© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Chest X-Ray Image Classification on Common Thorax Diseases using GLCM and AlexNet Deep Features

Tengku Afiah Mardhiah Tengku Zainul Akmal[1], Joel Chia Ming Than[1], Haslailee Abdullah[1], Norliza Mohd Noor[1*]

[1]Razak Faculty of Technology and Informatics,
Universiti Teknologi Malaysia, UTM Kuala Lumpur Campus, Jalan Sultan Yahya Petra, Kuala Lumpur, 51400, MALAYSIA

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2019.11.04.003
Received 21 April 2019; Accepted 8 July 2019; Available online 5 September 2019

Abstract: Image processing has been progressing far in medical as it is one of the main techniques used in the development of medical imaging diagnosis system. Some of the medical imaging modalities are the Magnetic Resonance Imaging (MRI), Computed Tomography (CT) Scan, X-Ray and Ultrasound. The output from all of these modalities would later be reviewed by the expert for an accurate result. Ensemble methods in machine learning are able to provide an automatic detection that can be used in the development of computer aided diagnosis system which can aid the experts in making their diagnosis. This paper presents the investigation on the classification of fourteen thorax diseases using chest x-ray image from ChestX-Ray8 database using Grey Level Co-occurrence Matrix (GLCM) and AlexNet feature extraction which are process using supervised classifiers: Zero R, k-NN, Naïve Bayes, PART, and J48 Tree. The classification accuracy result indicates that k-NN classifier gave the highest accuracy compare to the other classifiers with 47.51% accuracy for GLCM feature extraction method and 47.18% for AlexNet feature extraction method. The result shows that number of data by class and multilabelled data will influence the classifcation method. Data using GLCM feature extraction method has higher classification accuracy compared to AlexNet and required less processing step.

Keywords: Alexnet, classification, ensemble method, GLCM, supervised classification

1. Introduction

Image processing in medical has been progressing far than it ever did when it's one of the main techniques used in the biomedical imaging system and computer aided diagnosis systems. Few of the well-known medical imaging modalities are Magnetic Resonance Imaging (MRI), Computed Tomography (CT) Scan, X-Ray and Ultrasound. The output from these imaging modalities would later be reviewed by expert for an accurate result. Computer aided diagnosis not only save time as it can process thousand images just in a few minutes, it can also be used as tool to aid the expert as a second opinion, which in turn could save more lives.

Thorax consist of several organs which include heart, lungs, thymus gland, muscles and other various internal structures such as diaphragm. Thorax disease refer to diseases that affect the chest and the most common symptom is the chest pain. Some of the common thoracic disease are atelectasis, pneumothorax, pneumonia, nodule, mass, infiltration, effusion and cardiomegaly. Figure 1 shows the thorax diseases observed in chest X-ray [1]. Correct diagnosis and disease treatment are important especially during the early stage, which is the most difficult part of the process due to similarity in appearance.

*Corresponding author: norliza@utm.my

Mass detection [1] uses two stage approach, first stage is the Chest X-Ray Radiograph (CXR) is widely preferred for diagnosing of several lung disease as it non-invasive and relatively low cost compare to other diagnose method [2, 3]. Computer tomography (CT) scans can also be used for diagnosing but it is not recommended due to high radiation dose and high cost [4]. There are several researches has been done using CXR in detection and diagnosing of disease such as Tuberculosis (TB), nodule, mass tissue, pneumonia clouds detection[5].

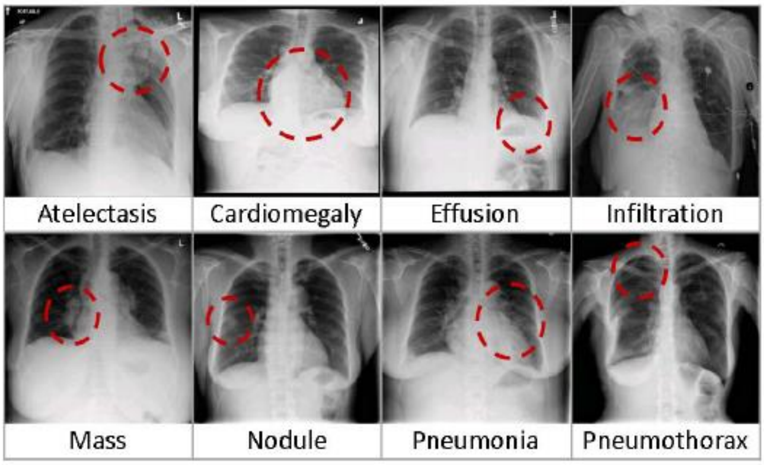

Fig. 1 - Chest X-ray Image with Thorax Diseases [6]

For the TB detection, [7] proposed three stages detection process which starts from lung segmentation, feature extraction to classification. The lung segmentation was done using adaptive threshold, feature extraction consists of intensity histogram (IH), gradient magnitude histogram (GM), shape descriptor histogram (SD), curvature descriptor histogram (CD), histogram of oriented gradients (HOG) and local binary patterns (LBP). K-Nearest Neighbour (k-NN) classifier is utilized to classify input image as abnormal or normal image. Another research on TB detection was also be done but the researcher uses graph cut based segmentation algorithms for the lung segmentation and hybrid combination of Artificial Neural Network (ANN) and Genetic Algorithm (GA) for classification [3].

A research done by [8] on thorax x-ray image using k-nearest neighbour (KNN) and GLCM feature extraction methods shows an accuracy of 97.83% for 46 images. While for nodule detection image classification done by [4] shows that support vector machine (SVM) trained with features managed to reduce high number of false positive. Initial pre-processing of image using contrast enhancing to detect a set of nodules and second stage consisted of classifying detected regions using pattern recognition technique were done by [2]. Utilized grey level co-occurrence matrix (GLCM) feature extraction for nodule detection achieved 75.6% accuracy. Thus, it proven that CXR image and machine learning or deep learning can be used in detection of lung disease.

In this paper, GLCM and AlexNet Deep Convolutional Neural Network (Deep CNN) features of chest X-ray images will be used for the classification of common thoracic disease. An open source software, Waikato Environment for Knowledge Analysis (WEKA) tool is used for the ensemble methods. “ChestX-Ray8” a database of hospital-scale chest X-ray image consists of fourteen common thoracic diseases [6] will be used for this paper dataset.

The objective of this paper is to identify classifier method and feature extraction method that gives higher accuracy for the detection of fourteen thorax diseases.

2. Material and method

This study uses 10000 x-ray images from “ChestX-Ray8” frontal chest x-ray image database where the image is labelled as single finding label, multi-label or no finding as per illustrated in **Error! Reference source not found.** [9]. The database also provide detail information for each x-ray images which are ‘image index’, ‘finding labels’, ‘follow-up number’, ‘patient unique ID’, ‘patient age’, ‘patient gender’, ‘view position’, ‘image width’, ‘image height’, ‘image pixel spacing x-axis’, ‘image pixel spacing y-axis’. This information is used as initial attributes for the supervise classification. The x-ray image feature is extracted using GLCM and AlexNet method in Matlab environment. GLCM feature extraction method extracts of four features whereas AlexNet feature extraction extracts 4096 features. The feature extracted from both methods are combined with the initial eleven attributes. Thus, the total attributes for GLCM data consisted of 15 attributes while AlexNet data consisted of 4107 attributes.

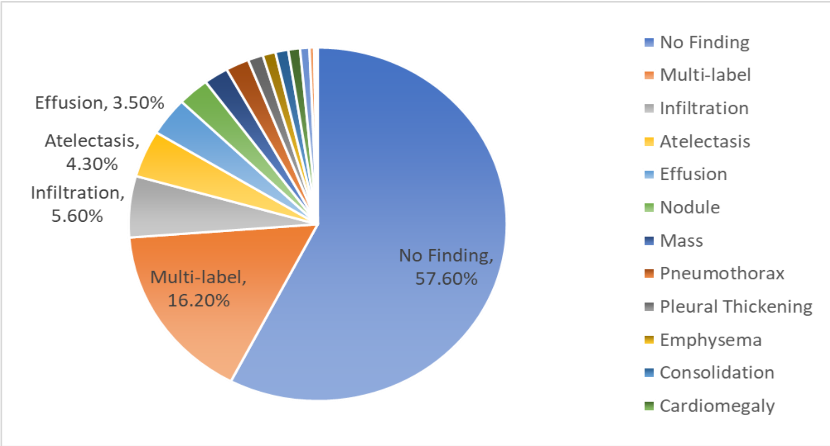

Fig. 3 - Input Data Record by Finding Labels

The overall process flow of this study is shown in Fig. *2*. Data pre-processing will be using Information Gain Attribute Evaluaton paired with Ranker for feature selections. It evaluates the worth of an attributes by measuring the information gain with respect to the class. The data images will be classify using supervised classifiers: Zero R, k-NN, Naïve Bayes, PART, and J48. The highest accuracy classifier will be selected based on the classification result and further accuracy enhancement is done through Ensemble Method.

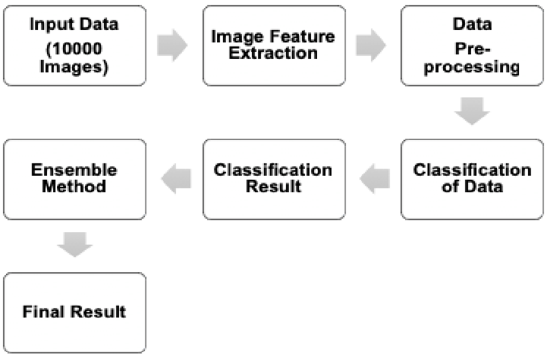

Fig. 2 - Flow process of research methodology

2.1 Grey Level Co-occurrence Matrix

Grey Level Co-occurrence Matrix (GLCM) is a matrix that characterizes the relative frequencies of a pair of grey levels present at certain distance apart and at a particular angle of an image. GLCM generates different features value based on the different pairs of angles and distance which are contrast, correlation, homogeneity and energy. Contrast is the grey level variation in a GLCM. Correlation gives information about how correlated a pixel is to its neighboring pixels. Energy measures the textural uniformity of an image and helps in determining disorders in texture while homogeneity measures the uniformity of the non-zero entries in the GLCM [10]. Various works had shown that GLCM features were useful in classification of lung disease severity [11-14], and in pattern recognition [15].

2.2 Alexnet

AlexNet is a deep convolutional neural network used to classify images into 1000 different classes. It contains eight layers with weights; the first five are convolutional and the remaining three are fully connected. The output of the last fully-connected layer consisted of 4096 dimensional features [16]. AlexNet had shown very useful in classification of medical imaging for diseases such as lung diseases [17], heart conditions [18, 19] as well as cancer [20].

2.3 Data pre-processing

As mentioned earlier, GLCM data consisted of 15 attributes thus it does not requires further feature selection process. However, the AlexNet data requires feature selection to minimize the number of attributes for the classification process. Information Gain Attribute Evaluation paired with Ranker is used to reduce the number of attributes using WEKA. The number of attributes is reduced to 30 attributes with respect to the finding label class from 4107 attributes.

3. Result and discussion

As stated previously, the chest x-ray images used for this study consisted of fifteen type (fourteen disease types and one label as no finding) of finding labels for thorax diseases which is then extracted using GLCM and AlexNet method. WEKA is utilized for pre-processing and classification process. Five supervised classifiers are used for classification which are Zero R, k-NN, Naïve Bayes, PART, and J48. The best classifier will be selected for the ensemble process.

3.1 Classification using GLCM data

The GLCM features data consists of 15 finding label classes which includes no finding label and data with multi-label classes. The GLCM features data is classified using all the identified classifier. It is run using 10-fold cross validation and the result of percent accuracy by classifier is displayed in Table 2. K-NN classifier shows highest percent accuracy, 47.51% followed by Zero R, PART and J48 with 47.21% percent accuracy. The different in accuracy percentage between these two classifiers is only 0.30%.

Table 2 - The results of different type classifier for GLCM data

Classifier	Percent Accuracy %
Zero R	47.21
k-NN (k=35)	47.51
Naïve Bayes	37.75
PART	47.21
J48	47.21

Due to the slight different in percent accuracy between Zero R and k-NN, both classifier results are compared. Table 3 shows the summary result of Zero R classifier and k-NN. The k-NN classifier was used with the k value of 35. K-NN performed slightly better compared to Zero R. Based on the detailed accuracy result of Zero R and k-NN classifier by finding label class in Table 4. It shows that Zero R can only classified 100% accuracy for no finding class and unable to classify other classes. Where as the k-NN classifier able to classify seven finding labels classes out of 15 classes. Highest true positive (TP) rate is 95.2% for no finding followed by 11% for infiltration and 7.3% for pneumothorax.

In order to further enhance the classification accuracy, ensemble method is carried out with the same dataset using k-NN classifier with k-value of 35. The result shows no significant improvement to the classification accuracy. The result of accuracy by ensemble method using GLCM data is summarize as per Table 5 below. Based on the detailed accuracy result of k-NN classifier, data with no finding label shows the highest accuracy which is relevant as 57.6% of the data is labelled with no finding. Thus, further classification is carried out by filtering out the no finding data, the dataset is now named as Filtered data. The result of the classifiers are summarize in Table 6. The k-NN classifier shows highest percent accuracy, 24.29% followed by Zero R, PART and J48 Decision Tree with 20.72%.

Table 3 - The result summary of Zero R Classifier and k-NN for GLCM data

Classifier	Zero R	k-NN
Correctly Classified Instances	47.2074%	47.5109%
Incorrectly Classified Instances	52.7926%	52.4891%
Kappa statistic	0	0.0772
Mean absolute error	0.0991	0.0933
Root mean squared error	0.225	0.2179
Relative absolute error	100%	94.1511%
Root relative squared error	100%	97.913%

Table 4 - Result of Zero R and k-NN Classifier for GLCM Data for various lung diseases

Classifier	k-NN			Zero R		
Finding Labels	**TP Rate**	**FP Rate**	**ROC**	**TP Rate**	**FP Rate**	**ROC**
Atelectasis	0.063	0.020	0.616	0.000	0.000	0.499
No Finding	0.952	0.841	0.678	1.000	1.000	0.500
Infiltration	0.11	0.038	0.642	0.000	0.000	0.499
Mass	0.021	0.004	0.660	0.000	0.000	0.497
Pneumothorax	0.073	0.007	0.774	0.000	0.000	0.497
Nodule	0.000	0.001	0.623	0.000	0.000	0.499
Emphysema	0.04	0.002	0.749	0.000	0.000	0.500
Pleural Thickening	0.000	0.000	0.636	0.000	0.000	0.496
Effusion	0.042	0.015	0.653	0.000	0.000	0.499
Consolidation	0.000	0.000	0.631	0.000	0.000	0.496
Pneumonia	0.000	0.000	0.601	0.000	0.000	0.492
Hernia	0.000	0.000	0.610	0.000	0.000	0.471
Cardiomegaly	0.000	0.000	0.646	0.000	0.000	0.493
Fibrosis	0.000	0.000	0.616	0.000	0.000	0.494
Edema	0.000	0.001	0.744	0.000	0.000	0.494
Weighted Average	**0.475**	**0.405**	**0.665**	**0.472**	**0.472**	**0.499**
Classified / Total Class		**7/15**			**1/15**	

Table 5 - The result using ensemble algorithm method for GLCM Data

Classifier	Percent Accuracy %
AdaBoost	47.5109%
Bagging	47.5109%
Stacking	31.9282%
Voting	47.2074%

Table 6 - The results of different type classifier for Filtered GLCM data

Classifier	Percent Accuracy %
Zero R	20.72
k-NN (k=21)	24.29
Naïve Bayes	10.95
PART	20.72
J48	20.72

The summary result of Zero R and k-NN classifier for the filtered GLCM data are shown in Table 7 while the detailed accuracy of finding label class is in Table 8. Zero R only able to classify Infiltration label with 100% TP Rate. k-NN classifier are able to classify nine finding labels classes out of 14 classes for filtered GLCM data . Highest TP Rate is 50.6% for no finding followed by 36.4% for atelectasis and 16.5% for pneumothorax. Classification using filtered GLCM data is further enhance by applying the ensemble method using k-NN classifier with k-value equals 35. The result shows deterioration of accuracy for all ensemble method except for AdaBoost. The accuracy percentage shows the same result for k-NN classifier with or without using AdaBoost method as shown in Table 9.

Table 7 - The result summary of Zero R and k-NN classifier for Filtered GLCM Data

Classifier	Zero R	k-NN
Correctly Classified Instances	20.7239%	24.3592%
Incorrectly Classified Instances	79.2761%	75.6408%
Kappa statistic	0	0.0948
Mean absolute error	0.1252	0.12
Root mean squared error	0.2502	0.2476
Relative absolute error	100%	95.8829%
Root relative squared error	100%	98.968%

Table 8 - Result of Zero R and k-NN Classifier for Filtered GLCM Data for various lung diseases

Classifier	k-NN			Zero R		
Finding Labels	**TP Rate**	**FP Rate**	**ROC**	**TP Rate**	**FP Rate**	**ROC**
Atelectasis	0.364	0.251	0.608	0.000	0.000	0.499
Infiltration	0.506	0.316	0.635	1.000	1.000	0.499
Mass	0.06	0.035	0.659	0.000	0.000	0.497
Pneumothorax	0.165	0.043	0.76	0.000	0.000	0.497
Nodule	0.117	0.056	0.646	0.000	0.000	0.498
Emphysema	0.092	0.007	0.742	0.000	0.000	0.500
Pleural Thickening	0.029	0.006	0.659	0.000	0.000	0.496
Effusion	0.284	0.188	0.606	0.000	0.000	0.499
Consolidation	0.000	0.001	0.605	0.000	0.000	0.496
Pneumonia	0.000	0.000	0.553	0.000	0.000	0.492
Hernia	0.000	0.000	0.641	0.000	0.000	0.471
Cardiomegaly	0.000	0.000	0.686	0.000	0.000	0.493
Fibrosis	0.000	0.000	0.685	0.000	0.000	0.494
Edema	0.027	0.003	0.736	0.000	0.000	0.494
Weighted Average	**0.244**	**0.149**	**0.645**	**0.207**	**0.207**	**0.498**
Classified / Total Class		**9/14**			**1/14**	

Table 9 - The result using ensemble algorithm method for filtered GLCM Data

Classifier	Percent Accuracy %
AdaBoost	24.3592%
Bagging	23.5358%
Stacking	21.4852%
Voting	23.4271%

Since the GLCM and Filtered GLCM data consists of multilabel data as well, another classification is done using single label image data only. The result of the classification for the single class GLCM data is shown in Table 10 where J48 and k-NN classifier has the highest percentage correct result. The summary result and detailed accuracy result of J48 and k-NN classifiers for the single class data are displayed in Table 11. Table 12 shows that the J48 classifier is able to classify 12 out of 14 finding label class. The highest TP rate is 40.6% for infiltration followed by 34.7% for Pneumothorax and 31.1% for Effusion. However, K-NN classifier is only able to classify 11 out of 14 finding label class as shown in Table 12. The highest TP rate is 47.2% for infiltration followed by 46% for Atelectasis and 33.7% for Pneumothorax. The k-NN classifier with 3 nearest neighbor is further enhance using the ensemble method and the result is summarize in Table 13 where Bagging method gave the highest accuracy results. The summary result for Bagging method is shown in Table 14.

Table 10 - The results of different type classifier for Single Class GLCM data

Classifier	Percent Accuracy %
Zero R	21.23
k-NN (k=3)	28.85
Naïve Bayes	14.28
PART	24.76
J48	26.20

Table 11 - The result summary of J48 classifier for Single Class GLCM Data

Classifier	J48	k-NN
Correctly Classified Instances	26.1814%	28.8491%
Incorrectly Classified Instances	73.8186%	71.1509%
Kappa statistic	0.1523	0.166
Mean absolute error	0.1094	0.1067
Root mean squared error	0.2949	0.2651
Relative absolute error	87.1375%	85.0054%
Root relative squared error	117.7147%	105.8446%

Table 12 - Result of J48 and k-NN classifier for Single Class GLCM Data for various lung diseases

Classifier	J48			k-NN		
Finding Labels	**TP Rate**	**FP Rate**	**ROC**	**TP Rate**	**FP Rate**	**ROC**
Atelectasis	0.288	0.165	0.564	0.46	0.283	0.622
Infiltration	0.406	0.189	0.621	0.472	0.225	0.678
Pneumothorax	0.347	0.062	0.676	0.337	0.057	0.719
Nodule	0.191	0.095	0.57	0.191	0.093	0.631
Effusion	0.311	0.116	0.612	0.249	0.078	0.65
Pleural Thickening	0.134	0.044	0.563	0.087	0.018	0.591
Mass	0.133	0.065	0.568	0.09	0.035	0.635
Emphysema	0.195	0.034	0.613	0.195	0.014	0.673
Cardiomegaly	0.122	0.024	0.593	0.133	0.008	0.583
Consolidation	0.134	0.021	0.594	0.152	0.013	0.624
Edema	0.225	0.009	0.709	0.175	0.007	0.75
Hernia	0.000	0.002	0.489	0.000	0.000	0.506
Fibrosis	0.039	0.018	0.535	0.000	0.005	0.576
Pneumonia	0.000	0.004	0.53	0.000	0.000	0.516
Weighted Average	**0.262**	**0.109**	**0.597**	**0.288**	**0.124**	**0.646**
Classified/Total Class		**12/14**			**11/14**	

Table 13 - The result using ensemble algorithm method for Single Class GLCM Data

Classifier	Percent Accuracy %
AdaBoost	28.8491%
Bagging	29.3826%
Stacking	27.3247%
Voting	28.9253

Table 14 - The result summary of Bagging Ensemble Method for Single Class GLCM Data

Ensemble Method	Bagging
Correctly Classified Instances	29.3826%
Incorrectly Classified Instances	70.6174%
Kappa statistic	0.191
Mean absolute error	0.1089
Root mean squared error	0.2554
Relative absolute error	86.7226%
Root relative squared error	101.9633%

By using the bagging method, the k-NN classifier able to classify 12 finding labels class as shown in Table 15. However, there are reduction of TP rate for few of the disease compared to k-NN classifier TP rate, such as Infiltration TP rate dropped from 47.2% to 44.7% and TP rate for Atelectasis dropped from 46% to 32.8 %. However, Pneumothorax TP rate increases from 33.7% to 35.6%.

Table 15: Result of Bagging Method for Single Class GLCM Data for various lung diseases

Finding Labels	TP Rate	FP Rate	ROC
Atelectasis	0.328	0.140	0.646
Infiltration	0.447	0.170	0.699
Pneumothorax	0.356	0.052	0.766
Nodule	0.240	0.094	0.664
Effusion	0.316	0.119	0.671
Pleural Thickening	0.142	0.039	0.651
Mass	0.194	0.077	0.657
Emphysema	0.239	0.025	0.716
Cardiomegaly	0.153	0.022	0.602
Consolidation	0.152	0.025	0.631
Edema	0.225	0.011	0.802
Hernia	0.000	0.005	0.478
Fibrosis	0.078	0.024	0.557
Pneumonia	0.000	0.003	0.505
Weighted Average	**0.294**	**0.101**	**0.671**

3.2 Classification using Alexnet Features Data

The AlexNet features data is classified using all the classifier used in the previous section. It is run using 10-fold cross validation and the result of percent accuracy by classifier is displayed in Table 15. Zero R, PART and J48 classifier shows the higher percent accuracy, 47.21% followed by k-NN classifier with 47.17 % percent accuracy. The different between these classifiers is only 0.04% which is relatively small. Thus k-NN classifier will be re-run further evaluation. Since Zero R, PART and J48 classifier shows the same result, only Zero R and k-NN classifier result will be discussed further. The summary result of Zero R and k-NN classifier is shown in Table 16 below. Based on the detailed accuracy result of Zero R and k-NN classifier by finding label class as shown in Table 17. It shows that Zero R can only classified 100% accuracy for no finding class and unable to classify other classes. The k-NN classifier is able to classify ten finding labels classes out of 15 classes. Highest TP Rate is 92.6% for no finding followed by 14.3% for infiltration and 9.5% for pneumothorax.

Table 15 - The result using different type classifier for AlexNet Data

Classifier	Percent Accuracy %
Zero R	47.21
k-NN (k=21)	47.17
Naïve Bayes	8.79
PART	47.21
J48	47.21

Table 16 - The result summary of Zero R and k-NN classifier for AlexNet Data

Classifier	Zero R	k-NN
Correctly Classified Instances	47.2074%	47.1828%
Incorrectly Classified Instances	52.7926%	52.8172%
Kappa statistic	0	0.1053
Mean absolute error	0.0991	0.092
Root mean squared error	0.2225	0.2178
Relative absolute error	100%	92.882%
Root relative squared error	100%	97.8698%

Table 17 - Result of Zero R classifier for AlexNet Data for various lung diseases

Classifier	k-NN			Zero R		
Finding Labels	**TP Rate**	**FP Rate**	**ROC**	**TP Rate**	**FP Rate**	**ROC**
Atelectasis	0.000	0.000	0.499	0.075	0.035	0.606
No Finding	1.000	1.000	0.500	0.926	0.760	0.703
Infiltration	0.000	0.000	0.499	0.143	0.054	0.634
Mass	0.000	0.000	0.497	0.025	0.006	0.669
Pneumothorax	0.000	0.000	0.497	0.095	0.012	0.760
Nodule	0.000	0.000	0.499	0.010	0.004	0.610
Emphysema	0.000	0.000	0.500	0.048	0.002	0.758
Pleural Thickening	0.000	0.000	0.496	0.007	0.001	0.665
Effusion	0.000	0.000	0.499	0.072	0.025	0.662
Consolidation	0.000	0.000	0.496	0.000	0.001	0.612
Pneumonia	0.000	0.000	0.492	0.000	0.000	0.571
Hernia	0.000	0.000	0.471	0.000	0.000	0.709
Cardiomegaly	0.000	0.000	0.493	0.000	0.000	0.652
Fibrosis	0.000	0.000	0.494	0.000	0.000	0.612
Edema	0.000	0.000	0.494	0.007	0.000	0.706
Weighted Average	**0.472**	**0.472**	**0.499**	**0.472**	**0.371**	**0.675**

The k-NN classifier shows more promising result compared to Zero R which can only classify data for one finding labels, the k-NN classifier is further enhanced using the ensemble method. The summarize result of ensemble method is shown in Table 18. Bagging method gave the highest accuracy result of 47.54%. Table 19 displayed the summary result using Bagging Method with k-NN classifier.As the no finding data affects the classification result due to high number of no finding label data records, the classification process is re-run by filtering out the no finding label data. The classifiers result is layed out in Table 20. K-NN classifier shows highest percent accuracy of 22.06%, followed by Zero R, PART and J48 with 20.72%. The k-NN classifier is re-run by changing the k-value from 21 to 35 nearest neighbours and shows slight improvement of accuracy from 22.06% to 23.1% as per Table 21. Based on the detailed accuracy result of k-NN classifier for the filtered AlexNet data, the highest TP rate is 51.2% for Infiltration followed by 39.4% for Atelectasis and 11.3% for nodule as displayed in Table 22. The k-NN classifier is further enhanced by using ensemble method. However the result shows no improvement in percent accuracy as per table 23.

Table 18 - The result using ensemble algorithm method for AlexNet data

Classifier	Percent Accuracy %
AdaBoost	47.1418%
Bagging	47.5355%
Stacking	46.6579%
Voting	47.3386%

Table 19 - The result summary of Bagging Ensemble Method for AlexNet Data

Ensemble Method	Bagging
Correctly Classified Instances	47.5355%
Incorrectly Classified Instances	52.4645%
Kappa statistic	0.0902
Mean absolute error	0.0929
Root mean squared error	0.2168
Relative absolute error	93.8055%
Root relative squared error	97.4288%

Table 20 - The result using different type classifier for Filtered AlexNet Data

Classifier	Percent Accuracy %
Zero R	20.72
k-NN (k=21)	22.06
Naïve Bayes	12.04
PART	20.72
J48	20.72

Table 21: The result summary of k-NN classifier for Filtered AlexNet Data where k = 35

Classifier	k-NN
Correctly Classified Instances	23.1008%
Incorrectly Classified Instances	76.8992%
Kappa statistic	0.0732
Mean absolute error	0.1212
Root mean squared error	0.2489
Relative absolute error	96.7898%
Root relative squared error	99.5067%

Table 22 - Result of k-NN classifier for Filtered AlexNet Data for various lung diseases

Finding Labels	TP Rate	FP Rate	ROC
Atelectasis	0.394	0.293	0.591
Infiltration	0.512	0.354	0.617
Mass	0.031	0.019	0.601
Pneumothorax	0.103	0.029	0.703
Nodule	0.113	0.055	0.642
Emphysema	0.012	0.002	0.671
Pleural Thickening	0.040	0.006	0.663
Effusion	0.224	0.167	0.574
Consolidation	0.000	0.000	0.588
Pneumonia	0.000	0.000	0.514
Hernia	0.000	0.000	0.652
Cardiomegaly	0.029	0.002	0.647
Fibrosis	0.000	0.000	0.670
Edema	0.000	0.000	0.716
Weighted Average	0.231	0.158	0.618

Table 23 - The result using ensemble algorithm method for Filtered AlexNet Data

Classifier	Percent Accuracy %
AdaBoost	23.1008%
Bagging	22.914%
Stacking	20.522%
Voting	22.3085%

Since the filtered AlexNet data still consists of data with multilabel class, thus another classification run is done for data with single label class only. The classifier results as shown in Table 24. K-NN classifier gave the highest percent accuracy of 25.13% followed by Zero R with 21.25%. The k-NN classifier is re-run using 35 nearest neighbor and the accuracy % increases from 25.13% to 26.52% as per Table 30. As shown in Table 26, k-NN classifier able to classify eight out of 14 finding label classes. The highest TP rate is 59.6% for infiltration followed by 42.8% for Atelectasis and 25.5% for Nodule. The k-NN classifier is further enhanced using Ensemble algorithm methods however there is no significant improvement can be seen by using these method as shown in Table 27.

Table 24 - The result using different type classifier for Single Class AlexNet Data

Classifier	Percent Accuracy %
Zero R	21.23
k-NN (k=21)	25.13
Naïve Bayes	19.45
PART	20.99
J48	12.63

Table 25 - The result summary of k-NN classifier for Single Class AlexNet Data

Classifier	k-NN
Correctly Classified Instances	26.5244%
Incorrectly Classified Instances	73.4756%
Kappa statistic	0.1189
Mean absolute error	0.1197
Root mean squared error	0.2472
Relative absolute error	95.3506%
Root relative squared error	98.6629%

Table 26 -3 Result of k-NN classifier for Single Class AlexNet Data for various lung diseases

Finding Labels	TP Rate	FP Rate	ROC
Atelectasis	0.428	0.298	0.606
Infiltration	0.596	0.328	0.680
Pneumothorax	0.109	0.031	0.684
Nodule	0.255	0.102	0.671
Effusion	0.218	0.098	0.631
Pleural Thickening	0.024	0.005	0.652
Mass	0.043	0.016	0.658
Emphysema	0.000	0.000	0.670
Cardiomegaly	0.000	0.001	0.606
Consolidation	0.009	0.002	0.683
Edema	0.000	0.000	0.753
Hernia	0.000	0.000	0.579
Fibrosis	0.000	0.000	0.722
Pneumonia	0.000	0.000	0.585
Weighted Average	0.265	0.146	0.655

Table 27 - The result using ensemble algorithm method for Single Class AlexNet Data

Classifier	Percent Accuracy %
AdaBoost	26.5244%
Bagging	25.4573%
Stacking	20.4268%
Voting	25.6098%

4. Conclusion

As a conclusion, both image feature extraction methods with k-NN classifier gave similar classification accuracy which is 47.51% for GLCM and 47.18 % for AlexNet. Although AlexNet has a slightly lower accuracy, the classifier able to classify ten finding labels compare to seven finding labels when using GLCM data. The classification is also carried out by removing no finding labels and the accuracy percentge was 24.28% for GLCM data and 23.20% for AlexNet Data. Since there was 16.2% multilabelled data, thus the data of no finding label was filtered to identify whether it can influenced the classification accuracy.The result shows increase of accuracy to 29.38% for GLCM and 26.52% for AlexNet data.This shows that the number of data by class and multilabelled data will influence the classification accuracy thus number of data should be increase and balance in terms of data per class to improve the classification accuracy. In term of processing time, GLCM data took 16 seconds to classify the data while AlexNet took 17 seconds. On top of the classification processing time, AlexNet requires additional 1 minutes and 30 seconds to pre-process the data before performing classification process. However, GLCM feature extraction processing time for 10000 images took 3.8 hours while AlexNet took 3.4 hours. Based on the classification results, GLCM feature extraction method shows better classification accuracy in terms of accuracy percentage compare to AlexNet.

Acknowledgement

This research is done by using the image from ChestX-ray8 database which a public database provided by NIH, Clinical Center, America's Research Hospital. This research is funded by Universiti Teknologi Malaysia and Ministry of Education Malaysia (Vot no.: Q.K130000.2540.18H45).

References

[1] X. Wang, et al. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.
[2] E.K. Dey and H.M. Muctadir. Chest X-ray analysis to detect mass tissue in lung. in 2014 International Conference on Informatics, Electronics & Vision (ICIEV). 2014.
[3] Z.Z. Htike et al. Computer-Aided Diagnosis of Pulmonary Nodules from Chest X-Rays Using Rotation Forest. in 2014 International Conference on Computer and Communication Engineering. 2014.
[4] A. Sridevi and G.K.D.P. Venkatesan. Tuberculosis malady recognition in chest radiographs via artificial neural networks. in 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE). 2017.
[5] T.S. Savithri and S.K.C. Devi. Nodule detection from posterior and anterior chest radiographs with different methods. in 2016 Future Technologies Conference (FTC). 2016.
[6] A. Sharma, D. Raju, and S. Ranjan. Detection of pneumonia clouds in chest X-ray using image processing approach. in 2017 Nirma University International Conference on Engineering (NUiCONE). 2017.
[7] B. Joykutty, K.G. Satheeshkumar, and B. Samuvel. Automatic tuberculosis detection using adaptive thresholding in chest radiographs. in 2016 International Conference on Emerging Technological Trends (ICETT). 2016.
[8] Riandini and M.K. Delimayanti. Feature extraction and classification of thorax x-ray image in the assessment of osteoporosis. in 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). 2017.
[9] ChestX-ray8 database: https://nihcc.app.box.com/v/ChestXray-NIHCC
[10] R.C. Gonzalez and R.E. Woods. Digital Image Processing, 2nd Ed. Prentice-Hall:New Jersey, 2002.
[11] J.C.M. Than, N.M. Noor, O.M. Rijal, R. M. Kassim, A. Yunos. Lung disease classification using GLCM and Deep Features from different deep learning architectures with principal component analysis. International Journal of Integrated Engineering, 10(7):76-89. 2018.
[12] M.B. Huber, M. Nagarajan, G. Leinsinger, L. A. Ray, A. Wismuller. Classification of interstitial lung disease patterns with topological texture features. in Proceedings SPIE Medical Imaging, 7624, 2–9. 2010.
[13] J. Wang, F. Li, Q. Li. Automated segmentation of lungs with severe interstitial lung disease in CT. Medical Physics, 36, 4592. 2009.
[14] J.T.C. Ming, N.M. Noor, O.M. Rijal, R. M. Kassim, A. Yunos, Lung disease classification using reticular pattern scoring and five class features with greedy stepwise based on GLCM. in IEEE Region 10 Annual International Conference (TENCON), 182-186. 2017
[15] S. Singh, D. Srivastava, and S. Agarwal. GLCM and its application in pattern recognition. in 2017 5th International Symposium on Computational and Business Intelligence (ISCBI). 2017.
[16] A. Krizhevsky, I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. in Advances in neural information processing systems. 2012.
[17] Ye, W., et al., Detection of pulmonary ground-glass opacity based on deep learning computer artificial intelligence. BioMedical Engineering Online, 2019. **18**(1).
[18] H. Sofian, J.C.M. Than, S. Mohammad, N.M. Noor. Calcification detection of coronary artery disease in intravascular ultrasound image: Deep feature learning approach. International Journal of Integrated Engineering, 10(7):43-57. 2018.
[19] Ruiz, J.T., J.D.B. Pérez, and J.R.B. Blázquez, Arrhythmia detection using convolutional neural models, in Advances in Intelligent Systems and Computing. 2019. p. 120-127.
[20] Hermessi, H., O. Mourali, and E. Zagrouba, Deep feature learning for soft tissue sarcoma classification in MR images via transfer learning. Expert Systems with Applications, 2019. **120**: p. 116-127.