IMPROVISED EXPLOSIVE DEVICE DETECTION USING CNN WITH X-RAY IMAGES

CHAKKAPHAT CHAMNANPHAN

MASTER OF SCIENCE
IN
INFORMATION TECHNOLOGY

SCHOOL OF INFORMATION TECHNOLOGY
MAE FAH LUANG UNIVERSITY
2022

IMPROVISED EXPLOSIVE DEVICE DETECTION USING CNN WITH X-RAY IMAGES

CHAKKAPHAT CHAMNANPHAN

THIS THESIS IS A PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
INFORMATION TECHNOLOGY

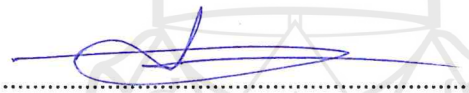
SCHOOL OF INFORMATION TECHNOLOGY
MAE FAH LUANG UNIVERSITY
2022

IMPROVISED EXPLOSIVE DEVICE DETECTION USING CNN WITH X-RAY IMAGES

CHAKKAPHAT CHAMNANPHAN

THIS THESIS HAS BEEN APPROVED
TO BE A PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE
IN
INFORMATION TECHNOLOGY
2022

EXAMINATION COMMITTEE



.....CHAIRPERSON
(Asst. Prof. Santichai Wicha, Ph. D.)

.....ADVISOR
(Assoc. Prof. Wg. Cdr. Tossapon Boongoen, Ph. D.)

..... EXAMINER
(Assoc. Prof. Natthakan Iam-on, Ph. D.)

.....EXAMINER
(Sujitra Arwatchananukul, Ph. D.)

.....EXTERNAL EXAMINER
(Assoc. Prof. Wg. Cdr. Thiansiri Luangwilai, Ph. D.)

ACKNOWLEDGEMENT

I would like to extend my sincere thanks to my advisor Assoc. Prof. Wg. Cdr. Dr. Tossapon Boongoen for giving me the opportunity to let me participate in a defense fund project. This project would not have been possible without the funding of Defense Technology Institute (DTI). Moreover, I would like to offer my special thanks to my advisor Assoc. Prof. Wg.Cdr. Dr. Tossapon Boongoen and my co-advisor Assoc. Prof. Dr. Natthakan Iam-on for their assistance at every stage of the research project.

Chakkaphat Chamnanphan

Thesis Title Improvised Explosive Device Detection Using CNN with X-Ray Images

Author Chakkaphat Chamnanphan

Degree Master of Science (Information Technology)

Advisor Assoc. Prof. Wg. Cdr. Tossapon Boongoen, Ph. D.

ABSTRACT

The idea of a smart city and other related services have been explored widely in terms of innovation development as well as applications of technological concepts. One of the major concerns to promote smart living is the security of personal life and asset, which have been put at risk by organized crime and acts of terrorism. A great deal of attention is paid to preventing terrorism incidence in public areas. In particular, to detect improvised explosive devices known as IEDs. This research focuses on developing analytical models that can recognize X-ray images of baggage contained with IEDs. It provides an alternative to conventional techniques that fail to disclose the truth with concealed or hidden devices. Specific to the current project, sample images are generated by experts to cover a number of cases found in operations during the past decade. These are then used to develop a deep learning model, with the help of several data augmentation methods to overcome the problem of small training datasets. As a result, models are more accurate by measured from sample datasets that never existed in training datasets, with the highest accuracy rate of 0.985. In addition, an empirical study is conducted herein to determine the size of the train set that exhibits good predictive performance.

Keywords: Improvised Explosive Device, Detection, X-Ray Image, Classification, CNN, Augmentation

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES (continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Background and Rational

Given a recent wave of development and technological applications for smart cities, different sensors and as well as artificial intelligence-led systems have been put forward to monitor and analyze behavioural patterns [1]. Based on the survey of [2], several previous works can be categorized under this umbrella. These include smart people [3], smart environment [4] and public safety [5], among others. In addition to road and transport safety, the last subject also highly overlaps with the domain of security, especially those related to organized crime and terrorism. To this end, various approaches have been introduced to tackle this threat to public and individual safety [6]. For instance, some researchers focus on discovering groups and communication trends from online and other relevant resources [7]. Some others direct their attention to cybersecurity, which has emerged as a significant theatre for terrorist attacks [8]. Besides, a few more have concentrated on the identification and detection of chemical, biological, and explosive (CBE) substances [9]. One of the most popular in terrorism is the use of improvised explosive devices hidden in baggage. The need for computational methods of detection and automation tools is in great demand. [10].

IED terror attacks have commonly occurred in densely populated and busy areas such as airports, and official and commercial buildings. Mostly, IEDs are concealed in a package or baggage such that common chemical-led detectors [11] become much less effective. This leads to an alternative of recognizing IEDs based on shape, density, and composition, which can be captured by an imaging sensor like an x-ray [12]. The aforementioned trend motivates the current research that aims to investigate the use of convolutional neural networks (CNN) to develop a recognition capability for an

accurate IED detection system. Details of the project and research work are given below.

Terrorist incidents in Thailand's three southern provinces have become a significant problem for national security, including the safety of life and property of government officials and people in the area. It also affects economic losses and investment. Of course, this initiates a negative impact on tourism, which is one of the main sources of national income [13]. Despite the fact that the government and Ministry of Defense have worked together to solve the problem continuously, which has been effective to a limited extent. There are still periodic bombings in public areas and ambushes of government officials. X-ray explosive detection systems that can work with ordnance-disposal robots [14] can be an essential decision-making aid and mechanism for preventing mishaps and minimizing losses. This work is a part of the collaborative project between The Defense Technology Institute (DTI), Mae Fah Luang University, and The Department of Ordnance, Royal Thai Air Force (RTAF) with the aim of reducing the import of high-value technology from abroad. Note that it follows the line of research studies previously conducted by this collaboration, e.g., face recognition [15-16] and suspect vehicle detection [17].

1.2 Objective

1.2.1 The CNN model development is based on x-ray images of baggage which contain improvised explosive devices.

1.2.2 The knowledge of using CNN models will be applied to robots for the detection of explosives.

1.3 Scope

The model was developed by using the image data captured through portable x-ray camera equipment installed on the robot system. As such, the image quality is lower than those produced by commercial units, which are normally seen in airports and noteworthy that these samples correspond to actual incidents found in the south of Thailand over a decade. 2-D images of cases are collected from representative scenarios designed by a team of experts from DTI and RTAF. The thesis is organized as presents concepts and related works of IED and detection techniques, with the focus being on image-based approaches. Details of the proposed method and materials exploited in this research are included. Also, findings related to the data augmentation process and parameter analysis are elaborated therein.

1.4 Research Plan

This study follows a convolutional neural network development process which starts from problem definition to data augmentation for model development and evaluation. These workflows can be summarized as follows.

Stage 1 – problem definition: They are conducted at the beginning of the project in collaboration work with experts from partner organizations, e.g., DTI and RTAF to clarify the research questions and scope of the investigation.

Stage 2 – The second stage is to find the progress made for the previously defined problem. This focuses on the methods and procedures used to detect explosives on the data science methods. It also provides a review of a way to add datasets in order to solve small datasets so that they can be used for model development and comparison with reported benchmarks.

Stage 3 – data collection and preparation: experts compiled the X-ray image dataset from the Royal Thai Air Force Ordnance Department in the form of improvised explosive devices hidden in baggage. Since the datasets provided by the experts are not enough to train the system, data manipulation methods are used to ensure that the

datasets are valid for future development, e.g., Data Augmentation will be applied to this information. to be standardized and ready for the following stages.

Stage 4 – development and evaluation of transformation feature approach: The primary classification model is built on the original feature set of the convolutional neural network and the specific dataset obtained with data augmentation. These were assessed using the confusion matrix for evaluation accuracy, recall, precision and F1.

Stage 5 – prepare publication and the final report.

1.5 Project Plan

This schedule consists of columns, which are the semester and rows, which are phrases that were defined in the methodology above. The black highlight is the stage of work that is performed that semester.

Given those stages within the proposed methodology, the next table summarizes the planning and timeline of this study in Table 1.1.

Table 1.1 Project time schedule divided by semester

Phases/ Semester	**1st Semester**	**2nd Semester**	**3rd Semester**	**4th Semester**	**5th Semester**	**6th Semester**
Stage 1						
Stage 2						
Stage 3						
Stage 4						
Stage 5						

CHAPTER 2

LITERATURE REVIEW

2.1 Related Works

Starting around 1990s, a rapid spread of terrorist attacks has been witnessed worldwide, with an explosion incident being the common theme [18-20]. Given the tight control of military explosives in most countries, improvised explosive devices (IEDs) have become the major tool for terrorism, as they are a lot easier to obtain [21]. They are normally concealed and hidden in a passenger's baggage, especially for air transportation, to avoid a positive detection [12]. Despite being effective for common explosives, many detection methods, for instance, mass spectrometry [22], ion mobility spectrometry [11], Raman spectrometry [23] and fluorescence sensor [24], are inherently less useful. As a result, since the event on September 11, 2001, many studies have shifted to explore the visual inspection of x-ray images, which can support a visual search and decision making rather well [25]. With this direction that is in line with the current research, Table 2.1 summarizes previous works that relate applications of artificial intelligence (AI) techniques to the visual inspection of IEDs.

Recognition of objects in an x-ray image is one of the difficult problems in the conjunction of computer vision and data science [26]. This is pretty much due to occlusion between the object of interest and others, self-occlusion, clutter in the background and variation of viewpoints, for instance. Hence, manual feature engineering approaches such as those implemented by [27-34] may be overfitting specific data sets exploited for their empirical studies. Even though they are compatible with most feature-based learning techniques, there is no guarantee to carry good predictive performance forward to a new image collection. To overcome such a burden, several investigations have made use of a deep learning model to determine informative features automatically [35-41]. It is assumed that a simple CNN can be effective for the

classification task with 2-D x-ray images, based on previous reports [35, 39] that focus on dissimilar contexts of 3-D imaging or object detection. In fact, it is reported by several works that the CNN with a standard layer set can be competitive to more complex counterparts [42-43]. However, both conditions of the number of samples and the diversity among them must be met for a model to deliver its potential [44]. This is demonstrated by the development of ImageNet [45] and Google Net [46] with more than 10 million images being available for model training. Unfortunately, most datasets of x-ray images are not publicly available, yet it is more difficult and expensive to generate one. It is true for many projects including the current one where the size of the dataset is quite small.

To overcome the aforementioned problem, several recent works have introduced a range of algorithms to optimize the quality of target model learned from a small sample set. At first, different techniques have been proposed to directly modify structural and algorithmic processes within a network. One famous example of the family is the introduction of dropout layers [47] to prevent overfitting through reducing parameters to update, thus solving the problem of insufficient training images. The second alternative is called transfer learning [48], by which a model trained in one domain is re-trained with samples specific to a new subject. The last category that is adopted by this research is the methodology of data augmentation that increases the number of samples by slightly changing original images using operations like flipping [49], rotation [50] and color space shifting [51]. For the task of x-ray image classification, this approach has been highly useful practice in a survey on Image Data Augmentation for Deep Learning [52] and the review of a Comprehensive Survey of CT Image Augmentation Techniques for classification based on convolutional neural networks [53-55].

Table 2.1 Summarization of related works on x-ray visual inspection of IEDs. Note that abbreviations used herein are listed as follows: SVM (support vector machine), SIFT (Scale-invariant feature transform), KNN (k-nearest neighbors), SURF (speeded-up robust feature), RF (random forest), and R-CNN (region-based CNN), respectively

Relate work & Reference	Context of visual inspection	Exploited AI techniques
Bag-of-visual-word feature [27-28]	2D x-ray image of baggage	Classification: SVM & SIFT
Shape-based modeling [29-30]	2D x-ray image of baggage	Classification: fuzzy KNN
Robust bag-of-visual-word [31-32]	2D x-ray image of baggage	Classification: SVM & SURF
Sparse texture descriptor [33]	2D x-ray image of baggage	Classification: SVM
3D object classification [34]	2D x-ray image of baggage	Classification: Clustering & RF
Transfer learning [35]	2D x-ray image of baggage	Object detection: CNN
Visual object detection [36-37]	2D x-ray image of baggage	Object detection: R-CNN
Active vision approach [38]	2D x-ray image of baggage	Object detection: Q-learning
Attention based detection [39]	2D x-ray image of baggage	Object detection: CNN
Object-wise anomaly detection [40]	2D x-ray image of baggage	Object detection: Dual-CNN
Change detection [41]	2D x-ray ground image	Object detection: CNN

2.2 Related Techniques

2.2.1 Supervised Learning Method

In this research, we also review the literature techniques of supervised learning classification to become knowledgeable to compare the method technique by using the basic classification techniques including K-Nearest Neighbor (KNN) [29-30], Random Forest [56] and Support Vector Machine (SVM) [27 - 33] and the project has developed the classification model by using Convolutional Neural Networks [35-41].

K-Nearest Neighbor (KNN) is the simplest and most efficient algorithm for finding object classes. It is a class method for categorizing data using the principle of comparing the data of interest to other data as being similar [57].

Random forest is a group called assembly learning. The principle is to practice multiple times in the same data set. In each training, they choose a part of their training data, and then they decide who is the most selective [57].

Support Vector Machine (SVM) is a statistical approach to classification that uses the principle of finding the coefficients of equations to create a separating line for data fed into the learning process. focusing on the lines that distinguish groups of data [58].

2.2.1.1 Architecture of feature learning with CNN

Convolutional Neural Network (CNN) is one of the bio-inspired neural networks, with CNN simulating human vision in fragmented areas and clustering them together to see what you see. When looking at subareas, Human attributes are different, such as contrasting lines and colors. Humans know that this area is a straight line or contrasting color. Because humans look at both the focus and the surrounding area together [59].

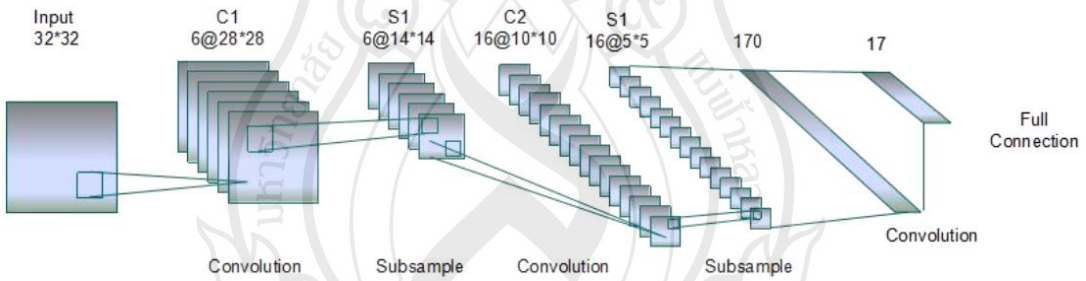


Figure 2.1 Architecture of CNN in training

2.2.1.2 Optimizer

Optimization has proven to be an essential step. There are several common optimization tools that are widely used, such as SGD, Adagrad, Adadelta, RMSprop, and Adam. These are mathematical functions that depend on the model's learned parameters, such as Weight and Bias. The optimizer used in the model calculates target values based on a set of forecast values. It can be accessed in learnable form parameters that can be learned inside the bias and weight of the neural network. Calculated output and optimization play a vital role to an essential role in reducing losses. Network

training is also in neural networks trained as a model. Several optimization problems are discussed in this article [60-61].

2.2.1.3 ReLU function

The ReLU function is a rectified linear function that is used to solve the Vanishing Gradient problem, allowing us to quickly train the model [62].

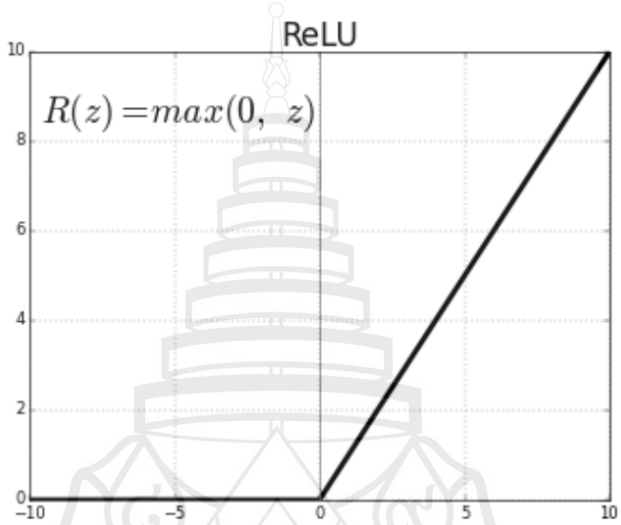


Figure 2.2 Logistic of the ReLU function

2.2.1.4 Sigmoid function

The sigmoid function is a simple S-curve function, and since the output of the Sigmoid function is between 0 and 1, it is suitable for use in applications that require a probabilistic output [62].

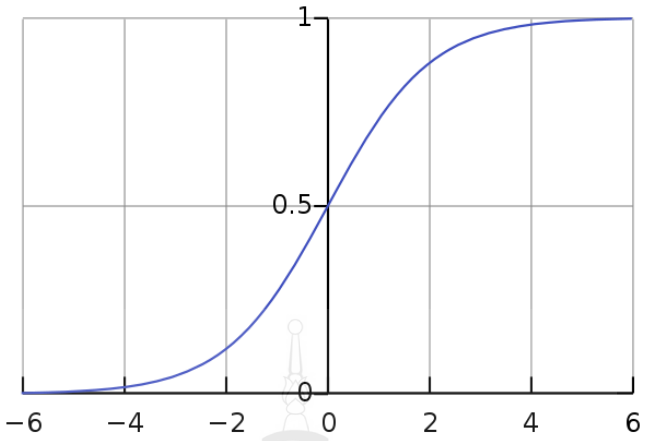


Figure 2.3 Logistic curve of the sigmoid function

2.2.1.5 Target Dropout

Dropout means selecting the target that will drop out only the winning ticket itself. The target dropout can either cut the weight off or the unit off. It was found that the dependency of the low magnitude weight or node was lost, which could be eliminated. It is a technique used to avoid overfitting by randomly ignoring some nodes from the network in each training sample. Figure 2.3 presents an example of its application [47, 63].

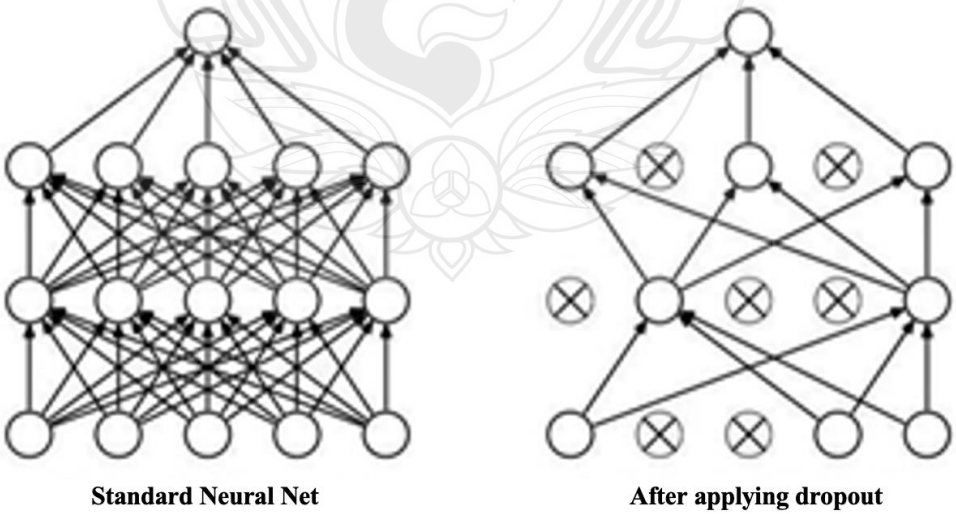


Figure 2.4 Before and After of Neural Network applying dropout

2.2.2 Data Augmentation

In this research, we augmented datasets from the original dataset by using data augmentation for converting original data into new data close to the original data to increase the amount of data used in model training. Computer Vision uses the method of bringing the original image to move, rotate, flip, cut some parts off, add noise or adjust the image's light, color, and sharpness to get slightly different information. but still have the same meaning [64-66].

Augmentation techniques in data analytic research have been widely used to leverage the problem with small data sets. For each of these techniques, researchers also specify the factor by which the size of your dataset would get increased. From many CNN studies, it can be seen that the machine learning program analyzes the image as a matrix structure with numerical values in each component of each layer. Structuring by these three approaches results in a new image with a changed structure and can be used as a training dataset. See Figure 2.4 for details.

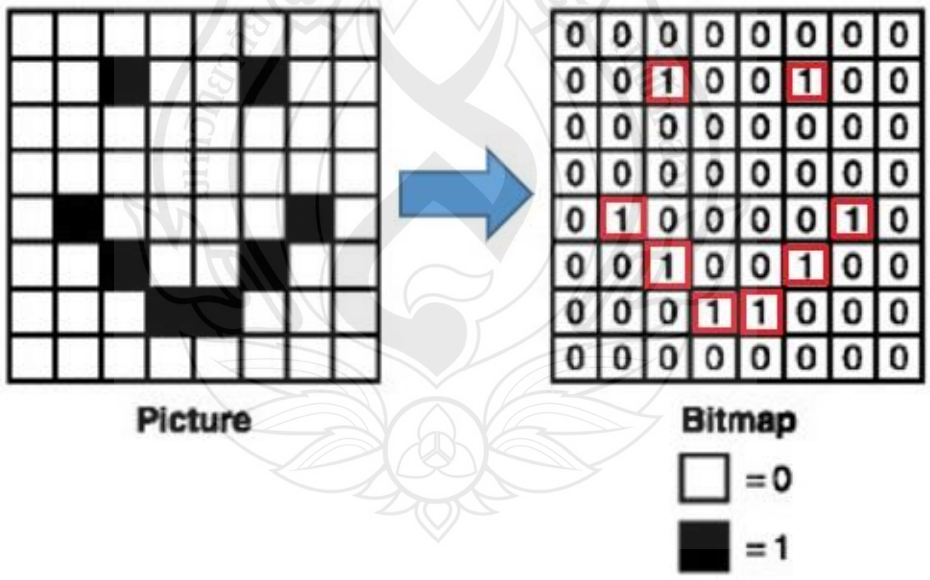


Figure 2.5 Randomized, label-preserving elastic distortions

2.2.2.1 Rotation

Rotation is a classical geometric format. The rotation process is through rotating the image. The axis, whether in the right direction or left, rotates from one to 359, can be used from a clear point of view, for example. A rotated image quality of 30 degrees creates an image [66].

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} x \begin{bmatrix} x \\ y \end{bmatrix}$$

Figure 2.6 Samples of rotated images

2.2.2.2 Flip

The technique of flipping a mirror image all around a vertical axis or horizontal axis or both horizontal-vertical axes. It's a way to help the developer maximize the number of images in the dataset apart from processing with artificial [66].

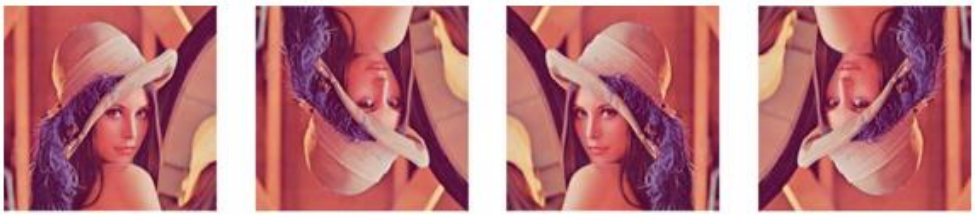


Figure 2.7 Flipping technique

1. Flip Vertical: The image is rotated overturn so that the x-axis is below and the y-axis is on top. The f_x and f_y values specify the current equal in rank or importance of each pixel after flipped, as shown.

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} x \begin{bmatrix} x \\ y \end{bmatrix}$$

Figure 2.8 Example Before and After of a flip vertical

2. Flip Horizontal: The image is rotated similar to reflection like a shadow in a mirror. where the right and left components f_x and f_y are the current equal in rank or importance of the pixel after the horizontal y-axis reflection.

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} x \begin{bmatrix} x \\ y \end{bmatrix}$$

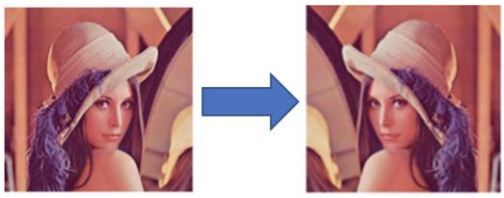


Figure 2.9 Example Before and After of a flip horizontal

3. Flip Horizontal-Vertical: The image is rotated vertically and horizontally. The f_x and f_y are the current equal in rank or importance of each pixel after reflection horizontal-vertical axis, as shown.

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} x \begin{bmatrix} x \\ y \end{bmatrix}$$

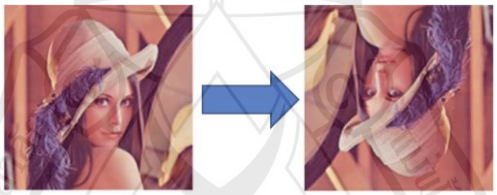


Figure 2.10 Example Before and After of a flip horizontal-vertical

2.2.2.3 Image filters

Image filters have many techniques for image processing such as adjusting histograms, increasing the brightness, decreasing the brightness, decreasing sharpness, blurring, and the flickering is a very common techniques. These filter techniques operate by shifting the n×m matrix from one side to the other of the image [66].

This technique for remodeling image intenseness to renew contrast, and white balance performs by changing the image such that a neutral source illuminates. Usually, special techniques are performed a separate together differently spectral region of the signal. A sharpening spatial filter is used to emphasize fine image details or enhance obscurely details. Whereas blurring is a process of averaging pixels, with neighbors being a merging process. Using sharpening and blurring filters may result in

misshapen images or horizontal or vertical edges having high contrast which will assist in recognizing the image details.

The filters mentioned above work by multiplying the source image matrix with a matrix of the filter. Figure 2.10 represents the output of the image filter using various filters such as histogram equalizer, white balance, and brightness enhancement.

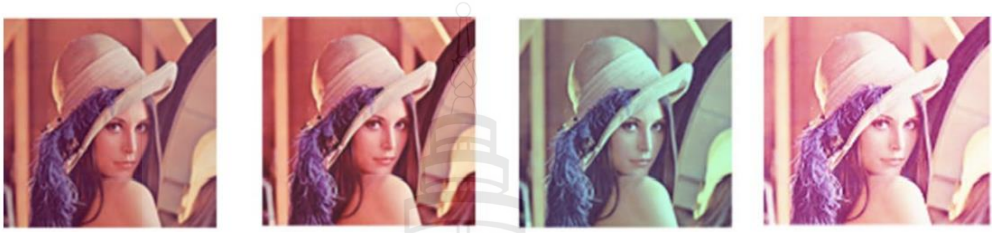


Figure 2.11 Example different filters, original, histogram, balance, brightness

2.2.2.4 Noise

Adding noise is a technique for generating images using an image normal distribution by inserting the noise matrix. The four famous features of noise can be used to enhance an image as such Gaussian, Poisson, Salt and Pepper, and Noise. Four of these noise features are selected and tracked. The first form of sound to be studied is the soundtrack, which is called soundtrack. Gaussian voices continued to change the color values that compose the image. The probability density function is contained in this article.

Figure 2.12 Example different noise, gaussian, Poisson, salt-pepper, speckle

2.2.2.5 Random erasing

The random deletion technique is one of the visual enhancement. This is not a fragment of the geometry transformation but it's the basic set of regulations of random subtraction by removing one square in a rectangular area of the image. which has proven to be effective as shown.

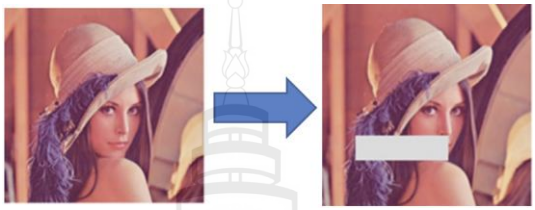


Figure 2.13 Example of removing random erasing image

2.2.2.6 Color space shifting

Color space shifting be the property of the classic optical augmentation. The color level is a mathematical tool by using for paint and construct shade. Humans analysis color by attribute of color, such as brightness and color shade. Colors can be displayed using the amount of red, green, and blue light generated by phosphorescent substance. For this reason this techniques have reveal features in the image inner under a exclusive color gamut by complement the classic photometric data Color space shifting is one of the key for increase the number of images. The most famous color space as such:

1. RGB or Additive : The color system obtained by combining colors or mixing colors together consists of 3 basic colors, Red, Green, Blue, which are used in the operation of the monitor.
2. CMYor Subtractive : It is the opposite of the additive color system in color mixing, which simulates the behavior of reflected light mostly used in printing. It consists of 4 basic colors that are used as such cyan, magenta, yellow, black.
3. YIQ: Color space is used by analog NTSC colors in a television.
4. HSL: Hue (degree on color 0 to 360), Saturation (percentage of value), Lightness (percentage of black to white).

Changing colors in space shift is a random brain teaser. Bright or dark images are a problem that can be seen by increasing the pixel values with constants. Another nice feature of color space management is the independent processing of common RGB color matrices and another method related to the definition of values for each pixel to a maximum or minimum. Advance the color of optical photographs without the need for advanced tools.

Figure 2.14 Example of color space as RGB, CMY, HSL, and YIQ

2.2.2.7 Translation

The translation is the technique for moving an image object from one location to another in image space. The translation is recommended to add geometric information so that part of the image space is black or white after translation to maintain image data or randomize data or gaussian noise. Translation can be performed on the X-axis or Y-axis or X&Y axes simultaneously on the right, left, up, or down image. translation directions can be very useful to avoid relating the position of bias in the f_x and f_y data as the new coordinates of each pixel. After translation x and y represent the original position on the image.

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix}$$

There are some disadvantages to using classic image enhancements and geometry transformations in particular such as the use of additional memory, and processing power. In addition, adding classic geometric data may remove some important characters from the image. for this reason, in some applications, such as medical image processing, Training data is not the same as testing data. This is due to

the classic retouching algorithm realizing it. Therefore, the scope of where and when to use the classic visual augmentation in practice is not enough.

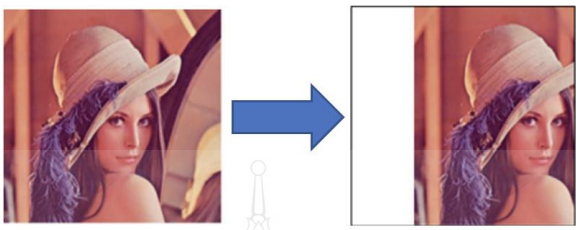


Figure 2.15 Translation in the direction of the X-axis

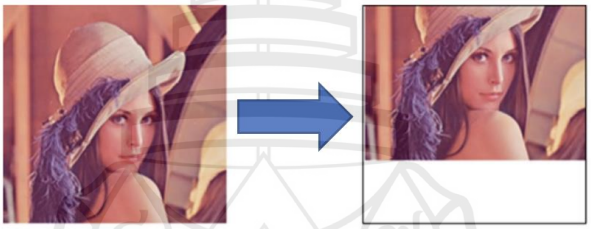


Figure 2.16 Translation in the direction of the Y-axis

Figure 2.17 Translation in the direction of the X-axis & Y-axis

2.2.2.8 Cropping

Cropping is labeled “zoom image” or “scaling image” in techniques of scientific research cropping is the process of enlarging the original image by adding these types of classic geometric shapes consisting approaches. The first action is to crop the image from the initial X, Y position to another X, Y position. For example, if the

image size is a width of 250 pixels and height of 250 pixels, the image may be cropped from the (0, 0) to (200, 200) position. or from (50, 50) to (150, 150). The second action is to resize the image to its original size. In the example above The image should be resized to the width of 250 pixels and height of 250 pixels after cropping. The equation shows the scale equation, f_x and f_y are the new coordinates of each pixel after the scale operation, and x and y represent the original position coordinates on the image. Figure 2.17 shows a cropping example.

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} Xscale & 0 \\ 0 & Yscale \end{bmatrix} x \begin{bmatrix} x \\ y \end{bmatrix}$$

Figure 2.18 Example of cropping result

2.2.2.9 Shearing

Shearing is the transformation of the original image in accordance with the x-axis and y-axis. It's a technique for changing the shape of objects in the image. There are two types of turning. The first type is inside the x-axis and the second type is within the y-axis. The first equation shows the machining in the x-direction, while the second equation shows the machining in the y-direction. f_x and f_y are the new positions of the x-axis for each pixel. After cutting, and the x and y of the image coordinate Figure 2.18 and 2.19 show an example of the cutting type.

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} 1 & shX \\ 0 & 1 \end{bmatrix} x \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ shY & 1 \end{bmatrix} x \begin{bmatrix} x \\ y \end{bmatrix}$$

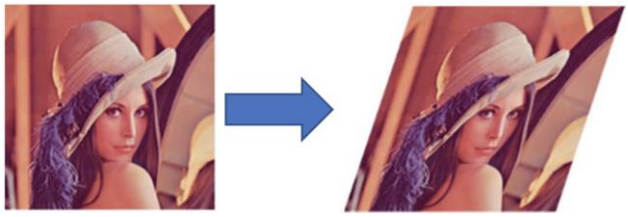


Figure 2.19 Shearing original image in the direction of the X-axis

Figure 2.20 Shearing original image in the direction of the Y-axis

2.2.3 Confusion Matrix

Confusion matrix, also known as an emergency table or error matrix, is a grid layout that allows visualization of the performance of supervised learning algorithms. Each column of the matrix represents an instance in the expected class. While each row represents an instance in the actual class, the following error message appears: All correct guesses are located diagonally of the table to easily identify errors with non-zero values outside the diagonal [67].

1. TP: True Positive mean the program indicates is an explosive and its reality is explosive

2. TN: True Negative mean the program indicates is nonexplosive, and its reality is nonexplosive

3. FP: False Positive mean the program indicates is explosive, and its reality is nonexplosive

4. FN: False negative mean the program indicates is nonexplosive, and its reality is explosive

C_1 – particular class C_2 – different class

		Predicted Class	
		C_1	C_2
Actual Class	C_1	True positive	False negative
	C_2	False positive	True negative

Figure 2.21 Confusion matrix table

CHEPTER 3

METHODOLOGY

3.1 Process Design

This project uses data augmentation techniques to add initial visual data for prediction and evaluation. which will bring each data set to be measured Accuracy and time taken for evaluation are then summarized as to how much dataset volume would be appropriate, with results being measured to be the highest comparative accuracy and least time-consuming. The process design and related tools that we used are shown in figure 3.1 as below.

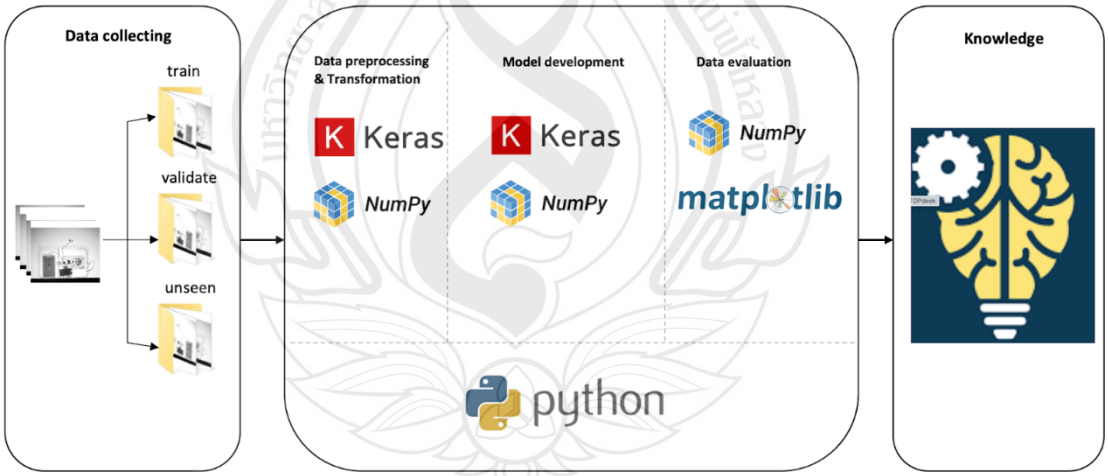


Figure 3.1 The overview of all process designs

Start with the data collecting process. Next, the programming phase consists of data preprocessing & transformation, model development, and data evaluation. Finally, the outcome of this is the optimal dataset volume and knowledge.

3.2 Data Collecting

As mentioned before, the image dataset X that will be used for the empirical study is generated by experts in the field following the generation approach of [55]. In particular, three subsets of x-ray images $X = \{X_{train}, X_{validate}, X_{unseen}\}$ have been prepared along with the conventional flow of model development and assessment. These include train, validation, and unseen data sets, respectively. The first X_{train} is employed to develop a classification model, while the second $X_{validate}$ allows the setting of parameters to be optimized. Following that, the unseen set X_{unseen} is exploited to evaluate the predictive quality of the resulting model. With each sample labeled as either explosive or nonexplosive for a binary classification problem, members of these three sets can be summarized as follows. Figures 3.2 and 3.3 show examples of both explosive and nonexplosive images.

1. Initial train dataset (X_{train}): originally consists of 39 explosive and 27 nonexplosive images,

2. Validation dataset ($X_{validate}$): has 200 explosive and another 200 nonexplosive 200 images,

3. Unseen dataset (X_{unseen}): is composed of 375 explosive and 375 nonexplosive images

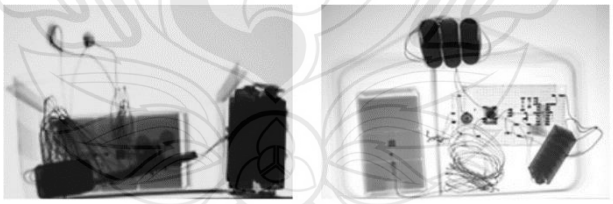


Figure 3.2 Examples of explosive images

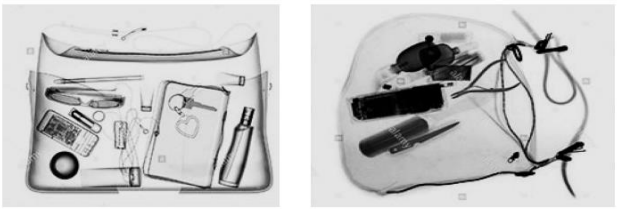


Figure 3.3 Examples of nonexplosive images

3.3 Data Preprocessing & Transformation

Preparation prior to the creation of a classification model can be summarized as follows. At first, image samples that have been acquired from the previous stage will go through a simple filtering process to reduce background noises. After that, data augmentation methods are exploited to increase the size of the training set from the original X_{train} with only 66 images (39 explosive and 27 nonexplosive). These techniques include flipping [49], and brightness adjustment [51], rotation [50], respectively. As a result, different variations of training sets have been created not only to overcome the problem of small-size training data but also to investigate the relationship between a range of sizes and the model's predictive performance. Since the problem of class imbalance is not covered in this research, augmented training sets contain the same amount of explosive and nonexplosive images.

All data augmentation methods are performed in three different ways: horizontal flip, and image brightness adjustment, 20-degree rotation. The reason for choosing only three processes is because other methods, such as cropping an image or adding noise, may cause the model to be carrying too many images, which may or may not be complete images that can be identified as exploding. This will make the model more time-consuming and resource intensive. Therefore, this research uses three main methods to enhance image properties instead. It starts with flipping the image in that direction to double the existing image. Thus, from the previous 66 images, there will be a total of 132 images, and in addition to this flipping, we can divide it into three. Groups are normal and light-up, as well as dimming-down images. This results in 396 total images now and after this, it starts to rotate using an angle of 20 degrees. The

reason for choosing 20 degrees is because the angle is not too small and not too much. If taken, another reason why this study didn't choose a vertical inversion is that a complete 180 rotation is equivalent to a vertical flip image. The data augment process design that we used is shown in figure 3.4 as below.

Augmentation techniques in data analytic research have been widely used to leverage the problem with small data sets. From many CNN studies, it can be seen that the machine learning program analyzes the image as a matrix structure with numerical values in each component of each layer. Structuring by these three approaches results in a new image with a changed structure and can be used as training data. Firstly, flipping is similar to a reflection of the image itself in one dimension, horizontal flips. Figure 3.5 illustrates an example of the horizontal flip. Figure 3.6 depicts two brightness operations that have been employed for this study. In addition, rotation is a change in the structure of an image. Specific to the current study, the right-hand rotation method is used, with the rotation degree $\in \{20\}$. See Figure 3.7 for an example. Note that the implementation of these operations and the model discussed in the nest section is provided via the project GitHub (https://github.com/KwinLook/ADEX-Project.git).

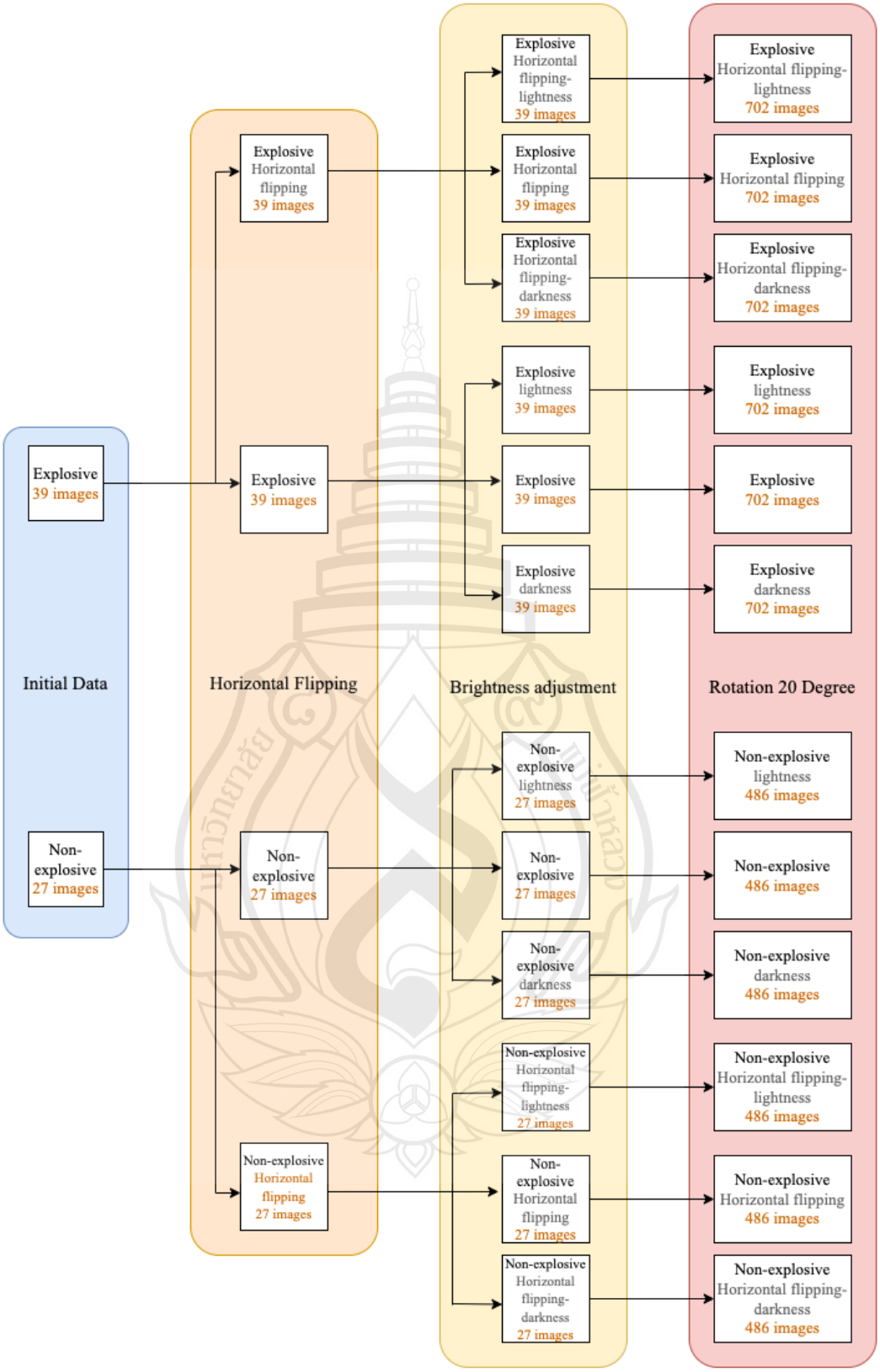


Figure 3.4 The overview of data augmentation process designs

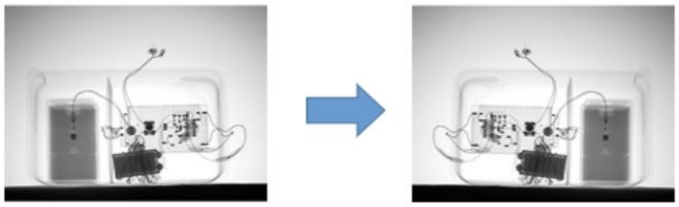


Figure 3.5 Example of a horizontal flipping

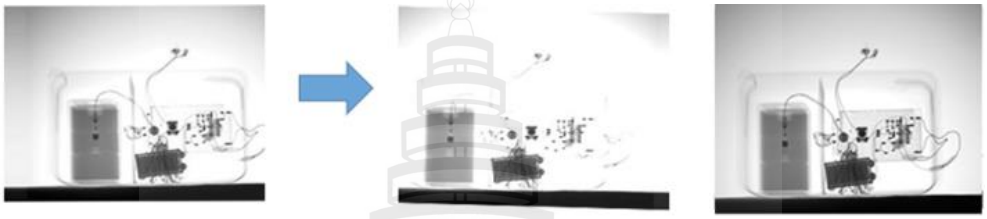


Figure 3.6 Examples of brightness adjustment: 25% lightness and 50% darkness

Figure 3.7 Example of rotating an image by 20 degrees

After obtaining the additional datasets, the next step begins dividing the datasets into five master datasets, where the basis for qualifying the datasets requires a flipping and brightness adjustment dataset, and if two of the above-mentioned stage are not enough to add images that have been choosing rotated from the importance of images such as original images, brightness adjustments, horizontal flipping, rotate images, rotate images obtained by brightness, rotate images obtained by flipping, rotate images obtained by flipping-brightness them in order to bring go through the testing process and collect the results. By starting to divide the data set as follows

1. Augmented set 500 ($X_{train-500}$): consists of 250 explosive and 250 nonexplosive images

1) 250 explosive images will consist of:
   A. Explosive 39 images
   B. Explosive lightness 39 images
   C. Explosive darkness 39 images
   D. Explosive Horizontal flipping 39 images
   E. Explosive Horizontal flipping-lightness 39 images
   F. Explosive Horizontal flipping-darkness 39 images
   G. Explosive Rotation 4 images
   H. Explosive lightness Rotation 2 images
   I. Explosive darkness Rotation 2 images
   J. Explosive Horizontal flipping Rotation 4 images
   K. Explosive Horizontal flipping-lightness Rotation 2 images
   L. Explosive Horizontal flipping-darkness Rotation 2 images
2) 250 nonexplosive images will consist of
   A. Nonexplosive 27 images
   B. Nonexplosive lightness 27 images
   C. Nonexplosive darkness 27 images
   D. Nonexplosive Horizontal flipping 27 images
   E. Nonexplosive Horizontal flipping-lightness 27 images
   F. Nonexplosive Horizontal flipping-darkness 27 images
   G. Nonexplosive Rotation 16 images
   H. Nonexplosive lightness Rotation 14 images
   I. Nonexplosive darkness Rotation 14 images
   J. Nonexplosive Horizontal flipping Rotation 16 images
   K. Nonexplosive Horizontal flipping-lightness Rotation 14 images
   L. Nonexplosive Horizontal flipping-darkness Rotation 14 images

2. Augmented set 1000 ($X_{train-1000}$): consists of 500 explosive and 500 nonexplosive images

1) 500 explosive images will consist of:
   A. Explosive 39 images
   B. Explosive lightness 39 images
   C. Explosive darkness 39 images
   D. Explosive Horizontal flipping 39 images
   E. Explosive Horizontal flipping-lightness 39 images
   F. Explosive Horizontal flipping-darkness 39 images
   G. Explosive Rotation 45 images
   H. Explosive lightness Rotation 44 images
   I. Explosive darkness Rotation 44 images
   J. Explosive Horizontal flipping Rotation 45 images
   K. Explosive Horizontal flipping-lightness Rotation 44 images
   L. Explosive Horizontal flipping-darkness Rotation 44 images
2) 500 nonexplosive images will consist of
   A. Nonexplosive 27 images
   B. Nonexplosive lightness 27 images
   C. Nonexplosive darkness 27 images
   D. Nonexplosive Horizontal flipping 27 images
   E. Nonexplosive Horizontal flipping-lightness 27 images
   F. Nonexplosive Horizontal flipping-darkness 27 images
   G. Nonexplosive Rotation 57 images
   H. Nonexplosive lightness Rotation 56 images
   I. Nonexplosive darkness Rotation 56 images
   J. Nonexplosive Horizontal flipping Rotation 57 images
   K. Nonexplosive Horizontal flipping-lightness Rotation 56 images
   L. Nonexplosive Horizontal flipping-darkness Rotation 56 images

3. Augmented set 2000 ($X_{train-2000}$): consists of 1000 explosive and 1000 nonexplosive images

1) 1000 explosive images will consist of:
   A. Explosive 39 images
   B. Explosive lightness 39 images
   C. Explosive darkness 39 images
   D. Explosive Horizontal flipping 39 images
   E. Explosive Horizontal flipping-lightness 39 images
   F. Explosive Horizontal flipping-darkness 39 images
   G. Explosive Rotation 129 images
   H. Explosive lightness Rotation 127 images
   I. Explosive darkness Rotation 127 images
   J. Explosive Horizontal flipping Rotation 129 images
   K. Explosive Horizontal flipping-lightness Rotation 127 images
   L. Explosive Horizontal flipping-darkness Rotation 127 images
2) 1000 nonexplosive images will consist of
   A. Nonexplosive 27 images
   B. Nonexplosive lightness 27 images
   C. Nonexplosive darkness 27 images
   D. Nonexplosive Horizontal flipping 27 images
   E. Nonexplosive Horizontal flipping-lightness 27 images
   F. Nonexplosive Horizontal flipping-darkness 27 images
   G. Nonexplosive Rotation 141 images
   H. Nonexplosive lightness Rotation 139 images
   I. Nonexplosive darkness Rotation 139 images
   J. Nonexplosive Horizontal flipping Rotation 141 images
   K. Nonexplosive Horizontal flipping-lightness Rotation 139 images
   L. Nonexplosive Horizontal flipping-darkness Rotation 139 images

4. Augmented set 3000 ($X_{train-3000}$): consists of 1500 explosive and 1500 nonexplosive images

1) 1500 explosive images will consist of:
   A. Explosive 39 images
   B. Explosive lightness 39 images
   C. Explosive darkness 39 images
   D. Explosive Horizontal flipping 39 images
   E. Explosive Horizontal flipping-lightness 39 images
   F. Explosive Horizontal flipping-darkness 39 images
   G. Explosive Rotation 211 images
   H. Explosive lightness Rotation 211 images
   I. Explosive darkness Rotation 211 images
   J. Explosive Horizontal flipping Rotation 211 images
   K. Explosive Horizontal flipping-lightness Rotation 211 images
   L. Explosive Horizontal flipping-darkness Rotation 211 images
2) 1500 nonexplosive images will consist of
   A. Nonexplosive 27 images
   B. Nonexplosive lightness 27 images
   C. Nonexplosive darkness 27 images
   D. Nonexplosive Horizontal flipping 27 images
   E. Nonexplosive Horizontal flipping-lightness 27 images
   F. Nonexplosive Horizontal flipping-darkness 27 images
   G. Nonexplosive Rotation 223 images
   H. Nonexplosive lightness Rotation 223 images
   I. Nonexplosive darkness Rotation 223 images
   J. Nonexplosive Horizontal flipping Rotation 223 images
   K. Nonexplosive Horizontal flipping-lightness Rotation 223 images
   L. Nonexplosive Horizontal flipping-darkness Rotation 223 images

5. Augmented set 4000 ($X_{train-4000}$): consists of 2000 explosive and 2000 nonexplosive images

1) 2000 explosive images will consist of:
   A. Explosive 39 images
   B. Explosive lightness 39 images
   C. Explosive darkness 39 images
   D. Explosive Horizontal flipping 39 images
   E. Explosive Horizontal flipping-lightness 39 images
   F. Explosive Horizontal flipping-darkness 39 images
   G. Explosive Rotation 295 images
   H. Explosive lightness Rotation 294 images
   I. Explosive darkness Rotation 294images
   J. Explosive Horizontal flipping Rotation 295 images
   K. Explosive Horizontal flipping-lightness Rotation 294 images
   L. Explosive Horizontal flipping-darkness Rotation 294 images
2) 2000 nonexplosive images will consist of
   A. Nonexplosive 27 images
   B. Nonexplosive lightness 27 images
   C. Nonexplosive darkness 27 images
   D. Nonexplosive Horizontal flipping 27 images
   E. Nonexplosive Horizontal flipping-lightness 27 images
   F. Nonexplosive Horizontal flipping-darkness 27 images
   G. Nonexplosive Rotation 307 images
   H. Nonexplosive lightness Rotation 306 images
   I. Nonexplosive darkness Rotation 306 images
   J. Nonexplosive Horizontal flipping Rotation 307 images
   K. Nonexplosive Horizontal flipping-lightness Rotation 306 images
   L. Nonexplosive Horizontal flipping-darkness Rotation 306 images

After five sets of data run experiments finished Makes another doubt that Between the 2000 and 3000 image datasets, is there a chance that it will have higher or nearer accuracy and, moreover, is it likely that the mean will be higher, or will it be close and take less time. Three additional datasets are added: 2250 2500 2750 Images.

1. Augmented set 2250 ($X_{train-2250}$): consists of 1125 explosive and 1125 nonexplosive images
    1) 1125 explosive images will consist of:
        A. Explosive 39 images
        B. Explosive lightness 39 images
        C. Explosive darkness 39 images
        D. Explosive Horizontal flipping 39 images
        E. Explosive Horizontal flipping-lightness 39 images
        F. Explosive Horizontal flipping-darkness 39 images
        G. Explosive Rotation 150 images
        H. Explosive lightness Rotation 148 images
        I. Explosive darkness Rotation 148 images
        J. Explosive Horizontal flipping Rotation 149 images
        K. Explosive Horizontal flipping-lightness Rotation 148 images
        L. Explosive Horizontal flipping-darkness Rotation 148 images
    2) 1125 nonexplosive images will consist of
        A. Nonexplosive 27 images
        B. Nonexplosive lightness 27 images
        C. Nonexplosive darkness 27 images
        D. Nonexplosive Horizontal flipping 27 images
        E. Nonexplosive Horizontal flipping-lightness 27 images
        F. Nonexplosive Horizontal flipping-darkness 27 images
        G. Nonexplosive Rotation 162 images
        H. Nonexplosive lightness Rotation 160 images
        I. Nonexplosive darkness Rotation 160 images
        J. Nonexplosive Horizontal flipping Rotation 161 images
        K. Nonexplosive Horizontal flipping-lightness Rotation 160 images
        L. Nonexplosive Horizontal flipping-darkness Rotation 160 images

2. Augmented set 2500 ($X_{train-2500}$): consists of 1250 explosive and 1250 nonexplosive images

1) 1250 explosive images will consist of:
   A. Explosive 39 images
   B. Explosive lightness 39 images
   C. Explosive darkness 39 images
   D. Explosive Horizontal flipping 39 images
   E. Explosive Horizontal flipping-lightness 39 images
   F. Explosive Horizontal flipping-darkness 39 images
   G. Explosive Rotation 170 images
   H. Explosive lightness Rotation 169 images
   I. Explosive darkness Rotation 169 images
   J. Explosive Horizontal flipping Rotation 170 images
   K. Explosive Horizontal flipping-lightness Rotation 169 images
   L. Explosive Horizontal flipping-darkness Rotation 169 images
2) 1250 nonexplosive images will consist of
   A. Nonexplosive 27 images
   B. Nonexplosive lightness 27 images
   C. Nonexplosive darkness 27 images
   D. Nonexplosive Horizontal flipping 27 images
   E. Nonexplosive Horizontal flipping-lightness 27 images
   F. Nonexplosive Horizontal flipping-darkness 27 images
   G. Nonexplosive Rotation 182 images
   H. Nonexplosive lightness Rotation 181 images
   I. Nonexplosive darkness Rotation 181 images
   J. Nonexplosive Horizontal flipping Rotation 182 images
   K. Nonexplosive Horizontal flipping-lightness Rotation 181 images
   L. Nonexplosive Horizontal flipping-darkness Rotation 181 images

3. Augmented set 2750 ($X_{train-2750}$): consists of 1375 explosive and 1375 nonexplosive images

1) 1375 explosive images will consist of:
   A. Explosive 39 images
   B. Explosive lightness 39 images
   C. Explosive darkness 39 images
   D. Explosive Horizontal flipping 39 images
   E. Explosive Horizontal flipping-lightness 39 images
   F. Explosive Horizontal flipping-darkness 39 images
   G. Explosive Rotation 191 images
   H. Explosive lightness Rotation 190 images
   I. Explosive darkness Rotation 190 images
   J. Explosive Horizontal flipping Rotation 191 images
   K. Explosive Horizontal flipping-lightness Rotation 190 images
   L. Explosive Horizontal flipping-darkness Rotation 190 images
2) 1375 nonexplosive images will consist of
   A. Nonexplosive 27 images
   B. Nonexplosive lightness 27 images
   C. Nonexplosive darkness 27 images
   D. Nonexplosive Horizontal flipping 27 images
   E. Nonexplosive Horizontal flipping-lightness 27 images
   F. Nonexplosive Horizontal flipping-darkness 27 images
   G. Nonexplosive Rotation 203 images
   H. Nonexplosive lightness Rotation 202 images
   I. Nonexplosive darkness Rotation 202 images
   J. Nonexplosive Horizontal flipping Rotation 202 images
   K. Nonexplosive Horizontal flipping-lightness Rotation 202 images
   L. Nonexplosive Horizontal flipping-darkness Rotation 202 images

3.4 Model Development

Following successful applications of CNN to the task of object detection in x-ray images [35, 39], the present work makes use of the common model for the classification of input images as either explosive or nonexplosive ones. In fact, CNN belongs to one of the special sorts of artificial neural networks or ANN. A standard ANN usually consists of an input, a hidden, and an output layer. Each layer contains a number of neurons that function as a computational unit, except those on the input layer. Every element in an image sample is represented by the vector $X = x_1, x_2, \ldots, x_n$ where x_n represents the value of the n^{th} pixel that is taken by a particular neuron. Therefore, the number of input neurons is equal to the resolution of the visual data. In the same way, the size of the output layer depends on the size of the classification problem. A binary classification requires two output neurons representing classes 0 and 1, with a probability of an input instance belonging to each class. Each neuron in a single layer is also connected to every neuron in adjacent layers. As such, a fully connected network is built with a set of connection-specific weights. The output of one-layer acts as the input of the next. For each computational neuron, the weighted sum of input vector x is assigned to an activation function, which injects nonlinear factors into the network to meet some of the specifications of the problem to be solved: $f(W^t + b)$ where W is the vector of weights and bis the bias. Two of famous activation functions are relu: $f(x) = max(0, x)$ and sigmoid: $f(x) = (1 + e^{-x})^{-1}$.

In a CNN model, one or more layers that alternate with the integration layers are added on top of the fully connected network. It can be conceptually illustrated by the layer-wise framework in Figure 3.8. Typically, a distorted layer consists of many property detectors, so-called filters or kernels, which are primarily intended to capture the properties of each input image. These filters can be perceived as a matrix of weights that scroll through an image. Distortion is achieved by calculating the dot product of the matrix and the corresponding part of the image at all positions. And every result is seen as an element in a new matrix called a feature tree. This is usually followed by an aggregation layer. where the bundled features received will be divided into sections of the same size (e.g. 2 × 2), then one input value is subsampled from all regions in terms

of maximum, average or even random values. Therefore, the subsampling function can continuously reduce the size of the input representation to reduce network complexity. This will help speed up the calculations. and makes the network strong against minor changes, distortions, and translations as well. In our work, a maximum integration strategy is used to preserve the most important attributes of the input. There can also be a dropout layer between convolutional layers that is the technique used to avoid overfit by randomly ignoring some nodes from the network in each training sample. For adapting the predefined CNN model to a given training data, the optimization has proven to be an important step. There are several common optimization tools that are widely used, such as SGD, Adagrad, Adadelta, RMSprop, and Adam. In this project will use adam optimizer. As it is the most popular technique today, it combines the advantages of both RMSProp and momentum, making it an attractive choice for this project.

These extend the basic optimizer in terms of batch size, momentum, adaptive learning rate, and more. The current work will experiment with all of the aforementioned activation and optimizer functions combined in order to investigate how each model responds to the classification problem in hand. Figure 3.9 summarizes details of the basic CNN model to be exploited in this empirical study.

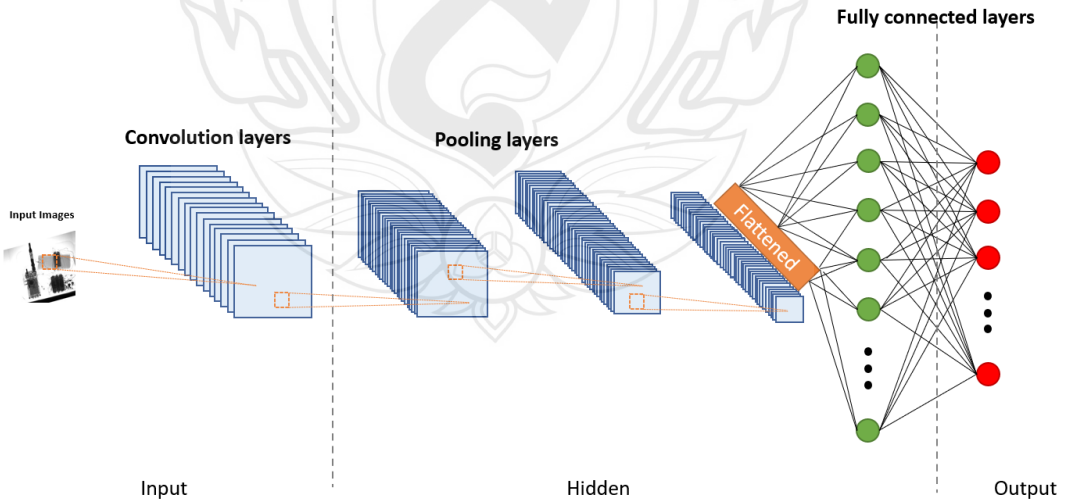


Figure 3.8 The layer-wise architecture of a common CNN model

Layer (type)	Output Shape	Param #
********** Pooling 1 **********		
Conv2D	(None, 148, 148, 128)	3584
MaxPooling2D	(None, 74, 74, 128)	0
Dropout	(None, 74, 74, 128)	0
********** Pooling 2 **********		
Conv2D	(None, 72, 72, 100)	115300
MaxPooling2D	(None, 36, 36, 100)	0
Dropout	(None, 36, 36, 100)	0
********** Pooling 3 **********		
Conv2D	(None, 34, 34, 64)	57664
MaxPooling2D	(None, 17, 17, 64)	0
Dropout	(None, 17, 17, 64)	0
********** Pooling 4 **********		
Conv2D	(None, 15, 15, 32)	18464
MaxPooling2D	(None, 7, 7, 32)	0
Dropout	(None, 7, 7, 32)	0
********** Pooling 5 **********		
Conv2D	(None, 5, 5, 16)	4624
MaxPooling2D	(None, 2, 2, 16)	0
Dense	(None, 2, 2, 16)	272
Flatten	(None, 64)	0
Dense	(None, 2)	130
SoftMax	(None, 2)	6

Figure 3.9 Architecture of the CNN model and its layer-specific parameter setting

3.5 Data Evaluation

From Figure 3.8 depicts the overall process of model development and evaluation. After those steps of training data preparation and augmentation, the CNN model is developed and initially assessed with the validation set to determine the setting of hyper-parameters. For this, patterns of accuracy and loss with respect to training and validation sets are investigated. Then, the desired model is evaluated against a test or unseen set to derive the conclusion, which is estimated from a confusion matrix. It represents measures of comparison between true classes and predictions made by the model of interest. Specific to the binary classification studied in this research (see

Figure 3.9), it contains the following four measures: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN), respectively.

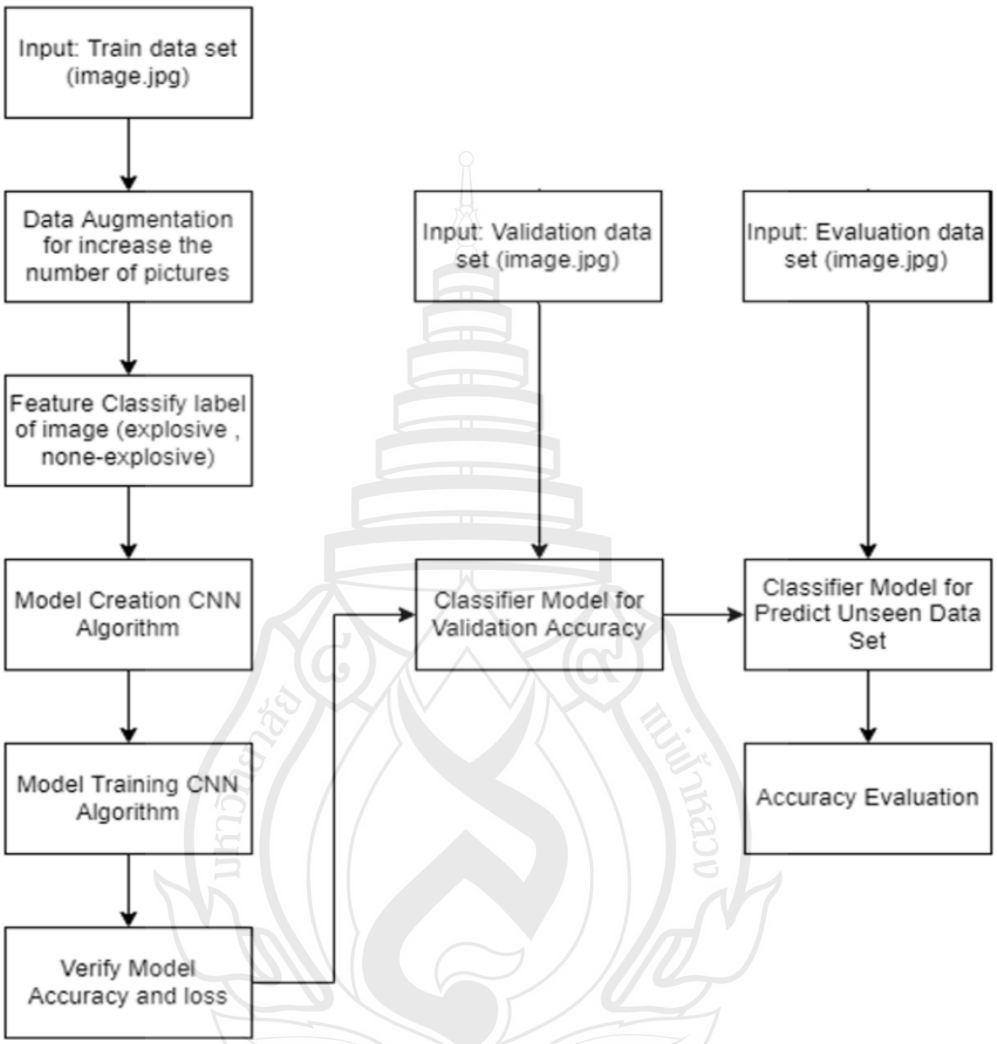


Figure 3.10 Overview of the over-system model

	Actual Values	
	Actually Positive (1) : explosive	Actually Negative (0) : none
Predicted Positive (1): explosive	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0):none	False Negatives (FNs)	True Negative (TNs)

Figure 3.11 Illustration of a confusion matrix for this binary classification problem

From the confusion matrix, accuracy, precision, recall and F1 measures are exploited to determine the predictive performance of the CNN model. At first, the accuracy of a given model is specified by

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

In addition, class-specific recall metrics can be defined as follows:

$$recall_1 = \frac{TP}{TP + FN} \quad , recall_0 = \frac{TN}{TN + FN}$$

Likewise, class-specific precision measures can also be estimated by the following.

$$precision_1 = \frac{TP}{TP + FP} \quad , precision_0 = \frac{TN}{TN + FN}$$

Having known those precision and recall measures, class specific F1 values are defined below.

$$F1_1 = \frac{2\ x\ precision_1\ x\ recall_1}{precision_1 + recall_1}, \quad F1_0 = \frac{2\ x\ precision_0\ x\ recall_0}{precision_0 + recall_0}$$

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSIONS

4.1 Results of Experimental by Using Initial Data

The first experiment part of this section is to report the results, with respect to model accuracy and loss during the training and validation stages, and using time for training is 368 seconds. Figure 4.1 displays these measures for the CNN model that is developed using the initial training set X_trainthat consists of only 66 images (39 explosive and 27 nonexplosive ones).

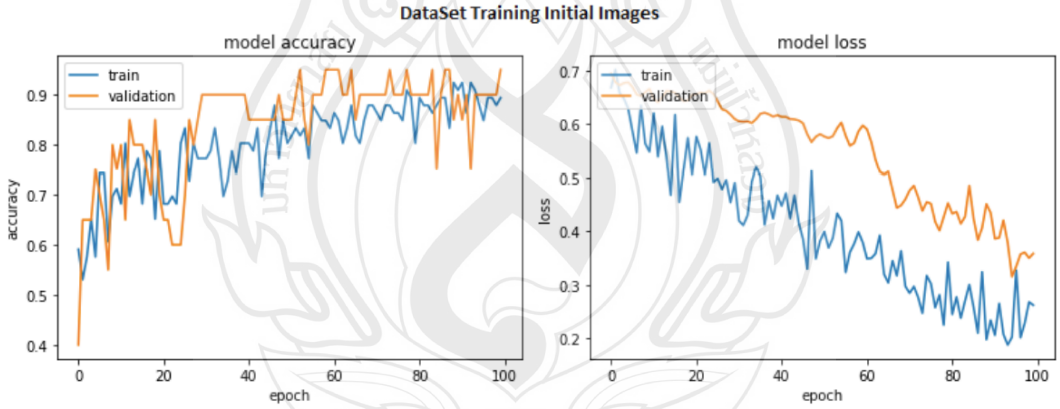


Figure 4.1 Comparison of accuracy and loss obtained from initial training set X_{train}

4.2 Results of Experimental by Using Augmentation Set 500 to 2000

The second experiment part of this section is along the repetition of the learning cycle that is specified as the number epoch (in the range between 1 and 100), the overall pattern of model accuracy conforms to the typical expectation as its gradually increases to the incline of epoch. This is also observed with the case of model loss that rapidly decreases as the model has learned more and more. However, both validation accuracy and loss are worse than those of the training counterpart, exhibiting that the model does not perform well with new cases not seen in the train set. Yet, those accuracy and loss curves shown in the figure are highly fluctuated, thus requiring more training samples to stabilize the underlying learning process [52]. As such, the same experiment is repeated with five different train sets, i.e., $X_{train-500}$, $X_{train-1000}$, $X_{train-2000}$, $X_{train-3000}$, $X_{train-4000}$ that have been generated using data augmentation techniques.

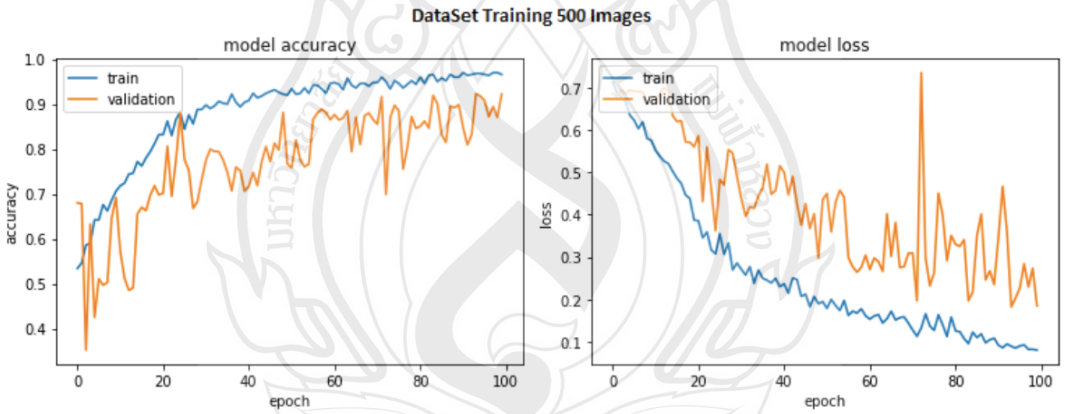


Figure 4.2 Comparison of accuracy and loss obtained from training set $X_{train-500}$

Figure 4.3 Comparison of accuracy and loss obtained from training set $X_{train-1000}$

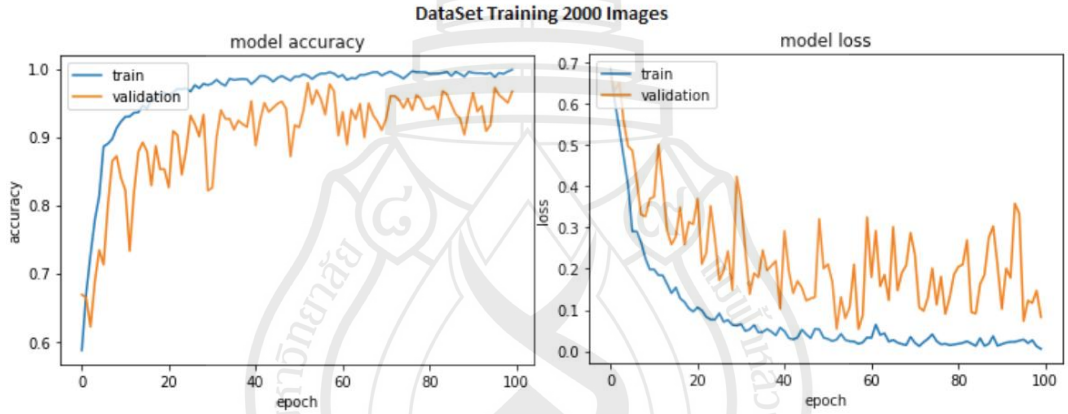


Figure 4.4 Comparison of accuracy and loss obtained from training set $X_{train-2000}$

Figure 4.5 Comparison of accuracy and loss obtained from training set $X_{train-3000}$

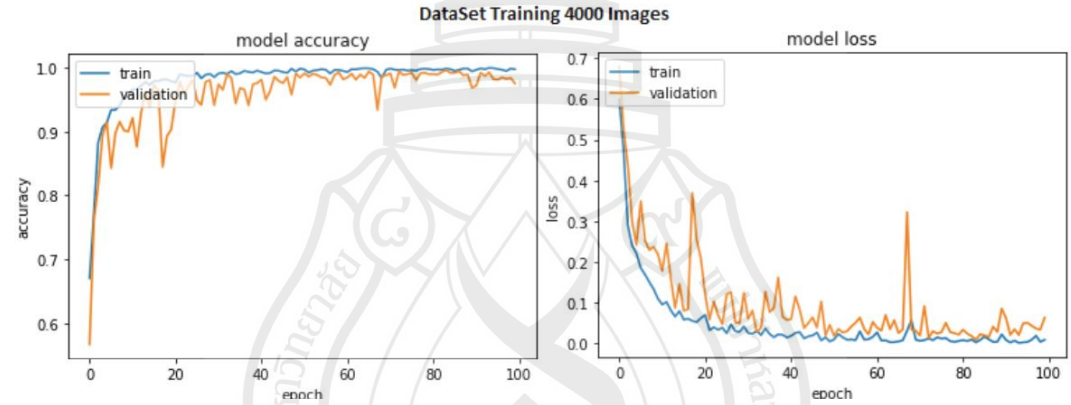


Figure 4.6 Comparison of accuracy and loss obtained from training set $X_{train-4000}$

From Figures 4.2 to 4.6 represent a graph of the observation with model accuracy and loss using augmented train sets of 500, 1,000, and 2,000 samples. As compared to the previous setting, training accuracy and loss patterns have become more stable, while validation measures slightly improve from those given in Figure 4.2. As shown in Figures 4.5 and 4.6, validation accuracy and loss have matched with training counterparts, pointing towards the stability of learned models. Of course, this new finding is acquired by increasing the size of the train set to 3,000 and 4,000 images, respectively. Despite the trends presented in these displays, there is a sign of overfitting as the model is trained with 4,000 instances to respond to this question, all the five

models created from augmented train sets are assessed against the unseen set, where the corresponding result is shown in Table 4.1

Table 4.1 Confusion matrices and corresponding measures of predictive performance obtained from five augmented train sets. Note that Precision/Recall/F1 are reported for class 1 only here, and the best two scores for each metric are highlighted in **bold font**

Size of train set	TP	FP	TN	FN	Precision	Recall	F1	Accuracy	Time
500 images	336	39	340	35	0.896	0.905	0.900	0.901	3,573 Sec
1,000 images	357	18	360	15	0.952	0.959	0.955	0.956	5,213 Sec
2,000 images	367	8	370	5	**0.978**	**0.986**	**0.982**	**0.982**	6,593 Sec
3,000 images	367	8	372	3	**0.978**	**0.991**	**0.985**	**0.985**	9,743 Sec
4,000 images	370	5	357	18	**0.986**	0.953	0.969	0.969	14,984 Sec

4.3 Results of Experimental by Using Dataset Between 2000 to 3000

The third experiment part of this section , the results of the previous experiments were observed, which were observed that the results calculating the precision, recall, F1, and accuracy were obtained by doing the confusion matrix have the values were high between 2000 and 3000, the researcher added three additional datasets, namely 2250 2500 2750, in order to test and calculate the accuracy of the number of datasets shown in Table 4.2.

Table 4.2 Confusion matrices and corresponding measures of predictive performance obtained from train set 2,000 2,250 2,500 2,750 3,000 images Note that Precision/Recall/F1 are reported for class 1 only here, and the best two scores for each metric are highlighted in **bold font**

Size of train set	TP	FP	TN	FN	Precision	Recall	F1	Accuracy	Time
2,000 images	367	8	370	5	0.978	0.986	0.982	0.982	6,593 Sec
2,250 images	368	7	370	5	0.981	0.986	0.983	0.982	7,682 Sec
2,500 images	372	3	285	90	0.992	0.805	0.889	0.876	8,437 Sec
2,750 images	369	6	369	6	**0.984**	0.984	**0.984**	0.984	**8,975 Sec**
3,000 images	367	8	372	3	0.978	**0.986**	0.982	**0.985**	9,743 Sec

After the third experiment part, From table 4.2 It has especially interesting after the experiment that the 2750 and 3000 datasets had similar accuracy, and although the 2750 datasets had a higher Precision, this Precision was a measure of how often the model answered the question matched with the same answer. therefore, in this research, The researcher will focus on the overall accuracy and the use of time instead, in the next section we will discuss it.

4.4 Discussion

According to Table 4.1, CNN models that are developed with $X_{train-3000}$ and $X_{train-2000}$ appear to provide the best predictive performance among five alternatives. The highest accuracy of 0.985 is acquired with the former, with a bit lower measure of 0.982 is obtained by the other. Similar tendencies are witnessed with other quality metrics of F1 and Recall that are specific to class 1. Nonetheless, the best precision value of 0.986 is reached by another CNN model that is trained with $X_{train-4000}$. This suggests that it has made a smaller number of false positives than the two mentioned earlier, but the recall capability has let it down. In other words, it is not as good as the other two in recognizing class-1 or explosive images that is the focus of this research, i.e., a higher FN value.

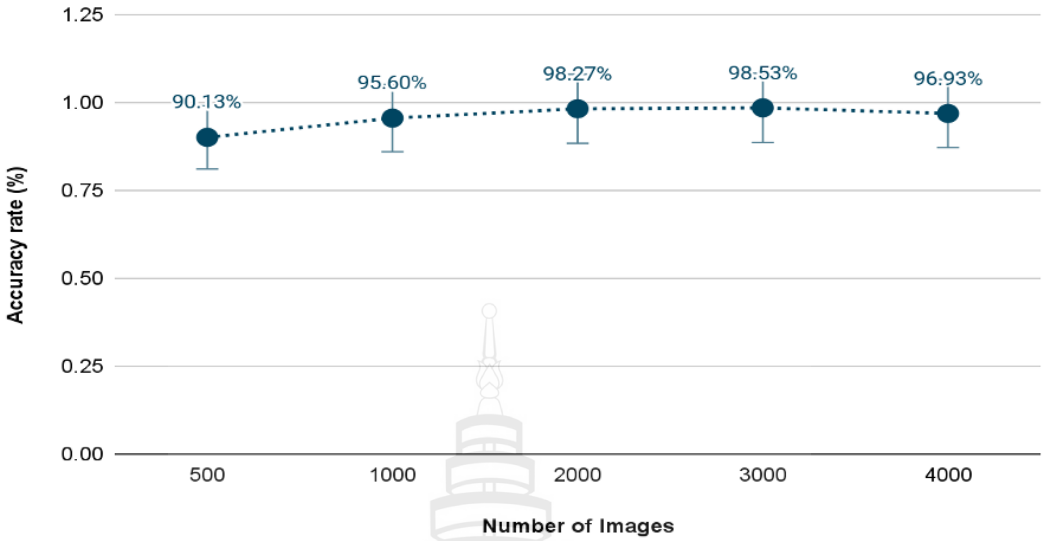


Figure 4.7 Comparison of accuracy measures obtained from five train sets

Given such a finding, the concept of data augmentation has proven effective for improving the model learning outcome, but too big of the augmented train set may not be necessary to achieve an optimal result. Hence, additional experiments with different sizes of train set, i.e., between 2,000 and 3,000 samples, are conducted. These new train sets are $X_{train-2250}$, $X_{train-2500}$, $X_{train-2750}$. This is to investigate the appropriate size that minimizes resource consumption whilst preserving a high level of predictive quality. Figure 5.2 shows accuracy measures obtained by the CNN model that has been trained with different sizes of augmented train sets. With these, it seems that the training process converges around a train set with 2,750 to 3,000 images, while other smaller sets might lead to an unstable outcome, especially in the case of 2,500 samples. Hence, the aforementioned range is recommended for anyone who is willing to extend the present work. Note that a bigger train set not only causes a training step to be expensive, it also generally leads to an overfitting scenario that degrades the prediction goodness of unseen instances.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

This article presents another new approach to explosives which is a collaborative study of Mae Fah Luang University and the Defense Technology Institute (Public Organization). This research relies on the process of the CNN model to classify explosive and nonexplosive images, which is the major component of the detection of IEDs. This is a difficult task because the number of images obtained is very small, forcing developers to find new methods. Help to increase the number of images to have a greater volume. In this case, Data Augmentation will be used by doing 3 processes, namely Flip images, Brightness, and Rotation. After getting more images, we will have to try again to see if the number of images that should be trained is how much. Optimal because if the number of images is too many, there will be overlearning, making training consumes a lot of resources and time. But the accuracy is not higher at all, so in addition, the researcher will have to increase the image volume, another important thing is to measure the size of the data set to be most accurate with the least training time and resources.

To summarize, the results of the empirical study suggest the exceptional accuracy gained by the proposed machine learning model. Based on the experiments, the best outcome of accuracy rate being 98.53% has been witnessed with the model trained with 3,000 training images. This is the case of training with only 2,750, which achieves a slightly lower rate of 98.4%. However, despite this achievement, there are several issues leading to worthwhile future work. Since the main problem of a small amount of initial image data set has been solved by data augmentation already.

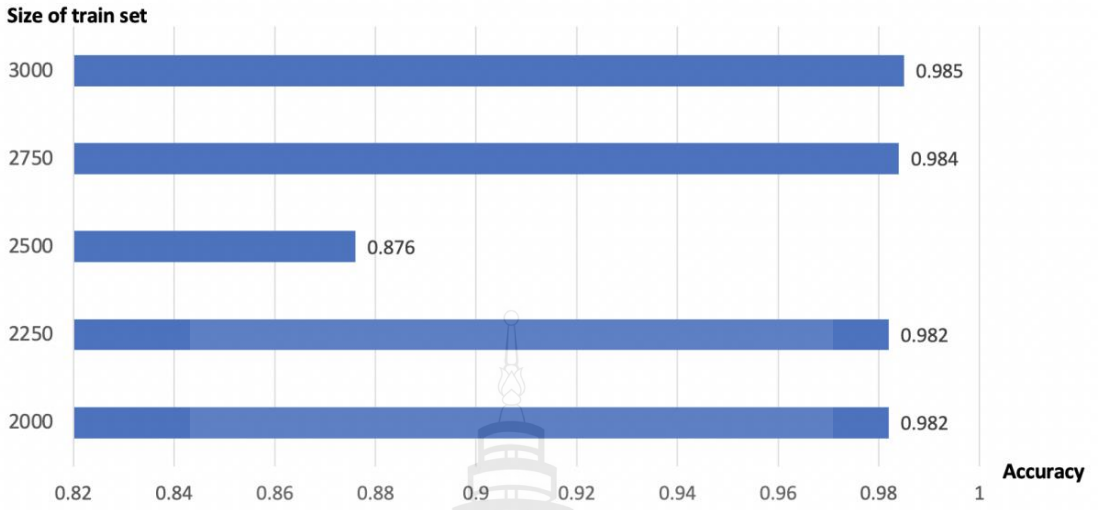


Figure 5.1 Comparison of accuracy measures obtained from different train sets

5.2 Future Work

Future work will focus on detection methods that focus on a specific area instead of focusing on the entire image as such.

1. Using technology like OpenCV to help focus images. It is possible to focus on a specific point before analyzing it.
2. Using the technique of transfer learning by choosing similar datasets such as X-ray images of various weapons to be attributed to current learning models.
3. Finally, Using the technique of Generative adversarial networks simulates reconstruction. This technique detects the strengths of the image to create a new image, which is a technique that is different from data augmentation.

As a future experiment, we will find out which techniques will be more accurate and will allow the model to use less time and resources to come to conclusions and answers for the development of ordnance-disposal robots.

REFERENCES

REFERENCES

[1] Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart cities. *Cities, 89,* 80-91. https://doi.org/10.1016/j.cities.2019.01.032

[2] Allam, Z., & Newman, P. (2018). Redefining the smart city: Culture, metabolism and governance. *Smart Cities, 1*(1), 4-25. https://doi.org/10.3390/smartcities1010002

[3] Petrolo, R., Loscrì, V., & Mitton, N. (2015). Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms. *Transactions on Emerging Telecommunications Technologies, 28*(1), e2931. https://doi.org/10.1002/ett.2931

[4] Dameri, R. P. (2012). Searching for smart city definition: A comprehensive proposal. *International Journal of Computers and Technology, 11*(5), 2544-2551. https://doi.org/10.24297/ijct.v11i5.1142

[5] Neirotti, P., De Marco, A. A., Cagliano, C., Mangano, G., & Scorrano, F. (2014). Current trends in smart city initiatives: Some stylised facts. *Cities, 38,* 25-36. https://doi.org/10.1016/j.cities.2013.12.010

[6] Boongoen, T., Shen, Q., & Price, C. (2010). Disclosing false identity through hybrid link analysis. *Artificial Intelligence and Law, 18*(1), 77-102.

[7] Allam, T. Jr., Bahmanyar, A., Biswas, R., Dai, M., Galbany, L., Hložek, R., Ishida, E. E. O., Jha, S. W., Jones, D. O., Kessler, R., Lochner, M., Mahabal, A. A., Mahabal, A. I., Mandel, K. S., Martínez-Galarza, J. R., McEwen, J. D., Muthukrishna, D., Narayan, G., Narayan, H., . . . Narayan, C. N. (2018). The Photometric LSST Astronomical Time-series Classification Challenge (PLAsTiCC): Data set. https://doi.org/10.48550/arXiv.1810.00001

[8] Archer, E. M. (2014). Crossing the rubicon: Understanding cyber terrorism in the European context. *The European Legacy*, *19*(5), 606-621. https://doi.org/10.1080/10848770.2014.943495

[9] Butler, H. J., Ashton, L., Bird, B., Cinque, G., . . . Martin, F. L. (2016). Using Raman spectroscopy to characterize biological materials. *Nature Protocols, 11*(4), 664-687. https://doi.org/10.1038/nprot.2016.036

[10] Bielecki, Z., Janucki, J., Kawalec, A., Mikołajczyk, J., . . . Stacewicz, T. (2012). Sensors and systems for the detection of explosive devices: An overview. *Metrology and Measurement Systems*, *19*(1), 3-28.

[11] Salinas, Y., Martinez-Manez, R., Marcos, M. D., Sancenon, F., . . . Gil, S. (2012). Optical chemosensors and reagents to detect explosives. *Chemical Society Reviews, 41*(3), 1261-1296.

[12] Hättenschwiler, N., Mendes, M., & Schwaninger, A. (2019). Detecting bombs in x-ray images of hold baggage: 2D versus 3D imaging. *Human Factors, 61*(2), 305-321. https://doi.org/10.1177/0018720818799215

[13] Jenne, N., & Chang, J. Y. (2019). Hegemonic distortions: The securitisation of the insurgency in Thailand's deep south. *TRaNS: Trans -Regional and -National Studies of Southeast Asia*, *7*(2), 209-232. https://doi.org/10.1017/trn.2018.13

[14] Tunwannarux, A., & Tunwannarux, S. (2008, May 27-30). *The explosive ordnance disposal robot: CEO mission EOD* [Paper presentation]. 10th WSEAS Int. Conf. on Automatic Control, Modelling & Simulation (ACMOS'08), Istanbul, Turkey.

[15] Boongoen, T., Iam-On, N., & Undara, B. (2016). Improving face detection with bi-level classification model. *NKRAFA Journal of Science and Technology, 12*, 52-63.

[16] Iam-On, N., & Boongoen, T. (2015). Improving face classification with multiple-clustering induced feature reduction. In *2015 International Carnahan Conference on Security Technology (ICCST)* (pp. 241-246). IEEE. https://doi.org/10.1109/CCST.2015.7389689

[17] Thongsatapornwatana, U., Lilakiatsakun, W., & Boongoen, T. (2017). Improvement of intelligent framework for suspect vehicle detection system. In *2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 419-422). IEEE. https://doi.org/10.1109/ECTICon.2017.8096263

[18] Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. *Computer Science Review, 40*, 100379. https://doi.org/10.1016/j.cosrev.2021.100379

[19] Ramprasath, M., Anand, M. V., & Hariharan, S. (2018). Image classification using convolutional neural networks. *International Journal of Pure and Applied Mathematics, 119*(17), 1307-1319.

[20] Williams, E. R., & Zare, R. N. (1990). Detection of concealed explosive. *Science, 248*(4962), 1471-1472.

[21] Liu, J., Si, T., & Zhang, Z. (2019). Mussel-inspired immobilization of silver nanoparticles toward sponge for rapid swabbing extraction and SERS detection of trace inorganic explosives. *Talanta, 204*, 189-197. https://doi.org/10.1016/j.talanta.2019.05.110

[22] Ifa, D. R., Manicke, N. E., Dill, A. L., & Cooks, R. G. (2008). Latent fingerprint chemical imaging by mass spectrometry. *Science, 321*(5890), 805-805. https://doi.org/10.1126/science.1157199

[23] Ding, Y., Wang, S., Li, J., & Chen, L. (2016). Nanomaterial-based optical sensors for mercury ions. *TrAC Trends in Analytical Chemistry, 82*, 175-190. https://doi.org/10.1016/j.trac.2016.05.015

[24] Wong, M. H., Giraldo, J. P., Kwak, S.-Y., Koman, V. B., Sinclair, R., Lew, T. T. S., . . . Strano, M. S. (2017). Nitroaromatic detection and infrared communication from wild-type plants using plant nanobionics. *Nature Materials, 16*(2), 264-272. https://doi.org/10.1038/nmat4771

[25] Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two, dissociable decision criteria in visual search. *Current Biology, 20*(2), 121-124. https://doi.org/10.1016/j.cub.2009.11.066

[26] Mery, D., Saavedra, D., & Prasad, M. (2020). X-ray baggage inspection with computer vision: A survey. *IEEE Access, 8*, 145620-145633. https://doi.org/10.1109/ACCESS.2020.3015014

[27] Baştan, M., Yousefi, M. R., & Breuel, T. M. (2011). Visual words on baggage x-ray images. In *Proceedings of International Conference on Computer Analysis of Images and Patterns* (pp. 360-368). Springer-Verlag Berlin Heidelberg.

[28] Lowe, D. G. (2004). Distinctive image features from scale-invariant key points. *International Journal of Computer Vision, 60*, 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[29] Mery, D. (2013). X-ray testing by computer vision. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 360-367). IEEE. https://doi.org/10.1109/CVPRW.2013.61

[30] Mansoor, M., & Rajashankari, R. (2012). Detection of concealed weapons in x-ray images using fuzzy K-NN. *International Journal of Computer Science, Engineering and Information Technology, 2*(2), 187- 196. https://doi.org/10.5121/ijcseit.2012.2216

[31] Turcsany, D., Mouton, A., & Breckon, T. (2013). Improving feature-based object recognition for x-ray baggage security screening using primed visual words. In *2013 IEEE International Conference on Industrial Technology (ICIT)* (pp. 1140-1145). IEEE. https://doi.org/10.1109/ICIT.2013.6505833

[32] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding, 110*(3), 346-359. https://doi.org/10.1016/j.cviu.2007.09.014

[33] Bastan, M., Byeon, W., & Breuel, T. (2013). Object recognition in multi-view dual energy x-ray images. In *Proceedings of British Machine Vision Conference* (pp. 1-11). BMVA Press. https://doi.org/10.5244/C.27.130

[34] Mouton, A., Breckon, T. P., Flitton, G. T., & Megherbi, N. (2014). 3D object classification in baggage computed tomography imagery using randomised clustering forests. In *Proceedings of 2014 IEEE International Conference on Image Processing (ICIP)* (pp. 5202-5206). IEEE. https://doi.org/10.1109/ICIP.2014.7026053

[35] Akçay, S., Kundegorski, M. E., Devereux, M., & Breckon, T. (2016). Transfer learning using convolutional neural networks for object classification within x-ray baggage security imagery. In *Proceedings of 2016 IEEE International Conference on Image Processing (ICIP)* (pp. 1057-1061). IEEE. https://doi.org/10.1109/ICIP.2016.7532519

[36] Akcay, S., Kundegorski, M. E., Willcocks, C. G., & Breckon, T. P. (2018). Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery. *IEEE Transactions on Information Forensics and Security, 13*(9), 2203-2215. https://doi.org/10.1109/TIFS.2018.2812196

[37] Mery, D., Svec, E., Arias, M., Riffo, V., Saavedra, J. M., & Banerjee, S. (2017). Modern computer vision techniques for x-ray testing in baggage inspection. *IEEE Transactions on Systems, Man and Cybernetics (System), 47*(4), 682-692. https://doi.org/10.1109/TSMC.2016.2628381

[38] Riffo, V., Flores, S., & Mery, D. (2017). Threat objects detection in x-ray images using an active vision approach. *Journal of Nondestructive Evaluation, 36*. https://doi.org/10.1007/s10921-017-0419-3

[39] Xu, M., Zhang, H., & Yang, J. (2018). Prohibited item detection in airport x-ray security images via attention mechanism-based CNN. In *Pattern Recognition and Computer Vision. PRCV 2018*. Springer. https://doi.org/10.1007/978-3-030-03335-4_37

[40] Gaus, Y. F. A., Bhowmik, N., Akçay, S., Guillén-Garcia, P. M., Barker, J. W., & Barker, T. P. (2019). Evaluation of a dual convolutional neural network architecture for object-wise anomaly detection in cluttered x-ray security imagery. In *Proceedings of International 2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE. https://doi.org/10.1109/IJCNN.2019.8851829

[41] Klomp, S. R., van de Wouw, D. W. J. M., & de With, P. H. N. (2019). Real-time small-object change detection from ground vehicles using a Siamese convolutional neural network. *Journal of Imaging Science and Technology, 63*, 060402–1-060402–16. https://doi.org/10.2352/J.ImagingSci.Technol.2019.63.6.060402

[42] Ba, L. J., & Caruana, R. (2014). Do deep nets really need to be deep?. In *Proceedings of International Conference on Neural Information Processing Systems* (pp. 2654–2662). https://doi.org/10.48550/arXiv.1312.6184

[43] Mosca, A., & Magoulas, G. D. (2019). Customised ensemble methodologies for deep learning: Boosted residual networks and related approaches. *Neural Computing and Applications, 31*, 1713-1731. https://doi.org/10.1007/s00521-018-3922-2

[44] Rahman, M. M., Islam, M. S., Sassi, R., & Aktaruzzaman. (2019). Convolutional neural networks performance comparison for handwritten Bengali numerals recognition. *SN Applied Sciences, 1*, 1660. https://doi.org/10.1007/s42452-019-1682-y

[45] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84-90. https://doi.org/10.1145/3065386

[46] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1-8). IEEE. https://doi.org/10.1109/CVPR.2015.7298594

[47] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskeer, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*(1), 1929-1958.

[48] Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1717-1724). IEEE. https://doi.org/10.1109/CVPR.2014.222

[49] Vyas, A., Yu, S., & Paik, J. (2018). Fundamentals of digital image processing. In *Multiscale transforms with application to image processing* (pp. 3-11). Springer Singapore.

[50] Sifre, L., & Mallat, S. (2013). Rotation, scaling and deformation invariant scattering for texture discrimination. In *Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1233-1240). IEEE. https://doi.org/10.1109/CVPR.2013.163

[51] Premaladha, J., & Ravichandran, K. S. (2016). Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms. *Journal of Medical Systems, 40*(4), 96. https://doi.org/10.1007/s10916-016-0460-2

[52] Yang, J., Zhao, Z., Zhang, H., & Shi, Y. (2019). Data augmentation for x-ray prohibited item images using Generative Adversarial Networks. *IEEE Access, 7*, 28894-28902. https://doi.org/10.1109/ACCESS.2019.2902121

[53] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*. https://doi.org/10.1186/s40537-019-0197-0

[54] Khalifa, N. E., Loey, M., & Mirjalili, S. (2022). A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artificial Intelligence Review, 55*, 2351-2377. https://doi.org/10.1007/s10462-021-10066-4

[55] Zhang, Y., Wang, S., Zhao, H., Guo, Z., & Sun, D. (2021). CT image classification based on convolutional neural network. *Neural Computing and Applications*, *33*(14), 8191-8200. https://doi.org/10.1007/s00521-020-04933-4

[56] Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification. In R. Meersman, Z. Tari, D. C. Schmidt (eds.), *On the move to meaningful internet systems 2003: CoopIS, DOA, and ODBASE. OTM 2003*. Springer, Berlin. https://doi.org/10.1007/978-3-540-39964-3_62

[57] Nedjar, I., El Habib Daho, M., Settouti, N., Mahmoudi, S., & Chikh, M. A. (2015). Random forest based classification of medical x-ray images using a genetic algorithm for feature selection. *Journal of Mechanics in Medicine and Biology*, *15*(02), 15400251-15400258. https://doi.org/10.1142/S0219519415400254

[58] Evgeniou, T., & Pontil, M. (1999). Support vector machines: Theory and applications. In G. Paliouras, V. Karkaletsis & C. D. Spyropoulos (eds.), *Machine learning and its applications* (pp. 249-257). Springer. https://doi.org/10.1007/3-540-44673-7_12

[59] Liu, T., Fang, S., Zhao, Y., Wang, P., & Zhang, J. (2015). Implementation of training convolutional neural networks. https://doi.org/10.48550/arXiv.1506.01195

[60] Vani, S., & Madhusudhana Rao, T. V. (2019). An experimental approach towards the performance assessment of various optimizers on convolutional neural networks. In *Proceeding of 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE. https://doi.org/10.1109/ICOEI.2019.8862686

[61] Mohammed Taqi, A., Awad, A., Al-Azzo, F., & Milanova, M. (2018). The impact of multi-optimizers and data augmentation on tensorflow convolutional neural network performance. In *Proceeding of 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE. https://doi.org/10.1109/MIPR.2018.00032

[62] Gomez, A. N., Zhang, I., Rao Kamalakara, S., Madaan, D., Swersky, K., Gal, Y., . . . Hinton, G. E. (2019). *Learning sparse networks using targeted dropout*. https://doi.org/10.48550/arXiv.1905.13678

[63] Ito, Y. (1991). Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory. *Neural Networks, 4*(3), 385-394. https://doi.org/10.1016/0893-6080(91)90075-G

[64] Khalifa, N. E., Loey, M., & Mirjalili, S. (2022). A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artificial Intelligence Review, 55*, 2351-2377. https://doi.org/10.1007/s10462-021-10066-4

[65] Bloice, M. D., Stocker, C., & Holzinger, A. (2017). *Augmentor: An image augmentation library for machine learning*. https://doi.org/10.48550/arXiv.1708.04680

[66] Patro, V. M., Patra, M. R. (2014). Augmenting weighted average with confusion matrix to enhance classification accuracy. *Transactions on Machine Learning and Artificial Intelligence, 2*(4), 77-91. https://doi.org/10.14738/tmlai.24.328

[67] Xu, M., Yoon, S., Fuentes, A., & Park, D. S. (2022). *A comprehensive survey of image augmentation techniques for deep learning*. https://doi.org/10.48550/arXiv.2205.01491