Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.4, August 2025, pp. 1406~1414
e-ISSN: 2548-6861

Application of Linear Discriminant Analysis Method With Gray Level Cooccurrence Matrix Method for Classification of Lung Disease Diagnosis Based on X-Ray Results

Nuriana [1]*, Zahratul Fitri [2]*, Ar Razi [3]*
* Teknik Informatika, Universitas Malikussaleh
nuriana.210170089@mhs.unimal.ac.id [1], zahratulfitri@unimal.ac.id mailto:mail2@polibatam.ac.id[2], ar.razi@unimal.ac.id [3]

Article Info

Article history:

Received 2025-06-22
Revised 2025-07-01
Accepted 2025-07-03

Keyword:

Lung Disease,
X-ray,
GLCM,
Linear Discriminant Analysis,
Texture Classification.

ABSTRACT

This study aims to classify lung diseases from X-ray images using a combination of Gray Level Cooccurrence Matrix (GLCM) and Linear Discriminant Analysis (LDA) methods. GLCM was used to extract texture features such as contrast, correlation, energy, and homogeneity from 300 lung X-ray images representing four categories: Normal, Pneumonia, Tuberculosis, and Bronchitis. The LDA method was then applied for classification based on these features. The results showed that Tuberculosis had the highest classification accuracy at 80%, while the overall model accuracy was 61.67%. Evaluation using precision, recall, F1-score, and confusion matrix confirmed that the GLCM and LDA combination performed best in identifying tuberculosis cases. However, overlapping features between Normal, Bronchitis, and Pneumonia classes reduced the classification performance. These findings suggest that the proposed method provides promising results and could be improved further with advanced feature extraction or classification techniques.

I. INTRODUCTION

The lungs are one of the vital organs in the human respiratory system. The lungs have an important role in the gas exchange process, namely taking oxygen from the air and removing carbon dioxide from the human body [1]. The lungs are located in the chest cavity and are protected by the ribs, the lungs are separated by the diaphragm from the abdominal cavity [2]. It is very important to keep the lungs healthy because healthy lungs can support optimal body function. Factors that can make the lungs unhealthy include smoking, air pollution, and exposure to chemicals [3]. Several of these factors can cause several lung diseases such as asthma, bronchitis, pneumonia, tuberculosis, and even lung cancer which can disrupt respiratory function and reduce quality of life [4].

Lung disease is one of the main causes of human death in the world, both in developed and developing countries, and has many impacts on world health and the world economy [5]. Lung diseases such as pneumonia, tuberculosis, and bronchitis are common lung diseases and require immediate detection and treatment to increase the patient's chances of recovery [6]. Based on data from the Indonesian Ministry of Health, the number of Tuberculosis cases in Indonesia has fluctuated since 2013. Starting from 327,103 cases in 2013, it decreased in 2014 to 324,539, then increased slowly until 2019. In 2020, cases decreased sharply to 393,323, possibly due to the pandemic. However, the number rose again in 2021 to 443,235 and jumped drastically in 2022 to 717,941 cases, an increase of around 61.98% from the previous year [7].

The use of artificial intelligence-based technology and image processing in the medical world shows great potential in supporting the diagnosis of diseases [8]. One of the most common ways used to detect lung conditions is with X-ray images [9]. One of the commonly used methods is the Gray Level Co-occurrence Matrix (GLCM) method. GLCM is used to extract texture features from X-ray images [10]. GLCM is used to identify complex texture patterns, patterns that are difficult to interpret manually by humans, the extracted features can provide an in-depth representation of image characteristics that are relevant for diagnosing disease [11]. In addition to the GLCM method, this study also uses

the Linear Discriminant Analysis (LDA) method. The LDA method is an effective statistical method for image classification, this method is also relatively simple and fast to use on small to medium datasets [12].

Research conducted by [13], This study proposes an approach to potato leaf disease classification by integrating three types of image features, namely color moments, texture using Gray-Level Co-occurrence Matrix (GLCM), and Renyi fractal dimension. The classification method used is Linear Discriminant Analysis (LDA), which is applied to classify potato leaves into three classes: healthy, early blight, and late blight. The dataset used consists of 456 images with a size of 256×256 pixels taken from the Kaggle platform. The results showed that when the features were tested separately, the color moment feature gave the highest accuracy of 97.58%, while the GLCM feature only achieved 60.75% and the Renyi fractal dimension was 62.07%. However, when the GLCM feature was combined with the color moment, the accuracy increased significantly to 98.03%. The highest accuracy of 98.46% was achieved when all three types of features were combined and classified using LDA. These results also show that the LDA method is able to outperform other classification methods such as Support Vector Machine (91.88%), K-Nearest Neighbor (84.44%), and Random Forest (92.34%). This study confirms that the integration of color, texture, and fractal features with LDA significantly improves the accuracy in potato leaf disease classification compared to using single features or classification methods.

Therefore, the purpose of this study is to implement the GLCM method in identifying lung disease detection patterns based on X-ray image results and also to implement the LDA method as a classification for grouping types of lung disease. The combination of GLCM and LDA methods is expected to provide a promising approach to improve the accuracy of lung disease diagnosis based on X-ray images.

Several machine learning algorithms such as Naïve Bayes and Support Vector Machine (SVM) have been widely applied in various classification tasks across different domains. For instance, Naïve Bayes has been successfully used to classify aquaculture survival rates based on environmental parameters with high accuracy [14], while SVM has shown its usefulness in pattern recognition problems such as traditional clothing image grouping based on color features [15]. These studies suggest the versatility of classification algorithms in different contexts, supporting the use of Linear Discriminant Analysis (LDA) in this research for medical image classification.

II. METHOD

This research was conducted through several structured stages to achieve the main objective, namely to build a lung disease diagnosis classification system based on X-ray images using the GLCM and LDA methods. The following are the stages carried out in the research which can be seen through the following flow diagram figure 1:

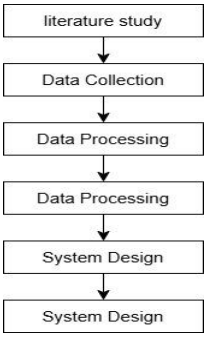

Figure 1. Research Flowchart

A. System Scheme

The following system scheme is a scheme for diagnosing lung disease based on X-ray results using the Linear Discriminant Analysis (LDA) and Gray Level Co-occurrence Matrix (GLCM) methods, which are an approach that combines texture feature extraction techniques from X-ray images and statistical classification methods to distinguish between normal and abnormal lung conditions. The system scheme can be seen in the figure 2 below:

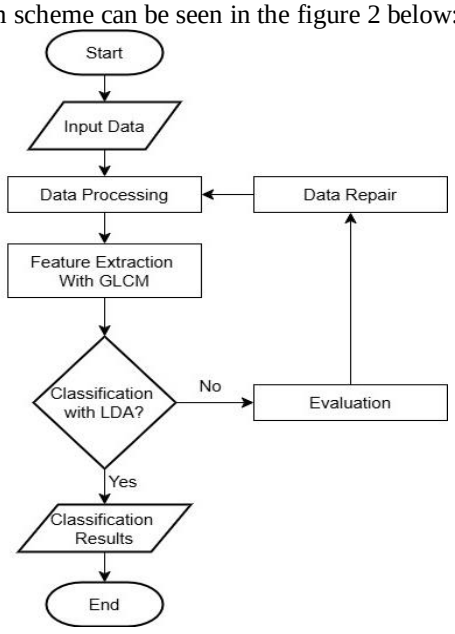

Figure 2. System Scheme

B. Gray Level Co-occurrence Matrix (GLCM)

Gray Level Co-occurrence Matrix (GLCM) is one of the methods used in image processing to extract texture features from an image [16]. GLCM is a frequency matrix that describes how often a pixel with a certain gray level appears together at a certain distance and direction [17]. This matrix is used to calculate the spatial relationship between pixels in a grayscale image. GLCM considers the spatial relationship between pixels where this relationship is determined by the distance (d) and direction (θ). Distance (d) determines how far two pixels are considered a pair, while direction (θ) is the

direction of the pixel pair expressed in degrees with angles of 0º, 90º, 45º, 135º [18].

In this study, the Gray Level Co-occurrence Matrix (GLCM) method was used to extract texture features from lung X-ray images. The GLCM was computed using the following parameters:

1. Pixel distance: 1 pixel
2. Directions (angles): 0°, 45°, 90°, and 135°
3. Gray level quantization: 256 levels
4. Mode: symmetric and normalized

Four main texture features were calculated in each direction: contrast, correlation, energy, and homogeneity, resulting in a total of 16 features per image (4 directions × 4 features).

The Gray Level Co-occurrence Matrix (GLCM) method has steps to build GLCM by creating a matrix that records the frequency of gray level pairs (i,j) according to the specified distance and direction. Matrix normalization to convert frequency values to probabilities by dividing each element by the total frequency [14]. The formula is as follows:

$$P(i,j) = \frac{M(i,j)}{\text{Total Pairs}}$$

Extracting texture features contained in the GLCM method such as [19]:

a. Contrast is used to measure the difference in intensity between a pixel and its neighbors.
$$\sum_{i,j} P(i,j)\,(i,j)^2$$
b. Correlation is used to measure the linear relationship between pixels.
$$\frac{\sum_{i,j} P(i,j)\,(i - \mu_i)\,(j - \mu_i)}{\sigma_i \sigma_j}$$
c. Energy that functions to measure the compactness or homogeneity of the matrix.
$$Energy = \sum_{i,j} P(i,j)^2$$
d. Homogeneity, which functions to measure the uniformity of value distribution.
$$Homogeneity = \sum_{i,j} \frac{P\,(i,j)}{1 + |i - j|}$$

Explanation:

i,j	:	The intensity values (gray-level) of two neighboring pixels (e.g.: 0, 1, 2, ..., 255)
P(i,j)	:	The probability of occurrence of a pair of values *i* and *j* in the GLCM matrix
∑i,j	:	Double addition over all pairs (*i*, *j*) in the matrix
μi	:	The average (mean) of the row intensity values (*i*) in GLCM
μj	:	The average (mean) of the intensity values of column (*j*) in GLCM
σi	:	Standard deviation of the row intensity values (*i*)
σj	:	Standard deviation of the row intensity values (*j*)
$M_{(i,j)}$	:	Frequency of occurrence of pixel value pairs i and j in GLCM
Total Pairs	:	The total number of elements in the GLCM matrix (total number of pixel pairs)

C. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a statistical method used to reduce data dimensions and optimally separate classes based on available features [20]. LDA is often used in classification and pattern analysis, as well as to understand how input variables affect the separation between classes. The goal of the LDA method is to classify objects into a number of classes according to the described characteristics. [21].

The working steps of the Linear Discriminant Analysis (LDA) method are as follows:

1. Data Preparation
   a. The dataset in the form of a feature vector $X = \{x_1, x_2, \ldots, x_m\}$ with each sample x_i given a class label C_k.
   b. Ensure that the dataset has a distribution that meets the LDA assumptions, namely a Gaussian distribution with equal covariance between classes.
2. Calculate the average for each class:
$$\mu_k = \frac{1}{N_k} \sum_{i \in C_k} x_i$$
3. Calculate the scatter matrix between classes (S_B) Between-Class Scatter which measures the spread of class averages against the global average:
$$S_B = \sum_{k=1}^{K} N_K (\mu_k - \mu)(\mu_k - \mu)^T$$
4. Calculate the scatter matrix in class (S_w) Within-Class Scatter which measures the spread of data within each class:
$$S_w = \sum_{k=1}^{K} \sum_{x \in C_k} (x_i - \mu_k)(x_i - \mu_k)^T$$
5. LDA finds the eigenvectors of the combined scatter matrix:
$$S_W^{-1} S_B$$
6. The data is transformed to a new dimension using the best eigenvectors:
$$Z = X \cdot W$$

Explanation:

$X=\{x1, x2, ..., xm\}$	:	Dataset with m feature vectors
x_i	:	Feature vector from sample to-*i*
C_k	:	Class to-*k*
N_k	:	Number of samples in class -*k*
K	:	Total number of classes
μk	:	The average of the feature vector (mean) in the class-*k*
μ	:	Global average of all data (total mean)
T	:	Transpose of matrix/vector
W	:	Eigenvectors resulting from LDA optimization (projection matrix to new dimensions)
Z	:	Projected data to new dimensions (feature reduction results)

D. Lung Disease

In the research conducted, lung diseases will be classified here into only 3 diseases as follows:

1. Pneumonia, A common acute respiratory infection is pneumonia. This infection affects the alveoli and distal bronchial tree of the lungs. It is caused by viruses, bacteria, fungi, or a combination of them, and causes inflammation and fluid accumulation in the lung parenchyma [4]. The results of the X-ray image for Pneumonia can be seen in figure 3 below.

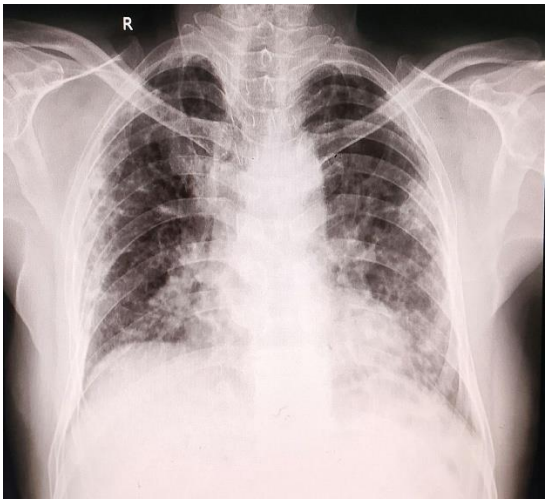

Figure 3. Pneumonia

2. Tuberculosis, is caused by the bacteria mycobacterium tuberculosis and is a long-term or chronic infection. This disease usually attacks the lungs regardless of age or gender. This bacillus bacteria is so strong that it takes a long time to treat it. This bacteria infects the lungs more often than other parts of the human body, so the most common tuberculosis in Indonesia is pulmonary tuberculosis [22]. The results of the X-ray image for Tuberculosis can be seen in figure 4 below.

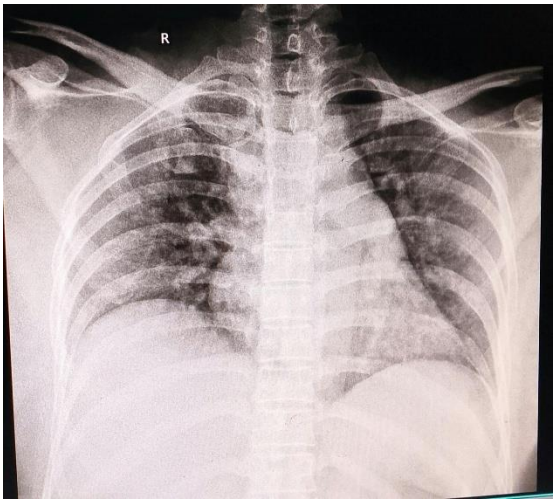

Figure 4. Tuberculosis

3. Bronchitis, is a form of Chronic Obstructive Pulmonary Disease (COPD) characterized by a productive cough for more than 3 months in 2 years. This condition involves inflammation of the bronchial tubes that connect the trachea to the lungs. Early symptoms of bronchitis often resemble an upper respiratory infection, such as a runny nose, sore throat, muscle aches, and a low-grade fever that usually does not exceed 39°C and resolves within 3–5 days. Fever is more common if the cause is a virus such as adenovirus, influenza, or Mycoplasma pneumoniae [23]. Acute bronchitis has symptoms such as a cough with phlegm accompanied by shortness of breath or chest pain when coughing, nasal congestion, headaches and persistent fever and can last for 2 weeks. Acute bronchitis can be caused by exposure to hazardous substances and cigarette smoke, as well as being influenced by allergies, a family history of lung disease, recurrent lower respiratory tract infections, and frequent exposure to pollution and chemicals in the form of smoke [24]. The results of the X-ray image for Bronchitis can be seen in figure 5 below.

Figure 5. Bronchitis

To ensure the validity and generalization of the model, the data is divided into two parts with an 80:20 ratio for training and testing. Additionally, model evaluation is performed using 5-fold cross-validation, accompanied by evaluation metrics such as confusion matrix, precision, recall, and F1-score to avoid overfitting and provide a more comprehensive view of performance.

III. RESULT AND DISCUSSION

The results of this study will implement the application of GLCM and LDA methods to classify lung diseases in Pidie Jaya Hospital based on X-ray results. Application of the GLCM method to analyze the characteristics of lung tissue texture by extracting relevant texture features such as contrast, correlation, energy and homogeneity. Application of the LDA method to classify the extraction results from GLCM so that the LDA method can separate and classify various types of lung diseases.

A. Preprocessing

The preprocessing stage is a very important initial step in the lung disease classification system to prepare X-ray images so that the data used has good value. The preprocessing process is carried out to improve image quality, eliminate noise, and standardize image formats so that they can be processed optimally by the feature extraction algorithm. The following are the stages of preprocessing:

1. Image Resizing
   The image resizing process is carried out to equalize the size of all X-ray images to 256x256 pixels. Images from Pidie Jaya Hospital have varying sizes due to differences in tools and settings when taking them. The bilinear interpolation method is used to maintain image quality and minimize distortion.
2. Grayscale Conversion
   RGB image conversion to grayscale is done because diagnostic information in X-ray images is in grayscale intensity. This process reduces data from 3 channels to 1 channel, thus speeding up the computation. The conversion uses a standard formula with different weights for each channel according to human visual perception.
3. Noise Reduction
   Noise reduction is performed using a 3x3 Gaussian filter to remove interference from various sources such as equipment or the digitizing process. This filter was chosen because it effectively reduces noise without removing important details and image edges.
4. Intensity Normalization
   Pixel intensity normalization converts image pixel values to the range 0–1 by dividing them by 255. The goal is to equalize the data scale, avoid bias between images, speed up convergence during training, and improve numerical stability.

The preprocessing pipeline did not include thresholding, cropping of the lung region, or image enhancement techniques. As such, some background noise or irrelevant regions may still be present in the input, which could affect feature extraction and classification performance.

B. Gray Level Co-occurrence Matrix

To learnings the texture characteristics of 300 preprocessed lung X-ray images, the Gray Level Co-occurrence Matrix (GLCM) method was used for feature extraction. The dataset consists of a balanced distribution with 75 images for each category (Bronchitis, Pneumonia, Tuberculosis, and Normal). The optimized parameters include 1 pixel spacing, four directions (0°, 45°, 90°, 135°), and 256 gray level quantization, with symmetry and normal settings enabled. In this process, 16 texture features (4 features × 4 directions) were generated for each image, which were used as input for the classification stage. The following table shows the complete descriptive statistics for the overall GLCM features extracted from 300 lung X-ray image samples with key textures such as contrast, correlation, energy, and homogeneity:

TABLE I
GLCM FEATURE STATISTICS BASED O N DIRECTION

Fiture	Direction	Min	Max	Mean	Standard Deviation
Contrast	0°	4,54	327,95	79,07	51,16
	45°	9,71	399,87	121	68,34
	90°	5,78	208,33	51,65	40,64
	135°	10,2	386,81	122,92	68,72
Correlation	0°	0,936	0,999	0,991	0,006
	45°	0,926	0,998	0,985	0,008
	90°	0,972	0,999	0,994	0,005
	135°	0,923	0,997	0,985	0,008
Energy	0°	0,014	0,185	0,042	0,023
	45°	0,013	0,172	0,039	0,022
	90°	0,018	0,177	0,045	0,023
	135°	0,013	0,171	0,039	0,022
Homogeneity	0°	0,161	0,623	0,383	0,091
	45°	0,135	0,545	0,331	0,086
	90°	0,205	0,621	0,421	0,092
	135°	0,133	0,538	0,331	0,086

The table I above shows that the contrast feature shows the highest variability, with a standard deviation of 68.72 in the 135° direction, indicating significant texture differences between samples. The correlation feature also shows high consistency, with an average value above 0.98 for all directions, indicating that the texture structure in the lung X-ray images is relatively regular. The 90-degree direction consistently shows more stable energy and homogeneous values.

The following table shows the characteristics of the GLCM features for each disease class in the 0° direction. This shows the texture differences between diagnosis categories:

TABLE II
GLCM FEATURE CHARACTERISTICS PER CLASS (DIRECTION 0°)

Class	Contrast	Correlation	Energy	Homogeneity
Normal	191,48	0,984	0,071	0,494
Tuberculosis	178,18	0,985	0,044	0,399
Bronchitis	143,9	0,987	0,051	0,397
Pneumonia	84,04	0,991	0,028	0,366

The table II above, it can be seen that the Normal class has the highest contrast value (191.48) and the highest energy (0.071), indicating a healthy lung structure with normal texture variations. In contrast, the Pneumonia class has the lowest contrast value (84.04) but the highest correlation value (0.991) indicating a more homogeneous texture due to the consolidation process. Bronchitis and tuberculosis show intermediate features, which can be distinguished by combining all of these features.

C. Linear Discriminant Analysis (LDA)

The implementation of the Linear Discriminant Analysis (LDA) method was carried out to classify lung diseases based on 16 GLCM features that have been extracted from 300 X-ray images. The classification process uses data division with a ratio of 80:20 for training data and test data, resulting in 240 samples for training and 60 samples for testing.

TABLE III
CLASS STATISTICS IN DATASET

Class	Number of samples	Average variance
Bronchitis	60	0,8322
Normal	60	0,7226
Pneumonia	60	1,374
Tuberculosis	60	0,764

From Table III, it can be seen that each class has a balanced sample distribution (60 samples per class), but with different variances. The Pneumonia class shows the highest variance (1.3740), indicating high feature variability within the class. In contrast, the normal class has the lowest variance (0.7226), indicating good consistency of feature characteristics for healthy lung conditions.

TABLE IV
SUMMARY OF PREDICTION ACCURACY AND DISTRIBUTION PER CLASS

Class	Total Sample	Correct Prediction	Wrong Prediction	Accuracy	Incorrect Distribution
Bronchitis	15	10	5	66,70%	3→Normal, 2→Pneumonia
Normal	15	7	8	46,70%	6→Bronchitis, 2→ Others
Pneumonia	15	8	7	53,30%	4→Tuberculosis, 3→Others
Tuberculosis	15	12	3	80%	2→ Pneumonia, 1→ Normal

From Table IV, it can be seen that Tuberculosis has the highest accuracy (80.0%), followed by Bronchitis (66.7%), Pneumonia (53.3%), and Normal (46.7%). The misclassification pattern shows that Bronchitis is often predicted as Normal, Pneumonia is often predicted as Tuberculosis, and Normal is often predicted as Bronchitis. Scatter matrix analysis shows the ability of LDA to separate lung disease classes. Within-class Scatter (Sw) measures variability within a class, while Between-class Scatter (Sb) measures separation between classes. LDA maximizes the Sb/Sw ratio for optimal separation, with the results showing that this method successfully creates an effective discriminant space to distinguish the four disease categories.

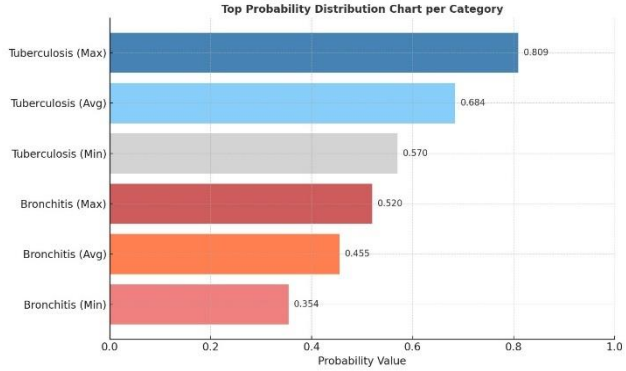

Figure 6. Highest Probability Distribution per Category

From Figure 6 above, it can be seen that the Tuberculosis class shows a higher and more consistent probability distribution than Bronchitis. Tuberculosis has a maximum probability of 0.809 and a minimum of 0.570, indicating relatively high confidence. In contrast, Bronchitis has a lower probability range (0.354-0.520), indicating higher uncertainty in the classification.

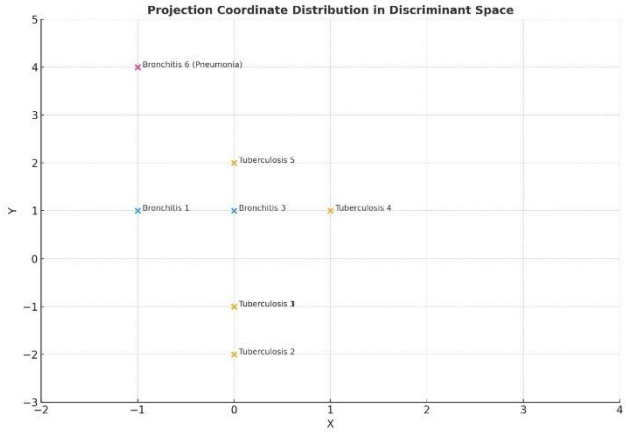

Figure 7. Distribution of Projection Coordinates in Discriminant Space

From Figure 7, it can be seen that Tuberculosis samples tend to be concentrated in the right quadrant (positive X), indicating good separability from other classes. Bronchitis samples show a wider distribution, with some misclassified samples in areas that overlap with other classes. This projection confirms that LDA has succeeded in creating a discriminant space that provides relatively good separation, although there is still overlap between certain classes.

D. Model Evaluation

The LDA model evaluation was conducted using various metrics to comprehensively measure the performance of lung disease classification. The evaluation included confusion matrix analysis, precision metrics, recall, F1-score, and cross-validation to ensure model robustness. A test dataset consisting of 60 samples (15 samples per class) was used to produce an objective evaluation of the model's ability to classify the four disease categories.

1. Confusion Matrix LDA Classification

The confusion matrix provides a detailed overview of the classification performance for each class and the error patterns that occur.

TABLE V
CONFUSION MATRIX LDA CLASSIFICATION RESULTS

Actual/Predicted	Bronkitis	Normal	Pneumonia	Tuberkulosis
Bronkitis	10	2	3	0
Normal	6	7	1	1
Pneumonia	1	0	8	6
Tuberkulosis	0	1	2	12

From Table V, it can be seen that the model shows the best performance for the Tuberculosis class with 12 correct predictions from 15 samples. The Bronchitis class has 10 correct predictions but is often predicted as Normal (2 cases) and Pneumonia (3 cases). The Normal class shows the greatest confusion with 6 samples predicted as Bronchitis, indicating significant overlap in characteristics. Pneumonia has moderate performance with 8 correct predictions, but 6 samples are predicted as Tuberculosis, indicating similarity in texture patterns between the two diseases.

In medical diagnosis, the clinical consequences of classification errors are critical. Based on the confusion matrix results, several important observations can be made. The model achieved 12 correct classifications out of 15 for Tuberculosis, but 3 samples were misclassified, which could include false negatives (e.g., Tuberculosis classified as Normal or other diseases). False negatives in TB detection are particularly dangerous, as they can lead to undiagnosed cases and potential spread of infection. Conversely, false positives—such as Normal cases misclassified as Tuberculosis—may cause unnecessary anxiety, treatment, and resource usage. Thus, reducing false negatives for TB should be prioritized in future improvements, even if it means tolerating slightly higher false positives. This emphasizes the need for sensitivity-optimized models in clinical applications.

2. Metric Analysis

Detailed evaluation using TP, FP, FN, and TN metrics provides an in-depth understanding of the error characteristics of each class.

TABLE VI
EVALUATION METRICS ANALYSIS PER CLASS

Class	TP	FP	FN	TN	Precision	Recall	Specificity	F1-Score
Bronchitis	10	7	5	38	0,5882	0,6667	0,8444	0,625
Normal	7	6	8	42	0,7	0,4667	0,9333	0,56
Pneumonia	8	7	7	39	0,5714	0,5333	0,8667	0,5517
Tuberculosis	12	3	3	38	0,6316	0,8	0,8444	0,7059
Overall	37	23	23	157				

From Table VI, it can be seen that Tuberculosis has the highest recall (0.8000) and the best F1-score (0.7059), indicating a very good model ability in detecting tuberculosis cases. Normal has the highest precision (0.7000) and the best specificity (0.9333), indicating a low false positive rate. Bronchitis shows good recall (0.6667) but moderate precision (0.5882) due to overlap with other classes. Pneumonia has a balanced but relatively low performance with an F1-score of 0.5517.

3. Overall Accuracy and Cross Validation

The LDA model achieved an overall accuracy of 61.67% (37 correct predictions out of 60 total samples) on the test dataset. Cross-validation was performed using 5-fold cross-validation to ensure model robustness and avoid overfitting.

TABLE VII
CROSS VALIDATION RESULTS (5-FOLD CROSS-VALIDATION)

Fold	Accuration	Precision	Recall	F1-Score
Fold 1	51,70%	0,55	0,52	0,53
Fold 2	43,30%	0,48	0,43	0,45
Fold 3	43,30%	0,49	0,43	0,46
Fold 4	40,00%	0,45	0,4	0,42
Fold 5	45,00%	0,51	0,45	0,48
Average	44,67%	0,5	0,45	0,47

From Table VII, it can be seen that the average cross-validation accuracy is 44.67% with a standard deviation of ±3.86%, indicating relatively good consistency 48 between folds. Fold 1 shows the best performance (51.7%) while Fold 4 has the lowest performance (40.0%). Moderate accuracy variation indicates that the model has fairly good stability but there is still room for improvement.

4. In-depth Analysis and Performance per Class

Individual performance evaluation for each class shows different characteristics in classification ability.

TABLE VIII
WEIGHTED AVERAGE PERFORMANCE SUMMARY

Metric	Bronchitis	Normal	Pneumonia	Tuberculosis	Weighted Average
Precision	0,59	0,7	0,57	0,63	0,62
Recall	0,67	0,47	0,53	0,8	0,62
F1-Score	0,62	0,56	0,55	0,71	0,61
Support	15	15	15	15	60

From Table 8, it can be seen that the weighted average for precision, recall, and F1-score shows consistent values around 0.61-0.62, confirming the overall accuracy of the model at 61.67%. Tuberculosis shows superior performance with an F1-score of 0.71 and a recall of 0.80, while Pneumonia has the lowest performance with an F1-score of 0.55.

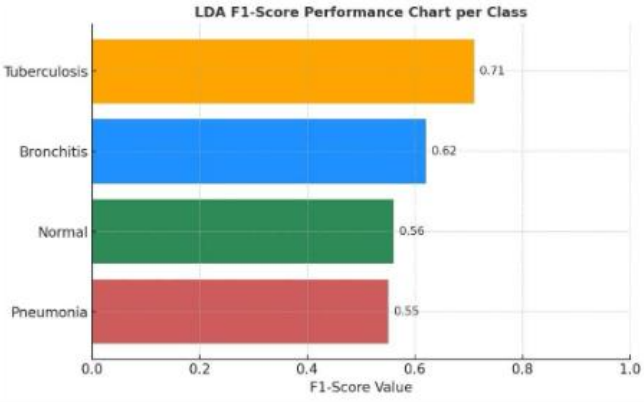

Figure 8. F1-Score LDA Performance per Class

From Figure 8, it can be seen that Tuberculosis has the highest F1-score, indicating the best balance between precision and recall. Bronchitis shows good performance, while Normal and Pneumonia have relatively similar but lower F1 scores. This pattern indicates that the characteristics of GLCM features for tuberculosis are most easily distinguished by the LDA algorithm.

IV. CONCLUSION

This study successfully implemented a combination of Gray Level Co-occurrence Matrix (GLCM) and Linear Discriminant Analysis (LDA) methods for lung disease classification based on X-ray images. The GLCM method was used to extract important texture features such as contrast, correlation, energy, and homogeneity, while LDA served as a classification method to distinguish between types of lung diseases. The experimental results showed that the Tuberculosis class had the highest accuracy of 80%, showing the most consistent and easily recognizable texture characteristics. The overall accuracy of the model reached 61.67%, with additional evaluation through 5-fold cross-validation showing an average accuracy of 44.67%. Evaluation using precision, recall, F1-score, and confusion matrix metrics showed that Tuberculosis had the best classification performance (F1-score: 0.71), while Pneumonia had the lowest performance (F1-score: 0.55). Although the model shows quite good performance, especially for Tuberculosis detection, there are still challenges in distinguishing the Bronchitis, Pneumonia, and Normal classes due to overlapping textures between the categories. This opens up opportunities for further development by adding more complex feature extraction methods or classification algorithms.

To improve classification performance, future developments can incorporate additional features such as shape statistics that represent the anatomical structure of the lungs, as well as wavelet transformations to capture frequency and spatial information from X-ray images. This research only uses LDA as a classification method without comparing its performance with other algorithms. To gain a more comprehensive understanding, future research could perform a performance comparison with other algorithms such as SVM, KNN, and Random Forest to determine the most optimal classification method for this dataset. Additionally, the application of dimensionality reduction techniques such as Principal Component Analysis (PCA) or t-Distributed Stochastic Neighbor Embedding (t-SNE) could help visualize the feature space, offering insights into the degree of overlap among classes and the separability of data.

REFERENCES

[1] F. Meila Azzahra Sofyan, A. Voutama, and Y. Umaidah, "Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Paru-Paru Menggunakan Rapidminer," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 2, pp. 1409–1415, 2023, doi: 10.36040/jati.v7i2.6810.

[2] S. Salmi, "Penggunaan Pemodelan Paru-Paru sebagai Upaya Meningkatkan Pemahaman Peserta Didik tentang Mekanisme Pernapasan di Kelas VIII.8 MTsN 2 Kota Bima," *Lamda J. Ilm. Pendidik. MIPA dan Apl.*, vol. 2, no. 2 SE-Articles, pp. 86–94, Nov. 2022, doi: 10.58218/lambda.v2i2.326.

[3] A. Naseh Khudori and M. Syauqi Haris, "Implementasi Decision tree Untuk Prediksi Kanker Paru-Paru," *J. Ris. Sist. Inf. Dan Tek. Inform. (JURASIK*, vol. 9, no. 1, pp. 94–106, 2024, [Online]. Available: https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

[4] S. Maharani and W. R. Aryanta, "Dampak Buruk Polusi Udara Bagi Kesehatan Dan Cara Meminimalkan Risikonya," *J. Ecocentrism*, vol. 3, no. 2, pp. 47–58, 2023, doi: 10.36733/jeco.v3i2.7035.

[5] L. B. Diantara, H. Hasyim, I. P. Septeria, D. T. Sari, G. T. Wahyuni, and R. Anliyanita, "Tuberkulosis Masalah Kesehatan Dunia: Tinjauan Literatur," *J. 'Aisyiyah Med.*, vol. 7, no. 2, pp. 78–88, 2022, doi: 10.36729/jam.v7i2.855.

[6] H. Fahmi and Sutisna, "Implementasi Data Mining Klasifikasi Gejala Penyakit TB Menggunakan Algoritma Naive Bayes pada Studi Kasus Puskesmas Pegangsaan Dua B," vol. 5, no. 3, pp. 2888–2898, 2024, doi: doi.org/10.35870/jimik.v5i3.970.

[7] M. Sabir and Sarifuddin, "Analisis Faktor Risiko Tingginya kasus Tuberkulosis Paru di Indonesia : Literatur Revieu," *J. Kolaboratif Sains*, vol. 6, no. 6, pp. 453–468, 2023, doi: 10.56338/jks.v6i6.3662.

[8] J. Scott, A. M. Biancardi, O. Jones, and D. Andrew, "Artificial Intelligence in Periodontology: A Scoping Review," *Dent. J.*, vol. 11, no. 2, 2023, doi: 10.3390/dj11020043.

[9] B. Nugroho and E. Y. Puspaningrum, "Kinerja Metode CNN untuk Klasifikasi Pneumonia dengan Variasi Ukuran Citra Input," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 3, pp. 533–538, 2021, doi: 10.25126/jtiik.2021834515.

[10] C. Wijaya, H. Irsyad, and W. Widhiarso, "Klasifikasi Pneumonia Menggunakan Metode K-Nearest Neighbor Dengan Ekstraksi Glcm," *J. Algoritm.*, vol. 1, no. 1, pp. 33–44, 2020, doi: 10.35957/algoritme.v1i1.431.

[11] W. N. Afifah and V. Lusiana, "Klasifikasi jenis batik semarangan menggunakan metode convolution neural network (cnn)," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 1, pp. 542–553, 2025, doi: doi.org/10.29100/jipi.v10i1.5873.

[12] A. Nurdiansyah, H. Erlanda, Y. B. Roza, and R. Sovia, "Klasifikasi Citra Dalam Identifikasi Kol Dan Wortel Menggunakan Algoritma LDA dan KNN," *J. Sci. Soc. Res.*, vol. 4307, no. May, pp. 1895–1902, 2025, doi: doi.org/10.54314/jssr.v8i2.2894.

[13] M. Muchtar and R. Arjaliyah Muchtar, "Integrasi Fitur Warna, Tekstur Dan Renyi Fraktal Untuk Klasifikasi Penyakit Daun Kentang Menggunakan Linear Discriminant Analysis," *J. Mnemon.*, vol. 7, no. 1, pp. 77–84, 2024, doi: 10.36040/mnemonic.v7i1.9258.

[14] A. Razi and D. Yulisda, "Klasifikasi Tingkat Keberhasilan Survival Rate (Sr) Pada Produksi Udang Vaname Menggunakan Algoritma Naïve Bayes," *Ekasakti J. Penelit. dan Pengabdi.*, vol. 4, no. 2, pp. 206–212, 2024, doi: doi.org/10.31933/ejpp.v4i2.

[15] E. Utaminingsih, R. Silviani, and Z. Fitri, "Pengelompokan Fitur Color Structure Descriptor (CSD) Menggunakan Support Vector Machine (SVM) Untuk Citra Busana Tradisional Aceh," vol. 4, no. 1, pp. 139–147, 2021.

[16] R. Amelia *et al.*, "Analisis Perbedaan Permukaan Buah Segar Dan Busuk Menggunakan Model Rgb Dan Gray-Level Co-Occurrence Matrix (GLCM)," vol. 9, no. 4, pp. 7219–7226, 2025, doi: doi.org/10.36040/jati.v9i4.14461.

[17] F. W. Nugraha, A. Harjoko, M. A. Muslim, and N. N. Alabid, "Mix histogram and gray level co-occurrence matrix to improve glaucoma prediction machine learning," *J. Soft Comput. Explor.*, vol. 4, no. 1, pp. 13–22, 2022, doi: 10.52465/joscex.v4i1.99.

[18] A. Ramola, A. K. Shakya, and D. Van Pham, "Study of statistical methods for texture analysis and their modern evolutions," *Eng. Reports*, vol. 2, no. 4, pp. 1–24, 2020, doi: 10.1002/eng2.12149.

[19] D. H. U. Ningsih, E. Zuliarso, M. R. Radyanto, and D. B. Santoso, "Feature Extraction Dengan Gray Level Co-Occurrence Matrix Warna Alami Dari Tanaman Ketapang Berbasis Geolokasi Dewi," *Featur. Extr. Dengan Gray Lev. Co-Occurrence Matrix Warn. Alami Dari Tanam. Ketapang Berbas. Geolokasi*, vol. 29, pp. 1–14, 2024, doi: doi.org/10.35315/dinamik.v29i2.9851.

[20] R. Destriana, D. Nurnaningsih, D. Alamsyah, and A. A. J. Sinlae, "Implementasi Metode Linear Discriminant Analysis (LDA) Pada Klasifikasi Tingkat Kematangan Buah Nanas," *Build. Informatics, Technol. Sci.*, vol. 3, no. 1 SE-Articles, Jun. 2021, doi: 10.47065/bits.v3i1.1007.

[21] I. Rashad, R. R. Isnanto, and C. E. Widodo, "Klasifikasi Penyakit Jantung Menggunakan Algoritma Analisis Diskriminan Linier," *J. Sist. Info. Bisnis*, vol. 13, no. 1, pp. 29–36, 2023, doi: 10.21456/vol13iss1pp29-36.

[22] A. Rasyid and L. Heryawan, "Klasifikasi Penyakit Tuberculosis (TB) Organ Paru Manusia Berdasarkan Citra Rontgen Thorax Menggunakan Metode Convolutional Neural Network (CNN)," *J. Manaj. Inf. Kesehat. Indones.*, vol. 11, no. 1, pp. 35–44, 2023, doi: 10.33560/jmiki.v11i1.484.

[23] D. A. Rizqiana and H. S. Indri, "Asuhan Keperawatan Bersihan Jalan NafasTidak Efektif Pada Pasien Bronkhitis Fisiotrapi Dada Di Ruang Edelweis Atas RSUD Kardinah kota Tegal," *J. Inov. Penelit.*, vol. 3, no. 3, pp. 1–4, 2022, doi: doi.org/10.47492/jip.v3i3.1881.

[24] G. Annisa and R. Khairani, "Rokok Dan Alergi Berhubungan Dengan Bronkitis Akut Pada Pasien Dewasa," *J. Akta Trimedika*, vol. 1, pp. 316–326, 2024, doi: doi.org/10.25105/aktatrimedika.v1i3.19974.