Adaptive image preprocessing and augmentation for disease screening on multi-source chest x-ray datasets

Miss Wasunan Chokchaithanakul

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2021
Copyright of Chulalongkorn University

กระบวนการปรับแต่งรูปภาพทางคอมพิวเตอร์สำหรับการคัดแยกโรคโดยใช้รูปภาพเอกซเรย์ทรวงอก
จากหลายแหล่งข้อมูล

น.ส.วสุนันท์ โชคชัยธนากุล

Thesis Title   Adaptive image preprocessing and augmentation for
                        disease screening on multi-source chest x-ray datasets
By   Miss Wasunan Chokchaithanakul
Field of Study   Computer Science
Thesis Advisor   Associate Professor Proadpran Punyabukkana, Ph.D.
Thesis Co Advisor   Ekapol Chuangsuwanich, Ph.D.

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

--- Dean of the FACULTY OF ENGINEERING

(Professor SUPOT TEACHAVORASINSKUN, D.Eng.)

THESIS COMMITTEE

--- Chairman

(Associate Professor ATIWONG SUCHATO, Ph.D.)

--- Thesis Advisor

(Associate Professor Proadpran Punyabukkana, Ph.D.)

--- Thesis Co-Advisor

(Ekapol Chuangsuwanich, Ph.D.)

--- External Examiner

(Sanparith Marukatat, Ph.D.)

วสุนันท์ โชคชัยธนากุล : กระบวนการปรับแต่งรูปภาพทางคอมพิวเตอร์สำหรับการคัดแยกโรคโดยใช้รูปภาพเอกซเรย์ทรวงอกจากหลายแหล่งข้อมูล. (Adaptive image preprocessing and augmentation for disease screening on multi-source chest x-ray datasets) อ.ที่ปรึกษาหลัก : รศ. ดร.โปรดปราน บุณยพุกกณะ, อ.ที่ปรึกษาร่วม : อ. ดร.เอกพล ช่วงสุนิช

งานวิจัยสำหรับการใช้การเรียนรู้เชิงลึกบนภาพรังสีทรวงอกได้รับความสนใจเพิ่มขึ้นในสาธารณะ อย่างไรก็ตามงานส่วนใหญ่มุ่งเน้นไปที่การพัฒนาแบบจำลองโดยใช้ข้อมูลแหล่งเดียวกับชุดฝึก ซึ่งข้อเสียเปรียบที่สำคัญเมื่อนำไปใช้ในสถานการณ์จริงคือข้อมูลที่ไม่ตรงกันกับชุดการฝึก ดังนั้นบางแบบจำลองจึงมีประสิทธิภาพต่ำกว่าเมื่อนำไปใช้งานจริง งานนี้มุ่งเน้นไปที่ผลกระทบของชุดข้อมูลที่ไม่ตรงกันบนภาพรังสีทรวงอกและวิเคราะห์วิธีการแก้ไขปัญหาการไม่ตรงกันของชุดข้อมูล เทคนิค Lung balance contrast enhancement technique (Lung BCET) จะระบุบริเวณปอดและปรับภาพเพื่อปรับปรุงความทนทานของแบบจำลองเมื่อใช้บนข้อมูลที่แหล่งข้อมูลต่างจากชุดฝึก นอกจากนั้น ยังได้สำรวจวิธีการเพิ่มจำนวนข้อมูลที่เหมาะสมกับภาพรังสีทรวงอกอีกด้วย ข้อมูลเกี่ยวกับวัณโรค, โควิด-19และปอดบวมจากชุดข้อมูลหลายชุดถูกรวบรวมเพื่อประเมินและเปรียบเทียบประสิทธิภาพของกระบวนการเตรียมข้อมูลและวิธีการเพิ่มจำนวนข้อมูลโดยใช้พื้นที่ใต้โค้งของกราฟ Receiver operating characteristic (AUC) และคุณภาพของ Heatmap สำหรับการทดสอบบนแหล่งข้อมูลที่ต่างจากชุดฝึก วิธีการเตรียมข้อมูลโดยใช้ Lung BCET ได้รับ AUC สูงสุดที่ 0.7978 และ 0.6240 สำหรับชุดข้อมูลแม่สอดและกองวัณโรค (BT) ตามลำดับ อย่างไรก็ตาม ไม่มีความแตกต่างสำหรับผลลัพธ์ของการทำนายผลบนโควิด-19และโรคปอดบวม การศึกษาของเรายังพบว่า Lung BCET สามารถใช้เพื่อเพิ่มข้อมูลร่วมกับวิธีการเพิ่มข้อมูลแบบทั่วไปเพื่อปรับปรุงประสิทธิภาพของผลการทดลองทั้งในข้อมูลจากแหล่งเดียวและต่างแหล่งจากชุดฝึกบนวัณโรค

สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่อนิสิต ....
ปีการศึกษา	2564	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

6272079921 : MAJOR COMPUTER SCIENCE

KEYWORD: Image preprocessing, Data augmentation, Chest radiography

Wasunan Chokchaithanakul : Adaptive image preprocessing and augmentation for disease screening on multi-source chest x-ray datasets. Advisor: Assoc. Prof. Proadpran Punyabukkana, Ph.D. Co-advisor: Ekapol Chuangsuwanich, Ph.D.

Research on deep learning models for chest radiology applications has increased attention by the public. However, most works focus on developing models using in-domain data, so the significant drawback, when applied in real-world scenarios, was the mismatched data with the training set. Consequently, some models perform inferior at the deployment stage. This work focused on the effects of dataset mismatch on chest radiography and analyzed the methods the overcome the mismatch issues. The lung balance contrast enhancement technique (lung BCET) automatically identifies the lung region and normalizes the image accordingly to improve the robustness of out-of-domain data developed. Additionally, augmentation methods that are suitable for chest radiography were explored. The data on Tuberculosis (TB), COVID-19, and pneumonia were compiled from multiple datasets to evaluate and compare the performance of the preprocessing and augmentation methods using the area under the receiver operating characteristic curve (AUC) and heatmap quality. For out-of-domain testing conditions, the lung BCET preprocessing method achieved the highest AUC scores of 0.7978 and 0.6240 for the Maesot and Bureau of TB (BT) datasets, respectively. However, there are no differences in performance on COVID-19 and pneumonia datasets. Our study also found that lung BCET can be used to perform data augmentation in conjunction with the standard augmentation techniques to improve the performance in both in- and out-of-domain conditions on the TB datasets.

Field of Study:	Computer Science	Student's Signature	
Academic Year:	2021	Advisor's Signature	
		Co-advisor's Signature	

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

LIST OF TABLES

LIST OF FIGURES

1. Introduction

1.1 Motivation

In recent years, machine learning has been increasingly used in medical computer vision studies, particularly image classification and segmentation. Typically, machine-learning models learn from a training set and evaluate on the test set. Nonetheless, medical image data of an appropriate size have some difficulties in collecting and using, compared to other image data, due to clinical data protection regulations and the deficiency of expert labeling. As a result, preprocessing and augmentation approaches have been developed to obtain an increasing number of high-quality and diverse data.

Image preprocessing is essential for enhancing image quality by adjusting brightness, contrast, noise, and other related factors, in order to reduce unwanted distortions and enhance some features required for model learning [1-5]. Data augmentation is a method that simulates the training dataset by flipping, rotating, and cropping existing data to avoid overfitting issues [6-8]. Advanced augmentation techniques, such as Generative adversarial network (GAN), are used to transform an image from one domain to another [9, 10].

Chest radiography is commonly used to diagnose lung diseases. However, different healthcare settings may have different X-ray machines, settings and systems, patients' demographics, and other related factors that could affect the image data. Therefore, applying the same preprocessing or augmentation technique may lead to over- and under-enhancement problems. Thus, the main objective of this study is to examine new preprocessing and augmentation techniques for chest radiography data to aid model learning when trained on non-diverse data.

The balance contrast enhancement technique (BCET) [11] is a popular method for normalizing the intensity values of an image. However, the BCET may fail with the presence of foreign objects or annotations, which can change the minimum and maximum values in the image. In this research, the BCET analyzed only the lung region, which is usually more uniform across different data sources, to perform the normalization so that the BCET can be more robust. Furthermore, the lung area can be identified by a lung segmentation model, so-called the lung BCET method. A set of augmentation techniques was also proposed by leveraging existing preprocessing techniques. Instead of removing all differences in the dataset via preprocessing, the preprocessing techniques were applied to generate slight variations by over- or under-correcting the images. Contrast limited adaptive histogram equalization (CLAHE) [12] and unsharp masking

(UM) [13] were selected in conjunction with our proposed preprocessing technique to produce such augmentations. Differential evolution (DE) [14], a genetic algorithm that employs a parallel direct search approach, was used to determine the suitable settings for the augmentation.

The area under the receiver operating characteristic curve (AUC) score was selected to compare the different techniques for both in- and out-of-domain data sources. Briefly, two separate experiments were performed. First, focusing on image preprocessing without augmentation, the lung balance contrast enhancement technique (lung BCET) method was compared with other related methods. Second, various augmentation methods, such as typical image augmentation techniques (random rotation, random brightness, and random contrast), the proposed augmentation (random CLAHE and random UM), and their combinations which require lung BCET preprocessing as a prerequisite, were compared. Results revealed that lung BCET could improve performance in both in- and out-of-domain settings. Moreover, different augmentations could further improve the performance.

1.2   Objective

This thesis aimed to evaluate the use of image preprocessing and augmentation techniques for the classification of chest X-rays image datasets. The main hypothesis of this study was:

The proposed image preprocessing techniques can improve the AUC score when tested on in- and out-of-domain data, and the augmentation techniques suitable for proposed image preprocessing techniques and chest X-ray images were searched for.

The objectives of this thesis were as follows:

1. To evaluate the output of the model using a new image preprocessing approach with both in- and out-of-domain data where no augmentation strategies are applied.

2. To compare augmentation strategies to learn more variants of chest X-ray images may help the model in learning more variations of augmentation.

1.3   Scope

The scope of this thesis is to use lung segmentation with the BCET to test a new image preprocessing technique and choose data augmentation techniques for machine learning of various chest X-ray datasets. The results were then evaluated and compared between techniques.

2. Related works

Owing to the variety of datasets, several researchers have been interested in comparing the performances of in- and out-of-domain data. Sathitratanacheewin et al. [15] used the Shenzhen TB dataset for training a model based on InceptionV3 and tested on the Shenzhen and other chest X-ray datasets using color space, crop flip, and rotation augmentation to determine the generalizability of the model. The AUC on the Shenzhen dataset was 0.8502; however, it dropped to 0.7054 on an out-of-domain data (Chest X-ray 8).

In a similar study, Zech et al. [16] tested the model generalization to detect pneumonia in chest X-ray images. The National Institutes of Health Clinical Center (NIH) and Mount Sinai Hospital (MSH) were selected as the training sets. Additionally, the Indiana University Network for Patient Care (IU) was selected as the out-of-domain data. The results showed that the out-of-domain data had lower AUC scores than the in-domain data.

The findings of these studies show that the model provides poor outcomes when utilizing out-of-domain data. As a result, most researchers have applied image enhancement and augmentation to increase the AUC scores of the models.

Image enhancement is a crucial image preprocessing step. Several enhancement techniques will emphasize important information and decrease or eliminate noise in the images in order to improve the quality of an image and the performance of models. Several researchers have focused on comparing preprocessing methods to determine the best ways to classify chest X-ray images.

Munadi et al. [17] compared the accuracy of three image enhancement techniques, including CLAHE, UM, and high-frequency emphasis filtering (HEF), for tuberculosis (TB) detection. The study revealed that UM approaches based on the EfficientNetB4 model had the highest accuracy through transfer learning (89.92%).

Zotin et al. [18] investigated lung border detection. A median filter was utilized to reduce image noise, and then the BCET was applied to improve image quality.

In a study by Rahman et al. [19], the covid-19 dataset and six distinct neural networks were used to evaluate the performance assessment matrix using numerous image improvement approaches, including histogram equalization (HE), CLAHE, image inversion, gamma correction,

and BCET. The study revealed the two best results for gamma correction using the ChexNet model and BCET utilizing DenseNet201.

As previously mentioned, BCET is an image preprocessing method used by researchers to normalize images from different sources. BCET uses a parabolic function in extending or compressing the image contrast without affecting the histogram pattern of the original images (see more detail in section 3).

The four basic densities on a chest X-ray are air (displayed in black), fat (gray), water/soft tissue/nodule (white to gray), and bone (white) [20]. Because the BCET works by adjusting the minimum, maximum, and mean values of the dataset, if the air (black) and nodule (white) colors are the same in all the datasets, the images will be similar. However, sometimes the most white and black parts of chest X-rays can be derived from other artifacts or distractions outside the region of interest. For example, some datasets might impose a box showing patient information onto the image, and the BCET may use wrong statistics to normalize the image. Thus, before modifying the minimum and maximum specified values, the colors of the air and nodules in the lung area of the image. Therefore, BCET with a lung segmentation model was applied as our new image preprocessing technique.

Data augmentation is a common technique used in training deep learning models in order to increase the amount of training data. The process usually involves changing images from the existing training data to avoid overfitting and improve the model performance.

Most medical studies that employ chest X-ray datasets used augmentation approaches to improve model performance. Fig. 1 presents the summary of several research using specific augmentation approaches to classify chest X-ray images [21-45] (see Table 1 for details). The most common augmentations are horizontal flip and random rotation.
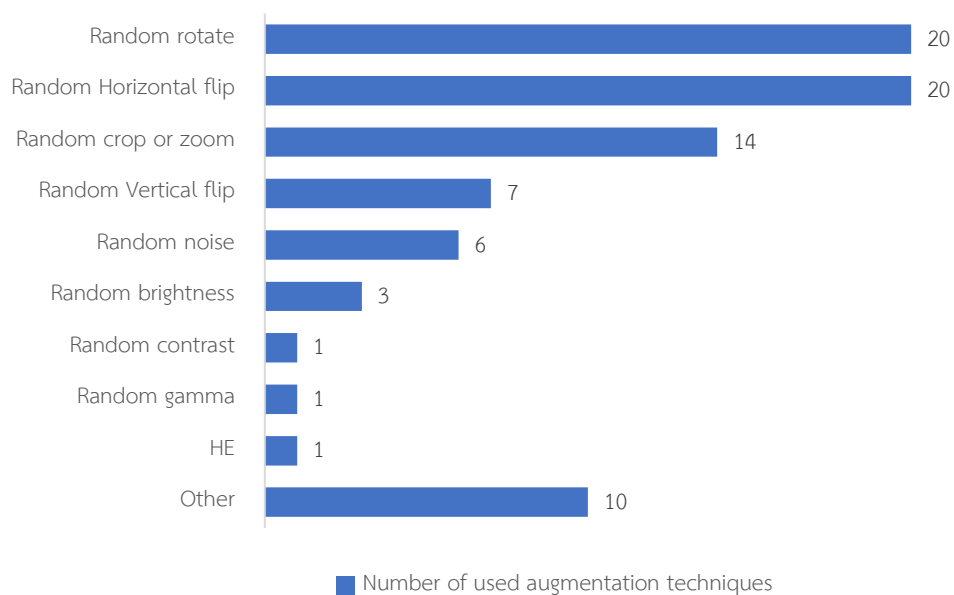
Figure 1: Summary research numbers sorted by augmentation strategies to classify chest X-ray images.

Table 1: Details of example papers using augmentation techniques for chest X-ray image analysis on classification model.

Citation	Augmentation techniques									
	Random rotate	Random crop or zoom	Horizontal flip	Vertical flip	Random noise	Random brightness	Random contrast	Random gamma	HE	Other
Abbas et al. (2018) [21]	✓		✓	✓	✓					✓
Yadav et al. (2018) [22]	✓			✓		✓				
Ahsan et al. (2019) [23]	✓	✓	✓			✓	✓			
Meraj et al. (2019) [24]	✓	✓	✓	✓						
Rohilla et al. (2017) [25]	✓							✓		
Hernández et al. (2019) [26]		✓			✓					
Nguyen et al. (2019) [27]	✓		✓							✓
Allaouzi et al. (2019) [28]			✓							

Citation	Augmentation techniques									
	Random rotate	Random crop or zoom	Horizontal flip	Vertical flip	Random noise	Random brightness	Random contrast	Random gamma	HE	Other
Duong et al. (2021) [29]	✓	✓	✓							✓
Sharma et al. (2020) [30]	✓	✓	✓							✓
De Moura et al. (2020) [31]			✓							
Misra et al. (2020) [32]	✓	✓	✓							
Sitaula et al. (2021) [33]	✓	✓	✓	✓						✓
Rahman et al. (2020) [34]	✓									✓
Asif et al. (2020) [35]	✓		✓	✓	✓					✓
Sirazitdinov et al. (2019) [36]	✓		✓	✓	✓	✓		✓		✓
Heidari et al. (2020) [37]	✓	✓	✓							✓
Nayak et al. (2021) [38]	✓	✓	✓		✓					
Seyyed-Kalantari et al. (2020) [39]	✓	✓	✓							
Basu et al. (2020) [40]	✓									
Minaee et al. (2020) [41]	✓		✓	✓	✓					
Wang et al. (2020) [42]	✓	✓	✓							✓
Chouhan et al. (2020) [43]		✓	✓							
Jain et al. (2021) [44]	✓	✓	✓							
Zhang et al. (2020) [45]		✓	✓							

Studies have compared various augmentation approaches to identify the best strategies for classifying chest X-ray images.

Sirazitdinov et al. [46] evaluated the AUC scores of diverse data augmentation methods, such as random gamma, contrast, brightness, rotation, horizontal flip, Gaussian noise, and Gaussian blur, to classify 14 classes of 14 chest X-ray datasets in the Inception-Resnet-v2 network. Their investigation suggested a combination of random horizontal flips, random rotations, random brightness, and random gamma, which could achieve an AUC score of 0.808.

Ogawa et al. [47] studied the effects of various augmentation approaches on three image resolutions, including 128 × 128, 192 × 192, and 256 × 156 pixels. The augmentation methods of random rotation, Gaussian blur, and brightness were used with or without the additional horizontal or vertical flip. For all image resolutions, random rotation with horizontal flip had the best accuracy. In contrast, Gaussian blur with and without flipping had the lowest accuracy because Gaussian blur reduced the image details.

Our studies attempted to apply new augmentation approaches in conjunction with popular augmentation methods. Image enhancement and preprocessing techniques were used as additional augmentations by introducing small offsets. CLAHE and UM were selected for random changes and then compared with the other techniques.

3. Background

3.1 Preprocessing techniques

3.1.1 Balance contrast enhancement technique

Balance contrast enhancement technique (BCET) [11] is a technique for preparing images by extending or compressing the image contrast without affecting the original histogram pattern of the images. The function to perform the BCET is a parabolic function, i.e.,

$$y = a(x - b)^2 + c \tag{1}$$

where x is an input image and y is an output image.

The parabolic function is controlled by the following three coefficients:

$$a = \frac{H - L}{(h - l)(h + l - 2b)} \tag{2}$$

$$b = \frac{h^2(E - L) - s(H - L) + l^2(H - E)}{2[h(E - L) - e(H - L) + l(H - E)]} \tag{3}$$

$$c = L - a(l - b)^2 \tag{4}$$

where l and h represent the minimum and maximum values of the input image, respectively;

L and H represent the minimum and maximum values of the output image, respectively;

e and E represent the mean value of the input and output image, respectively; and

s represents the mean square sum of the input image derived from the following equation:

$$s = \frac{1}{N} \sum_{i=1}^{N} x_i^2 \tag{5}$$

where N represents the pixel number of the input image.

The histograms of the original image and the image following the BCET are shown in Fig. 2.

Original image and
histogram of original image

BCET image and
histogram of BCET image
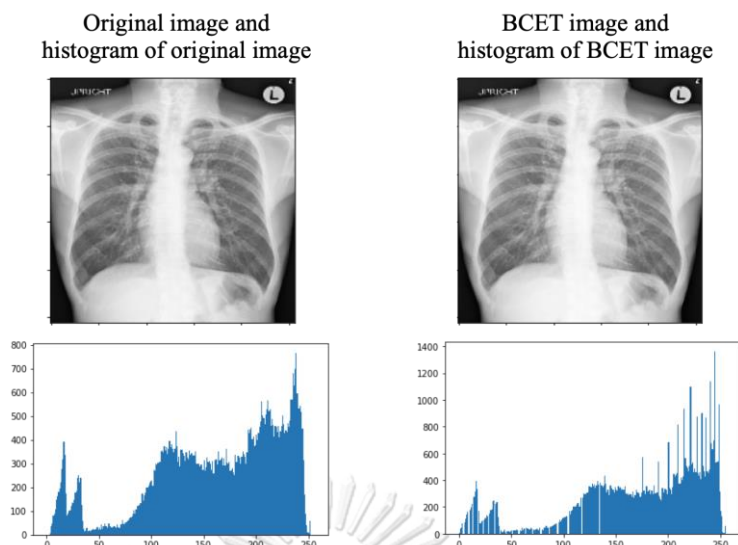
Figure 2: Example of an images and histograms: **Left -** *original image and* **Right -** *BCET image.*

3.2 Augmentation techniques

Data Augmentation is a method for reducing model overfitting by expanding the training dataset using existing data. The following is a list of augmentation strategies utilized in this paper:

3.2.1 Random brightness

Random brightness is an augmentation technique that generates images with varying brightness. Using the Albumentation library, a random integer in the limited range of the limited parameter will be applied to the original image and adjust the image to either brighter or darker.

3.2.2 Random contrast

The difference in brightness or color that distinguishes one object from other objects in the same field of vision is described as a contrast. Most of the images are classified into two types: low contrast and high contrast. The histogram of the low contrast image will be clustered in a limited intensity range, whereas the histogram of the high contrast image will have a larger gap between intensity values.

3.2.3 Random rotate

The random rotation technique means the image will be rotated in a clockwise direction using a random angle from the range. The range for random angle is defined as one limited parameter in the Albumentation library.

3.2.4   Horizontal flip

Horizontal flip is a technique that involves horizontally reversing the complete rows and columns of image pixels. After applying a horizontal flip, the left side of the original image will be switched to the right side.

3.2.5   Gamma correction

Gamma correction is a technique that uses a non-linear operation on input image pixels to adjust the brightness of an image. The following is the gamma correction function:

$$O = \left(\frac{I}{255}\right)^{\gamma} \times 255 \tag{6}$$

where O represents the output image, I represents the input image, and γ represents the gamma parameter. When $\gamma < 1$, the dark regions will be brighter, and the histogram will be shifted to the right, and vice versa if $\gamma > 1$. In the Albumentation library, γ uses tens digits instead of decimals.

3.2.6   Contrast limited adaptive histogram equalization

Contrast limited adaptive histogram equalization (CLAHE) [12] is a technique for improving image contrast based on adaptive histogram equalization (AHE). Originally, HE and AHE improve the image by extending out the intensity range and spreading out frequent pixel intensity values. While HE extends out the entire image intensity, AHE divides the image into sub-images and stretches out the intensity in each sub-image. Both approaches discovered an over-enhancement problem and increased image noise. Therefore, CLAHE is developed to decrease noise in the HE and AHE methods by clipping the histogram and spreading out frequent pixels over all intensity values before performing the AHE process.

3.2.7   Unsharp Masking

Unsharp Masking (UM) [13] is one of the techniques most used for image processing. This technique is used to sharpen the image by the following equation:

$$enhanced_{img} = original_{img} + amount(original_{img} - blurred_{img}) \tag{7}$$

where $original_{img}$ represents the input image, $blurred_{img}$ represents the input image after applied Gaussian blur, $enhanced_{img}$ represents the output image and $amount$ represents the number for adjusting contrast is added to the edges of the image.

From the scikit-image library, the radius parameter is referred to the sigma parameter of the Gaussian blur function in equation 8.

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

(8)

where x is the distance from the origin in the horizontal axis, y is the distance from the origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution.

3.3   Parameter selection

3.3.1   Differential Evolution

Differential Evolution (DE) [14] is a sort of genetic algorithm that uses a parallel direct search technique via the stages below:

1. Create an initial NP vector in the population by randomly choosing and covering the entire parameter space.
2. Calculate each initial vector in the population and collect the result.
3. For each vector in the population:

    3.1. Create a mutant vector using the following equation to calculate the value of each parameter:

$$p_i^{mut} = p_i^{best} + F \cdot (p_i^{r1} - p_i^{r2})$$

(9)

   where p_i^{mut} represents the parameter value of mutant vector,

   p_i^{best} represents the parameter value of the best vector (vector from the previous vector that has the best calculation result),

   F represents the mutation rate in a range of $[0, 2]$, and

   p_i^{r1} and p_i^{r2} represent the difference between two randomly chosen vector, which are $r1$ and $r2$.

    3.2. Create a trial vector by defining the number of recombination rates and selecting a current parameter or mutant parameter at random from a range of $[0, 1]$. If R is less

than the rate of recombination, the system will choose the mutant parameter. If R is greater than the rate of recombination, the system will use the current value.

    3.3. Calculate and gather the results for each trial vector, and then compare the results of the current and trial vectors. If the outcome of the trial vector is better than the current vector, the system will replace the trial vector with the current vector.

4. Repeat step 3 until the total number of iterations or standard deviations for the whole population is less than a specific percentage of the mean value of the result.

3.4   Deep Convolution neural network

3.4.1   Resnet

The deep convolution neural network (DCNN) is the most employed in studies. CNN's architecture comprises a convolutional layer, a ReLu activation layer, a pooling layer, and a fully connected layer. When training DCNN, the vanishing gradient problem can occur as the gradient will be vanishingly tiny until it reaches zero, effectively preventing the weight from changing its value. The vanishing gradient may stop the DCNN training.

To solve the vanishing gradient problem, Microsoft has designed the Resnet, or Deep Residual Network [48], by including shortcut connections into DCNN. The construction block is used to create a shortcut connection for Resnet18 and Resnet34. The bottleneck is added into the neural network for Resnet50 and other Resnets with more than 50 layers. The distinct architecture between a construction component and a bottleneck is displayed in Fig. 3.
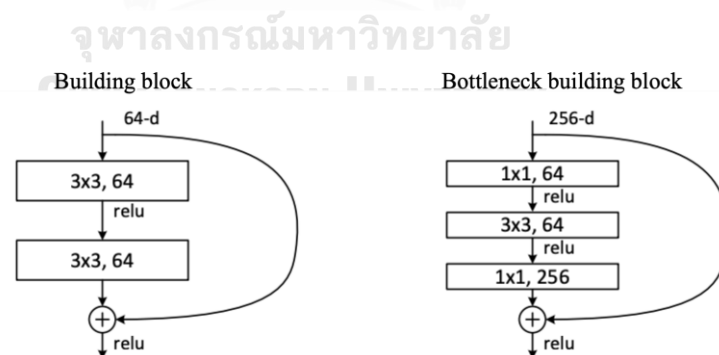
Figure 3: Different architecture between building block and bottleneck building block.

3.4.2 Efficient Net

Many researchers have attempted to improve the accuracy of neural networks by using a deeper or larger network. In 2020, Google demonstrated how balancing network depth, breadth, and resolution might improve performance and introduced the Efficient Net model [49]. Compound scaling was used to improve the breadth, depth, and resolution of the Efficient Net while keeping the ratio constant. With the constant ratio value, each version of Efficient Net (B0 through B7) is different.

3.4.3 Pyramid Localization Network

The Pyramid Localization Network (PYLON) [50] is a neural network developed by the Department of Computer Engineering, Chulalongkorn University. It was intended to increase the quality of the CAM heatmap and to detect the lesions correctly. PYLON architecture is divided into three parts: an encoder that can be replaced with other neural networks similar to the Resnet; a decoder that includes the Pyramid Attention (PA) and Upsampling (UP) modules; and a prediction head for the classification and the heatmap output. Fig. 4 depicts the PYLON architecture and the differences between the CAM and PYLON heatmaps.
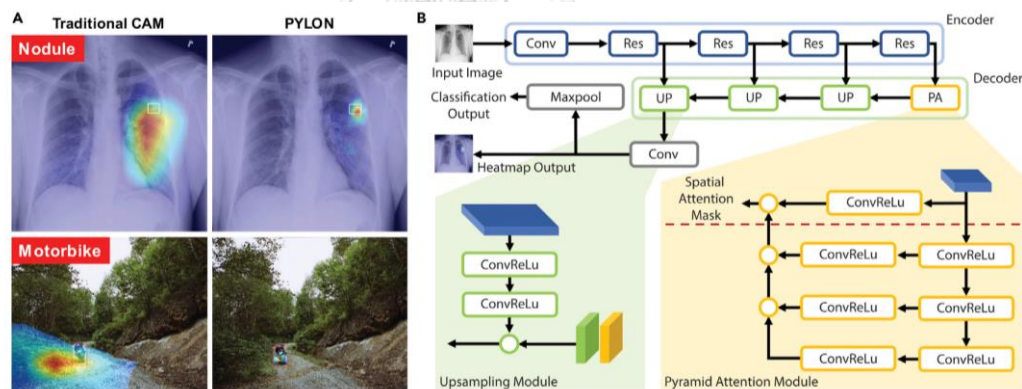
Figure 4: (A) Comparison between the CAM and PYLON heatmaps and (B) The architecture of the PYLON.

4. Proposed methods

4.1   Dataset

Tuberculosis (TB), COVID-19, and pneumonia labels were obtained from both public and private sources. The in-domain data were separated into training and validation sets (80%), and test sets (20%). Subsequently, the training and validation sets were randomly split in an 80:20 ratio. Out-of-domain data were used as the extramural test sets to evaluate the robustness of the model.

4.1.1   Tuberculosis

There are two main classes in this study: TB and non-TB. The entire dataset contains 6,168 frontal-view chest radiographs obtained from five different sources: Montgomery County, Maryland's Department of Health and Human Services [51] (Montgomery; 138 radiographs); Shenzhen No.3 Hospital in Shenzhen, Guangdong province, China [51] (Shenzhen; 662 radiographs); Banglamung hospital, Thailand (BLM; 3,540 radiographs); the Bureau of Tuberculosis, Thailand (BT; 1,604 radiographs); and Maesot hospital, Thailand (Maesot; 224 radiographs). Montgomery, Shenzhen, and BLM were treated as in-domain data and used in the model training, whereas BT and Maesot were used as out-of-domain data. Details of the dataset splitting are listed in Table 2.

Table 2: Details of training, validation, and test set from the TB datasets.

Source	Label	Number of images			
		Training set	Validation set	Test set	Total
Montgomery	TB	40	6	12	58
	non-TB	56	8	16	80
Shenzhen	TB	234	34	68	336
	non-TB	228	33	65	326
BLM	TB	50	7	14	71
	non-TB	2,428	347	694	3,469
BT	TB	-	-	1,295	1,295

Source	Label	Number of images			
		Training set	Validation set	Test set	Total
	non-TB	-	-	309	309
Maesot	TB	-	-	9	9
	non-TB	-	-	215	215
Total number of images		3,036	435	2,697	6,168

4.1.2 COVID-19

The COVID-19 dataset comprises 26,097 frontal view chest radiographs labeled as COVID-19 and non-COVID. These data were collected from the following publicly accessible datasets, online sources, and published papers: Valencian Region Medical ImageBank COVID19+ [52] (BIMCV; 2,474 radiographs); Hospital Universitario Clínico San Cecilio, Granada, Spain [53] (CG; 852 radiographs); COVID-CXR [54] (CC; 782 radiographs); German medical school [55], Italian Society of Medical Radiology [56], GitHub, Kaggle & Twitter and another source (Other; 1,142 radiographs); and King Chulalongkorn Memorial hospital, Thailand (KCMH-C; 5,984 radiographs). The out-of-domain data was gathered from the KCMH-C dataset by selecting only admitted cases with chest radiograph images within four days before admission. Details of the dataset splitting are listed in Table 3.

.

Table 3: Details of training, validation, and test set from the COVID-19 datasets.

Source	Label	Number of images			
		Training set	Validation set	Test set	Total
BIMCV	COVID-19	1,781	198	495	2,474
	non-COVID	-	-	-	-
CG	COVID-19	272	69	85	426
	non-COVID	272	68	86	426
CC	COVID-19	306	76	96	478

Source	Label	Number of images			
		Training set	Validation set	Test set	Total
	non-COVID	194	49	61	304
Other	COVID-19	819	93	230	1,142
	non-COVID	-	-	-	-
KCMH-C	COVID-19	-	-	888	888
	non-COVID	-	-	5,096	5,096
Total number of images		3,644	553	7,037	11,234

4.1.3 Pneumonia

The RSNA Pneumonia Detection Challenge (2018) dataset (RSNA) [57], which contains 26,684 frontal view chest radiographs labelled as pneumonia and non-pneumonia, was used to train and evaluate this study. Data from the King Chulalongkorn Memorial Hospital in Thailand (KCMH-P; 7,654 radiographs) were obtained as the out-of-domain data. Table 4 lists the details of the dataset splitting.

Table 4: Details of training, validation, and test set from the pneumonia datasets.

Source	Label	Number of images			
		Training set	Validation set	Test set	Total
RSNA	Pneumonia	11,412	2,854	3,567	17,833
	non-pneumonia	5,665	1,416	1,770	8,851
KCMH-P	Pneumonia	-	-	1,194	1,194
	non-pneumonia	-	-	6,460	6,460
Total number of images		17,077	5,337	12,991	34,338

4.2 Proposed preprocessing

Images in DICOM or JPEG formats were converted to a NumPy array of type uint8 and down sampled to 256 × 256 pixels before applying any preprocessing techniques (lung BCET method, gamma correction, and UM).

The lung BCET method was created to reduce the differences between data sources by combining lung segmentation [58] and BCET. The study objective was to determine the statistics of the lung area for the BCET to operate only on the main area of interest and have less prone to encounter artifacts. Automatic lung segmentation via a deep learning model was used to determine the lung area of an image in order to gather the minimum and maximum values. Subsequently, the values of all the image pixels were updated. If the pixel value was less than the minimum value, the pixel value was changed to the minimum value. Similarly, the pixel value was changed to the maximum value if the pixel value was greater than the maximum value. Then, the images were subjected to the BCET.

However, the automatic lung segmentation process may encounter some problems in the images with unclear lung surroundings, resulting in an incapability to determine the lung area or a wrong segmentation outside the lung area. Thus, a heuristic was applied to flag whether the lungs are properly segmented. If not, the original BCET was applied instead of the lung BCET. To determine whether the lung was properly segmented, the percentage of the lung segmentation area was calculated by dividing the segmented area by the entire image area. If the percentage is less than a cut-off threshold, the BCET, without any special processing, will be applied to the image. A percentage cut-off threshold was treated as a hyperparameter and found via tuning on the validation set. Fig. 5 shows the diagram of the lung BCET process and an example of an image after lung BCET was applied to different sources of dataset.

The gamma correction parameter was chosen from the maximum limit default in the Albumentation library, and the UM from the parameter form [17]. Parameter values are presented in Table 5.
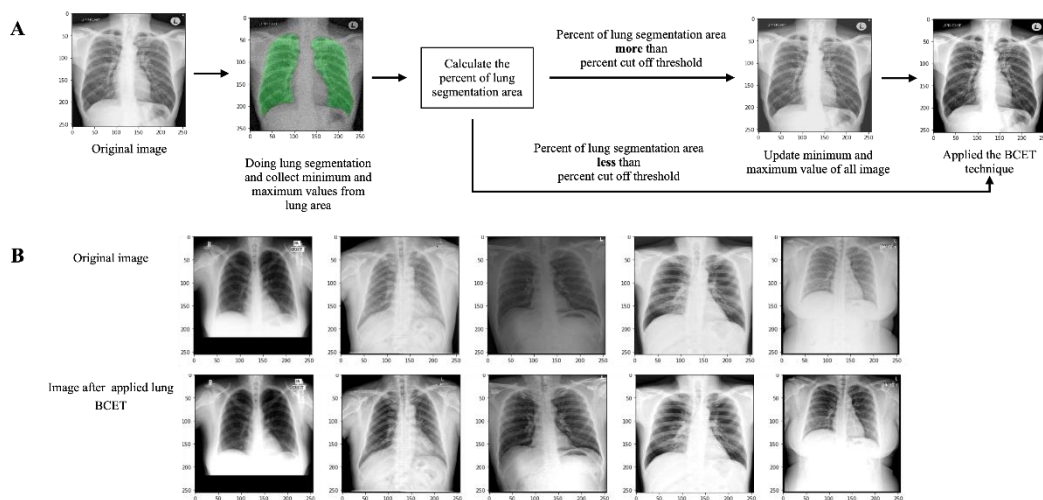
Figure 5: Processing details and the example images from the lung BCET: (A) Diagram of the lung BCET process and (B) Comparison of the original images from different sources of datasets with the images obtained after applying the lung BECT.

Table 5: Parameter settings for the preprocessing techniques.

Preprocessing technique	Parameter	Value
Lung BCET	percent_cutoff	10 (TB), 40 (COVID-19) and 15 (pneumonia)
Gamma correction	gamma_value	120
UM	radius	5
	amount	2

4.3   Proposed augmentation

The augmentation methods used in previous studies and our proposed preprocessing-based augmentation method were compared in this study. Basic augmentation techniques were chosen from the most common augmentation approaches used in medical image research (Fig. 1), and the default values of the Albumentation library were used as the parameters.

Techniques that adjust the color of the image were particularly focused in our study because images from different sources had diverse intensities. CLAHE and UM are two methods for

modifying image intensity that have been used for image preparation in several chest X-ray studies [17, 26, 27]. In this study, the augmentation techniques of random CLAHE and random UM, which randomly tweak the parameter values slightly within certain boundaries, were adopted.

Since many parameters are presented in different preprocessing and augmentation techniques, hyperparameter tuning via grid search is not feasible. Instead, differential evolution (DE) was used to discover the best parameter values. First, ten population groups and ten rounds were chosen to run DE using the ResNet50 model. Then, the best parameter values were selected by finding the best AUC scores on the validation set. See Table 6 for more details of the augmentation parameter values.

Table 6: Parameters for augmentation techniques.

Data augmentation	Parameter	Value
Random brightness	limit	0.2
Random contrast	limit	0.2
Random rotate	limit	90
Random horizontal flip	-	-
Random gamma	gamma_limit	(80,120)
Random CLAHE	clip_limit	2
Random UM	tile_grid_size	(n, n) where n is an integer number and selected randomly from (4,5)
	radius	n where n is an integer number and selected randomly from (4,5)
	amount	n where n is an integer number and selected randomly from (2,4)

Related studies [46, 47] observed that combining multiple augmentation methods yielded better results than any single augmentation technique. Therefore, in this study, the baseline augmentation methods were combined with the proposed augmentation method using the OneOf function from the Albumentation library, which randomly selected only one augmentation technique from the random UM, random CLAHE, and random gamma. Then, the horizontal flip was applied with a probability of 0.5, and random rotation was applied with a probability of 0.5.

The experimental details for the augmentation comparison were the same as for the preprocessing comparison. However, all the experiments applied lung BCET as the preprocessing approach. Previous works by Sirazitdinov et al. [46] and Ogawa et al. [47] were selected to compare the AUC score with the baseline and our proposed augmentation methods.

4.4   Detail of the experiments

PyTorch lightning was used for model training and evaluation, and ResNet50 and EfficientNetB0 for base models. Additionally, PyTorch and Torchvision were used for experiments with PylonResNet50. All experiments applied normalization augmentation from the Albumentation library, with mean and standard deviation values of 0.4984 and 0.248, respectively. The initial weights of ImageNet were used to train all the neural networks. The training used a batch size of 64 chest X-rays with a learning rate of 0.0001 and the Adam optimizer. The training was done for 100 epochs with early stopping if there was no improvement for two consecutive epochs. For the PyTorch lightning, a deterministic flag was used to ensure full reproducibility during training. To compare different preprocessing techniques (Fig. 6A), our models were trained using five different random seeds and reported the average. All data were subjected to the proposed preprocessing for the augmentation comparison (Fig. 6B).
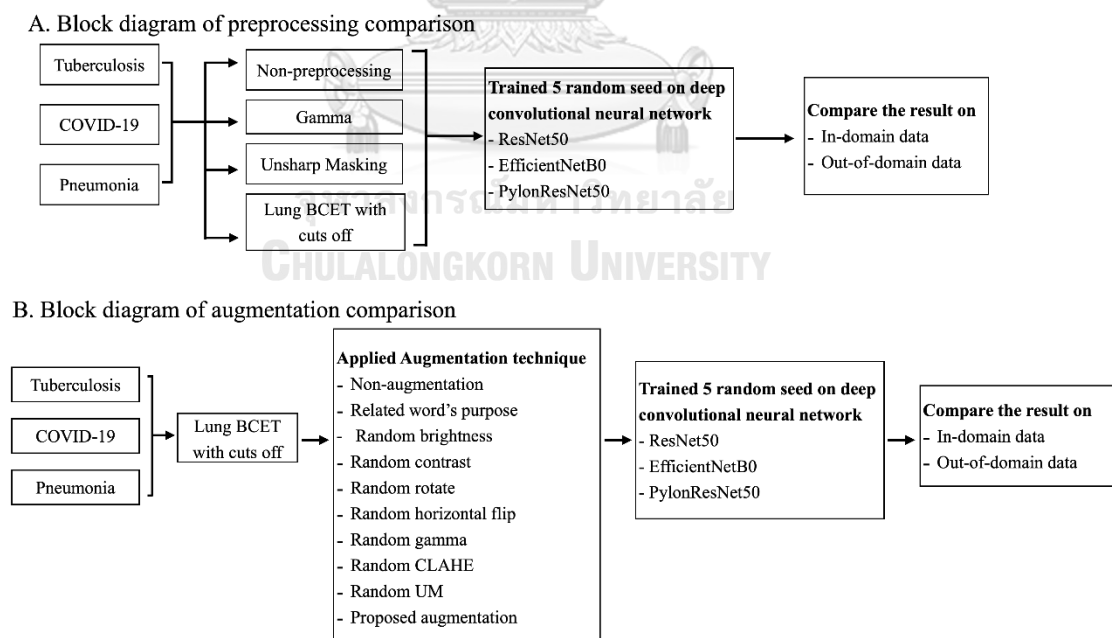
Figure 6: Block diagram of the methodology: (A) Preprocessing comparison and (B) Augmentation comparison.

4.5   Performance testing

For in-domain and out-of-domain data performance testing, the weight from the best epoch (i.e., the epoch with the least loss) was chosen. The average AUC across five different random seeds was reported. Besides the accuracy measurements, the methods were qualitatively compared by analyzing whether the heatmaps from the model were rational. Since the datasets had no bounding box annotation, the accuracy of the heatmaps was measured via a simple heuristic; whether or not the heatmap was located in the lungs. To be specific, only positive images were selected and computed the heatmaps of each model were. The results were considered correct if the maximum value of the heatmaps contained within the lungs (as defined by the automatic lung segmentation). The percentage of correct results over the entire positive images in the test set was then calculated. Grad-CAM [59] was used for the heatmap computation. Furthermore, the ResNet50 model from one of the random seeds was selected for heatmap comparisons.

5. Experimental results

5.1   Preprocessing performance comparison
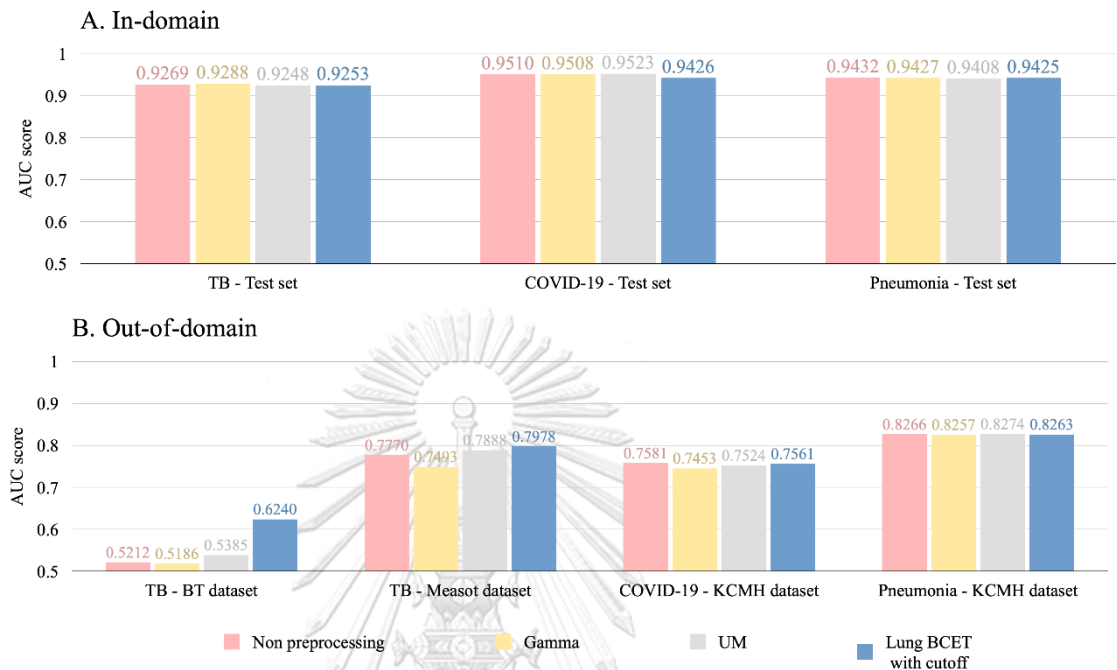
Figure 7: Comparison of the average AUC scores from various preprocessing approaches: (A) In-domain data and (B) Out-of-domain data.

As shown in Fig. 7, the lung BCET received the highest AUC scores of 0.6240 and 0.7978 in two out-of-domain TB data. Gamma received an AUC score of 0.9288, which was the best score, whereas lung BCET received an AUC score of 0.9253, which was 0.0035 lower than the best score.

The outcomes for COVID-19 datasets revealed that the AUC scores for all preprocessing techniques were similar for both in- and out-of-domain data. The UM had the best AUC score for the in-domain data (0.9523), whereas the lung BCET received 0.9426. For the out-of-domain data, the AUC score of the lung BCET was 0.7561, which was the second rank score and lowered only 0.002 from the best score. The lung BCET barely improved performance on the COVID-19 datasets because 1,563 photos (31% of the training sets) in the COVID-19 image segmentation contained segmentation areas of less than 15% of the total image pixels. This suggested that many COVID-19 images did not segment the lung area, thereby affecting the lung BCET process.

For the pneumonia labels, all results had almost identical AUC. Therefore, our study observed that preprocessing strategies had little impact on the performance when training on a large dataset.

Tables 9-11 in appendix A provide more details on the results of the TB, COVID-19, and pneumonia labels and the AUC scores of different backbone networks.
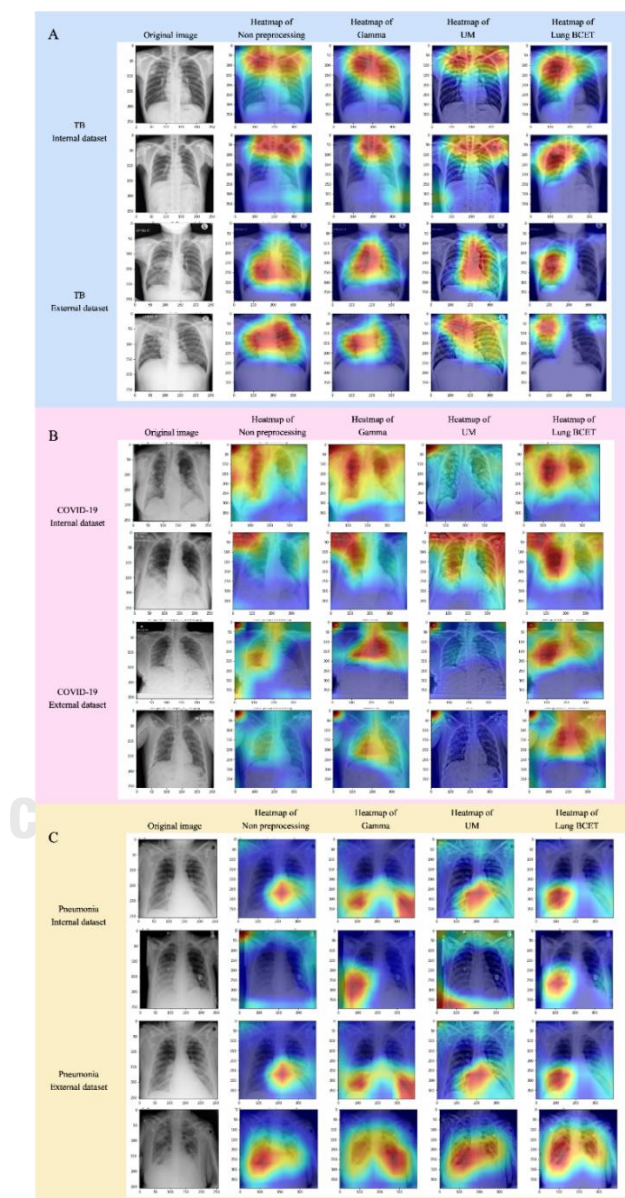
Figure 8: Example of true positive images and their heatmaps for various preprocessing approaches on in- and out-of-domain data: (A) Heatmaps of TB-labeled chest X-ray, (B) Heatmaps of COVID-19-labeled chest X-ray, and (C) Heatmaps of pneumonia-labeled chest X-ray.

According to Fig. 8, the lung BCET method improved the heatmaps of all labels by highlighting the correct regions with a more specific (smaller) area than other methods.

Table 7 demonstrates that the lung BCET enhanced the heatmap accuracy on both in- and out-of-domain settings on the TB label, except for the BT dataset, in which the accuracy of the lung BCET was slightly lower than no preprocessing. For pneumonia labels, the lung BCET enhanced the heatmap accuracy on both in- and out-of-domain data. For COVID-19, the lung BCET received the best percentage on the in-domain data. However, this percentage calculation was less reliable due to many COVID-19 images did not fully segment the lung area.

Table 7: Comparing the percentages of positive image maximum heatmaps using various preprocessing methods.

Label	Preprocessing techniques	percentages of the positive image maximum heatmaps inside the lung area		
		Test set	BT	Maesot
TB	-	54.26%	**63.94%**	55.56%
	Gamma	45.74%	32.12%	33.33%
	UM	27.66%	28.11%	22.22%
	Lung BCET with a 10% cut off	**62.77%**	58.76%	**66.67%**

Label	Preprocessing techniques	percentages of the positive image maximum heatmaps inside the lung area	
		Test set	KMCH-C
COVID-19	-	13.11%	**10.70%**
	Gamma	13.35%	4.17%
	UM	11.12%	2.93%
	Lung BCET	**14.64%**	3.38%

Label	Preprocessing techniques	percentages of the positive image maximum heatmaps inside the lung area	
	with a 40% cut off		
		Test set	KMCH-P
Pneumonia	-	14.24%	11.14%
	Gamma	19.88%	24.20%
	UM	17.27%	16.08%
	Lung BCET with a 15% cut off	**30.56%**	**35.85%**

5.2 Augmentation performance comparison
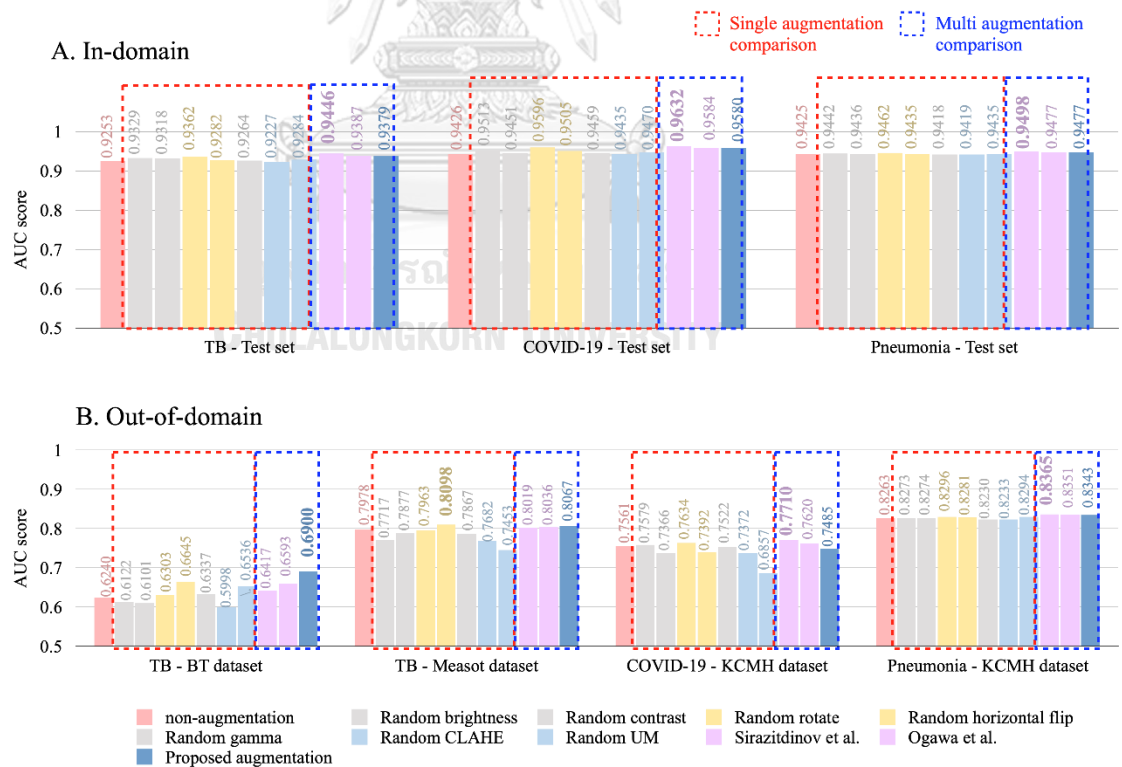
Figure 9: Comparison of the average AUC scores for various augmentation approaches: (A) In-domain data and (B) Out-of-domain data.

For the TB label, multiple augmentations provided better results than a single augmentation for the in-domain datasets. (Fig. 9A) For the in-domain data, random rotation had the greatest AUC score compared to any single augmentation technique. However, the previous work by Sirazitdinov et al. [46] obtained the AUC score of 0.9447 which was the best for in-domain data. Fig. 9B shows that the proposed method achieved the first- and second-rank AUC scores of 0.69 and 0.8067 for the BT and Maesot datasets, respectively.

For the results of the COVID-19 label, previous work of Sirazitdinov et al. [46] demonstrated the best AUC scores for both in- and out-of-domain data. Random rotation achieved the best AUC score in all experiments for single augmentation. Our proposed method achieved the AUC scores of 0.9580 and 0.7458 for the in- and out-of-domain datasets, respectively.

For the pneumonia label, the results demonstrated that all augmentation techniques had similar AUC scores for in- and out-of-domain data. A previous study by Sirazitdinov et al. [46] also achieved the best AUC score in all experiments on both in- and out-of-domain data. Among single augmentation, random rotation achieved the best AUC score. Our proposed method received AUC scores of 0.9477 and 0.8343 for the in- and out-of-domain, respectively, which were 0.0022 lower than the best score.

See Tables 12-14 in appendix A for the augmentation comparison results on the TB, COVID-19, and pneumonia labels, using different backbone models.

For the heatmap comparison of the augmentation techniques, non-augmentation, multiple augmentations (i.e., two related works and the proposed augmentations), and commonly-used augmentations, such as random rotation and random horizontal flip, were selected. Fig. 10 shows that the heatmap from all augmentation approaches reveals similarities in the visualization of the TB and pneumonia labels. For COVID-19, the heatmap proposed by Ogawa et al. [47] highlighted lesions and was better than other augmentation methods.

According to the results in table 8, the method previously proposed by Ogawa et al. [47] received the highest percentages of the positive image maximum heatmaps in the TB (except for the Maesot dataset) and COVID-19 labels. The proposed augmentation had the highest percentages in the pneumonia label (38.10% and 47.32% in the in- and out-of-domain data, respectively).
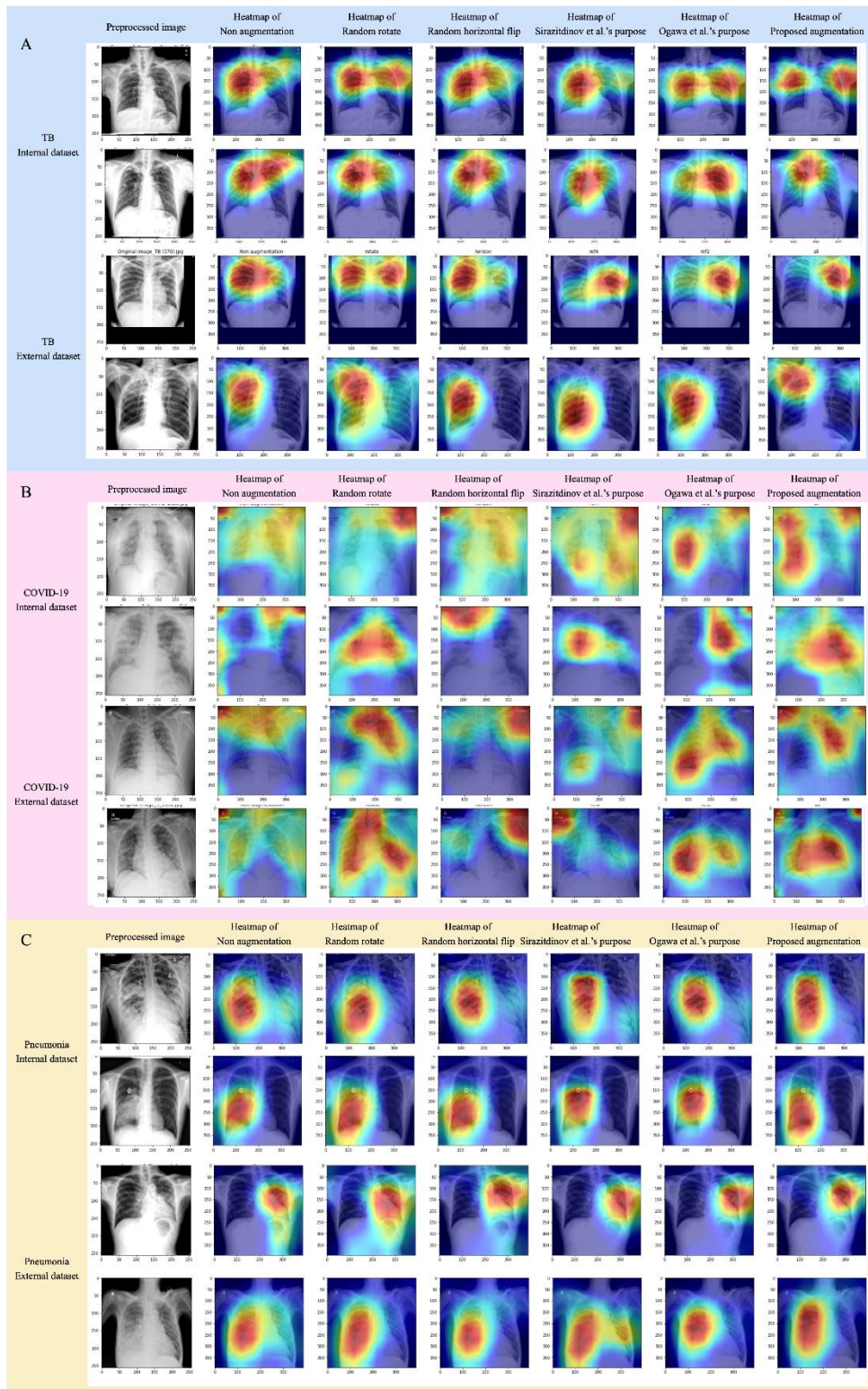
Figure 10: Example of a true positive image and its heatmap from various augmentation approaches on in-domain and out-of-domain data: (A) Heatmaps from the TB label, (B) Heatmaps from the COVID-19 label, and (C) Heatmaps from the pneumonia label.

Table 8: Comparing the percentages of positive image maximum heatmaps using various augmentation techniques.

Label	Augmentation techniques	The percentages of positive image maximum heatmaps inside the lung area		
		Test set	BT	Maesot
TB	-	62.77%	58.76%	**66.67%**
	Random rotate	56.38%	54.98%	44.44%
	Random horizontal flip	50.00%	59.92%	33.33%
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	59.57%	45.71%	22.22%
	Random rotate, Random horizontal flip [Ogawa et al.]	**75.53%**	**67.03%**	44.44%
	Proposed augmentation	23.40%	29.19%	11.11%

Label	Augmentation techniques	The percentages of positive image maximum heatmaps inside the lung area	
		Test set	KMCH-C
COVID-19	-	14.64%	3.38%
	Random rotate	8.67%	11.82%
	Random horizontal flip	2.34%	2.25%
	Random brightness, Random gamma,	12.30%	12.61%

	Random rotate, Random horizontal flip [Sirazitdinov et al.]		
	Random rotate, Random horizontal flip [Ogawa et al.]	**29.39%**	**24.44%**
	Proposed augmentation	4.57%	9.01%

Label	Augmentation techniques	The percentages of positive image maximum heatmaps inside the lung area	
		Test set	KMCH-P
Pneumonia	-	30.56%	35.85%
	Random rotate	24.14%	25.04%
	Random horizontal flip	25.04%	25.71%
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	24.95%	31.41%
	Random rotate, Random horizontal flip [Ogawa et al.]	33.33%	35.68%
	Proposed augmentation	**38.10%**	**47.32%**

6. Discussion

The proposed image preprocessing method showed that lung BCET significantly enhances the model performance on out-of-domain data by reducing the gap between data sources. On the contrary, lung BCET was less likely to improve in-domain data.

This study discovered that lung BCET could improve the model performance, particularly on a small number of training sets, when comparing the results from different sizes of the training sets. Related studies indicated that the size of the training set and the quality of the images affect the model robustness [60] and [61], respectively. According to our findings, the RSNA dataset with the pneumonia label, which contained a large amount of data with good image quality, obtained excellent results across all preprocessing approaches for both in- and out-of-domain data. However, the BT and Maesot datasets with the TB label, which contain a small dataset, had better performance when lung BCET was used as a preprocessing method. Lung BCET was likely to improve the performance on a large training set compared with the other preprocessing methods.

The results from the COVID-19 label demonstrated that image quality significantly influenced lung BCET because lung BCET segments the lung area before adjusting the image. If the image quality is poor (low brightness, blur, or a high level of noise), the segmentation mode had difficulty in segmenting the lung area and was unable to collect the actual minimum and maximum intensity values for the lung area. Consequently, if inaccurate values are used to update the image, an over-enhancement problem or even more detail loss in the image will occur. Consistent with other studies, the quality of the dataset was important for developing models and ensuring their robustness.

Fig. 8 and Table 7 show that lung BCET enhances the heatmaps. Heatmaps from gamma, UM, and no preprocessing described the lesions with more extensive areas than the lung BCET heatmaps. Moreover, lung BCET received the highest percentages of positive images with the maximal heatmaps placed inside the lung area, indicating lung BCET models were more accurately focused on the lung regions than the other methods. Our hypothesis was that the lung segmentation used in the method could better normalize the lung region.

Taken together, our study reported that the lung BCET was a proposed image preprocessing approach that helped improve model performance on small training sets and, apparently, on out-of-domain data with high-quality images.

Another observation worth noting was how the heatmaps are abysmal without lung BCET. This was consistent with the previous study [50], which pointed out that a larger amount of training data (100000+) was required in a weakly-supervised manner to obtain high-quality heatmaps. Without sufficient data, the model cannot detect adequate data variance to display the most important parts. However, our lung BCET can contribute to the model learning to achieve better heatmap quality.

Regarding the augmentation comparisons, Fig. 9 shows that random rotation and random horizontal flip, commonly used in chest radiography research (Fig. 1), had better AUC scores than the other augmentations. When considering a single augmentation technique, random CLAHE, and random UM, the proposed augmentation received a lower AUC score than other popular approaches. When single and multiple augmentations were compared, multiple augmentations yielded a higher AUC score than single augmentation in almost all cases, consistent with earlier studies [17, 26, 27].

For in-domain data, previous work by Sirazitdinov et al. [46] reported the best performance for all the labels. Our multiple augmentations achieved a slightly lower AUC score. However, our proposed augmentation performed well on the TB label in out-of-domain situations. Sirazitdinov et al. [46] obtained the best score for out-of-domain data of the COVID-19 and pneumonia labels. On the contrary, the augmentation of the pneumonia label rarely improved the performance of both in- and out-of-domain data, which indicated that augmentation did not improve the performance on a sizable and high-quality training set.

The heatmap from all augmentation methods highlights the similarities in the visualization. However, Ogawa et al. [47] augmented received the best percentage of positive images in which the maximal heatmap was placed inside the lung area on the test set and BT dataset while no augmentation received the best percent on Maesot dataset.

The heatmaps from all augmentation methods highlighted the similarities in the visualization. However, Ogawa et al. [47] augmentation received the best percentages of positive images because the maximal heatmaps were placed inside the lung area on the test and BT data. In contrast, no augmentation received the best percentages on the Maesot dataset. Our proposed augmentation received the highest percentages on both in- and out-of-domain data in the pneumonia label. The heatmaps of the COVID-19 label show that augmentation affected the model attention. Augmentation by Ogawa et al. [47] obtained the maximal heatmaps placed

inside the lung area and the best percentages of positive images, reflecting that its heatmap was highlighted more correctly than other methods.

As mentioned above, previous studies by Sirazitdinov et al. [46] and Ogawa et al. [47] achieved better performances than our proposed method. A limitation of this study was that random CLAHE and UM parameters were selected by running DE and identifying the parameters with the highest AUC scores from the validation set. Due to limited image resources, ten epochs were used to run the DE, whereas more epochs could be run to obtain the optimal parameter value. This parameter value was not necessary for the COVID-19 and pneumonia labels, which contained larger training sets than the TB label, which required more epochs for optimal performance.

7. Conclusion and future work

7.1 Conclusion

Lung BCET, which performs color normalization only on the lung region, could improve the AUC score in out-of-domain data and small training set with a high-quality dataset while improve heatmap visualizations in two following scenarios: in- and out-of-domain settings. The proposed multi-augmentations could improve the AUC score of the TB labels but achieved a performance lower than augmentations in related studies on the COVID-19 and pneumonia labels. However, the lower heatmap scores inspire further investigation for improving the performance and the turning parameter to improve the performance of the proposed augmentation on the COVID-19 and pneumonia labels.

7.2 Future work

Poor-quality data must be enhanced to increase its quality before lung BCET is applied. In future work, experiments to identify image enhancements (such as adjusting the brightness and sharpness) that provide higher quality are beneficial. For the proposed augmentation, it could not demonstrate better results with our parameter values than other augmentations due to resource limitations. Therefore, running more differential evolution or turning training hyperparameter may be necessary to discover an appropriate parameter value.

8. Appendix A. Comparing the average AUC score of the different backbone networks

Table 9: Comparing the average AUC score of different preprocessing techniques on the tuberculosis label using the different backbone networks.

Model	Preprocessing techniques	AUC score		
		Test set	BT	Maesot
ResNet50	-	0.9336	0.6330	0.7988
	Gamma	0.9365	0.6271	0.8048
	UM	0.9426	0.5518	**0.8464**
	Lung BCET with a 10% cut off	**0.9444**	**0.7618**	0.8269
EfficientNetB0	-	0.9200	0.3514	0.7616
	Gamma	**0.9276**	0.3908	0.7704
	UM	0.9043	0.4034	**0.7857**
	Lung BCET with a 10% cut off	0.9208	**0.4810**	0.7497
PylonResNet50	-	0.9272	0.5792	0.7706
	Gamma	0.9221	0.5380	0.6726
	UM	**0.9275**	**0.6601**	0.7344
	Lung BCET with a 10% cut off	0.9106	0.6294	**0.8167**

Table 10: Comparing the average AUC score of different preprocessing techniques on the COVID-19 label using the different backbone networks.

Model	Preprocessing techniques	AUC score	
		Test set	KMCH-C
ResNet50	-	0.9567	0.7599
	Gamma	0.9550	0.7732
	UM	**0.9585**	0.7782
	Lung BCET with a 40% cut off	0.9516	**0.8000**
EfficientNetB0	-	0.9460	**0.7703**
	Gamma	**0.9466**	0.7632
	UM	0.9422	0.7334
	Lung BCET with a 40% cut off	0.9308	0.7240
PylonResNet50	-	0.9504	0.7442
	Gamma	0.9507	0.6996
	UM	**0.9562**	**0.7457**
	Lung BCET with a 40% cut off	0.9454	0.7442

Table 11: Comparing the average AUC scores of different preprocessing techniques on the pneumonia label using the different backbone networks.

Model	Preprocessing techniques	AUC score	
		Test set	KMCH-P
ResNet50	-	0.9454	0.8296
	Gamma	0.9451	0.8280
	UM	0.9447	0.8318
	Lung BCET with a 15% cut off	**0.9473**	**0.8319**
EfficientNetB0	-	**0.9415**	0.8199
	Gamma	0.9409	0.8179
	UM	0.9407	**0.8214**
	Lung BCET with a 15% cut off	0.9388	0.8147
PylonResNet50	-	**0.9428**	0.8304
	Gamma	0.9421	0.8312
	UM	0.9370	0.8290
	Lung BCET With a 15% cut off	0.9415	**0.8324**

Table 12: Comparing the average AUC scores of different augmentation techniques on the tuberculosis label using the different backbone networks.

Model	Augmentation techniques	AUC score		
		Test set	BT	Maesot
ResNet50	-	0.9444	0.7618	0.8269
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	0.9511	0.7246	0.8393
	Random rotate, Random horizontal flip [Ogawa et al.]	**0.9561**	0.7447	0.8545
	Random brightness	0.9415	0.6996	0.8001
	Random contrast	0.9454	0.7416	0.8366
	Random rotate	0.9468	0.6988	**0.8593**
	Random horizontal flip	0.9390	0.7575	0.8403
	Random gamma	0.9372	0.7214	0.8312
	Random CLAHE	0.9343	0.6942	0.8201
	Random UM	0.9355	**0.7801**	0.8074
	Select one of Random UM, Random CLAHE, Random gamma, and Random horizontal flip, Random rotate [combinations between	0.9415	0.7354	0.8333

Model	Augmentation techniques	AUC score		
		Test set	BT	Maesot
	proposed and baseline augmentation]			
EfficientNetB0	-	0.9208	0.4810	0.7497
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	**0.9364**	0.5126	0.7947
	Random rotate, Random horizontal flip [Ogawa et al.]	0.9239	**0.6305**	0.7884
	Random brightness	0.9283	0.5027	0.7759
	Random contrast	0.9266	0.4271	0.7529
	Random rotate	0.9242	0.5464	0.7883
	Random horizontal flip	0.9278	0.5152	**0.7990**
	Random gamma	0.9222	0.4650	0.7664
	Random CLAHE	0.9219	0.4791	0.7260
	Random UM	0.9196	0.5257	0.7504
	Select one of Random UM, Random CLAHE, Random gamma, and Random horizontal flip, Random rotate	0.9299	0.5867	0.7959

Model	Augmentation techniques	AUC score		
		Test set	BT	Maesot
	[combinations between proposed and baseline augmentation]			
PylonResNet50	-	0.9106	0.6294	**0.8167**
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	**0.9463**	0.6878	0.7717
	Random rotate, Random horizontal flip [Ogawa et al.]	0.9363	0.6027	0.7678
	Random brightness	0.9290	0.6343	0.7391
	Random contrast	0.9234	0.6617	0.7735
	Random rotate	0.9376	0.6457	0.7413
	Random horizontal flip	0.9179	0.7209	0.7902
	Random gamma	0.9196	0.7146	0.7625
	Random CLAHE	0.9119	0.6261	0.7584
	Random UM	0.9302	0.6549	0.6779
	Select one of Random UM, Random CLAHE, Random gamma, and Random horizontal flip, Random rotate	0.9424	**0.7479**	0.7909

Model	Augmentation techniques	AUC score		
		Test set	BT	Maesot
	[combinations between proposed and baseline augmentation]			

Table 13: Comparing the average AUC scores of different augmentation techniques on the COVID-19 label using the different backbone networks.

Model	Augmentation techniques	AUC score	
		Test set	BT
ResNet50	-	0.9516	**0.8000**
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	0.9616	0.7729
	Random rotate, Random horizontal flip [Ogawa et al.]	**0.9628**	0.7604
	Random brightness	0.9567	0.7820
	Random contrast	0.9507	0.7613
	Random rotate	0.9627	0.7559
	Random horizontal flip	0.9560	0.7512
	Random gamma	0.9541	0.7778
	Random CLAHE	0.9507	0.7834
	Random UM	0.9527	0.6360
	Select one of Random UM, Random CLAHE, Random gamma, and Random horizontal flip, Random rotate [combinations between	0.9562	0.7417

Model	Augmentation techniques	AUC score	
		Test set	BT
	proposed and baseline augmentation]		
EfficientNetB0	-	0.9308	0.7240
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	**0.9647**	0.7872
	Random rotate, Random horizontal flip [Ogawa et al.]	0.9541	0.7845
	Random brightness	0.9450	0.7301
	Random contrast	0.9376	0.7224
	Random rotate	0.9581	**0.8059**
	Random horizontal flip	0.9460	0.7483
	Random gamma	0.9376	0.7211
	Random CLAHE	0.9375	0.7270
	Random UM	0.9459	0.7087
	Select one of Random UM, Random CLAHE, Random gamma, and Random horizontal flip, Random rotate	0.9610	0.7790

Model	Augmentation techniques	AUC score	
		Test set	BT
	[combinations between proposed and baseline augmentation]		
PylonResNet50	-	0.9454	0.7442
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	**0.9633**	0.7530
	Random rotate, Random horizontal flip [Ogawa et al.]	0.9583	0.7410
	Random brightness	0.9521	**0.7616**
	Random contrast	0.9471	0.7261
	Random rotate	0.9580	0.7285
	Random horizontal flip	0.9494	0.7180
	Random gamma	0.9461	0.7576
	Random CLAHE	0.9422	0.7011
	Random UM	0.9424	0.7123
	Select one of Random UM, Random CLAHE, Random gamma, and Random horizontal flip, Random rotate	0.9568	0.7247

Model	Augmentation techniques	AUC score	
		Test set	BT
	[combinations between proposed and baseline augmentation]		

Table 14: Comparing the average AUC score of different augmentation techniques on the pneumonia label using the different backbone networks.

Model	Augmentation techniques	AUC score	
		Test set	BT
ResNet50	-	0.9473	0.8319
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	0.9489	0.8379
	Random rotate, Random horizontal flip [Ogawa et al.]	**0.9503**	**0.8406**
	Random brightness	0.9469	0.8284
	Random contrast	0.9460	0.8300
	Random rotate	0.9465	0.8318
	Random horizontal flip	0.9471	0.8327
	Random gamma	0.9449	0.8260
	Random CLAHE	0.9450	0.8250
	Random UM	0.9453	0.8325
	Select one of Random UM, Random CLAHE, Random gamma, and Random horizontal flip, Random rotate [combinations between	0.9487	0.8357

Model	Augmentation techniques	AUC score	
		Test set	BT
	proposed and baseline augmentation]		
EfficientNetB0	-	0.9388	0.8147
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	**0.9480**	**0.8279**
	Random rotate, Random horizontal flip [Ogawa et al.]	0.9457	0.8275
	Random brightness	0.9430	0.8198
	Random contrast	0.9424	0.8187
	Random rotate	0.9454	0.8211
	Random horizontal flip	0.9433	0.8195
	Random gamma	0.9397	0.8165
	Random CLAHE	0.9410	0.8180
	Random UM	0.9428	0.8229
	Select one of Random UM, Random CLAHE, Random gamma, and Random horizontal flip, Random rotate	0.9456	0.8273

Model	Augmentation techniques	AUC score	
		Test set	BT
	[combinations between proposed and baseline augmentation]		
PylonResNet50	-	0.9415	0.8324
	Random brightness, Random gamma, Random rotate, Random horizontal flip [Sirazitdinov et al.]	**0.9527**	**0.8438**
	Random rotate, Random horizontal flip [Ogawa et al.]	0.9471	0.8371
	Random brightness	0.9426	0.8337
	Random contrast	0.9426	0.8336
	Random rotate	0.9467	0.8359
	Random horizontal flip	0.9402	0.8321
	Random gamma	0.9410	0.8266
	Random CLAHE	0.9396	0.8268
	Random UM	0.9423	0.8328
	Select one of Random UM, Random CLAHE, Random gamma, and Random horizontal flip, Random rotate	0.9489	0.8398

Model	Augmentation techniques	AUC score	
		Test set	BT
	[combinations between proposed and baseline augmentation]		

REFERENCES

1. Kushol, R., et al., *Contrast enhancement of medical x-ray image using morphological operators with optimal structuring element.* arXiv preprint arXiv:1905.08545, 2019.

2. More, L.G., et al. *Parameter tuning of CLAHE based on multi-objective optimization to achieve different contrast levels in medical images*. in *2015 IEEE International Conference on Image Processing (ICIP)*. 2015. IEEE.

3. Zhao, Z. and Y. Zhou. *Comparative study of logarithmic image processing models for medical image enhancement*. in *2016 IEEE international conference on systems, man, and cybernetics (SMC)*. 2016. IEEE.

4. Rajaraman, S. and S.K. Antani, *Modality-specific deep learning model ensembles toward improving TB detection in chest radiographs.* IEEE Access, 2020. **8**: p. 27318-27326.

5. Oyelade, O.N., A.E.-S. Ezugwu, and H. Chiroma, *CovFrameNet: An enhanced deep learning framework for COVID-19 detection.* Ieee Access, 2021. **9**: p. 77905-77919.

6. Elgendi, M., et al., *The effectiveness of image augmentation in deep learning networks for detecting COVID-19: A geometric transformation perspective.* Frontiers in Medicine, 2021. **8**: p. 629134.

7. Monshi, M.M.A., et al., *CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR.* Computers in biology and medicine, 2021. **133**: p. 104375.

8. Chlap, P., et al., *A review of medical image data augmentation techniques for deep learning applications.* Journal of Medical Imaging and Radiation Oncology, 2021. **65**(5): p. 545-563.

9. Sundaram, S. and N. Hulkund, *Gan-based data augmentation for chest X-ray classification.* arXiv preprint arXiv:2107.02970, 2021.

10. Ganesan, P., et al. *Assessment of data augmentation strategies toward performance improvement of abnormality classification in chest radiographs*. in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2019. IEEE.

11. Guo, L.J., *Balance contrast enhancement technique and its application in image colour composition.* Remote Sensing, 1991. **12**(10): p. 2133-2151.

12. Pizer, S.M., et al., *Adaptive histogram equalization and its variations.* Computer vision, graphics, and image processing, 1987. **39**(3): p. 355-368.

13. Polesel, A., G. Ramponi, and V.J. Mathews, *Image enhancement via adaptive unsharp*

masking. IEEE transactions on image processing, 2000. **9**(3): p. 505-510.

14. Storn, R. and K. Price, *Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces.* Journal of global optimization, 1997. **11**(4): p. 341-359.

15. Sathitratanacheewin, S., P. Sunanta, and K. Pongpirul, *Deep learning for automated classification of tuberculosis-related chest X-Ray: dataset distribution shift limits diagnostic performance generalizability.* Heliyon, 2020. **6**(8): p. e04614.

16. Zech, J.R., et al., *Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study.* PLoS medicine, 2018. **15**(11): p. e1002683.

17. Munadi, K., et al., *Image Enhancement for Tuberculosis Detection Using Deep Learning.* IEEE Access, 2020. **8**: p. 217897-217907.

18. Zotin, A., et al., *Lung boundary detection for chest X-ray images classification based on GLCM and probabilistic neural networks.* Procedia Computer Science, 2019. **159**: p. 1439-1448.

19. Rahman, T., et al., *Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images.* Computers in biology and medicine, 2021. **132**: p. 104319.

20. Pezzotti, W., *Chest X-ray interpretation: not just black and white.* Nursing2020, 2014. **44**(1): p. 40-47.

21. Abbas, A. and M.M. Abdelsamea. *Learning transformations for automated classification of manifestation of tuberculosis using convolutional neural network.* in *2018 13th International Conference on Computer Engineering and Systems (ICCES)*. 2018. IEEE.

22. Yadav, O., K. Passi, and C.K. Jain. *Using deep learning to classify X-ray images of potential tuberculosis patients.* in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2018. IEEE.

23. Ahsan, M., R. Gomes, and A. Denton. *Application of a Convolutional Neural Network using transfer learning for tuberculosis detection.* in *2019 IEEE International Conference on Electro Information Technology (EIT)*. 2019. IEEE.

24. Meraj, S.S., et al., *Detection of pulmonary tuberculosis manifestation in chest x-rays using different convolutional neural network (CNN) models.* Int. J. Eng. Adv. Technol.(IJEAT), 2019. **9**(1): p. 2270-2275.

25. Rohilla, A., R. Hooda, and A. Mittal. *TB detection in chest radiograph using deep learning architecture.* in *Proceeding of 5th International Conference on Emerging Trends in Engineering, Technology, Science and Management (ICETETSM-17)*. 2017.

26. Hernández, A., Á. Panizo, and D. Camacho. *An ensemble algorithm based on deep learning for tuberculosis classification*. in *International conference on intelligent data engineering and automated learning*. 2019. Springer.

27. Nguyen, Q.H., et al. *Deep learning models for tuberculosis detection from chest x-ray images*. in *2019 26th International Conference on Telecommunications (ICT)*. 2019. IEEE.

28. Allaouzi, I. and M.B. Ahmed, *A novel approach for multi-label chest X-ray classification of common thorax diseases.* IEEE Access, 2019. **7**: p. 64279-64288.

29. Duong, L.T., et al., *Detection of tuberculosis from chest X-ray images: boosting the performance with vision transformer and transfer learning.* Expert Systems with Applications, 2021. **184**: p. 115519.

30. Sharma, H., et al. *Feature extraction and classification of chest x-ray images using cnn to detect pneumonia*. in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. 2020. IEEE.

31. De Moura, J., et al., *Deep convolutional approaches for the analysis of covid-19 using chest x-ray images from portable devices.* IEEE Access, 2020. **8**: p. 195594-195607.

32. Misra, S., et al., *Multi-channel transfer learning of chest X-ray images for screening of COVID-19.* Electronics, 2020. **9**(9): p. 1388.

33. Sitaula, C. and M.B. Hossain, *Attention-based VGG-16 model for COVID-19 chest X-ray image classification.* Applied Intelligence, 2021. **51**(5): p. 2850-2863.

34. Rahman, T., et al., *Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray.* Applied Sciences, 2020. **10**(9): p. 3233.

35. Asif, S., et al. *Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Network*. in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*. 2020. IEEE.

36. Sirazitdinov, I., et al., *Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database.* Computers & electrical engineering, 2019. **78**: p. 388-399.

37. Heidari, M., et al., *Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms.* International journal of medical informatics, 2020. **144**: p. 104284.

38. Nayak, S.R., et al., *Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study.* Biomedical Signal Processing and Control, 2021. **64**: p. 102365.

39. Seyyed-Kalantari, L., et al. *CheXclusion: Fairness gaps in deep chest X-ray classifiers*. in *BIOCOMPUTING 2021: Proceedings of the Pacific Symposium*. 2020. World Scientific.

40. Basu, S., S. Mitra, and N. Saha. *Deep learning for screening covid-19 using chest x-ray*

images. in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2020. IEEE.

41.   Minaee, S., et al., *Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning.* Medical image analysis, 2020. **65**: p. 101794.

42.   Wang, L., Z.Q. Lin, and A. Wong, *Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images.* Scientific Reports, 2020. **10**(1): p. 1-12.

43.   Chouhan, V., et al., *A novel transfer learning based approach for pneumonia detection in chest X-ray images.* Applied Sciences, 2020. **10**(2): p. 559.

44.   Jain, R., et al., *Deep learning based detection and analysis of COVID-19 on chest X-ray images.* Applied Intelligence, 2021. **51**(3): p. 1690-1700.

45.   Zhang, J., et al., *Covid-19 screening on chest x-ray images using deep learning based anomaly detection.* arXiv preprint arXiv:2003.12338, 2020. **27**.

46.   Sirazitdinov, I., et al. *Data Augmentation for Chest Pathologies Classification*. in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 2019. IEEE.

47.   Ogawa, R., T. Kido, and T. Mochizuki, *Effect of augmented datasets on deep convolutional neural networks applied to chest radiographs.* Clinical radiology, 2019. **74**(9): p. 697-701.

48.   He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

49.   Tan, M. and Q. Le. *Efficientnet: Rethinking model scaling for convolutional neural networks*. in *International Conference on Machine Learning*. 2019. PMLR.

50.   Preechakul, K., et al., *Improved image classification explainability with high accuracy heatmaps.* Iscience, 2022: p. 103933.

51.   Jaeger, S., et al., *Two public chest X-ray datasets for computer-aided screening of pulmonary diseases.* Quantitative imaging in medicine and surgery, 2014. **4**(6): p. 475.

52.   Vayá, M.D.L.I., et al., *Bimcv covid-19+: a large annotated dataset of rx and ct images from covid-19 patients.* arXiv preprint arXiv:2006.01174, 2020.

53.   Tabik, S., et al., *COVIDGR dataset and COVID-SDNet methodology for predicting COVID-19 based on chest X-ray images.* IEEE journal of biomedical and health informatics, 2020. **24**(12): p. 3595-3605.

54.   Cohen, J.P., et al., *Covid-19 image data collection: Prospective predictions are the future.* arXiv preprint arXiv:2006.11988, 2020.

55.   *covid-19-image-repository [Online]*. Available from: https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png.

56.   *the Italian Society of Medical Radiology COVID-19 dataset* Available from: https://sirm.org/category/senza-categoria/covid-19/.

57. RSNA. *RSNA Pneumonia Detection Challenge*. 2018; Available from: https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pneumonia-detection-challenge-2018.

58. Ovcharenko, I. *Lung-segmentation*. Available from: https://github.com/IlliaOvcharenko/lung-segmentation.

59. Selvaraju, R.R., et al. *Grad-cam: Visual explanations from deep networks via gradient-based localization*. in *Proceedings of the IEEE international conference on computer vision*. 2017.

60. Lei, S., et al., *How training data affect the accuracy and robustness of neural networks for image classification.* 2018.

61. Ren, S. and C.Q. Li, *Robustness of transfer learning to image degradation.* Expert Systems with Applications, 2022. **187**: p. 115877.

VITA

NAME   Wasunan Chokchaithanakul

DATE OF BIRTH   6 December 1995

PLACE OF BIRTH   Bangkok

INSTITUTIONS ATTENDED   Chulalongkorn University

HOME ADDRESS   29/52(114) Wararom Phetkasem 81 Village, Bangbon4 Rose,
                            Bangbon, Bangbon, Bangkok, 10150