Proceeding of 2nd International Conference on Education, Society and Humanity
e-ISSN : 2986-5832
p-ISSN : 2986-6979

Vol. 02 No. 01 (2024)
Available online at https://ejournal.unuja.ac.id/index.php/icesh

CLASSIFICATION OF K-NEAREST NEIGHBOR (K-NN) AND CONVOLUTIONAL NEURAL NETWORK (CNN) FOR THE IDENTIFICATION OF BRONCHITIS DISEASE IN TODDLERS USING GLCM FEATURE EXTRACTION BASED ON THORAX X-RAY IMAGES

M.Fachrurrozi Nasution[1], Wanayumini[2], Rika Roesnelly[3]
[1,2,3]Universitas Potensi Utama, North Sumatra, Indonesia
Email: fahrurrozi994@gmail.com[1], wanayumini@gmail.com[2], rika@potensi-utama.ac.id[3]

Abstract:

K-Nearest Neighbor (K-NN) is a classification method that seeks the majority class from the k-nearest neighbors of a sample to be classified. Meanwhile, Convolutional Neural Network (CNN) is a type of artificial neural network specifically designed to recognize patterns in image data. The features are then extracted using GLCM (Gray Level Co-occurrence Matrix) from Thorax X-Ray images. This research aims to develop two classification approaches, namely K-Nearest Neighbor (K-NN) and Convolutional Neural Network (CNN), to detect bronchitis disease in toddlers based on Thorax X-Ray images. Feature extraction based on the Gray Level Co-occurrence Matrix (GLCM) is used to transform images into numerical features that can be processed by classification algorithms. The results from both methods will be combined based on various evaluation metrics, such as accuracy, precision, recall, F1-score, etc.

Keywords: *KNN, CNN, GLCM, Bronkitis.*

INTRODUCTION

Bronchitis is an infectious disease affecting the bronchi, which are the air passages in the lungs. This disease commonly affects children living in environments with high pollution levels, such as homes where parents smoke, exposure to motor vehicle exhaust, and smoke from cooking using wood fuel (Imiliati et al., 2021) (Balamurali et al., 2021). The causes of bronchitis include infections from viruses like Rhinovirus, Respiratory Syncytial Virus (RSV), influenza viruses, parainfluenza viruses, and Coxsackie virus. Additionally, bronchitis can be caused by parasites such as roundworms (ascariasis) and fungi. Besides infectious causes, non-infectious factors like physical or chemical agents and other risk factors can also contribute to bronchitis.

In medical literature, the implementation of artificial intelligence and machine learning in the medical industry has been shown to effectively enhance the quality, effectiveness, and efficiency of medical services (Komalasari, 2022). Classification algorithms, including K-Nearest Neighbor (K-NN) and Convolutional Neural Network (CNN), have proven capable of solving various classification problems. Two possible approaches using K-NN with Euclidean and K-NN with Manhattan distances are considered (Hidayati & Hermawan, 2021) (Guo et al., 2003). Previous research on the early detection of Hepatitis C using the K-Nearest Neighbor (KNN) classification method resulted in a testing accuracy of 92%, precision of 92%, and recall of 99% with parameter metrics using the Manhattan distance calculation method, distance weighting, and a K value of 11.

A total of 144 trials were conducted to produce the best parameters with a training accuracy of 95%. So it can be concluded that the best parameter combination for the classification of Hepatitis C disease data using the K-Nearest Neigbor (KNN) algorithm is the distance calculation metric using the Manhattan method, weighting using distance, and the value of K = 11. This research has shortcomings, namely only for early detection of Hepatitis C disease.

Convolutional Neural Network (CNN) is a type of Deep Learning algorithm that consists of multiple layers to process data and extract features. This type of algorithm is commonly used for image processing and object detection (Arrofiqoh & Harintaka, 2018) (Jayalakshmy et al., 2020). In the literature, it is known that by using Convolutional Neural Network

Testing on images can get a precision value of 98%, recall 94% (Rahim et al., 2019). Based on previous research entitled "Classification of Human Activity Based on Radar Signal Using 1-D Convolutional Neural Network" The experimental results show that the classification accuracy of the proposed method is 96.1% (Chen & Ye, 2019). And the method used is only 1 method, namely CNN, it is hoped that combining Machine Learning and Deep Learning can increase the accuracy of a study so that results can be obtained that can be useful in helping to analyze bronchitis disease. The research concluded that K-Nearest Neighbors and Convolutional Neural Network can be used for research in the field of Health and the development of Arficial Intelligence-based systems. It can also be used to assist health agencies in providing better services. Based on the research conducted (Made, et al. 2022) which has been described above, the author develops by adding the CNN method to take the title "Classification of K-Nearest Neighbor (K-NN) and Convolutional Neural Network (CNN) for Disease Identification in Toddlers Using GLCM Feature Extraction Based on Thorax X-Ray Images".

LITERATURE REVIEW

K-Nearest Neighbor (K-NN)

The K-Nearest Neighbor (K-NN) algorithm is a classification method that determines the label (class) of a new object based on the majority class of the nearest k distances in the training data (Srivastava et al., 2021) (Naufal et al., 2021). K-Nearest Neighbor (K-NN) algorithm is a classification algorithm that works by taking a certain number K of the nearest data points as a reference to determine the class of new data. The classification in KNN is based on similarity or proximity to other data. The data points closest to each other are called "neighbors" (Isnain et al., 2021) (Gazalba & Reza, 2017).

Deep Learning

Deep Learning is a branch of artificial intelligence that mimics the way the human brain processes data and creates patterns for decision-making. Deep Learning is part of Machine Learning and includes various types, with some of the most commonly used being Convolutional Neural Network (CNN), Multilayer Perceptrons (MLP), and Recurrent Neural Network (RNN). The advantages of deep learning include its excellent performance in solving complex problems, reducing the need for feature engineering, and having architectures that can easily adapt to new challenges. However, deep learning requires a large amount of data, time-consuming training processes, and is susceptible to overfitting (Naufal et al., 2021) (GM et al., 2021). The image below illustrates a comparison between a simple neural network and a deep learning neural network.

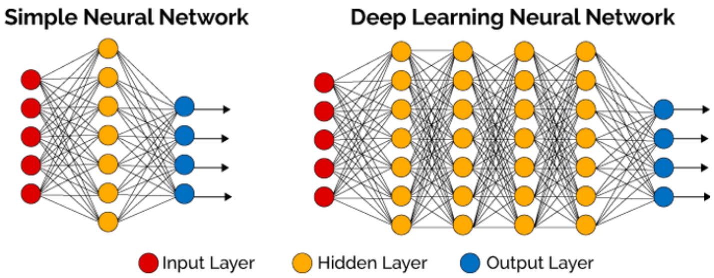


Figure 1. Comparison of a Simple Neural Network with Deep Learning Neural Network

Convolutional Neural Network

Convolutional Neural Network (CNN) is one class of deep learning capable of image recognition and classification. The CNN method is a class of neural networks specialized in processing data with a grid-like topology, such as images. CNN can be used in face recognition, document analysis, image classification, video classification, etc. Below is an illustration of the Convolutional Neural Network architecture.

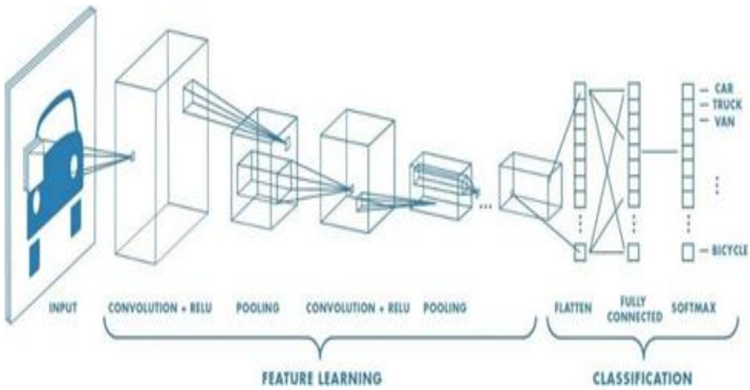


Figure 2. Illustration of Convolutional Neural Network Architecture

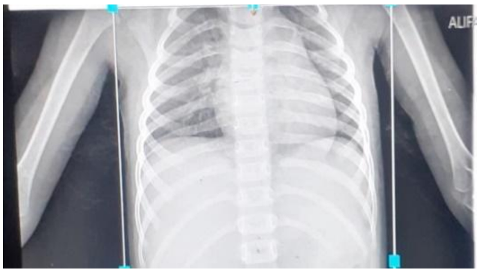
Gray Level Co-Occurrence Matrix (GLCM)

GLCM is a feature extraction method that utilizes second-order texture calculations, taking into account pairs of two pixels in the original image. In the first order, it involves statistical calculations based on the pixel values of the original image without considering pixel adjacency. Co-occurrence can be interpreted as events occurring together, meaning the frequency of events at one pixel level adjacent to the values of other pixels based on distance (d) and the orientation of a specific angle (θ). Distance is represented in pixels, while orientation is represented in degrees. Orientation is formed from four angular directions at intervals of 0, 45, 90, and 135 degrees, and the pixel distance is determined to be 1 pixel (Surya et al., 2017).

X-Ray Thorax Image

X-Ray Thorax Image is a radiographic projection used to assess the anatomical condition using X-ray radiation within the thorax (chest), which includes the heart and lungs. Lung examinations commonly use X-ray radiography due to its relatively quick, easy, and cost-effective nature compared to MRI and CT-Scan tests (Sapata & Juniati, 2019) (Prasetyo, 2020).

A lung image can be considered normal if it lacks spots, has a smooth texture, and the shape and size of the lungs remain unchanged. Conversely, a lung image is considered abnormal if it contains spots, an uneven texture, or a change in the shape and size of the lungs. Changes in the shape and size of lung images can be observed from the pixel intensity on the object (lungs). The sample result of the X-ray used in this study can be seen in the following image.



Source: Radiology of Bronchitis

Figure 3. Result of X-Ray Thorax Exposed to Bronchitis. Radiological Depiction in Bronchitis Shows an Increase in Coarse Bronchovascular Markings

Source: Radiology of Bronchitis

Figure 4. Result of X-Ray Thorax in a Healthy Condition.

RESEARCH METHODS

In this stage, the steps involved consist of 9 stages. Visual explanations of these stages can be seen in the following image.

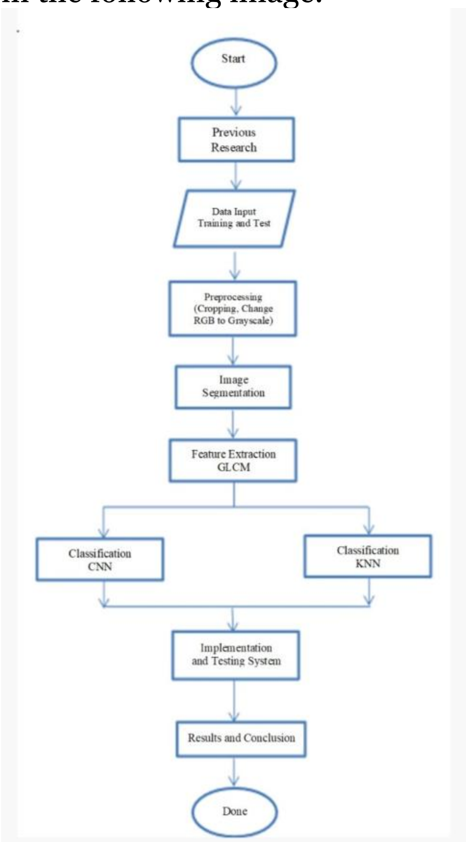


Figure 5. Research Methodology

The following image explains that the research process begins with several steps:

1. Input Training and Test Data
   This stage involves inputting training and test data from X-Ray Thorax images exposed to bronchitis and healthy conditions at Bunda Thamrin Hospital.
2. Preprocessing (Cropping, RGB to Grayscale Conversion)
   This stage includes cropping images to eliminate irrelevant parts and converting RGB images to grayscale.
3. Image Segmentation
   Segmentation is performed on X-ray thorax images to separate the background from the foreground and isolate pixels with different characteristics.
4. GLCM Feature Extraction
   GLCM feature extraction involves forming a matrix of X-Ray Thorax images by creating a co-occurrence matrix for each image. Features such as Entropy, Energy, Homogeneity, and Contrast are then extracted at angles of 0, 45, 90, and 135 degrees.
5. KNN Classification
   The K-Nearest Neighbor (KNN) method is applied to evaluate the performance of the dataset. Hyperparameter tuning and repeated testing of k values are used to train the dataset and determine the optimal k parameter.
6. CNN Classification
   The Convolutional Neural Network (CNN) method is implemented for bronchitis identification, utilizing a specially designed neural network architecture for image processing and pattern recognition in X-Ray Thorax images.
7. System Implementation and Testing
   This stage involves testing and training X-Ray Thorax images using the K-Nearest Neighbor (KNN) and Convolutional Neural Network (CNN) algorithms.
8. Conclusion of Research Results
   The research findings are summarized by observing the measurement results of KNN and CNN algorithms with both complete and selected features, as well as GLCM feature extraction after feature selection.

RESULTS AND DISCUSSION

Results

In this stage, data from toddlers affected by bronchitis is required for processing using the KNN and CNN methods, which involve several attributes to generate accuracy values.

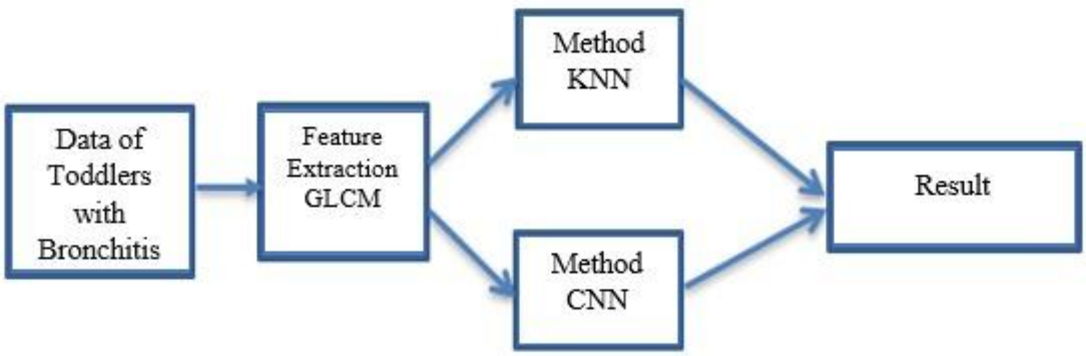


Figure 6. Data Analysis Stage

Here are the stages of data analysis explained:

1. Data of Toddlers with Bronchitis and Healthy Condition

In this stage, the process involves inputting data obtained, specifically data related to toddlers suffering from bronchitis.
2. GLCM Feature Extraction
GLCM (Gray Level Co-Occurrence Matrix) method is applied to the X-Ray Thorax images in this stage, extracting relevant features.
3. KNN Method
The K-Nearest Neighbor (KNN) method is implemented at this stage using the X-Ray Thorax images.
4. CNN Method
The Convolutional Neural Network (CNN) method is applied in this stage to process the X-Ray Thorax images.
5. Results
In this final stage, the research has obtained results, specifically the Classification of K-Nearest Neighbor (K-NN) and Convolutional Neural Network (CNN) for the Identification of Bronchitis in Toddlers Using GLCM Feature Extraction Based on X-Ray Thorax Images.

Discussion

A. Implementation

In this phase, the Classification of K-Nearest Neighbor (K-NN) and Convolutional Neural Network (CNN) for the Identification of Bronchitis in Toddlers Using GLCM Feature Extraction Based on X-Ray Thorax Images is implemented. The implementation is carried out using MATLAB to generate numerical data.

B. Testing Data

This stage involves testing data to assess the alignment between theoretical and practical aspects using MATLAB. Table 1 represents the data that will be tested using MATLAB.

Table 1. Testing Data

Name	Symptoms	X-Ray	Diagnosis
Alifa	1. Rough Bronchovascular Markings 2. Both right and left hilum are normal 3. Both lung fields are normal		Bronchitis_ Chronic
Suriyatno	1. No enlargement of the markings 2. Sinuses and diaphragm are normal 3. Bronchovascular markings are good 4. Both right and left hilum are normal 5. Both lung fields are normal 6. Soft tissue within the chest x-ray area is good 7. There is visible thoracic vertebral scoliosis		Bronchitis_ Chronic
Boini	1. The markings are not enlarged 2. Sinuses and diaphragm are normal 3. Bronchovascular markings are good 4. Both right and left hilum are normal 5. Both lung fields are normal 6. Peribronchial thickening is observed 7. Soft tissue and skeletal structures within the chest x-ray area are good		Bronchitis_ Chronic

Sutomo	1. Markings are not enlarged. 2. Sinuses and diaphragm are normal. 3. Bronchovascular markings are good. 4. Both right and left hilum are normal. 5. Both lung fields are normal. 6. Peribronchial thickening is observed. 7. Soft tissue and skeletal structures within the chest x-ray area are good		Bronchitis_ Chronic
Atik Siregar	1. CTR > 50% 2. Sinuses and diaphragm are normal. 3. Bronchovascular markings are good. 4. Both right and left hilum are normal. 5. Both lung fields are normal		Kardiomegaly

C. Testing Stage

After the initial Matlab display, the Matlab Command Window will appear for data testing, as seen in Figure 7.

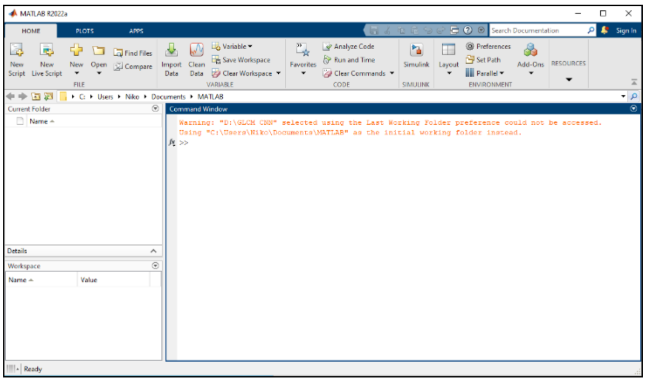


Figure 7. Command Window View

Next, the user can press the "Open" button to open the encoding results in Matlab, as shown in Figure 8.

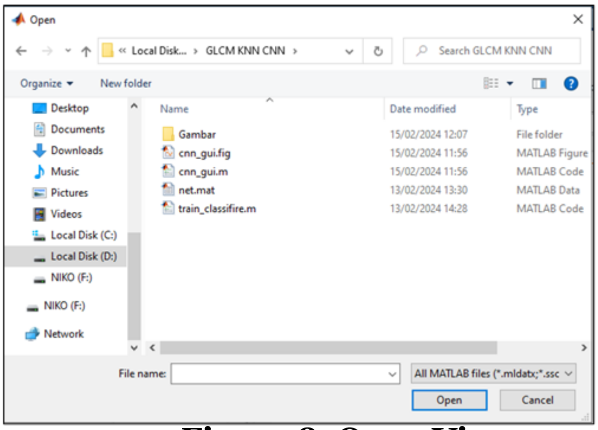


Figure 8. Open View

The next step is to select the Matlab encoding results, as seen in Figure 9.

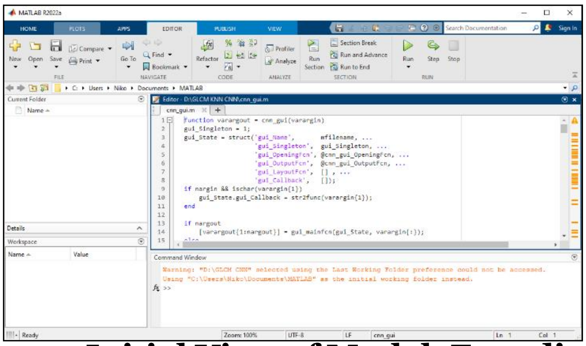


Figure 9. Initial View of Matlab Encoding Results

The following step is to press the "Run" button in the upper right, as shown in Figure 10.

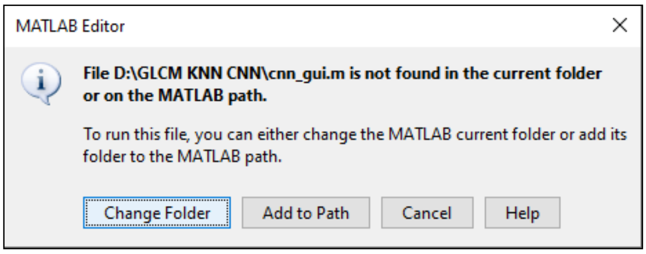


Figure 10. View After Pressing the Run Button

Next, select "Change Folder," and Matlab will display the created GUI, as seen in Figure 11.

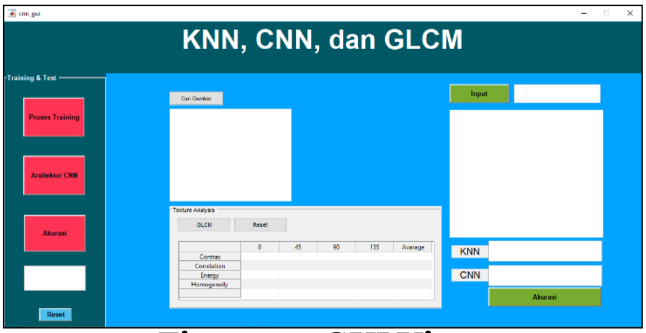


Figure 11. GUI View

The next step is to press the "Search Image" button at the top, as seen in Figure 12.

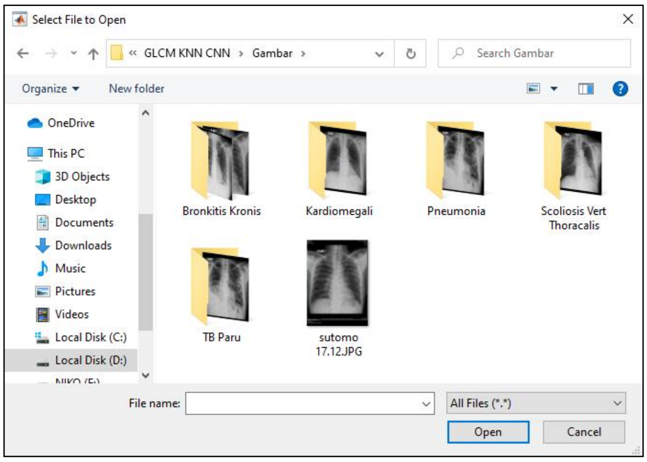


Figure 12. View After Searching for Image

Then, choose an image and click the "Open" button, as seen in Figure 13.

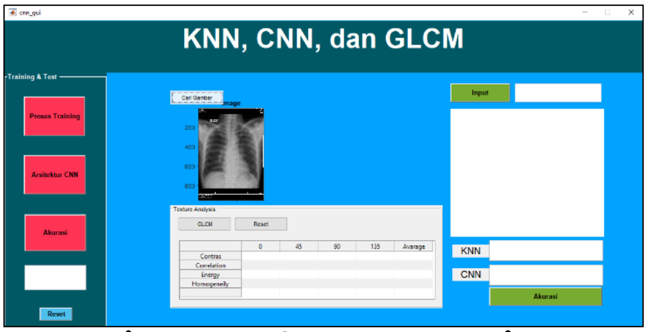
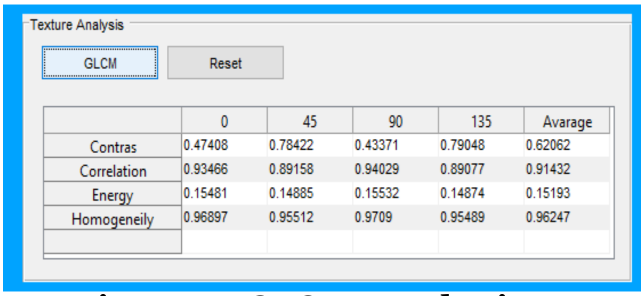


Figure 13. Open Image View

The next step is to press the "GLCM" button, and the result is shown in Figure 14.


Texture Analysis

GLCM Reset

	0	45	90	135	Avarage
Contras	0.47408	0.78422	0.43371	0.79048	0.62062
Correlation	0.93466	0.89158	0.94029	0.89077	0.91432
Energy	0.15481	0.14885	0.15532	0.14874	0.15193
Homogeneily	0.96897	0.95512	0.9709	0.95489	0.96247

Figure 14. GLCM Result View

Figure 15. represents the view of saving the extracted image from GLCM.

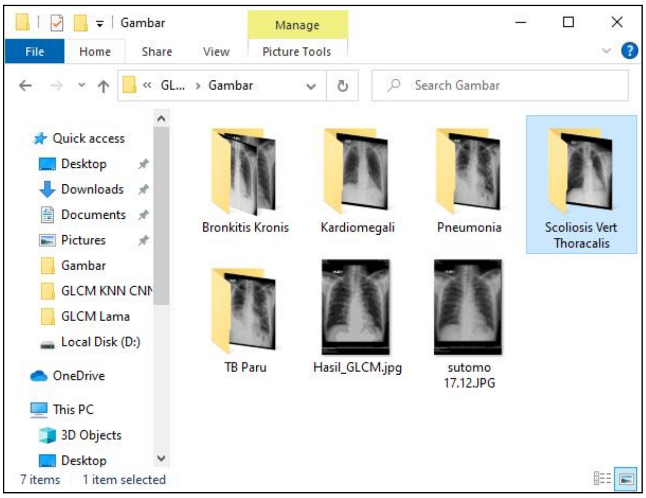


Figure 15. View of Saving Extracted Image Result

Then the newly extracted image is input into the program by pressing the "Input" button as shown in Figure 16.

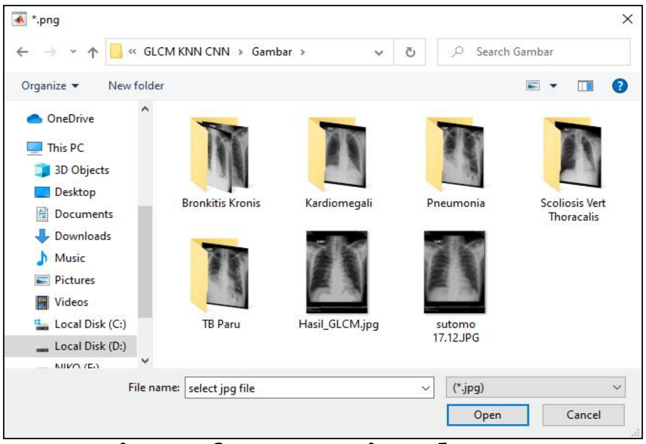


Figure 16. View After Pressing the Input Button

At this stage, the result is the correlation matrix of GLCM (Gray Level Co-occurrence Matrix), representing the relationship between various GLCM features. GLCM shows the correlation between four different features: Contrast, Correlation, Energy, and Homogeneity.

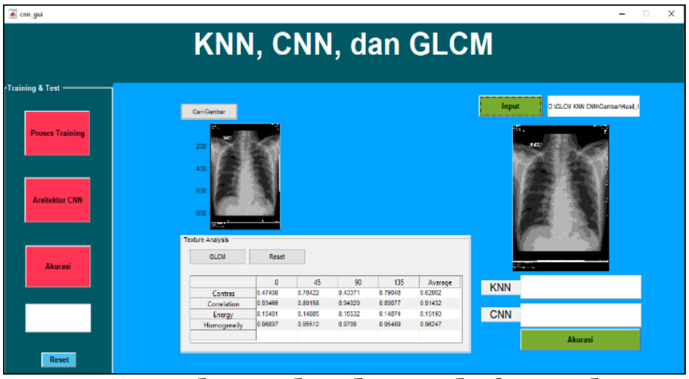


Figure 17. Represents the Upload Result from the Newly Extracted Image

Next, press the "Accuracy" button, which will display the results of KNN and CNN as shown in Figure 18.

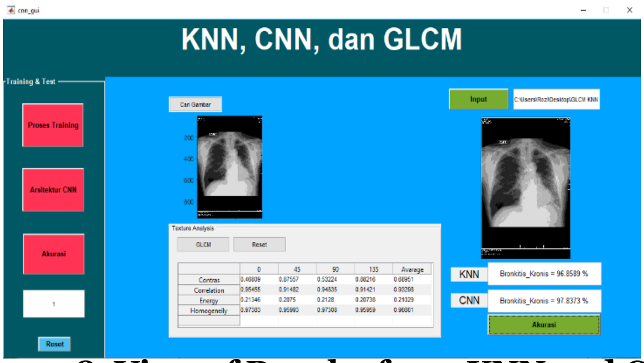


Figure 18. View of Results from KNN and CNN

From the obtained results, it can be concluded that the results of both methods are as follows:

1. The accuracy of the KNN method in identifying Chronic Bronchitis is 96.8589%, and the accuracy of the CNN method in identifying Chronic Bronchitis is 97.8373%.
2. Therefore, the results of the KNN and CNN methods are accurate, with a Training and Test Process Accuracy Value of 1 (100%), indicating the identification of Chronic Bronchitis.

In conclusion, both the KNN and CNN methods produce accurate results, with a Training and Test Process Accuracy Value of 1 (100%), indicating the identification of Chronic Bronchitis.

CONCLUSION

Based on the previous discussion, several conclusions can be drawn:

1. The classification of K-Nearest Neighbor (K-NN) and Convolutional Neural Network (CNN) for the identification of bronchitis in toddlers using GLCM feature extraction based on X-Ray Thorax images achieved accurate results. The accuracy of the KNN method in identifying Chronic Bronchitis is 96.8589%, while the accuracy of the CNN method is 97.8373%. Both methods show a Training and Test Process Accuracy Value of 1 (100%), indicating the identification of Chronic Bronchitis.
2. The classification of K-Nearest Neighbor (K-NN) and Convolutional Neural Network (CNN) for the identification of bronchitis in toddlers using GLCM feature extraction based on X-Ray Thorax images can be implemented using Matlab programming.

REFERENCES

Arrofiqoh, E. N., & Harintaka, H. (2018). Implementasi metode convolutional neural network untuk klasifikasi tanaman pada citra resolusi tinggi. *Geomatika*, *24*(2), 61.

Balamurali, B. T., Hee, H. I., Kapoor, S., Teoh, O. H., Teng, S. S., Lee, K. P., Herremans, D., & Chen, J. M. (2021). Deep neural network-based respiratory pathology classification using cough sounds. *Sensors*, *21*(16), 5555.

Chen, H., & Ye, W. (2019). Classification of human activity based on radar signal using 1-D convolutional neural network. *IEEE Geoscience and Remote Sensing Letters*, *17*(7), 1178–1182.

Gazalba, I., & Reza, N. G. I. (2017). Comparative analysis of k-nearest neighbor and modified k-nearest neighbor algorithm for data classification. *2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 294–298.

GM, H., Gourisaria, M. K., Rautaray, S. S., & Pandey, M. (2021). Pneumonia detection using CNN through chest X-ray. *Journal of Engineering Science and Technology (JESTEC)*, *16*(1), 861–876.

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN model-based approach in classification. *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, 986–996.

Hidayati, N., & Hermawan, A. (2021). K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation. *Journal of Engineering and Applied Technology*, *2*(2), 86–91.

Imiliati, C., Nugroho, N. B., & Syaifuddin, M. (2021). SISTEM PAKAR DALAM MENDIAGNOSA PENYAKIT BRONCHITIS PADA ANAK MENGGUNAKAN METODE CERTAINTY FACTOR DAN TEOREMA BAYES DI UPT. RSKPARU PROVENSI SUMATERA UTARA. *Jurnal Cyber Tech*, *3*(7).

Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(2), 121–130.

Jayalakshmy, S., Priya, B. L., & Kavya, N. (2020). CNN based Categorization of respiratory sounds using spectral descriptors. *2020 International*

Conference on Communication, Computing and Industry 4.0 (C2I4), 1–5.
Komalasari, R. (2022). Pemanfaatan Kecerdasan Buatan (Ai) Dalam Telemedicine: Dari Perspektif Profesional Kesehatan. *Jurnal Kedokteran Mulawarman*, *9*(2), 72–81.
Naufal, M. F., Kusuma, S. F., Tanus, K. C., Sukiwun, R. V., Kristiano, J., & Lieyanto, J. O. (2021). Analisis Perbandingan Algoritma Klasifikasi Citra Chest X-ray Untuk Deteksi Covid-19. *Teknika*, *10*(2), 96–103.
Prasetyo, B. D. (2020). *Klasifikasi Citra X-Ray Paru-paru Anak Pneumonia dan Non-Pneumonia Menggunakan Metode Segmentasi dan Deteksi Tepi*.
Rahim, A., Luthfi, E. T., & Learning, D. (2019). *Convolutional Neural Network Untuk Kalasifikasi*.
Sapata, B. M., & Juniati, D. (2019). Klasifikasi Penyakit Paru Berdasarkan Citra X-Ray Thorax Menggunakan Metode Fraktal Box Counting. *MATHunesa: Jurnal Ilmiah Matematika*, *7*(3), 228–235.
Srivastava, A., Jain, S., Miranda, R., Patil, S., Pandya, S., & Kotecha, K. (2021). Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease. *PeerJ Computer Science*, *7*, e369.
Surya, R. A., Fadlil, A., & Yudhana, A. (2017). Ekstraksi ciri metode Gray Level Co-Occurrence Matrix (GLCM) dan Filter Gabor untuk klasifikasi citra batik pekalongan. *Jurnal Informatika: Jurnal Pengembangan IT*, *2*(2), 23–26.