p-ISSN 2089 – 0133 e-ISSN 2477-6416
URL : https://jurnal.uns.ac.id/ijap/article/view/77120
Indonesian Journal of Applied Physics (IJAP) Vol. 13 No. 2 page 325
DOI : https://doi.org/10.13057/ijap.v13i2.77120

# PNEUMONIA CLASSIFICATION BASED ON GLCM FEATURES EXTRACTION USING K-NEAREST NEIGHBOR

**Suharyana[1], Fuad Anwar[1], Armylia Chandra Dewi[1], Mohtar Yunianto[1*], Umi Salamah[2], Rifai Chai[3]**

[1]Physics Department, Universitas Sebelas Maret, Surakarta, Indonesia
[2]Informatics Department, Universitas Sebelas Maret, Surakarta, Indonesia
[3]Engineering Technologies, Swinburne University of Technology, Melbourne, Australia
[*]mohtaryunianto@staff.uns.ac.id

Received 24-07-2023, Revised 16-10-2023, Accepted 20-10-2023
Available Online 20-10-2023, Published Regularly October 2023

## ABSTRACT

Pneumonia has been detected using Machine learning. The stages in this study began with preprocessing in 4 stages: resizing, cropping, filtering using a high pass filter, and Adaptive Histogram Equalization. The feature extraction process continued with 22 Gray Level Co-occurrence Matrix (GLCM) features and classification using K-Nearest Neighbor (KNN). The image used was 150 data sets for training on the classification of 3 classes with a ratio of 50:50:50, while training on two classes was 50 bacterial pneumonia and 50 viral pneumonia. The most optimal training data accuracy results were obtained using the angle direction on the GLCM, namely 135o with the KNN classification (k = 3). For the classification of two classes Using 40 data sets, an accuracy of 91% was obtained, while testing for three classes with 60 data sets was 83.3%.

Keywords: pneumonia; adaptive histogram equalization; GLCM; KNN

**Cite this as:** Suharyana., Anwar, F., Dewi, A. C., Yunianto, M., Salamah, U., & Chai, R. 2023. Pneumonia Classification Based on GLCM Features Extraction Using K-Nearest Neighbor. *IJAP: Indonesian Joirnal of Applied Physics, 13*(2), 325-338. doi: https://doi.org/10.13057/ijap.v13i2.77120

## INTRODUCTION

Artificial Intelligent has the potential to detect disease through images (images) generated from X-rays. AI has been applied to diagnose chest X-ray images in pediatric pneumonia and is proven to help human experts classify the disease [1]. Pneumonia is an inflammation of the lungs caused by bacteria, fungi, viruses, or parasites that clog the alveoli so that the lungs become inflamed and filled with fluid [2]. According to the World Health Organization (WHO), the highest death rate due to pneumonia occurs in toddlers and children. Children's body systems are still weak and susceptible to pneumonia. However, detecting the type of pneumonia is challenging for radiologists [3].

Doctors cannot distinguish between normal lungs, bacterial pneumonia, and viral pneumonia without the naked eye. Not a few of the resulting images from CXR look blurry or lack contrast, making it difficult for radiologists to diagnose the images [4]. One of the ways to predict the difference between bacterial pneumonia and viral pneumonia is by taking a swab to obtain a sample. This technique requires high costs, special materials, and equipment [5].

Previous research that has performed image processing on pneumonia results from examinations using imaging techniques, [6] used 5,863 X-ray images of the lungs consisting of

images of normal lungs, bacterial pneumonia, and viral pneumonia. The method used is feature extraction of 4 Gray Level Co-Occurrence Matrix (GLCM) features and K-Nearest Neighbor (KNN) classification with k=3, 5, and 7 values. However, the best accuracy results obtained from this method are 66.20% with k=5.

For different classification purposes, [7] classified pneumonia images with details of 69 images, including 30 normal lung images, 19 common pneumonia images, 11 fever without pneumonia images, and 9 Covid-19 images. This study uses the Thermal Imaging of the back method to extract high-temperature regions and feature selection. The classification technique for lung images uses Machine Learning, with the classification of Support Vector Machine (SVM), KNN, Decision Tree, Gaussian Naïve Bayes, Linear or Quadratic Discriminant (LDA and QDA). The accuracies obtained from the normal and pneumonia lung image classification for each method were 93%, 91%, 90%, 86%, and 85%. Classification of pneumonia images and fever without pneumonia obtained the accuracy of each of the above methods is 86%, 78%, 78%, 83%, 71%, and 75%. In the classification into three classes, namely normal lungs, fever without pneumonia, and pneumonia, each of the above methods obtained an accuracy of 81%, 68%, 70%, 75%, 67%, and 65%. The highest accuracy obtained from the classification of 2 normal and pneumonia lung image classes was achieved by the SVM method at 93%.

Istianah et al. conducted research, namely the classification of 9 CXR results images consisting of 3 images of fungal pneumonia, three bacterial pneumonia, and lipoid pneumonia. This study used the texture characteristics method of histograms and GLCM with the Multi-Layer Perceptron (MLP) classification method. The accuracy obtained from this method is 100%. However, in this case, the image dataset used as a test is only a small number to produce perfect accuracy [8].

Alhudhaif et al. conducted research, namely the classification of Covid-19 Pneumonia. The dataset is 1,218 chest X-ray images of Covid-19 pneumonia and 850 other pneumonia cases. This study used the SVM method, and the accuracy obtained reached 84% [9].

Based on research [6], the accuracy results obtained could be more optimal using more than 4 GLCM features and the K-Nearest Neighbor classification with k=5. Classification using the K-Nearest Neighbor method has the advantage of forming a classification model from training data [10]. For the same case, this study will improve the accuracy of the system in reading CXR image results with the same method using the addition of features in GLCM and K-Nearest Neighbor classification with the right k selection and also an image enhancement process so that the image will be easier to identify. Identified the presence of bacterial pneumonia, viral, and normal lung images by increasing the accuracy of the system in reading the image results from CXR using the addition of features to the GLCM and K-Nearest Neighbor classification with the correct selection of k and also an image enhancement process so that the image will be easier to identify.

## METHOD

The method material used in this study is Chest X-Ray (CXR) image data with the existence of .jpeg, which can be obtained from https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. X-Ray Image Data after the resizing process has a size of 200 x 200 pixels. The data used in this study is 250 image data sets, with 150 training data and 100 test data. This research was conducted by classifying into two classes with 50 bacterial pneumonia images and 50 viral pneumonia images with a test data of 20:20 for each type of image. In the classification, three data classes were used, namely 50 normal lung images, 50 bacterial pneumonia, and 50 viral pneumonia, with test data 20:20:20 for each type of image.

This method uses MATLAB R2018a software. The initial step before performing image processing is image acquisition by preparing a data set of medical images of patients with normal lungs, lung images of viral pneumonia, and bacterial pneumonia. The preprocessing stage is done by resizing to standardize the dimensions of each image and cropping to remove unimportant parts in the area outside the chest cavity or outside the lungs that affect the process.

Then filtering is done on the image using a high pass filter to sharpen the contrast of an image to help get information in the next stage. After that, the image histogram is uniform using the adaptive histogram equalization technique. The next stage is feature extraction to recognize patterns in the image, namely using the Gray Level Co-Occurrence Matrix (GLCM) with angle variations of $0^0$, $45^0$, $90^0$, and $135^0$.

Image textures can be captured using the GLCM matrix using several features. In [11-13], 22 features can be used for feature extraction of GLCM. The equations used from several GLCM features are:

1. Autocorrelation: Autocorrelation is a measure of smoothness and roughness in texture [12].
$$autocorrelation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)ij \quad (1)$$
2. Contrast: Contrast is a measure of local variation in intensity, which supports the value of the diagonal (i=j) [12].
$$contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 p(i,j) \quad (2)$$
3. Correlation 1: Correlation is used to determine how pixels correlate with their surroundings [14].
$$correlation\ 1 = \sum_{i,j} \frac{(i-\mu i)(j-\mu i)p(i,j)}{\sigma_i \sigma_j} \quad (3)$$
4. Correlation 2:
$$correlation\ 2 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)ij - \mu_x \mu_y}{\sigma_x(i)\sigma_y(j)} \quad (4)$$
5. Cluster prominence measures GLCM slope and asymmetry [12].
$$cluster\ prominence = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left((i+j-\mu_x-\mu_y)\right)^4 p(i,j) \quad (5)$$
6. Cluster shade: Cluster shade is a measure of GLCM slant and uniformity [12].
$$cluster\ shade = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i+j-\mu_x-\mu_y)^2 p(i,j) \quad (6)$$
7. Dissimilarity: Dissimilarity is used to show dissimilarity in the image [12].
$$dissimilarity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i-j| p(i,j) \quad (7)$$
8. Energy: Energy gives the number of square elements in the GLCM matrix [12].
$$energy = \sum_{ij} P(i,j)^2 \quad (8)$$
9. Entropy: Entropy is used to calculate the irregularity of the image [12].
$$entropy = -\sum_{i=1}^{N_g} p(i) log_2(p(i)+\epsilon) \quad (9)$$
10. Homogeneity 1: Homogeneity is used to show another measure of the homogeneity of the image [12].
$$homogeneity\ 1 = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \quad (10)$$
11. Homogeneity 2:
$$homogeneity\ 2 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i,j)}{1+|i-j|^2} \quad (11)$$
12. Maximum probability: Maximum probability is the emergence of the most dominant neighbor intensity value pair [12].
$$maximum\ probability = max_{i=1,\dots q; j=1,\dots q}(P(i,j)) \quad (12)$$

13. Sum of squares: the sum of squares or variance measures the distribution of pairs of neighboring intensity levels about the average intensity level in GLCM. [12].

$$sum\ of\ squares = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-\mu_x)^2 p(i,j) \quad (13)$$

14. Sum average: Sum average is the relationship between lower and higher intensity values based on the average number.[12].

$$sum\ average = \sum_{k=2}^{2N_g} P_{x+y}(k)k \quad (14)$$

15. Sum variance: Sum variance is the value of the total color intensity variation in the image [12].

$$sum\ variance = \sum_{k=2}^{2N_g} (k-SA)^2 P_{x+y}(k) \quad (15)$$

16. Sum entropy: Sum entropy is the total of the intensity values of the different neighbors [12].

$$sum\ entropy = \sum_{k=2}^{2N_g} p_{x+y}(k) log_2(p_{x+y}(k)+\epsilon) \quad (16)$$

17. Difference variance: Difference variance is the difference in the size of the color intensity variation in an image [12].

$$difference\ varience = \sum_{k=0}^{N_g-1} \left(k - \sum_{k=0}^{N_g-1} kp_{x-y}(k)\right)^2 P_{x-y}(k) \quad (17)$$

18. Different entropy: Different entropy is the difference in the size of the variability in the intensity value of an image [12].

$$different\ entropy = \sum_{k=0}^{N_g-1} P_{x-y}(k) \times log_2(P_{x-y}(k)+\epsilon) \quad (18)$$

19. Information measure of correlation 1: Information measure of correlation shows the results of an invariant distribution [12].

$$IMC\ 1 = \frac{HXY-HXY1}{max\{HX,HY\}} \quad (19)$$

20. Information measure of correlation 2: Information measure of correlation 2 is the correlation value between the probability distribution i and j (measuring texture complexity) [12].

$$IMC\ 2 = \sqrt{1-e^{-2(HXY2-HXY)}} \quad (20)$$

21. Inverse Difference Normalized (INN): Inverse Difference Normalized (INN) normalizes the difference between neighboring intensity values by dividing the number of discrete intensity values [12].

$$inverse\ difference\ normalized = \sum_{k=0}^{N_g-1} \frac{p_{x-y}(k)}{1+\left(\frac{k}{N_g}\right)} \quad (21)$$

22. Inverse Difference Moment Normalized (IDMN): Inverse Difference Moment Normalized (IDMN) normalizes the difference between neighboring intensity values by dividing the total number of discrete intensity values [12].

$$IDMN = \sum_{k=0}^{N_g-1} \frac{p_{x-y}}{1+\left(\frac{k^2}{N_g^2}\right)} \quad (22)$$

The last step was identifying two classes consisting of bacterial pneumonia images and viral pneumonia, then identifying three classes consisting of normal lung images, bacterial pneumonia lung images, and viral pneumonia lung images using the K-Nearest Neighbor classification.

The K-Nearest Neighbor algorithm is a non-parametric learning model where input prediction is only determined by the closest data label to the original image. This model uses the standard

Euclidean distance to calculate the variance between the training and test data. The k value indicates the number of nearest neighbors that helps predict the test data class from the training data results. The standard Euclidean equation for distance d(x,y) is as follows [15].

$$d(x_i, y_j) = \sqrt{(a_i(x_i) - a_i(x_{j)})^2} \quad (23)$$

The feature extraction results are used to classify using the K-Nearest Neighbor method. So that it can be known the presence of bacterial pneumonia, viral pneumonia, and normal lungs, this classification process is carried out in two stages, namely conducting training on training data, each of which already has a class. This method uses the k parameter to determine the nearest neighbors in an image. The k parameter in KNN greatly influences the classification results. In this study, the k value will be determined from the highest accuracy produced by the training data with variations in the k = 1,3,5 and 7 values.

The results of the highest accuracy obtained from training with variations in the value of k will be used to test the image data set. The K-Nearest Neighbor classification decision results are determined by the most classification of data that enters the k value. The results of the best training process from angle variations in GLCM will be used as a reference for testing the test data that has been determined in the data acquisition stage.

### Data Analysis

The analysis carried out on classification data using K-Nearest Neighbor can be identified by comparing the results of the classification program with the classification by the database. In this study, percentage calculations were carried out to determine the accuracy of the classification results for two classes with the following formulation [16]:

$$Sensitivity = \frac{TP}{TP+FN} \quad (24)$$

$$Spesificity = \frac{TN}{TN+FP} \quad (25)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (26)$$

Where TP (True Positive) is the amount of bacterial pneumonia data classified correctly, TN (True Negative) is the amount of viral pneumonia data classified correctly, FP (False Positive) is the amount of bacterial pneumonia data classified incorrectly, FN (False Negative) is the amount of viral pneumonia data classified incorrectly.

In this study, percentage calculations were carried out to determine the accuracy of the classification results of 3 classes with the following formulation [17].

$$Sensitivity = \frac{TP}{TP+FN} \quad (27)$$

$$Spesificity = \frac{TP}{TP+FP} \quad (28)$$

$$Accuracy = \frac{TP(A)+TP(B)+TP(C)}{TP+FP} \quad (29)$$

TP (True Positive) is the sum of the data of the three classes that are classified as true in the system, FN (False Negative) is the number of classes that are classified as wrong in the correct class, FP (False Positive) is the correct class that is classified as wrong.

## .RESULTS AND DISCUSSION

The stages of image processing before the training and testing process begin with preprocessing, then pattern recognition with feature extraction. Initial processing or preprocessing is carried out in 4 stages: resizing, cropping, filtering using a high pass filter, and Adaptive Histogram Equalization. The differences between the images before and after cropping are shown in Figure 1 for the image before cropping and Figure 2 for the image after cropping. The cropping results show the cropped lung image on the right and left sides. This process is carried out to obtain images of the lungs without any right and left background, affecting the information obtained in the next process.

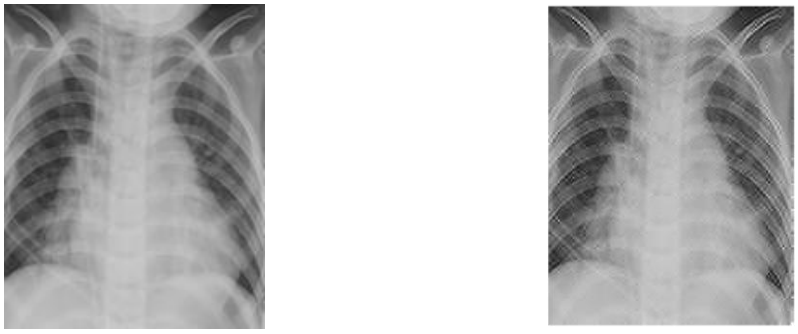

(a) (b) (c)

**Figure 1.** Input image before cropping (a) bacterial pneumonia, (b) viral pneumonia, and (c) normal lungs

(a) (b) (c)

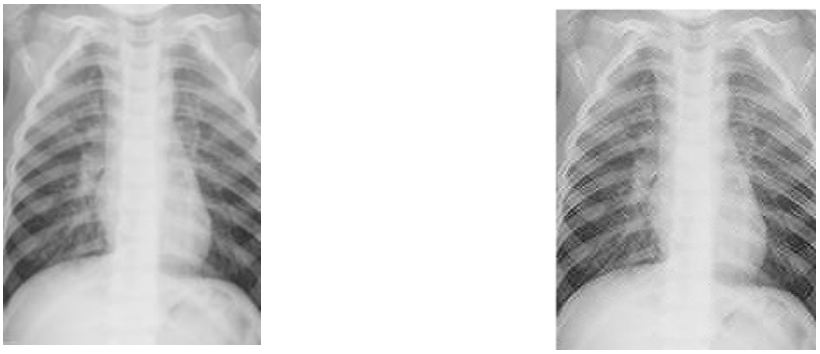
**Figure 2.** Input image after cropping (a) bacterial pneumonia, (b) viral pneumonia, and (c) normal lungs.

The results of the high-pass filtering process can be seen in Figure 3 for images of bacterial pneumonia, Figure 4 for images of viral pneumonia, and Figure 5 for images of normal lungs.



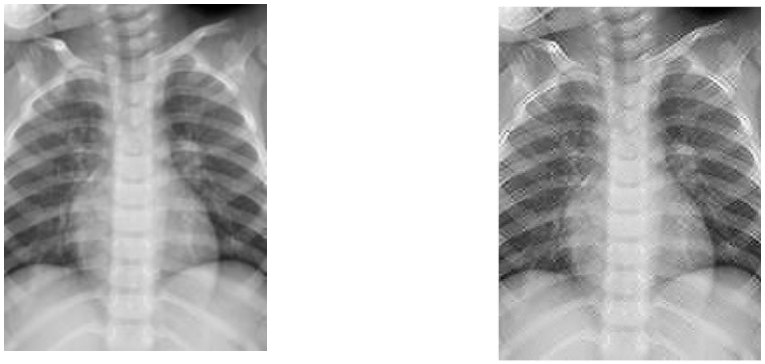
(a) (b)

**Figure 3** (a) Bacterial pneumonia input image, (b) high pass filtering result image.



(a) (b)

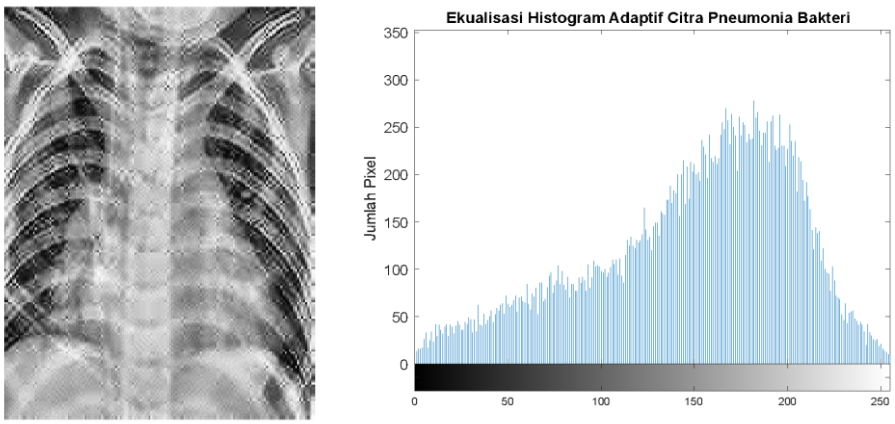
**Figure 4** (a) Pneumonia virus input image, (b) high pass filtering result image.



(a) (b)

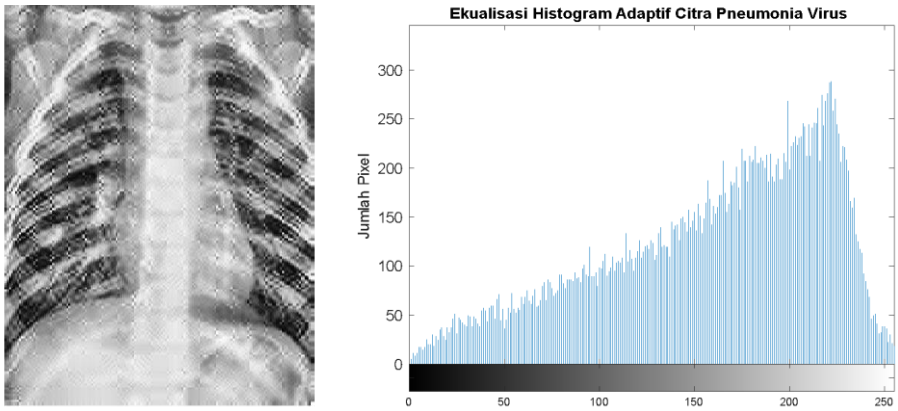
**Figure 5** (a) Normal lung entry image, (b) high pass filtering image.

Improving image quality using a high pass filter produces a sharper image. From the high pass filter process, it can be seen that this filter will filter out the number of pixels in the part of the image that have a high-intensity value and pass pixels with a high-intensity value to the number of neighboring pixels that have a low-intensity value [18]. For the resulting image to have more contrast, image correction is performed using Adaptive Histogram Equalization. It is done to increase the visual effect and intensity of an image. The results of this process can be seen in Figure 6, Figure 7, and Figure 8.



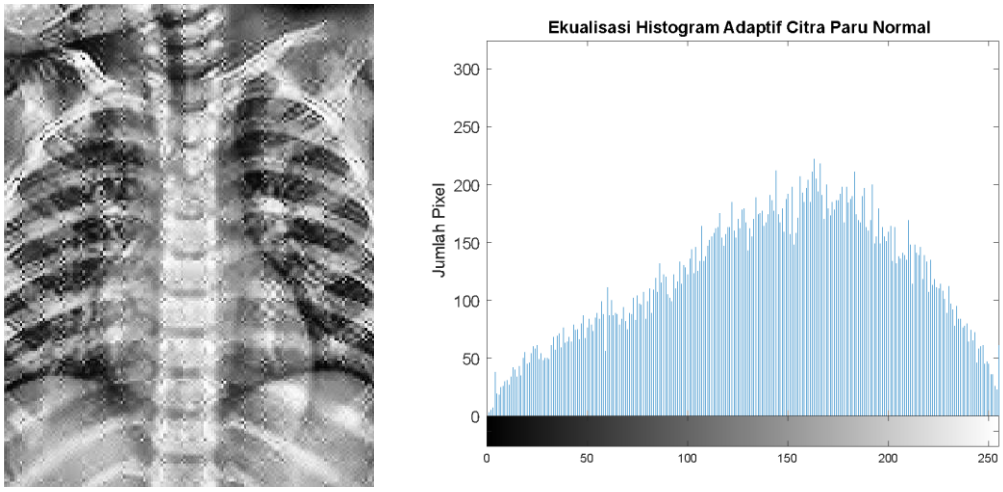

(a) (b)

**Figure 6.** (a) Image of bacterial pneumonia resulting from adaptive histogram equalization process, (b) Histogram of bacterial pneumonia image resulting from adaptive histogram equalization




(a) (b)

**Figure 7.** (a) Adaptive histogram equalized pneumonia virus image, (b) Adaptive histogram equalized viral pneumonia image histogram




(a) (b)

**Figure 8.** (a) Normal lung image from adaptive histogram equalization, (b) Histogram image of normal lung from adaptive histogram equalization

In this process, the algorithm can automatically increase the gray level, and contrast in an image can be limited. Adaptively, this image can adjust the distance of two adjacent gray levels in the new histogram [19].

## GLCM Feature Extraction Results

In this study, the feature extraction used is GLCM feature extraction with 22 features, 1-pixel distance, and variations in the direction of the angles $0^0$, $45^0$, $90^0$, and $135^0$. The results obtained from the GLCM feature extraction using 22 parameters are then averaged for each feature in the overall image with variations in the direction of the angles $0^0$, $45^0$, $90^0$, and $135^0$. GLCM feature extraction was carried out in 2 groups, the first for images of bacterial pneumonia and viral pneumonia and the second for feature extraction of GLCM on images of bacterial pneumonia, viral pneumonia, and normal lungs. The feature extraction results for the first group are the average results for each feature extraction feature for bacterial pneumonia and viral pneumonia, as shown in Table 1. Table 2 shows the average results for each feature extraction for bacterial pneumonia, Table 3 shows the average results for each feature extraction for viral

pneumonia, and Table 4 shows the average results for each feature extraction feature for normal lungs.

**Table 1.** The average results of each GLCM feature extraction for bacterial pneumonia classes and viral pneumonia in the two classes

| GLCM | bacterial pneumonia | | | | viral pneumonia | | | |
|---|---|---|---|---|---|---|---|---|
| | 0° | 45° | 90° | 135° | 0° | 45° | 90° | 135° |
| Autocorrelation | 27.420 | 27.570 | 27.518 | 27.473 | 29.391 | 29.689 | 29.504 | 29.584 |
| Contrast | 1.394 | 1.314 | 1.292 | 1.347 | 1.250 | 1.194 | 1.191 | 1.214 |
| Correlation 1 | 0.757 | 0.770 | 0.777 | 0.767 | 0.792 | 0.799 | 0.802 | 0.797 |
| Correlation 2 | 0.757 | 0.770 | 0.777 | 0.767 | 0.775 | 0.799 | 0.802 | 0.797 |
| Cluster Prominence | 263.285 | 262.943 | 271.318 | 265.096 | 300.273 | 294.716 | 304.186 | 296.368 |
| Cluster Shade | -14.276 | -14.332 | -14.464 | -14.048 | -20.282 | -19.754 | -20.207 | -19.741 |
| Dissimilarity | 0.714 | 0.717 | 0.681 | 0.727 | 0.678 | 0.683 | 0.654 | 0.688 |
| Energy | 0.062 | 0.062 | 0.064 | 0.061 | 0.063 | 0.064 | 0.065 | 0.063 |
| Entropy | 3.226 | 3.231 | 3.198 | 3.247 | 3.201 | 3.204 | 3.182 | 3.215 |
| Homogeneity 1 | 0.718 | 0.714 | 0.729 | 0.711 | 0.725 | 0.722 | 0.735 | 0.721 |
| Homogeneity 2 | 0.702 | 0.696 | 0.713 | 0.693 | 0.710 | 0.706 | 0.721 | 0.704 |
| max probability | 0.141 | 0.143 | 0.143 | 0.140 | 0.149 | 0.154 | 0.151 | 0.151 |
| Varience | 27.959 | 28.100 | 28.041 | 28.025 | 30.061 | 30.159 | 29.974 | 30.067 |
| Sum average | 10.029 | 10.056 | 10.036 | 10.034 | 10.429 | 10.443 | 10.400 | 10.423 |
| Sum variance | 67.617 | 68.031 | 67.787 | 67.707 | 73.898 | 74.179 | 73.512 | 73.806 |
| Sum entropy | 2.477 | 2.475 | 2.482 | 2.480 | 2.492 | 2.488 | 2.497 | 2.493 |
| Difference variance | 1.394 | 1.314 | 1.292 | 1.347 | 1.236 | 1.194 | 1.191 | 1.214 |
| Difference entropy | 1.101 | 1.102 | 1.076 | 1.112 | 1.068 | 1.075 | 1.054 | 1.081 |
| IMC 1 | -0.278 | -0.272 | -0.294 | -0.268 | -0.298 | -0.293 | -0.311 | -0.290 |
| IMC 2 | 0.802 | 0.795 | 0.815 | 0.793 | 0.820 | 0.815 | 0.830 | 0.814 |
| INN | 0.927 | 0.926 | 0.930 | 0.925 | 0.930 | 0.929 | 0.933 | 0.929 |
| IDM | 0.981 | 0.982 | 0.982 | 0.981 | 0.983 | 0.983 | 0.984 | 0.983 |

**Table 2** The average results of each GLCM feature extraction feature for the class of bacterial pneumonia in the three classes

| GLCM | bacterial pneumonia | | | |
|---|---|---|---|---|
| | 0° | 45° | 90° | 135° |
| Autocorrelation | 27.536 | 27.673 | 27.587 | 27.541 |
| Contrast | 1.568 | 1.301 | 1.286 | 1.336 |
| Correlation 1 | 2.368 | 0.770 | 0.775 | 0.767 |
| Correlation 2 | 0.792 | 0.770 | 0.775 | 0.767 |
| Cluster Prominence | 256.176 | 260.425 | 266.926 | 262.602 |
| Cluster Shade | -14.087 | -14.620 | -14.772 | -14.338 |
| Dissimilarity | 0.692 | 0.711 | 0.677 | 0.721 |
| Energy | 0.077 | 0.063 | 0.065 | 0.062 |
| Entropy | 3.172 | 3.218 | 3.187 | 3.234 |
| Homogeneity 1 | 0.724 | 0.716 | 0.730 | 0.714 |
| Homogeneity 2 | 0.703 | 0.698 | 0.715 | 0.695 |
| maximum probability | 0.143 | 0.145 | 0.145 | 0.143 |
| Sum of sqaures: Varience | 27.991 | 28.198 | 28.107 | 28.090 |
| Sum average | 10.037 | 10.079 | 10.055 | 10.051 |

| | | | | |
|---|---|---|---|---|
| Sum variance | 67.752 | 68.406 | 68.061 | 67.973 |
| Sum entropy | 2.473 | 2.469 | 2.475 | 2.474 |
| Difference varience | 1.377 | 1.301 | 1.286 | 1.336 |
| Difference entropy | 1.098 | 1.098 | 1.073 | 1.108 |
| IMC 1 | -0.279 | -0.273 | -0.295 | -0.270 |
| IMC 2 | 0.802 | 0.796 | 0.815 | 0.794 |
| INN | 0.927 | 0.927 | 0.931 | 0.926 |
| IDM | 0.981 | 0.982 | 0.983 | 0.981 |

**Table 3**. The average results of each GLCM feature extraction for viral pneumonia classes in 3 classes

| **GLCM** | **viral pneumonia** | | | |
|---|---|---|---|---|
| | 0° | 45° | 90° | 135° |
| Autocorrelation | 29.608 | 27.673 | 29.504 | 29.584 |
| Contrast | 1.236 | 1.301 | 1.191 | 1.214 |
| Correlation 1 | 0.792 | 0.770 | 0.802 | 0.797 |
| Correlation 2 | 0.792 | 0.770 | 0.802 | 0.797 |
| Cluster Prominence | 297.396 | 260.425 | 304.186 | 296.368 |
| Cluster Shade | -20.282 | -14.620 | -20.207 | -19.741 |
| Dissimilarity | 0.678 | 0.711 | 0.654 | 0.688 |
| Energy | 0.063 | 0.063 | 0.065 | 0.063 |
| Entropy | 3.201 | 3.218 | 3.182 | 3.215 |
| Homogeneity 1 | 0.725 | 0.716 | 0.735 | 0.721 |
| Homogeneity 2 | 0.710 | 0.698 | 0.721 | 0.704 |
| maximum probability | 0.149 | 0.145 | 0.151 | 0.151 |
| Sum of sqaures: Varience | 30.061 | 28.198 | 29.974 | 30.067 |
| Sum average | 10.429 | 10.079 | 10.400 | 10.423 |
| Sum variance | 73.898 | 68.406 | 73.512 | 73.806 |
| Sum entropy | 2.492 | 2.469 | 2.497 | 2.493 |
| Difference varience | 1.236 | 1.301 | 1.191 | 1.214 |
| Difference entropy | 1.068 | 1.098 | 1.054 | 1.081 |
| IMC 1 | -0.298 | -0.273 | -0.311 | -0.290 |
| IMC 2 | 0.820 | 0.796 | 0.830 | 0.814 |
| INN | 0.930 | 0.927 | 0.933 | 0.929 |
| Inverse difference moment normalized | 0.983 | 0.982 | 0.984 | 0.983 |

**Table 4** The average results of each GLCM feature extraction for normal Lung classes in 3 classes

| **GLCM** | **Normal Lung** | | | |
|---|---|---|---|---|
| | 0° | 45° | 90° | 135° |
| Autocorrelation | 26.612 | 26.567 | 26.512 | 26.570 |
| Contrast | 1.912 | 1.966 | 1.978 | 1.958 |
| Correlation 1 | 0.729 | 0.720 | 0.720 | 0.721 |
| Correlation 2 | 0.729 | 0.720 | 0.720 | 0.721 |
| Cluster Prominence | 344.053 | 337.629 | 340.511 | 338.505 |
| Cluster Shade | -12.706 | -12.141 | -11.780 | -12.238 |

| | | | | |
|---|---|---|---|---|
| Dissimilarity | 0.870 | 0.933 | 0.892 | 0.931 |
| Energy | 0.042 | 0.038 | 0.041 | 0.038 |
| Entropy | 3.500 | 3.557 | 3.520 | 3.555 |
| Homogeneity 1 | 0.678 | 0.654 | 0.673 | 0.655 |
| Homogeneity 2 | 0.654 | 0.626 | 0.648 | 0.627 |
| maximum probability | 0.090 | 0.084 | 0.089 | 0.084 |
| Sum of sqaures: Varience | 27.405 | 27.413 | 27.374 | 27.433 |
| Sum average | 9.805 | 9.802 | 9.790 | 9.802 |
| Sum variance | 64.156 | 64.038 | 63.889 | 64.044 |
| Sum entropy | 2.599 | 2.599 | 2.600 | 2.599 |
| Difference varience | 1.912 | 1.966 | 1.978 | 1.958 |
| Difference entropy | 1.235 | 1.271 | 1.255 | 1.270 |
| IMC 1 | -0.238 | -0.208 | -0.228 | -0.209 |
| IMC 2 | 0.780 | 0.749 | 0.771 | 0.749 |
| INN | 0.913 | 0.906 | 0.911 | 0.906 |
| IDM | 0.975 | 0.973 | 0.974 | 0.973 |

The feature extraction results of bacterial pneumonia and viral pneumonia with angle variations of $0^0$, $45^0$, $90^0$, and $135^0$ were used to train in the classification process using the K-Nearest Neighbor method. The same thing was done on the results of image feature extraction for three classes, namely images of bacterial pneumonia, viral pneumonia, and normal lung. The results of this feature extraction are compared with the input values for the classification of two image classes, namely 0 for bacterial pneumonia images and 1 for viral pneumonia images, while the input values used in the 3 class classification are 0 for bacterial pneumonia images, 1 for viral pneumonia images and 2 for normal lung images.

## Training Stage

In the classification training stage using the K-Nearest Neighbor method, the k value approach has been carried out with k = 1,3,5,7. The best k-value approach is used to consider more accurate classification results. Predictions from training data are based on obtaining the highest accuracy by using the value of k in the K-Nearest Neighbor classification.

The most optimal results from training for two classes were obtained at an angle of 135° with a value of k = 3 with the acquisition of accuracy, sensitivity, and specificity of 93%, 96%, and 91%. The results of this training were used as a test on 40 test data sets with 20 images of bacterial pneumonia and 20 images of viral pneumonia. The results of the comparison of angle variations with experimental values of k = 1,3,5,7 obtained from training for two classes can be seen in Table 5

**Table 5** Comparison of angle variations and k values for 2 class training

| Angle | Comparison Angle variation and K-value | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Accuracy (%) | | | | Sensitivity (%) | | | | Specificity (%) | | | |
| | 1 | 3 | 5 | 7 | 1 | 3 | 5 | 7 | 1 | 3 | 5 | 7 |
| 0° | 100 | 92 | 91 | 87 | 100 | 96 | 94 | 88 | 100 | 89 | 89 | 86 |
| 45° | 100 | 92 | 80 | 74 | 100 | 96 | 81 | 80 | 100 | 89 | 79 | 70 |
| 90° | 100 | 91 | 90 | 89 | 100 | 94 | 93 | 93 | 100 | 89 | 85 | 81 |
| 135° | 100 | 93 | 90 | 90 | 100 | 96 | 95 | 93 | 100 | 91 | 86 | 87 |

In training for three classes with angle variations and values of k = 1,3,5,7. The highest accuracy was obtained at k=1. However, this special case shows that the classification is predicted based on the nearest neighbor only, in other words, using the Nearest Neighbor algorithm itself [20]. The most optimal results from training for three classes were obtained at an angle of 135° with a value of k = 3 with the acquisition of accuracy, sensitivity, and specificity of 91%, 92%, and 91%. The results of the comparison of angle variations with experimental values of k = 1,3,5,7 obtained from training for three classes can be seen in Table 6.

**Table 6.** Comparison of angle variations and k values for 3 class training

| Angle | Comparison Angle variation and K-value | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Accuracy (%) | | | | Sensitivity (%) | | | | Specificity (%) | | | |
| | 1 | 3 | 5 | 7 | 1 | 3 | 5 | 7 | 1 | 3 | 5 | 7 |
| 0° | 100 | 90 | 83 | 79 | 100 | 91 | 86 | 82 | 100 | 91 | 87 | 82 |
| 45° | 99 | 89 | 85 | 85 | 98 | 89 | 87 | 85 | 99 | 89 | 86 | 85 |
| 90° | 99 | 86 | 84 | 83 | 99 | 89 | 84 | 84 | 99 | 89 | 84 | 84 |
| 135° | 99 | 91 | 86 | 85 | 98 | 92 | 86 | 86 | 98 | 91 | 86 | 85 |

## Testing Stage

The testing process for test data in the three-class classification of 60 sets of image data entered into the program resulted in an accuracy of 83.3%, a specificity of 83.5%, and a sensitivity of 83.3%. The results of the confusion matrix for two classes and three classes can be seen in Table 7 and Table 8

**Table 7.** Confusion matrix for two classes

| | | bacterial pneumonia | viral pneumonia |
|---|---|---|---|
| Predicted Class | bacterial pneumonia | 17 | 1 |
| | viral pneumonia | 3 | 19 |

**Table 8** Confusion matrix for three classes

| | bacterial pneumonia | viral pneumonia | Normal |
|---|---|---|---|
| bacterial pneumonia | 14 | 1 | 2 |
| viral pneumonia | 2 | 18 | 0 |
| Normal | 4 | 1 | 18 |

Based on the comparison results shown in Table 9, it can be shown that the proposed program can distinguish between two classes, namely bacterial pneumonia and viral pneumonia, and for three classes, namely bacterial pneumonia, viral pneumonia, and normal lungs. It can also be due to adding features to the GLCM and using the most appropriate angle when conducting training. Adding many features is proven to increase accuracy in testing images.

**Table 9** Comparison of research program performance

| No | Literatur | Data | Method | Accuracy |
|---|---|---|---|---|
| 1 | [6] | 5863 | *Contrast Stresching*, GLCM (4 fiture), KNN | 66.20 % |
| 2 | [8] | 9 | GLCM and *Multi Layer Perceptron* (MLP) | 100 % |
| 3 | [9] | 1218 | SVM | 84 % |
| 4 | [7] | 69 | *Extract high temperature region, feature selection*, classification SVM | 93 % |
| 5 | This research | 250 | *High pass filter, Histogram equalization adaptive*, GLCM (22 fiture), KNN | bacterial pneumonia and viral pneumonia, training accuracy : 93%, testing accuracy : 90 %<br><br>bacterial pneumonia dan viral pneumonia, normal, training accuracy: 91,33 % testing accuracy: 83,3 % |

## CONCLUSION

The K-Nearest Neighbor algorithm and the addition of 22 GLCM features can be applied to the pneumonia classification using image data of bacterial pneumonia, viral pneumonia, and normal lungs as input. The testing system with 22 GLCM features at an angle of 135o and KNN classification k=3 resulted in optimal accuracy in training for two classes, namely bacterial pneumonia and viral pneumonia, by 93% and training for three classes, namely bacterial pneumonia, viral pneumonia, and normal lung by 91.33%. The accuracy obtained from two classes is 90% while testing for three classes is 83.3%.

## REFERENCES

1 Kermany, D. S., Goldbaum, M., Cai., Valentim, C. C. S., Liang, H., Baxter, S.L., Zhang, K. 2018. Identifying Medical Diagnosis and Treatable Diseases by Image-Based Deep Learning. *Cell*, 172(5), 1122-1131.

2 Maysanjaya, I. M. D. 2020. Klasifikasi Pneumonia pada Citra X-rays Paru-paru dengan Convolutional Neural Network. *Jurnal Nasional Teknik Elektro dan teknologi Informasi*, 9(2), 190-195.

3 Meng, X. 2018. Digital Image Processing Technology Based on MATLAB. *Proceedings of the 4th International Conference on Virtual Reality*, 79-82.

4 Pratiwi, E. H., & Juniati, D. 2022. Clustering Penyakit Paru-Paru Berdasarkan Rontgen Dada Menggunakan Dimensi Fraktal Box Counting Dan K-Medoids. *Jurnal Riset dan Aplikasi Matematika (JRAM)*, *6*(1), 1-12.

5 El-Dahshan, E. S. A., Bassiouni, M. M., Hagag, A., Chakrabortty, R. K., Loh, H., & Acharya, U. R. 2022. RESCOVIDTCNnet: A residual neural network-based framework for COVID-19 detection using TCN and EWT with chest X-ray images. *Expert Systems with Applications*, 117410.
6 Wijaya, I. W. A., & Kusumadewi, A. 2015. Penerapan Algoritma K-Means Pada Kompresi Adaptif Citra *Medis MRI. Informatika*, 11(2).
7 Qu, Y., Meng, Y., Fan, H., & Xu, R. X. 2022. Low-cost thermal imaging with machine learning for non-invasive diagnosis and therapeutic monitoring of pneumonia. *Infrared Physics & Technology*, 104201.
8 Istianah, L., & Sumarti, H. 2020. Classification of Pneumonia in Thoracic X-Ray images based on texture characteristics using the MLP (Multi-Layer Perceptron) method. *Journal Of Natural Sciences And Mathematics Research*, *6*(2), 78-84.
9 Alhudhaif, A., Polat, K., & Karaman, O. 2021. Determination of COVID-19 pneumonia based on generalized convolutional neural network model from chest X-ray images, .*Expert Systems with Applications*,180, 115141.
10 Putra, P., Pardede, A. M., & Syahputra, S. 2022. Analisis Metode K-Nearest Neighbour (Knn) Dalam Klasifikasi Data Iris Bunga, *Jtik (Jurnal Teknik Informatika Kaputama)*, *6*(1), 297-305.
11 Soh, L., & Tsatsoulis, C. 1999. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices, *IEEE Transactions on Geoscience and Remote Sensing*, 37(2).
12 Haralick, R. M., Shanmugam, K., and Dinstein, I. 1973. Textural Features of Image Classification. *IEEE Transactions on Systems, Man and Cybernetics* (vol. SMC-3, no. 6).
13 Clausi, D. A. 2002. An analysis of co-occurrence texture statistics as a function of grey level quantization, Can. J. *Remote Sensing*, 28(1), 45-62.
14 Kumar, D. 2020. Feature extraction and selection of kidney ultrasound images using GLCM and PCA. *Procedia Computer Science*, 167, 1722-1731.
15 Hasoon, J. N., Fadel, A. H., Hameed, R. S., Mostafa, S. A., Khalaf, B. A., Mohammed, M. A., & Nedoma, J. 2021. COVID-19 anomaly detection and classification method based on supervised machine learning of chest X-ray images. *Results in Physics*, *31*, 105045.
16 Haukat, F., Raja, G., Ashraf, R., Khalid, S., Ahmad, M., & Ali, A. 2019. Artificial Neural Network Based Classification of Lung Nodules in CT Images Using Intensity, Shape and Texture Features. *Journal of Ambient Intelligence and Humanized Computing*, *10*(10), 4135-4149.
17 Hidayah, N., & Sahibu, S. 2021. Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *5*(4), 820-826.
18 Mulyana, T. M. S. 2017. Efek high pass filtering dengan koefesien nol pada citra biner. *Jurnal Muara Sains, Teknologi, Kedokteran dan Ilmu Kesehatan*, *1*(1), 75-83.
19 Zhu, Y., & Huang, C. 2012. An adaptive histogram equalization algorithm on the image gray level mapping. *Physics Procedia*, *25*, 601-608.
20 Lubis, Z. 2019. Optimasi Nilai K pada Algoritma K-NN dalam Clustering Menggunakan Algoritma Expectation Maximization. (Doctoral dissertation, Universitas Sumatera Utara).