obtained from the detection results with YOLOv5s_ds03_aug3x_DIPN_640, with the smallest FP of 60. The highest recall 0.966 was obtained from the results with YOLOv5s_ds01_DIPN_640 with the smallest FN of 68. The results from the trained FasterRCNN + ResNet-50 from 10 epoch training also produce

a high recall value with 0.916 and the second smallest FN with 165. Meanwhile, the results from the trained RetinaNet model from 5 epoch training and loss 0.485, generates the smallest recall with 0.328 due to many nodules undetected. Thus, retraining the RetinaNet model with more epochs are necessary, to generate smaller to obtain more detected nodule.

In real application of lung nodule detection in CXR image to assist the radiologist in diagnosing lung cancer, there are two points of view that need to be considered, which were in terms of precision or sensitivity level. In higher sensitivity but lower precision, most of the total real nodule could be detected with many false detections. In higher precision but lower sensitivity, most of the detected nodules were correct, but many of the real nodules were not detected. In case of early detection, prevention is more important to avoid late treatment for the patient, thus the lung nodule detection with high sensitivity level was necessary. Meanwhile, for therapy, monitoring of the patient was conducted by observing the number of lung nodules. In this case, lung nodule detection algorithm with high precision level was required. Therefore, two trained model YOLOv5s_ds01_DIPN_640 and YOLOv5s_ds03_aug3x_DIPN_640 could be used to detect nodule with its own priority purposes. The model YOLOv5s_ds00_640 with high precision and sensitivity level could also be used for regular checking. So far, the lung nodule detection on the CXR test image was evaluated with same threshold value, which was 0.5. Thus, comparison of the nodule detection results by means of the threshold value variation were also conducted to understand its effect for detection performance. The results in Table 8 and Figure 9 shows that as the threshold value increased, the precision value was increasing, but the recall value was decreasing. The optimum value of precision and recall value can be obtained when the threshold value was 0.35.

Table 8. Summary of TP, FP, FN, and Metric Evaluation of Nodule Detection Results

No	Model	∑Real	∑Detect	TP	FP	FN	P	R
1	YOLOv5s_ds00_640 with conf 0.1	1975	2589	1907	682	68	0.737	0.966
2	YOLOv5s_ds00_640 with conf 0.2	1975	2252	1849	403	126	0.821	0.936
3	YOLOv5s_ds00_640 with conf 0.3	1975	2051	1781	270	194	0.868	0.902
4	YOLOv5s_ds00_640 with conf 0.4	1975	1881	1711	170	264	0.910	0.866
5	YOLOv5s_ds00_640 with conf 0.5	1975	1735	1624	111	351	0.936	0.822
6	YOLOv5s_ds00_640 with conf 0.6	1975	1551	1477	74	498	0.952	0.748
7	YOLOv5s_ds00_640 with conf 0.7	1975	1325	1293	32	682	0.976	0.655
8	YOLOv5s_ds00_640 with conf 0.8	1975	945	938	7	1037	0.993	0.475
9	YOLOv5s_ds00_640 with conf 0.9	1975	104	104	0	2	1.000	0.053

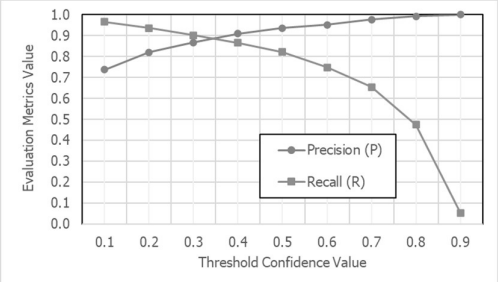


Figure 9. Trade-off Between Precision and Recall Based on Threshold Value Variation

For comparison purpose with previous study, cross validation with different CXR image dataset was also conducted. The different public dataset used was the JSRT (Japanese Society of Radiological Technology) dataset that has been widely used **(Raddar, 2020)**. It has a total number of 247 CXR images, with 154 of them with nodule, while the rest 93 of them without nodule or healthy. The results were described in Table 9.

Table 9. Nodule Detection Metric Result on Cross Validation with JSRT Public Dataset

No	Model	∑Real	∑Detect	TP	FP	FN	P	R	FPPI
1	YOLOv5s_ds00_640	154	121	88	33	66	0.727	0.571	0.273
2	YOLOv5s_ds01_DIPN_640	154	139	67	72	87	0.482	0.435	0.518
3	YOLOv5s_ds02_aug3x_640	154	135	97	38	57	0.719	0.630	0.281
4	YOLOv5s_ds03_aug3x_DIPN_640	154	246	95	151	59	0.386	0.617	0.614
5	YOLOv5x_ds00_640	154	105	83	22	71	0.790	0.539	0.209

From 154 CXR image with nodules, the trained YOLOv5s model with three times data augmentation generates 135 nodules detected, with 97 of them were TP, 38 FP, and FPPI 0,281. It could increase the TP value and reduce the FN value, and increased recall value compared with the results from the trained model with original dataset. Meanwhile, the YOLOv5 model detection results with image processing technique, negative image, as well as both combination of them, shows unsatisfactory result due to the high FP and FN, that yields low precision value higher FPPI value. Thus, implementation of appropriate image processing techniques in the object detection algorithm could enhance the detection result if the dataset used were the same. Meanwhile, the implementation of data augmentation could improve the detection results, even though using different dataset with the same study case.

4. CONCLUSION

Based on simulation results, it can be concluded that the YOLOv5 model is appropriate object detection algorithm to be used for lung nodule detection in chest X-Ray images. This is based on the high object detection evaluation metrics generated by the model, even though by using original dataset without any data augmentation and image processing technique. The results also shows that the most appropriate image processing technique to be used was the negative image technique, which inverts the color intensity from dark to bright, vice versa, so that the nodule can be easier to be identified with the highest recall 0.804 during training phase. Combination of intense data augmentation technique and negative image to each of dataset image generates the highest precision of 0,83 during training phase. The mAP value in this study were generally increased with the increasing size of the dataset image size, along with the heavier computation load as the constraint. The test results on the different CXR image with JSRT public dataset shows that the trained YOLOv5 model also shows good results with data augmentation technique. Further study can be done with the implementation of bone suppression technique for the higher pixel size of the CXR image dataset. The implementation of the higher size pre-trained YOLOv5 model (medium, large, extra-large) could also be used to improve the lung nodule detection with more precise, sensitive, accurate results.

REFERENCES

Ausawalaithong, W., Marukatat, S., Thirach, A., & Wilaiprasitporn, T. (2018). Automatic Lung Cancer Prediction from Chest X-ray Images Using the Deep Learning Approach. *2018 11th Biomedical Engineering International Conference (BMEiCON)*, (pp. 1–5).

Berg, V. J., Young, S., Carolus, H., Wolz, R., Saalbach, A., Hidalgo, A., Giménez, A., & Franquet, T. (2015). A novel bone suppression method that improves lung nodule detection : Suppressing dedicated bone shadows in radiographs while preserving the remaining

signal. *International Journal of Computer Assisted Radiology and Surgery*, *11*(4), 641–655.

Chen, S., Suzuki, K., & MacMahon, H. (2011). Development and evaluation of a computer-aided diagnostic scheme for lung nodule detection in chest radiographs by means of two-stage nodule enhancement with support vector classification. *Medical Physics*, *38*(4), 1844–1858. https://doi.org/10.1118/1.3561504

Ding, Y., Li, Z., & Yastremsky, D. (2021). *Real-time Face Mask Detection in Video Data*. http://arxiv.org/abs/2105.01816

Gang, P., Zhen, W., Zeng, W., Gordienko, Y., Kochura, Y., Alienin, O., Rokovyi, O., & Stirenko, S. (2018). Dimensionality reduction in deep learning for chest X-ray analysis of lung cancer. *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*, (pp. 878–883). https://doi.org/10.1109/ICACI.2018.8377579

Gordienko, Y. G., Gang, P., Hui, J., Zeng, W., Kochura, Y., Alienin, O., Rokovyi, O., & Stirenko, S. G. (2017). Deep Learning with Lung Segmentation and Bone Shadow Exclusion Techniques for Chest X-Ray Analysis of Lung Cancer. *ArXiv*, *abs/1712.0*.

Hayati, M., Muchtar, K., Roslidar, Maulina, N., Syamsuddin, I., Elwirehardja, G. N., & Pardamean, B. (2023). Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning. *Procedia Computer Science*, *216*, 57–66. https://doi.org/10.1016/j.procs.2022.12.111

Humbarwadi, S. (2020). *Object Detection with RetinaNet*. Keras.Io. https://keras.io/examples/vision/retinanet/

Jin, Z. (2018). *ChestX-ray nodule detection | Kaggle*. https://www.kaggle.com/datasets/zhangjin4415/chestxray-nodule-detection

Jung, H., Kim, B., Lee, I., Yoo, M., Lee, J., Ham, S.-Y., Woo, O., & Kang, J. (2018). Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. *PLoS ONE*, *13*.

Knight, S. B., Crosbie, P. A., Balata, H., Chudziak, J., Hussell, T., & Dive, C. (2017). Progress and prospects of early detection in lung cancer. *Open Biology*, *7*(9). https://doi.org/10.1098/rsob.170070

Li, X., Shen, L., & Luo, S. (2018). A Solitary Feature-Based Lung Nodule Detection Approach for Chest X-Ray Radiographs. *IEEE Journal of Biomedical and Health Informatics*, *22*(2), 516–524. https://doi.org/10.1109/JBHI.2017.2661805

Li, X., Shen, L., Xie, X., Huang, S., Xie, Z., Hong, X., & Yu, J. (2020). Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection.

Artificial Intelligence in Medicine, *103*, 101744. https://doi.org/https://doi.org/10.1016/j.artmed.2019.101744

Luo, Y., Zhang, Y., Sun, X., Dai, H., & Chen, X. (2021). Intelligent Solutions in Chest Abnormality Detection Based on YOLOv5 and ResNet50. *Journal of Healthcare Engineering*, *2021*, 2267635. https://doi.org/10.1155/2021/2267635

Maindola, G. (2021). *Introduction to YOLOv5 Object Detection with Tutorial - MLK - Machine Learning Knowledge*. Machine Learning Knowledge. machinelearningknowledge.ai/introduction-to-yolov5-object-detection-with-tutorial/

Matsumoto, T., Doi, K., Kano, A., Nakamura, H., & Nakanishi, T. (1993). [Evaluation of the potential benefit of computer-aided diagnosis (CAD) for lung cancer screenings using photofluorography: analysis of an observer study]. *Nihon Igaku Hoshasen Gakkai zasshi. Nippon acta radiologica*, *53*(10), 1195–1207.

Oğul, B. B., Koşucu, P., Özçam, A., & Kanik, S. D. (2015). Lung nodule detection in X-ray images: A new feature set. *IFMBE Proceedings*, *45*, 150–155. https://doi.org/10.1007/978-3-319-11128-5_38

Orbán, G., & Horváth, G. (2011). *IFMBE Proceedings 37 - A Hybrid Lung Nodule Detection Scheme on Chest X-ray Images* (Vol. 37). www.springerlink.com

Raddar. (2020). *Nodules in Chest X-rays (JSRT) | Kaggle*. https://www.kaggle.com/datasets/raddar/nodules-in-chest-xrays-jsrt

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.48550/arxiv.1506.01497

Schultheiss, M., Schober, S. A., Lodde, M., Bodden, J., Aichele, J., Müller-Leisse, C., Renger, B., Pfeiffer, F., & Pfeiffer, D. (2020). A robust convolutional neural network for lung nodule detection in the presence of foreign bodies. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-69789-z

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Udeshani, K. A. G., Meegama, G., & Fernando, T. G. I. (2011). Statistical Feature-based Neural Network Approach for the Detection of Lung Cancer in Chest X-Ray Image. *International Journal of Image Processing (IJIP)*, *5*, 425–434.