

Clustering of Tuberculosis and Normal Lungs Based on Image Segmentation Results of Chan-Vese and Canny with K-Means

¹Fayza Nayla Riyana Putri, ²Nur Cahyo Hendro Wibowo, ³Hery Mustofa

^{1,2,3}Departement of Information Technology, Walisongo State Islamic University

Email: ¹fayzanayla_1908096020@student.walisongo.ac.id, ²nur_cahyo@walisongo.ac.id, ³herymustofa@walisongo.ac.id

Article Info

Article history:

Received Jan 11th, 2023

Revised Feb 13th, 2023

Accepted Mar 16th, 2023

Keyword:

Canny

Chan-Vese

Tuberculosis

K-Means

Lungs

ABSTRACT

The lungs are a vital organ in the human body. If there is interference with lung function, the health of the human body as a whole can be affected. Examination by medical workers needs to be done when there is interference with lung function. This examination can usually be done in various ways, one of which is through a chest X-ray radiographic examination procedure. The application of Artificial Intelligence is growing rapidly in the medical field, especially in diagnostics and treatment management. Artificial intelligence in the medical world can also be applied in processing image data in radiology to analyze X-ray results as supporting diagnostic information. Operators Chan-Vese and Canny are two edge detection operators in digital image processing in an effort to obtain the necessary information based on the shape and size of the object. This study was conducted for clustering of normal and tuberculosis lung conditions based on the results of chest X-ray image segmentation from Chan-Vese and Canny using K-Means Clustering. The results of clustering using K-Means obtained an accuracy value of 73%.

Copyright © 2023 Puzzle Research Data Technology

Corresponding Author:

Hery Mustofa,

Departement of Information Technology,

Walisongo State Islamic University,

Prof.Dr Hamka Street, Ngaliyan, Semarang

Email: herymustofa@walisongo.ac.id

DOI: <http://dx.doi.org/10.24014/ijaidm.v6i1.21835>

1. INTRODUCTION

The lungs are a vital organ in the human body. This organ is a respiratory organ that is responsible for the human respiratory and circulatory systems. This organ that belongs to the respiratory system functions as a transmission medium for the exchange between (O₂) oxygen from the air and (Co₂) carbon dioxide and (H₂O) water vapor from the blood. This process is called external respiration or breathing [1]. The lungs are organs of the human body that function in managing incoming air and sorting O₂ (Oxygen) and Co₂ (Carbon Dioxide). The main function of the lungs is to change oxygen from the air with carbon dioxide from the blood. If there is interference with lung function, the health of the human body as a whole can be affected. Examination of medical workers needs to be done when there is interference with lung function. Several diseases of the lungs have similar symptoms such as fever, cough with phlegm, weight loss, loss of appetite, and bad breath, therefore it is necessary to carry out additional examinations [2]. This examination can usually be done in various ways, one of which is through a chest X-ray radiographic examination procedure.

The chest X-ray procedure provides an image based on the diffraction of electromagnetic waves, namely X-rays [1]. With this light, the condition of the organs in the human body can be seen visually without surgery (non-invasive procedure). Of course, this examination must be carried out carefully, but basically, the reading of X-ray results tends to be done through visual observation, so an appropriate analysis is needed to distinguish the results of a chest X-ray from the condition of the organs [3].

One of the conveniences is the implementation of technological developments, namely Artificial Intelligence which has been implemented and helps humans in everyday life. Artificial Intelligence is a branch of computer science that is concerned with creating smart machines that can do jobs like humans. According

to John McCarthy (1956), Artificial Intelligence is a computer system that is formed to identify and model human thought processes and design machines so that they can mimic human behavior. Artificial Intelligence Technology in various forms has been used in various fields such as education, banking and financial markets, manufacturing, agriculture and even now the implementation of artificial intelligence technology developments has been found in the world of health [4].

Artificial intelligence in the world of health not only provides reliable support for medical practitioners and facilities, but also maximizes the efficiency of practitioners. The use of Artificial Intelligence is growing rapidly in the medical field, especially in diagnostics and treatment management [5]. Artificial Intelligence (AI) methods from Machine Learning to Deep Learning have important functions in various domains related to well-being, including improving new clinical systems, patient information and records, and treatment of various diseases [4]. Artificial intelligence in the medical world can also be applied in processing image data in radiology. In radiology, trained doctors visually assess medical images and report findings to detect, characterize, and monitor disease. Therefore, artificial intelligence technology is needed to analyze X-ray results as supporting diagnostic information. The use of artificial intelligence in processing imaging data has been researched in a previous study by Bayu Dwi Prasetyo in 2020 Classification of X-Ray Image of the Lungs of Pneumonia and Non-Pneumonia Children Using the Segmentation and Edge Detection Methods and produced trial results from Sardjito Hospital data with accuracy value of 73.3%, sensitivity 62.5%, and specificity 77.2%. Test results from kaggle data with an accuracy of 83.3%, a sensitivity of 70%, and a specificity of 90%.

The Chan-Vese method is a developmental technique from Edge-Based Model and Level Set and is an active contourless method [6]. The Active Contour Model is an active model in a segmentation technique that utilizes energy and force constraints on images for grouping regions [7]. The active contour model is commonly used in various image processing applications, especially medical image processing, such as images from X-Ray Thorax, CT-Scan of the brain, or MRI images of other organs of the body [8]. Active contour models are used in digital image processing by means of pixel separation. The Chan-Vese method is used to detect image objects with evolutionary energy. The development of the Mumford-Shah model and this model is based on the energy minimization problem, which can be reformulated in the formulation of set levels, leading to an easier way to solve the problem [9]

Canny detection is a technique in Edge-Based Segmentation discovered by John F. Canny in 1986 [10]. Basically, the Canny method works by using grayscale images as input and producing intensity discontinuity locations as output or edge images [11]. In its application, there are steps for the canny edge detection method, namely smoothing or noise filtering, calculates the magnitude of the gradient or the direction of the gradient, Implement non-maximum suppression, set the Hysteresis Thresholding process [10]. Clustering analysis is a multivariate technique that aims to group similar observations into clusters based on the observed values of several variables for each individual. The goal in making clustering is to minimize variation within a cluster and maximize variation between clusters [12]. There are many algorithms that are commonly used, one of the most widely used algorithms is K-means clustering. K-Mean clustering is one of the unsupervised learning algorithms included in non-hierarchical cluster analysis which is used to group data based on variables or features. The goal of K-Means Clustering, like other clustering methods, is to obtain cluster data by maximizing the similarity of characteristics within the clusters and maximizing the differences between the existing clusters. The advantages of the K-Means algorithm are as follows: Relatively simple and easy to implement, Scalable for large datasets, Easy to adapt to new instances, Commonly implemented to clusters of different shapes and sizes. The K-means clustering algorithm groups data based on the distance between the data and the cluster centroid points obtained through an iterative process. The performance of the K-means clustering algorithm depends on the highly efficient clusters it forms. In accordance with the aims and objectives of this study, two clusters will be created, namely Tuberculosis and Normal.

This research was conducted as an extension of the previous method, namely lung classification into tuberculosis lungs and normal lungs based on chest X-rays using K-means clustering. The method used in this study is a combination of the Chan-Vese Segmentation Method for image segmentation and the Edge Detection Method, namely Canny Edge-Based Segmentation for detecting edges that work by using grayscale images as input and producing intensity discontinuity locations as output or edge images. Smart detectors are the most powerful edge detectors produced by edge functions [13]. With the method used, it is hoped that it can provide accurate lung classification results including the lungs of tuberculosis sufferers and normal lungs based on chest X-ray images.

2. RESEARCH METHOD

In this study, the data used is a type of secondary data. Secondary data is data obtained or collected by people who conduct research from existing sources. This data is used to support the primary information that has been obtained, namely from library materials, literature, previous research, books, and so on. The

secondary data used in this research was obtained through the Literature Study method, which is a method for collecting data obtained from a number of books, magazines, or articles that have an attachment to the problem and research objectives they have [14]. Literature study can also be defined as activities in collecting data by reading, recording and processing research materials or instruments. This method is used by collecting secondary data in research in the form of x-ray images of lung disease, especially tuberculosis, sourced from kaggle.com.

This research is a type of quantitative research, namely a type of research that produces findings that can be achieved (obtained) using statistical procedures or other means of quantification (measurement). In this study, numerical data was used in the form of a range of values or the range of the number of white pixels in the X-Ray image which would then be processed for classification of lung conditions, especially tuberculosis and normal lungs. The steps in this study are represented through a flowchart diagram in Figure 1. A flowchart functions as a diagram to visually describe a sequence of operations by displaying a number of symbols [15].

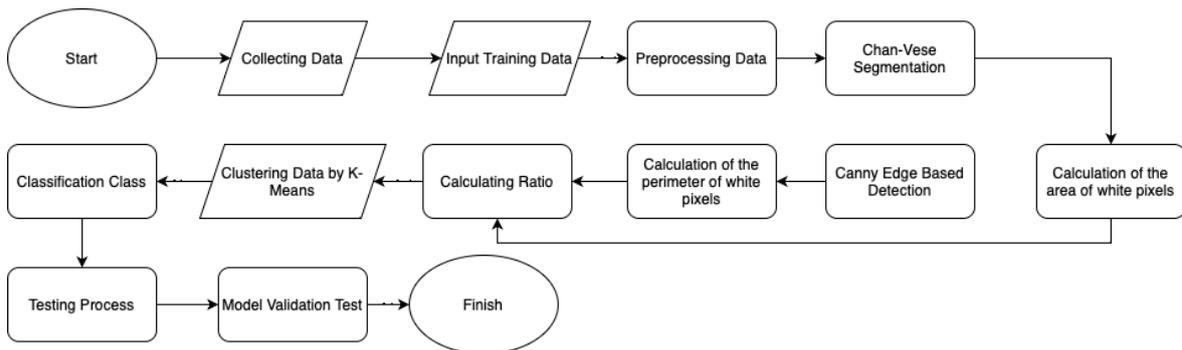


Figure 1. Research flowchart

The segmentation process for images in this study aims to obtain features from the image so that it can be used as a reference for producing clusters, namely for Tuberculosis lung classes and normal lungs. The segmentation process goes through two stages, namely the chan-veese and the canny edge detection method. The Chan-Vese method is a developmental technique from Edge-Based Model and Level Set and is an active contourless method (Dwi Prasetyo, 2020). The Active Contour Model is an active model in a segmentation technique that utilizes energy and force constraints on images for grouping regions (Getreuer, 2012). The active contour model is commonly used in various image processing applications, especially medical image processing, such as images from X-Ray Thorax, CT-Scan of the brain, or MRI images of other organs of the body (Hemalatha et al., 2018). Active contour models are used in digital image processing by means of pixel separation. The Chan-Vese method is used to detect image objects with evolutionary energy [16]. The development of the Mumford-Shah model and this model is based on the energy minimization problem, which can be reformulated in the formulation of set levels, leading to an easier way to solve the problem (Getreuer, 2012).

Canny detection is a technique in Edge-Based Segmentation discovered by John F. Canny in 1986[17]. Basically, the Canny method works by using grayscale images as input and producing intensity discontinuity locations as output or edge images [11]. There are criteria which are advantages in the application of Canny Edge Detection, namely [18] :

1. Detection Criteria

Canny detection has edge detection capabilities according to defined convolution parameters and is flexible in determining the level of edge thickness detection.

2. Localization Criteria

Canny detection makes it possible to provide a minimum distance between the detected edge and the original edge.

3. Response Criteria

Each edge has only one response so as to minimize clutter in the next image processing stage.

After the segmentation process is carried out on the image, all the feature values will be processed and the K-Means Clustering process will be carried out. In the clustering process especially with the K-Means method, there are several stages that need to be passed, including [19]:

Determination of the value of k or cluster to be created

1. Initialize centroid values randomly

$$v = \frac{\sum_{i=1}^n x_i}{n}; i=1,2,3,\dots,n \quad (1)$$

Information :

v : centroid on cluster

x_i : i-th x object

n : the number of objects / number of objects that are members of the cluster.

2. Assign each data point to the nearest centroid by calculating the distance for each data to the centroid that has been made using the Euclidean distance.

$$(x,y)=\|x-y\|=\sqrt{\sum_{i=1}^n (x_i-y_i)^2}, i=1,2,3,\dots,n \quad (2)$$

Information :

x_i : i-th x object

y_i : i-th y power

n : number of objects

3. Recalculate the centroid value of the newly formed cluster by calculating the mean value of each data point in the cluster.
4. Perform optimization so that the criteria are met by repeating the last two steps.

3. RESULTS AND ANALYSIS

In The first step in this study was to collect data related to the X-ray image used as the object of this study. Data in the form of digital images will be processed in the research in Portable Network Graphics (*.png) format. The image used is an X-ray image of the lungs for people with tuberculosis and normal lungs.

The processed data is divided into 2 folders, namely Training Data and Testing Data which contain subfolders for each category of Tuberculosis and Normal Lung imagery. Description of the amount of data is shown in Table 1 below:

Table 1. The Processed Data

	Total Data
Normal Lungs	1341
Tuberculosis Lungs	650

All data was obtained through the kaggle.com website which is a platform that accommodates a collection of datasets as research material. This dataset collection is done by downloading the required dataset folder as a research object. The data in the form of this image already has a name tag so that the type of disease is known in each X-ray image. The image is the result of an X-ray of a patient with tuberculosis and normal lungs. The dataset obtained for this study is data that already has a label and is divided into two classes, namely Tuberculosis Lungs Data and Normal Lungs Data.

The next step after inputting the image is image preprocessing. This step aims to improve the image before analysis so that the raw data is ready to be processed. In this study there are 3 stages in the preprocessing process, namely changing the size of the image so that all images have the same size, namely 256x256 pixels, changing the RGB image to a Grayscale image, and applying filtering using the Histogram Equalization method.

3.1. Image Size Normalization Process.

The normalization process is carried out by changing the size of the image to 256x256 pixels as shown in the result in Figure 2 by using a function from MATLAB, namely `imresize`.

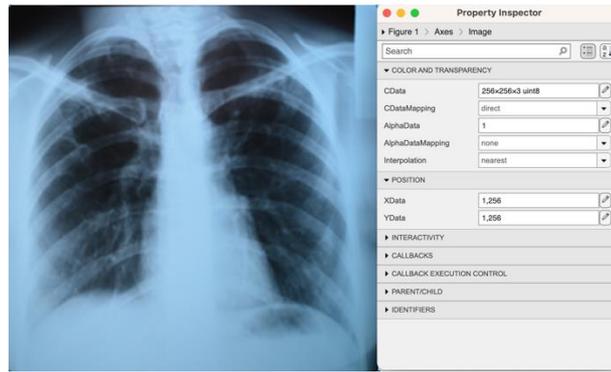


Figure 2. Input and Output Image Size Normalization Process

3.2. Grayscale Process.

After resizing the image, the next step is the Grayscale process to change the type of image that was previously RGB to a Grayscale image with the `rgb2gray` function so that it can produce a Grayscale image in Figure 3.



Figure 3. Output Grayscale process

3.3. Filtering Process with Histogram Equalization.

Filtering process is done by Histogram Equalization. This process is carried out by evenly distributing all pixels within the entire range of values so that the image has a flatter histogram.

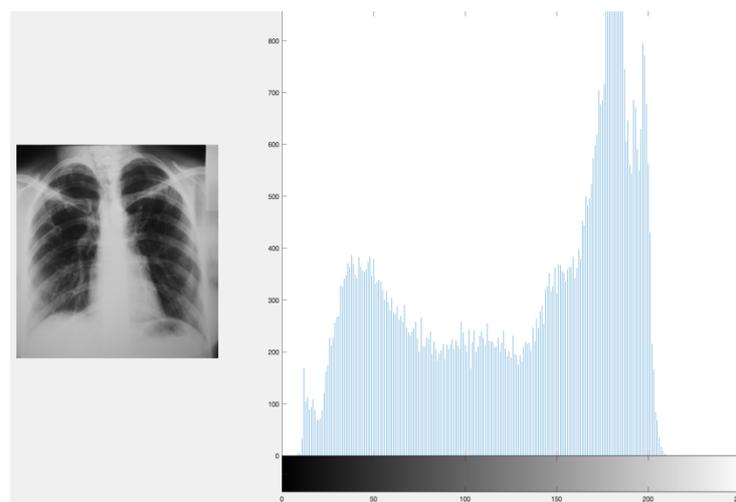


Figure 4. Distribution of All Pixels Before The Filtering Process

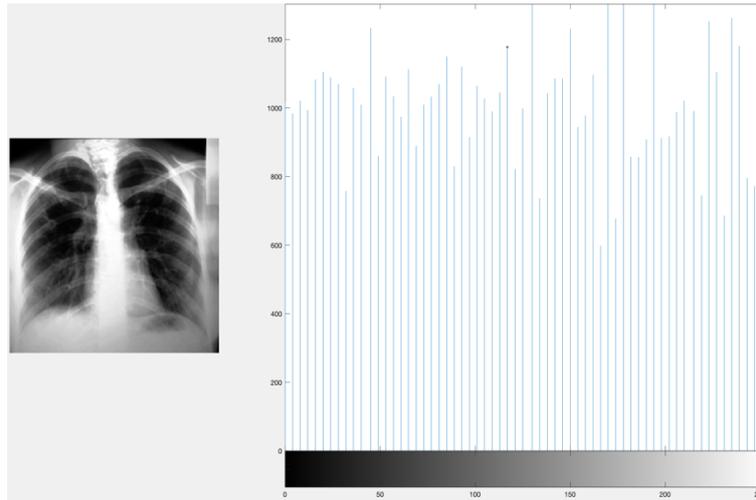


Figure 5. Distribution of All Pixels After The Filtering Process

From Figure 5 it can be seen that the color of the image has more contrast and has a flatter histogram so that it can produce an image with a higher contrast than the image in Figure 4.

3.4. Calculation of The Area of White Pixels Chan-Vese Segmentation

In the image resulting from Chan-Vese segmentation, the area of the white pixel area is processed as a variable for calculating X-ray image classification. In this process, the regionprops function is used to measure the properties of a binary image.

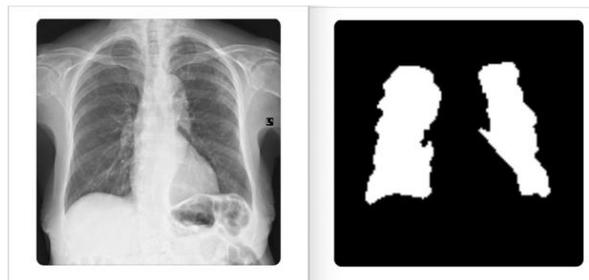


Figure 6. The Image Resulting from Chan-Vese Segmentation

Before calculating the area of segmented image, the image resulting from the previous chan-veve segmentation is converted into a binary image with the im2bw function and the process of filling in image areas that have holes with the imfill function is carried out.

The results of the calculation of one of the images can be seen in Figure 7. In Figure 7, there are area results for two segmentation areas in one image. Each image can have a varying number of areas, therefore the sum of all areas is carried out so that it can be applied to all images [20].

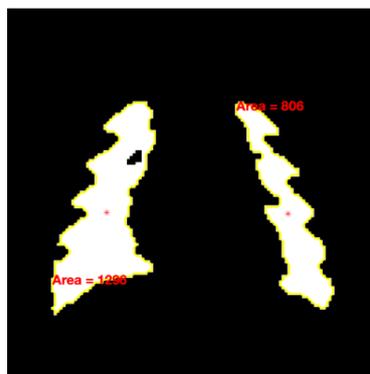


Figure 7. The Results of the Calculation from Chan-Vese Segmentation

The calculation results for all images are translated into the workspace provided by MATLAB as shown in Table 2 for Tuberculosis Lungs and Table 3 for Normal Lungs.

Table 2. The calculation results for Tuberculosis Lungs

Images	Area
'Tuberculosis-1.png'	4342
'Tuberculosis-10.png'	3700
'Tuberculosis-100.png'	4265
'Tuberculosis-101.png'	4426
'Tuberculosis-102.png'	3328
'Tuberculosis-103.png'	3356
'Tuberculosis-104.png'	3728

Table 3. The calculation results for Normal Lungs

Images	Area
'IM-0115-0001.png'	2823
'IM-0117-0001.png'	2097
'IM-0119-0001.png'	2269
'IM-0122-0001.png'	2067
'IM-0125-0001.png'	1990
'IM-0127-0001.png'	2297
'IM-0128-0001.png'	2629

3.5. Canny Edge Based Segmentation

After performing Chan-Vese segmentation on the image, the next step is the Canny edge detection process. This process is implemented so that it can find the edge line and produce the circumference of the image as other variables needed in the classification calculation process. The edges themselves are found by detecting changes in the brightness level of the image [10].

The results of the implemented Canny operator are shown in Figure 8 as a comparison of the results of Chan-Vese Segmentation and Canny Edge Detection.



Figure 8. Comparison of The Results of Chan-Vese Segmentation and Canny Edge Detection.

The calculation results for all images are translated into the workspace provided by MATLAB as shown in Table 4 for Tuberculosis Lungs and Table 5 for Normal Lungs.

Table 4. Calculation Results with Canny Edge Detection for Tuberculosis Lungs

Images	Circumference
'Tuberculosis-1.png'	418
'Tuberculosis-10.png'	486
'Tuberculosis-100.png'	485
'Tuberculosis-101.png'	513
'Tuberculosis-102.png'	487
'Tuberculosis-103.png'	592
'Tuberculosis-104.png'	468

Table 5. Calculation Results with Canny Edge Detection for Normal Lungs

Images	Circumference
'IM-0115-0001.png'	539
'IM-0117-0001.png'	473
'IM-0119-0001.png'	491
'IM-0122-0001.png'	375
'IM-0125-0001.png'	495
'IM-0127-0001.png'	414
'IM-0128-0001.png'	506

3.6. Calculating Ratio

The classification process is carried out to distinguish between tuberculosis and normal lung conditions based on the number of white pixels in the image of each category. This process is carried out by applying equation (3)

Here is an example calculation:

1. Number of Pixel Image Detection of Canny ‘Tuberculosis-1.png’ : 418
2. Number of Pixel Image Segmentation Chan-Vese ‘Tuberculosis-1.png’ : 4342

Calculation:

$$\frac{\text{The number of pixels of the canny detection image}}{\text{The number of pixels of the chan-veese segmentend image}} = \frac{418}{4342} = 0.096269 \tag{3}$$

So from the results above it can be concluded, the number of white pixels in the ‘Tuberculosis-1.png’ image is 0.096269. This process is implemented in all images.

All calculation data is stored in the MATLAB workspace which is then processed through Jupyter Notebook with the Python Programming Language for the K-Means Clustering process as a reference for class division in the classification process.

Table 6. Ratio Calculation Results for Tuberculosis Lungs

Images	Area	Circumference	Ratio
'Tuberculosis-1.png'	4342	418	0,096269
'Tuberculosis-10.png'	3700	486	0,13135135
'Tuberculosis-100.png'	4265	485	0,1137163
'Tuberculosis-101.png'	4426	513	0,11590601
'Tuberculosis-102.png'	3328	487	0,14633413
'Tuberculosis-103.png'	3356	592	0,17640048
'Tuberculosis-104.png'	3728	468	0,12553648

Table 7. Ratio Calculation Results for Normal Lungs

Images	Circumference	Area	Ratio
'IM-0115-0001.png'	539	2823	0,19093163
'IM-0117-0001.png'	473	2097	0,22556032
'IM-0119-0001.png'	491	2269	0,21639489
'IM-0122-0001.png'	375	2067	0,18142235
'IM-0125-0001.png'	495	1990	0,24874372
'IM-0127-0001.png'	414	2297	0,18023509
'IM-0128-0001.png'	506	2629	0,19246862

3.7. Clustering Data by K-Means

The first step before clustering is to calculate the average of all data, namely Tuberculosis and Normal data. This average value is the value obtained from the previous process, namely the ratio result.

Table 8. The Average Value of The Ratio

Data	Mean (Average)
Tuberculosis	0,137366886
Normal	0,18574854

This clustering process aims to group data into two classes or clusters, namely Cluster 0 and Cluster 1. Cluster 0 has the closest value to the average value in the Tuberculosis data and so does Cluster 1 in Normal data, so it can be concluded that Cluster 0 defines Cluster for the Tuberculosis class and Cluster 1 for the Normal class. The clustering process is carried out through the Jupyter Notebook application and the centroid points and results are obtained as follows:

Table 9. The Centroid Value of Each Cluster

Cluster Class	Centroid
Cluster 0	0,144031
Cluster 1	0,203558

Table 10. K-Mean Cluster Results

Images	Result	Cluster
'Tuberculosis-1.png'	0,096269	0
'Tuberculosis-10.png'	0,131351	0

Images	Result	Cluster
'Tuberculosis-100.png'	0,113716	0
'Tuberculosis-101.png'	0,115906	0
'Tuberculosis-102.png'	0,146334	0
'Tuberculosis-104.png'	0,125536	0
.....
'NORMAL2-IM-1360-0001.png'	0,190455	1
'NORMAL2-IM-1365-0001.png'	0,25779	1
'NORMAL2-IM-1371-0001.png'	0,279753	1
'NORMAL2-IM-1396-0001.png'	0,220749	1
'NORMAL2-IM-1400-0001.png'	0,176865	1
'NORMAL2-IM-1401-0001.png'	0,177606	1
'NORMAL2-IM-1406-0001.png'	0,20443	1
'NORMAL2-IM-1412-0001.png'	0,226816	1
'NORMAL2-IM-1419-0001.png'	0,182996	1
'NORMAL2-IM-1422-0001.png'	0,279835	1
'NORMAL2-IM-1423-0001.png'	0,224031	1

3.8. Testing Process

All test data already has a label so that the testing process can be carried out easily, namely by matching whether the value of the number of white pixels in each test image belongs to one of the predefined cluster classes, namely in Table. The following are the steps in the testing process :

1. Grouping data in each category included in the cluster class. Based on the centroid points that have been obtained in the table, categories can be grouped. Determination of the classification class is done by matching the calculation values of the number of pixels closest to the predetermined cluster centroid points.
2. Counting the amount of data in each category to do the validation test.

After grouping the data, the next step is to calculate the amount of true and false data included in the predetermined cluster class with the results as shown in the Table 11.

Table 11. The Amount of Clustered and Labeled Data

Clus-tered Class	Classes	Labeled Class	
		Tuberculosis	Normal
	Tuberculosis	629	21
	Normal	521	820

3.9. Model Validation Test

Table 11 defines the amount of data labeled as tuberculosis lung that belongs to the clustered tuberculosis lung class of 629 out of a total of 650 data while the amount of data labeled as normal lung that belongs to the normal lung class that is clustered is 820 out of a total of 1341 data. The accuracy of applying the algorithm model to the data can be determined from the number of labeled data that corresponds to the clustering class as follows [21]:

$$\begin{aligned}
 \text{Accuracy} &= \frac{\text{The Number Of Clustered Data That Corresponds To The Data Label}}{\text{Total}} \times 100\% \\
 &= \frac{629+820}{1991} \times 100\% = 73\% \tag{4}
 \end{aligned}$$

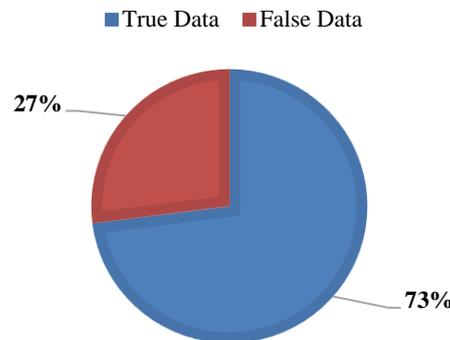


Figure 9. Data Validity Diagrams

4. CONCLUSION

This research was conducted as a refinement of the previous method, namely the classification of lung conditions which include tuberculosis and normal lungs based on chest X-ray images using the K-Means Clustering method. From the results of the discussion, it was found that the validation test of applying the algorithm model to the data can be determined from the number of labeled data that corresponds to the clustering class. The implementation of the Chan-Vese Segmentation Method and the Edge Detection Method, namely Canny Edge-Based Segmentation, produced an accuracy value of 73% and this shows that the model is feasible to use. Based on the analysis of the results, in order to increase the accuracy of the results, it is necessary to select a more balanced number of sample data between chest X-ray images of normal lungs and tuberculosis lungs.

REFERENCES

- [1] T. D. Wulan, E. Purwanti, and M. Yasin, "Deteksi Kanker Paru-Paru Dari Citra Foto Rontgen Menggunakan Jaringan Saraf Tiruan Backpropagation," *UNAIR REPOSITORY*, Feb. 11, 2023. <https://repository.unair.ac.id/25693/> (accessed Feb. 11, 2023).
- [2] K. H. Rahouma, S. M. Mabrouk, and M. Aouf, "Lung Cancer Diagnosis Based on Chan-Vese Active Contour and Polynomial Neural Network," *Procedia Comput Sci*, vol. 194, pp. 22–31, Jan. 2021, doi: 10.1016/J.PROCS.2021.10.056.
- [3] P. Andhi *et al.*, "Deteksi COVID-19 Berdasarkan Hasil Rontgen Dada (Chest Xray) Menggunakan Python," *Journal of Information Technology and Computer Science*, vol. 1, no. 1, pp. 58–67, Jun. 2021, doi: 10.47111/JOINTECOMS.V1I1.2956.
- [4] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," *J Ambient Intell Humaniz Comput*, 2022, doi: 10.1007/S12652-021-03612-Z.
- [5] D. Vilimek *et al.*, "A Quantitative and Comparative Analysis of Edge Detectors for Biomedical Image Identification Within Dynamical Noise Effect," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12034 LNAI, pp. 90–101, 2020, doi: 10.1007/978-3-030-42058-1_8/COVER.
- [6] D. Prasetyo, "Klasifikasi Citra X-Ray Paru-Paru Anak Pneumonia dan Non-Pneumonia Menggunakan Metode Segmentasi dan Deteksi Tepi," 2020, Accessed: Sep. 22, 2022. [Online]. Available: <https://dspace.uui.ac.id/handle/123456789/28404;jsessionid=A40324F8FED71BB89D98E5C98FE00327>
- [7] E. S. Brown, T. F. Chan, and X. Bresson, "Completely convex formulation of the Chan-Vese image segmentation model," *Int J Comput Vis*, vol. 98, no. 1, pp. 103–121, May 2012, doi: 10.1007/S11263-011-0499-Y/METRICS.
- [8] R. J. Hemalatha *et al.*, "Active Contour Based Segmentation Techniques for Medical Image Analysis," *Medical and Biological Image Analysis*, Jul. 2018, doi: 10.5772/INTECHOPEN.74576.
- [9] E. Sert and A. Alkan, "Image Edge Detection Based on Neutrosophic Set Approach Combined with Chan-Vese Algorithm," <https://doi.org/10.1142/S0218001419540089>, vol. 33, no. 3, Feb. 2019, doi: 10.1142/S0218001419540089.
- [10] A. Shihab Ahmed, "COMPARATIVE STUDY AMONG SOBEL, PREWITT AND CANNY EDGE DETECTION OPERATORS USED IN IMAGE PROCESSING," *J Theor Appl Inf Technol*, vol. 15, p. 19, 2018, Accessed: Feb. 14, 2023. [Online]. Available: www.jatit.org
- [11] E. A. Sekehrevani, E. Babulak, and M. Masoodi, "Implementing canny edge detection algorithm for noisy image," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1404–1410, Aug. 2020, doi: 10.11591/EEL.V9I4.1837.
- [12] M. Darwis, L. H. Hasibuan, M. Firmansyah, N. Ahady, and R. Tiaharyadini, "Implementation of K-Means clustering algorithm in mapping the groups of graduated or dropped-out students in the Management Department of the National University," *JISA(Jurnal Informatika dan Sains)*, vol. 4, no. 1, pp. 1–9, Jun. 2021, doi: 10.31326/JISA.V4I1.848.
- [13] P. Amoako-Yirenykyi, J. K. Appati, I. K. Dontwi, P. Amoako-Yirenykyi, J. K. Appati, and I. K. Dontwi, "Performance Analysis of Image Smoothing Techniques on a New Fractional Convolution Mask for Image Edge Detection," *Open Journal of Applied Sciences*, vol. 6, no. 7, pp. 478–488, Jul. 2016, doi: 10.4236/OJAPPS.2016.67048.
- [14] J. Riyono, S. D. Puspa, and C. E. Pujiastuti, "Simulasi Clustering Provinsi di Indonesia dalam Penyebaran Covid-19 Berdasarkan Indikator Kesehatan Masyarakat Menggunakan Algoritma Gaussian Mixture Model," *Majamath: Jurnal Matematika dan Pendidikan Matematika*, 2022. <http://ejurnal.unim.ac.id/index.php/majamath/article/view/1699/756> (accessed Feb. 11, 2023).
- [15] W. D. Yuniarti, "Dasar-dasar pemrograman dengan Python," *Deepublish*, Nov. 2019. <https://opac.perpusnas.go.id/DetailOpac.aspx?id=1279704> (accessed Nov. 23, 2022).
- [16] M. Nasor and W. Obaid, "Segmentation of osteosarcoma in MRI images by K-means clustering, Chan-Vese segmentation, and iterative Gaussian filtering," *IET Image Process*, vol. 15, no. 6, pp. 1310–1318, May 2021, doi: 10.1049/IPR2.12106.
- [17] S. T. H. Kieu, A. Bade, M. H. A. Hijazi, and H. Kolivand, "COVID-19 Detection Using Integration of Deep Learning Classifiers and Contrast-Enhanced Canny Edge Detected X-Ray Images," *IT Prof*, vol. 23, no. 4, pp. 51–56, Jul. 2021, doi: 10.1109/MITP.2021.3052205.

- [18] S. Kieu *et al.*, “Ensemble deep learning for tuberculosis detection using chest X-Ray and canny edge detected images,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 4, pp. 429–435, 2019, doi: 10.11591/ijai.v8.i4.pp429-435.
- [19] E. Umargono, J. E. Suseno, and S. K. V. Gunawan, “K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula,” pp. 121–129, Oct. 2020, doi: 10.2991/ASSEHR.K.201010.019.
- [20] E. E. Nithila and S. S. Kumar, “Segmentation of lung from CT using various active contour models,” *Biomed Signal Process Control*, vol. 47, pp. 57–62, Jan. 2019, doi: 10.1016/J.BSPC.2018.08.008.
- [21] E. Hany Pratiwi and D. Juniati, “Clustering Penyakit Paru-Paru Berdasarkan Rontgen Dada Menggunakan Dimensi Fraktal Box Counting Dan K-Medoids,” *J. Ris. & Ap. Mat*, vol. 06, no. 01, pp. 1–12, 2022.

BIBLIOGRAPHY OF AUTHORS



Fayza Nayla Riyana Putri is an active undergraduate student from the Information Technology Department, Walisongo State Islamic University Semarang. The focus of her research is the application of artificial intelligence in digital image processing.



Nur Cahyo Hendro Wibowo is an active Head of the Associate's Degree and Bachelor's Degree Departments at the Faculty of Science and Technology, UIN Walisongo Semarang. He received Bachelor's Degree in Indonesian Islamic University and Master's Degree in Dian Nuswantoro University.



Hery Mustofa is an active Expert Assistant and Lecturer at the Information Technology Department, Faculty of Science and Technology, UIN Walisongo Semarang. He received Bachelor's Degree and Master's Degree in Dian Nuswantoro University.