The Implementation of Channel Area Thresholding in Early Detection System of Acute Respiratory Infection

Zilvanhisna Emka Fitri[1, a)], Fabrizal Adam Pramudya[2] and Arizal Mujibtamala Nanda Imron[3, b)]

[1,2] *Informatics Engineering, Information Technology Department, Jember Polytechnic, Jember, Indonesia*
[3] *Electronics Engineering, Department of Electrical Engineering, Faculty of Engineering, Jember University, Jember, Indonesia*

[a)]Corresponding author: zilvanhisnaef@polije.ac.id
[b)]arizal.tamala@unej.ac.id

Abstract. Acute respiratory infections (ARI) are infectious diseases that affect both children and adults, particularly in the context of climate change. Bacteria are one of the causes of ARI. According to the government, the discovery of the bacteria that cause ARI is an indicator of successful management of infectious diseases. The current obstacle is the limited number of medical analysts, which results in longer microscopic examination times and requires a high level of objectivity. Therefore, a system for the early detection of ARI-causing bacteria was developed using digital image processing techniques, specifically channel area thresholding as one of the segmentation methods. This research employs four shape features for bacterial classification: the number of bacterial colonies, area, perimeter, and shape. The Naïve Bayes intelligent system method is used for the classification process. The system had an accuracy rate of 86.84% in the classification of four types of bacteria: *S. aureus, S. pneumoniae, C. diphteriae* and *M. tuberculosis.*

ARTICLE INFO

Article history:

Received 1 March 2024

Revised 16 May 2024

Accepted 31 May 2024

Available online 1 June 2024

Keywords: Acute Respiratory Infection, Detection System, Segmentation

Cite this as:
Fitri, Z. E., & Imron, A. M. N. The Implementation of Channel Area Thresholding in Early Detection System of Acute Respiratory Infection (ARI). Indonesian Applied Physics Letters, 5(1). https://doi.org/10.20473/iapl.v5i1.55626

Indonesian Applied Physics Letters
e-ISSN: 2745-3502
DOI: 10.20473/iapl.v5i1.55626

INTRODUCTION

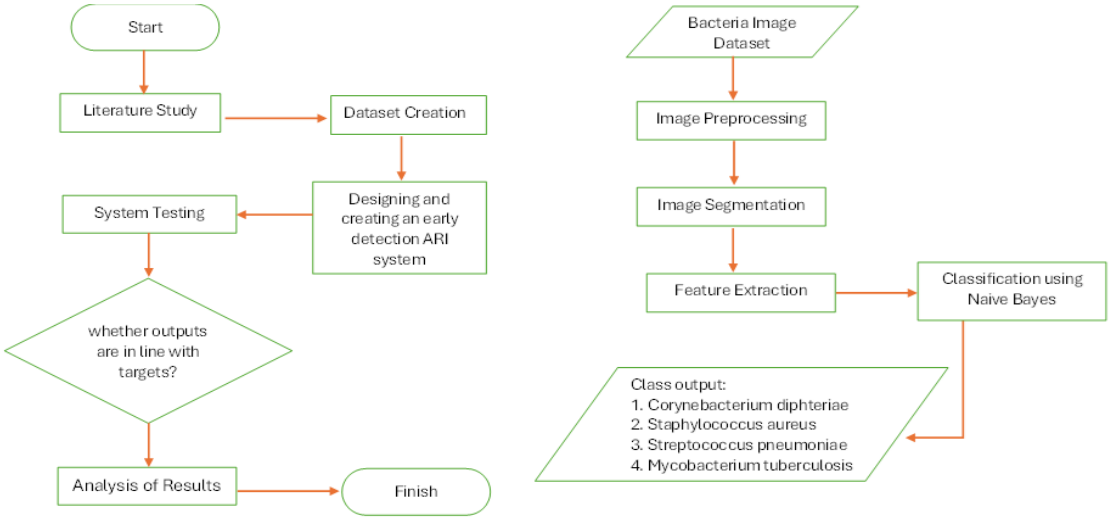
In recent years, global warming has caused climate change, particularly in tropical countries such as Indonesia. Indonesia is experiencing increased rainfall, especially during the rainy season, while it experiences high temperatures and extreme heat during the dry season (Malihah, 2022). Furthermore, air pollution's environmental damage also affects the current climate, which in turn impacts the emergence of acute respiratory infections (ARI) (Herawati et al., 2023). Acute Respiratory Infection (ARI) is a highly contagious disease that affects the upper or lower respiratory tract. It is a major cause of morbidity and mortality worldwide, particularly in children and adults (Indhira and Hendrik, 2023). ARI can be caused by viruses and bacteria, but the focus of this study is on bacteria, including pharyngitis-causing bacteria such as gonorrhea, diphtheria, mycoplasma, and chlamydia, as well as bacteria that cause pneumonia and tuberculosis (Islam, 2023).

Tuberculosis and pneumonia are the two most prevalent infectious diseases in East Java. Jember is one of the districts with the highest number of TB cases, with 5,244 reported cases. Pneumonia is also a leading cause of death among children under five, and its case finding rate in East Java is above 70%. East Java also reported 163 cases. (Dinas Kesehatan Provinsi Jawa Timur, 2023). One effort to control the disease is to detect cases of coughing, sneezing, vomiting, and shortness of breath; patients should seek immediate medical attention and provide sputum or oral specimens for microbiologic examination by a medical analyst (Fitri et al., 2021). However, a limited number of analysts can hinder the early detection of diseases due to longer processing times (Fitri et al., 2022). To develop an early detection system for bacteria that cause ARI, researchers can use computer vision.

Microscopic image analysis is used to identify the presence of Mycobacterium tuberculosis (M. tuberculosis) by using features like area, eccentricity, aspect ratio, compactness, circularity, and roughness (Reshma and Beegum, 2017) and segmentation using Channel Area Thresholding (CAT) (Mithra and Emmanuel, 2018). Additionally, the LVQ accurately detects bacilli that cause acute respiratory infections such as tuberculosis and diphtheria with 97% accuracy (Fitri et al., 2022). Bacteria causing ARI were classified not only by LVQ, but also by KNN method with value K = 3, 5, 7. The best accuracy was 91.7% (Fitri et al., 2021). This research aims to develop a system by comparing it to other methods, such as Naive Bayes, due to its simplicity and high accuracy (Mahran et al., 2020).

RESEARCH METHODOLOGY

The research is divided into six stages (a) : literature study, creation of bacterial image datasets, system design and development, system testing and analysis. The system design and development stage is further divided into four stages (b): data preprocessing, segmentation, feature extraction, and bacterial classification, as shown in Figure 1.



(a) (b)

FIGURE 1. (a) The research phase and (b) the system development phase.

Generally, bacteria have three forms, namely cocci, bacilli and sprochetes (Mahon and Lehman, 2019), In this study, bacteria were analyzed, including chain-shaped cocci bacteria (streptococci), grape-like clustered bacteria (staphylococci), and bacillus bacteria such as club-shaped pleomorphic rods and aerobic acid-fast rods obtained from the Balai Besar Laboratorium Kesehatan (BBLK) Surabaya (Figure 2) (Fitri et al., 2021). This study analyzed bacterial preparations by Gram and Ziehl-Neelson (ZN) staining.

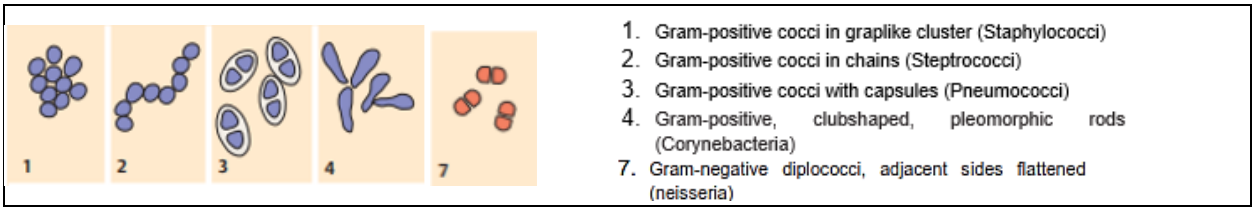
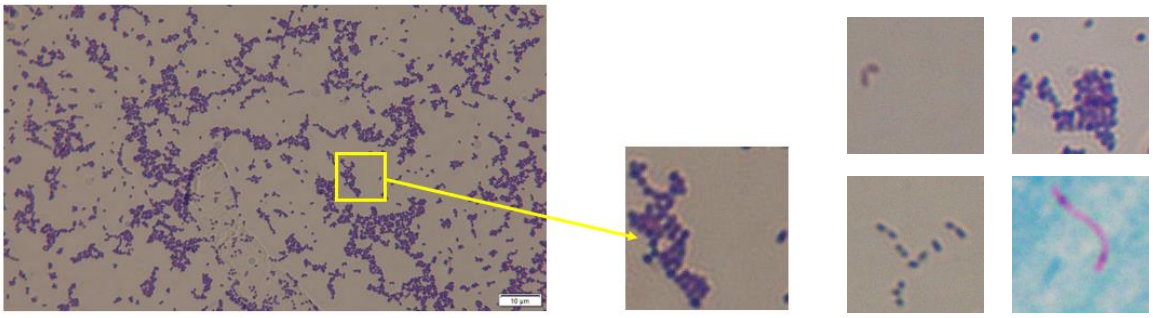


FIGURE 2. Bacterial morphology (Kayser, 2005)

Dataset Creation

The bacterial preparation image is 1920x1080 pixels, then cropped to 151x151 pixels to reduce computational load and eliminate unnecessary background, as shown in Figure 3. The cropping process is based on the shape of the bacteria as shown in Figure 2. A total of 378 images were used, including 94 images of *Corynebacterium diphteriae*, 90 images of *Staphylococcus aureus*, 107 images of *Streptococcus pneumoniae*, and 87 images of *Mycobacterium tuberculosis*.



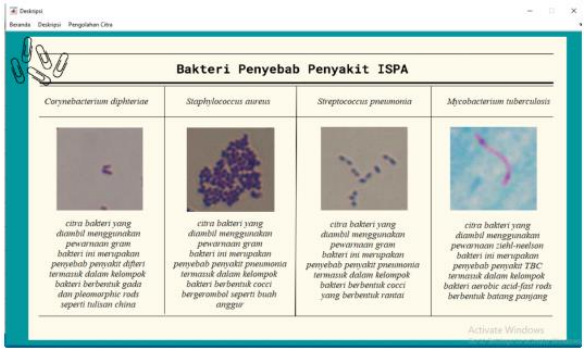
(a) (b)

FIGURE 3. (a) ARI bacteria cropping process and (b) 4 bacteria image dataset.

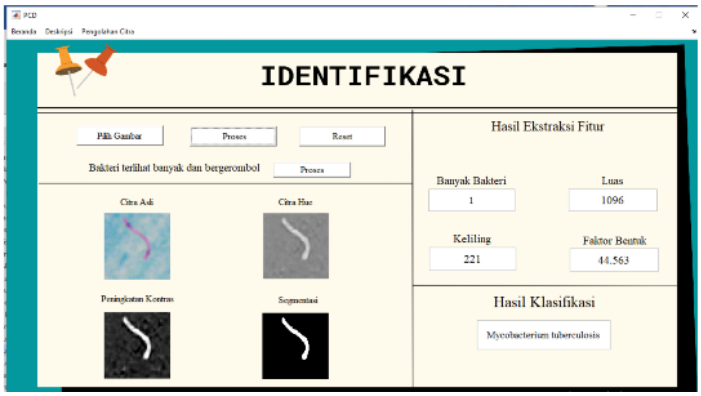
Designing and Creating System

There are two main stages in this phase: designing the system and building the system. The system design includes three menus: home, description, and image processing (see Figure 4). It consists of four steps: data pre-processing, segmenting, feature extraction, and bacterial classification.

(a) homepage



(b) description page



(c) image processing page

FIGURE 4. Display of some app menus

Image Preprocessing

Cropping and converting the image to HSV color space is the first step of preprocessing. This is because the RGB image has a high value, which makes segmenting difficult (Fitri and Imron, 2021). The HSV color space uses a model similar to human vision, specifically cone cells, using the formula equation:

$$hue = tan\left(\frac{3x(G-B)}{(R-G)+(R-B)}\right)$$

$$saturation = 1 - \frac{min\,(R,G,B)}{V}$$

$$Value = \frac{R+G+B}{3}$$

Furthermore, the image should also increase contrast to achieve a higher contrast value. Increasing contrast affects the histogram graph, which is stretched after determining the minimum and maximum values as limits (Erwin and Ningsih, 2020) using the formula equation:

$$Contrast\ (x,y) = \frac{f(x,y) - min}{\max - min} x1$$

Contrast (x,y) represents the output image after contrast stretching, while f(x,y) represents the input image.

Image Segmentation

Aim of this stage is to use threshold value to separate object from background and obtain a binary image. We use the equation based on the gray value of the image histogram to determine the threshold (T).

$$Segmentation\ (x,y) = \begin{cases} 1, if\ T_1 \leq S(x,y) \leq T_2 \\ 0, if\ S(x,y) < T_1 \\ 0, if\ S(x,y) > T_2 \end{cases}$$

In addition to the thresholding segmentation process to obtain the shape of the bacteria as shown in Figure 2, the segmentation process is performed based on the area threshold, known as channel area thresholding. (Destarianto et al., 2022).

$$Area_{New} = T_{area1} \leq Area_{old} \leq T_{area2}$$

Feature Extraction

(Fitri et al., 2021) said that to classify 4 types of bacterial images that cause ARI using 4 shape features, namely the number of bacterial colonies, area, perimeter and shape obtained using the formula equation :

$$Area = Number\ of\ pixels\ in\ row - 1 + row\ to - 2 + \cdots + row\ to - 8$$

$$Perimeter = \sum Even\ code + \sqrt{2}\,x \sum odd\ code$$

$$Shape = \frac{Perimeter^2}{Area}$$

Classification using Naïve Bayes

This research uses the Naive Bayes algorithm, which is based on probabilities and has the advantage of requiring only a small amount of data while maintaining high accuracy (Asmara et al., 2018). The Naïve Bayes algorithm begins with the following steps:

1. Calculate the prior probability of each existing class.
2. Calculate the mean value of each feature using the formula equation :

$$\mu = \frac{\sum n}{k}$$

   Where k = number of data and n = data value
3. Calculate the standard deviation (sd) value of the feature using the formula equation:

$$Sd = \sqrt{\frac{n\sum_{i=1}^{n}(xi - \bar{x})^2}{(n-1)}}$$

4. Calculate the probability density using the equation :

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

5. Calculate the probability of each class using the provided formula after obtaining the probability density and prior values:

$$P = P(X|Ci)\ x\ P(Ci)$$

RESULT AND DISCUSSION

The research begins by converting the RGB color space to the HSV color space, which allows the bacteria to appear in the Hue, Saturation, and Value channels, as shown in Figure 5. The RGB color space is difficult to segment because of its wide 24bit color range, so you need to decompose into RGB components or convert to another color space. HSVs are chosen because their cone model is like that of human eye cone cells, making it an appropriate choice for representing bacterial shape.

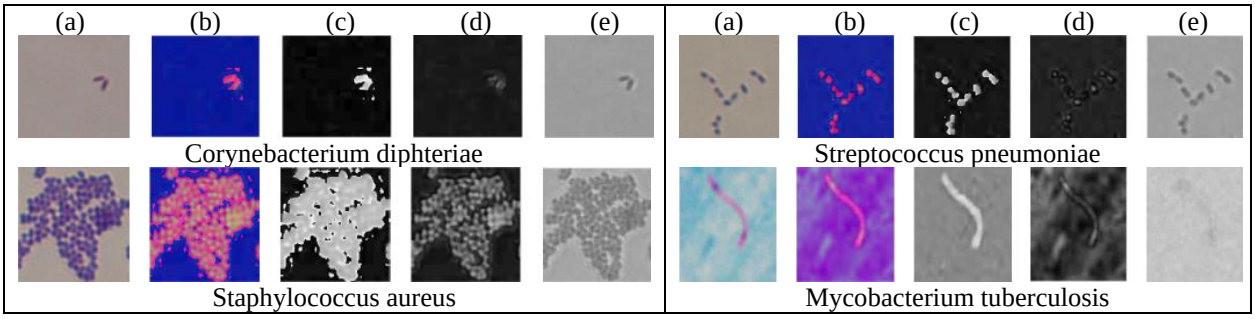
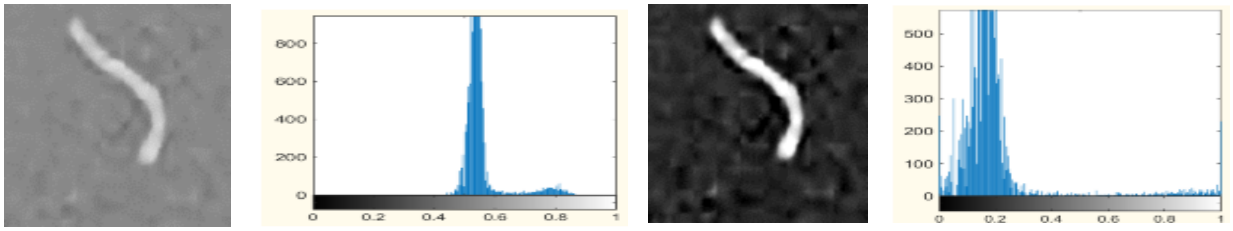


FIGURE 5. (a) Original image, (b) HSV image, (c) Hue image, (d) Saturation image and (e) Value image

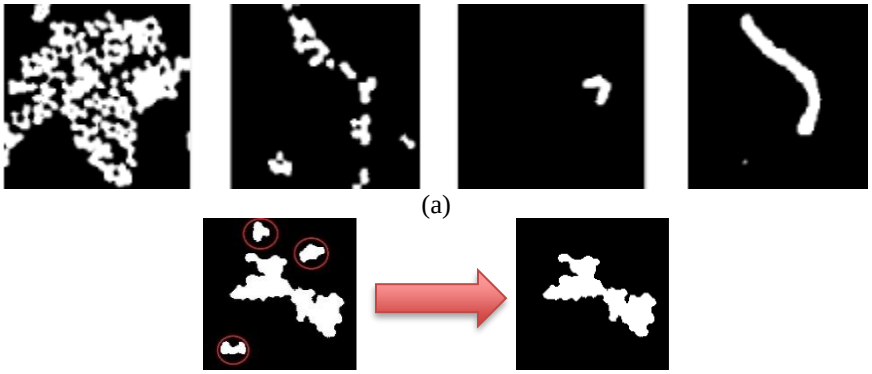
Based on the results shown in Figure 5, the hue channel images provide the best representation of the shape of the four bacterial species and are used as the input for the enhancement operation. The main goal of this process is to improve the clarity of the shape of the bacteria, especially that of the M. tuberculosis bacteria. Importantly, this process also affects the histogram, resulting in an even distribution of gray values across all intensity scales, as shown in Figure 6.



(a) (b)

FIGURE 6. (a) Original image and its histogram, and (b) Contrast stretching image and its histogram

Fitri et al. (2021) stated that the segmentation process requires two threshold values: $T_1 = 0.4$ and $T_2 = 0.7$. The segmentation process produces a binary image with two values: 1 (white) and 0 (black) (see Figure 7a). The resulting image may contain noise, such as small spots (from Gram's stain or ZN) and bacterial colonies, which can lead to ambiguity in the identification of clustered forms of bacteria, such as *S. pneumoniae* and *S. aureus*. Therefore, this method uses segmentation based on area, also known as Channel Area Thresholding (CAT). Two thresholds are used, but each bacterium is given a different treatment, as in previous studies. Figure 7 shows that $T_{area1} = 70$ and $T_{area2} = 5000$ were used for three bacteria, S. pneumoniae, C. diphtheriae and M. tuberculosis, whereas $T_{area1} = 80$ and $T_{area2} = 8000$ were used for S. aureus.



(a)

(b)

FIGURE 7. Segmentation is performed using two methods, (a) thresholding and (b) channel area thresholding (CAT).

The next step in the segmentation process is feature extraction, which involves the identification of the most prominent and distinguishing features among the four bacteria, such as the number of colonies, the area, the perimeter, and the shape (see Table 1). The minimum, maximum, and mean values for each feature are used as input values for the Naïve Bayes algorithm. The table shows that the largest number of colonies for S. aureus bacteria is 25, while for S. pneumoniae it is 17. However, the largest area for S. aureus is 6166 pixels, while for *S. pneumoniae* it is 3078 pixels. The perimeter and shape feature values are largest for S. aureus and smallest for *M. tuberculosis* bacteria.

TABLE 1. Morphological feature values of each bacterium

Features		C. diphteriae	S. aureus	S. pneumoniae	M. tuberculosis
Colony Number	Min	1	4	2	1
	Max	6	25	17	2
	Mean	2	13	8	1
Area	Min	175	851	331	447
	Max	1136	6166	3078	1545
	Mean	386	2615	629	868
Perimeter	Min	50	313	127	89
	Max	293	1978	741	297
	Mean	119	948	302	166
Shape	Min	12,30	115,12	47,86	17,72
	Max	109,58	787,06	308,50	61,37
	Mean	38,27	355,21	146,12	32,56

The probability value for each class must be calculated, followed by the mean (see Table 1) and standard deviation value of each class (see Table 2), according to the steps of the Naive Bayes algorithm. Classification is performed by comparing training and test data with different percentages of the total data of 378 images, as described in Table 3.

TABLE 2. Morphological characteristic values for each bacterium

The bacteria	P(Ci| bacteria)	Standar deviasi			
		Colony Counts	Area	Perimeter	Shape
C. diphteriae	0,249	1,14	200,77	56,62	21,95
S. aureus	0,24	4,88	975,60	306,89	128,59
S. pneumoniae	0,284	2,92	414,58	106,85	53,82
M. tuberculosis	0,228	0,16	250,43	41,54	10,65

TABLE 3. The results of accuracy, precision and recall of the system on each comparison of training data and test data.

Comparison of training data and testing data	Accuracy (%)	Precision	Recall
50 : 50	86,77	0,875	0,868
60 : 40	86,76	0,876	0,868
70 : 30	84,96	0.857	0,850
80 : 20	86,84	0,891	0,868
90 : 10	86,84	0,884	0,868

Table 3 shows that the Naïve Bayes method produces the highest system accuracy of 86.84% in a data comparison of either 80:20 or 90:10. However, there is a difference in the precision values between the two comparisons. The system precision for the 80:20 comparison is 0.891, while for 90:10 it is 0.868. The precision value indicates the proportion of correctly classified positive category data out of the total positive classified data, while recall indicates the percentage of positive category data that is correctly classified by the system. Based on these results, the optimal data comparison for classifying the four bacterial cells that cause ARI is 80% training data and 20% test data from 378 bacterial image data.

One example of the calculation of testing a random unclassified data, known features number of colonies = 1, area = 1096, perimeter = 221 and shape = 44.56 then to classify the data must be calculated probability density and probability of each class as shown in Table 4. The data are classified by determining the maximum probability value between classes. Based on the table, the maximum value is 0.0000059771718, so the random data is classified into the *M. tuberculosis* class.

TABLE 4. Calculation of probability density and inter-class probability.

The bacteria	Probability Density				Probability (P)
	Colony Counts	Area	Perimeter	Shape	P = P(X|Ci) x P(Ci)
C. diphteriae	0,227	0,0001	0,0120	0,0826	0,0000000054186
S. aureus	0,0105	0,0041	0,0016	0,0023	0,0000000000391
S. pneumoniae	0,0121	0,0135	0,0256	0,0096	0,0000000114034
M. tuberculosis	0,9876	0,0159	0,0251	0,0666	0,0000059771718

Comparison with the previously used K-Nearest Neighbor method is the next step. However, in this study we compared with the 80:20 with the value of K=3, 5, 7 as shown in Table 5.

TABLE 5. K-Nearest Neighbor accuracy, precision and recall results.

	Comparison of the data 80 : 20			
	K = 3	K = 5	K = 7	K = 9
Accuracy (%)	93,42	92,11	92,11	92,11
Precision	0,934	0,921	0,921	0,921
Recall	0,934	0,921	0,921	0,921

As shown in Table 5, the KNN method has the highest accuracy rate, 93.42 %, when comparing 80:20 data, which indicates that the KNN method is better than the Naive Bayesian method, which has the highest accuracy rate, 86.84 %. KNN methods classify data based on Euclidean distance, and then assign priority classes to K-values, while Bayesian methods classify data based on probability values between classes. Compared to previous research, using Channel Area Thresholding in the segmentation process improves the shape of bacterial cells, thus increasing the accuracy of the system.

CONCLUSION

This research aims to apply the Channel Area Thresholding method as one of the segmentation processes for early detection of bacteria causing Acute Respiratory Infection (ARI). This method uses the Naive Bayes algorithm to classify bacteria. The method could classify four bacteria causing ARI with an accuracy rate of 86.84%. When compared with the K-NN method, the best accuracy rate was 93.42%.

REFERENCE

1. R.A. Asmara, B.S. Andjani, U.D. Rosiani, and P. Choirina, "Klasifikasi Jenis Kelamin Pada Citra Wajah Menggunakan Metode Naive Bayes," *Jurnal Informatika Polinema* 4: 3 (2018) 212.
2. P. Destarianto, A.N. Noviana, Z.E. Fitri, and A.M.N. Imron, "Detection of Essential Thrombocythemia based on Platelet Count using Channel Area Thresholding," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)* 6: 1 (2022) 9–15.
3. Dinas Kesehatan Provinsi Jawa Timur, *Profil Kesehatan Provinsi Jawa Timur Tahun 2022* (Surabaya: Dinas Kesehatan Provinsi Jawa Timur, 2023) <https://dinkes.jatimprov.go.id/userfile/dokumen/PROFIL KESEHATAN JATIM 202022.pdf> Accessed.
4. Erwin and D.R. Ningsih, "Improving Retinal Image Quality Using the Contrast Stretching, Histogram Equalization, and CLAHE Methods with Median Filters," *International Journal of Image, Graphics and Signal Processing* 12: 2 (2020) 30–41.
5. Z.E. Fitri and A.M.N. Imron, "Classification of White Blood Cell Abnormalities for Early Detection of Myeloproliferative Neoplasms Syndrome Using Backpropagation," in Triwiyanto, H.A. Nugroho, A. Rizal, and W. Caesarendra, eds., *Proceedings of the 1st International Conference on Electronics, Biomedical Engineering, and Health Informatics* ,Lecture Notes in Electrical Engineering (Singapore: Springer Singapore, 2021).
6. Z.E. Fitri, L.N. Sahenda, P.S.D. Puspitasari, P. Destarianto, D.L. Rukmi, and A.M.N. Imron, "The The Classification of Acute Respiratory Infection (ARI) Bacteria Based on K-Nearest Neighbor," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi* 12: 2 (2021) 91.
7. Z.E. Fitri, L.N. Sahenda, P.S.D. Puspitasari, and A.M.N. Imron, "Identification of Bacilli Bacteria in Acute Respiratory Infection (ARI) using Learning Vector Quantization:," in presented at the *2nd International Conference on Social Science, Humanity and Public Health (ICOSHIP 2021)* (Jember, Indonesia: 2022).

8. A.T. Herawati, S.W. Megawati, V.J. Anggraeni, S. Kartadarma, A. Kandanisa, and N.L.B. Astuti, "Optimalisasi Peran Ibu dalam pencegahan ISPA (Infeksi Saluran Pernafasan Akut) Anggota Keluarga di Musim Hujan," *Jurnal Abdimas (Journal of Community Service): Sasambo* 5: 2 (2023) 368–373.
9. S. Indhira and B. Hendrik, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit ISPA," *Journal of Information System and Education Development (JISED)* 1: 3 (2023) 31–35.
10. S. ul Islam, *Infectious diseases smart study guide for medical students, residents, and clinical providers* (Amsterdam: Academic Press, 2023).
11. F.H. Kayser, ed., *Medical microbiology* ,Thieme Flexibook (Stuttgart New York: Thieme, 2005).
12. C.R. Mahon and D.C. Lehman, eds., *Textbook of diagnostic microbiology* (Sixth edition) (St. Louis, Missouri: Elsevier Saunders, 2019).
13. A.A. Mahran, R.K. Hapsari, and H. Nugroho, "Penerapan Naive Bayes Gaussian Pada Klasifikasi Jenis Jamur Berdasarkan Ciri Statistik Orde Pertama," *Network Engineering Research Operation* 5: 2 (2020) 91.
14. L. Malihah, "Tantangan Dalam Upaya Mengatasi Dampak Perubahan Iklim Dan Mendukung Pembangunan Ekonomi Berkelanjutan: Sebuah Tinjauan," *Jurnal Kebijakan Pembangunan* 17: 2 (2022) 219–232.
15. K.S. Mithra and W.R.S. Emmanuel, "Segmentation of Mycobacterium Tuberculosis Bacterium From ZN Stained Microscopic Sputum Images," in *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)* presented at the *2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (Tirunelveli, India: IEEE, 2018).
16. S.R. Reshma and T.R. Beegum, "Microscope image processing for TB diagnosis using shape features and ellipse fitting," in *2017 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)* presented at the *2017 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)* (Kollam: IEEE, 2017).