https://doi.org/10.3311/PPee.37572

Periodica Polytechnica Electrical Engineering and Computer Science, 69(1), pp. 99–109, 2025

Optimization of Bacterial Image Processing for Early Detection of Acute Respiratory Infection (ARI) Disease

Arizal Mujibtamala Nanda Imron[1*], Zilvanhisna Emka Fitri[2], Alfian Pramudita Putra[3]

[1] Electrical Engineering, Department of Electrical Engineering, Faculty of Engineering, Universitas Jember, Kalimantan Street No. 37, Bumi Tegalboto Campus, P. O. B. 159, 68121 Jember, East Java, Indonesia
[2] Informatics Engineering, Department of Information Technology, Politeknik Negeri Jember, Mastrip Street, P. O. B. 164, 68121 Jember, East Java, Indonesia
[3] Biomedical Engineering, Department of Physics, Faculty of Science and Technology, Universitas Airlangga, Dr. Ir. H. Soekarno Mulyorejo Street, Merr C Campus, 60115 Surabaya, East Java, Indonesia
[*] Corresponding author, e-mail: arizal.tamala@unej.ac.id

Received: 31 May 2024, Accepted: 22 November 2024, Published online: 13 December 2024

Abstract

Acute respiratory infection (ARI) is one of the most prevalent infectious diseases in Indonesia. ARI persists as a significant global health concern in the post-Covid-19 era. ARI – a respiratory infection – is categorized into four distinct types: pneumonia, diphtheria, pharyngitis, and tuberculosis. These diseases are caused by different bacteria. To control the transmission of these bacteria, the Ministry of Health of the Republic of Indonesia has identified the discovery of disease-causing bacteria as a crucial step. However, the issue that arises is that the limited number of medical analysts causes an extended period for the early detection of the disease, and the process is still dependent on the experience of the medical analyst. To help these issues, researchers developed a system that applies computer vision techniques to classify the types of bacteria that cause ARI using digital images. The research employs digital image processing techniques, including image quality improvement and the use of two segmentation methods (thresholding and Channel Area Thresholding). The parameters employed for classification comprise colony count, area, perimeter, and shape. This research compares the accuracy of five intelligent systems by applying them to the same data set. The SVM method achieves the highest system accuracy rate of 94.06%, while the accuracy rate of multi-layer perceptron and KNN is 93.07%, RBF is 91.09% and Naïve Bayes is 86.84% on a data comparison of 80:20 of 504 data.

Keywords

acute respiratory infection, bacteria, image processing, computer vision, KNN, SVM

1 Introduction

Acute respiratory infection (ARI) is one of the most prevalent infectious diseases in Indonesia. Respiratory infections are categorized into two types based on their location: upper respiratory tract infections and lower respiratory tract infections caused by viruses and bacteria [1]. In December 2019, a respiratory infection caused by the SARS-CoV-2 virus occurred in Wuhan, China [2]. The disease eventually became a global pandemic, leading the Indonesian government to declare it a national disaster due to its rapid spread and high mortality rate [3]. Besides viruses, ARI can also be caused by bacteria such as pharyngitis (gonorrhea, diphtheria, mycoplasma and chlamydia), peritonsillar, epiglottitis, sinusitis, bronchitis, pneumonia, meningitis and tuberculosis [1]. ARIs caused by bacteria consist of pneumonia, diphtheria, pharyngitis and tuberculosis [4]. The Ministry of the Republic of Indonesia has identified tuberculosis (TB), pneumonia and the novel coronavirus disease (COVID-19) as direct infectious diseases. Meanwhile, diphtheria has been classified as an extraordinary event (KLB) or outbreak-causing pathogen, manifesting in symptoms of pharyngitis, tonsillitis, laryngitis and tracheitis. However, the incidence of diphtheria can be reduced through vaccination [5].

The two most prevalent infectious diseases in East Java are tuberculosis (TB) and pneumonia. TB is caused by *Mycobacterium tuberculosis* bacteria and represented 78.799 cases (73.3%) in 2022. Surabaya City had the highest number of TB cases with 10.382, followed by

Cite this article as: Imron, A. M. N., Fitri, Z. E., Putra, A. P. "Optimization of Bacterial Image Processing for Early Detection of Acute Respiratory Infection (ARI) Disease", Periodica Polytechnica Electrical Engineering and Computer Science, 69(1), pp. 99–109, 2025. https://doi.org/10.3311/PPee.37572

Jember District with 5.244 cases and Sidoarjo District with 5.141 cases [6]. Pneumonia is also caused by bacteria such as *Haemophilus influenzae*, *Mycoplasma pneumoniae*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Legionella pneumophila* and *Chlamydophila pneumoniae* [7]. Pneumonia is a common disease in toddlers, with high mortality rates. Early diagnosis and intervention are crucial for reducing the mortality associated with this disease in this age group. The case detection rates in East Java were above 70% in 14 districts, while the remaining 21 districts failed to reach the target due to the impact of the epidemic. This resulted in a reduction in the number of infant visits to health services and the capacity of officers to conduct early case detection. Diphtheria is a contagious disease which can be prevented efficaciously through vaccination. The disease is caused by the *Corynebacterium diphtheriae* bacterium, which is endemic to East Java. There were 163 diphtheria cases reported from 34 districts/cities in 2022, including both laboratory-confirmed and clinically compatible cases [6].

The government employs disease control measures with the objective of reducing the incidence, prevalence, morbidity, mortality, and impact of diseases on public health [5]. One method of controlling the disease is through early detection and identification of cases. It is recommended that individuals exhibiting symptoms such as coughing, sneezing, vomiting and difficulty breathing proceed to their nearest healthcare facility for sputum or oral specimen collection for microbiological analysis, which is performed exclusively by medical analysts. Nevertheless, a significant challenge in the provision of health services is the insufficient number of medical analysts [8], contrary to the rising number of patients, especially in the post-COVID-19 era.

The evaluation process is designed to be comprehensive and meticulous to prevent misdiagnosis. The symptoms of respiratory infections, particularly those caused by bacteria, are like those of SARS-CoV-2 infection, which presents with a range of symptoms that can overlap with those of other conditions [9]. The preliminary diagnosis of patients presenting with symptoms consistent with COVID-19 is based on the examination of nasal or nasopharyngeal swabs. The virus's presence in the respiratory tract is temporary; thus, additional tests, such as those examining viral RNA and the virus's molecular structure, are necessary [10]. Although the diagnosis of pneumonia employs the conventional culture of sputum swabs to verify the presence or absence of pathogens [11], it remains challenging to identify the specific type of pathogen responsible for the pneumonia [12].

For this reason, it is necessary to create a system that can assist medical analysts in detecting, classifying and identifying pathogens for the early detection of acute respiratory infections. the focus of the pathogens we studied was on the bacteria that cause ARI, not viruses such as the cause of COVID-19.

2 Related works

Previously, we developed an expert system for the early detection of acute respiratory infections based on patient symptoms using a combination of forward chaining and certainty factor methods under the guidance of dr. Angga Madro Raharjo, Sp.P. [9]. The findings of recent research indicate that the early detection of acute respiratory infections (ARI) cannot be based on the symptoms experienced by patients themselves. The presence of specific pathogens causing ARI infections also needs to be identified. The system employs image processing techniques and artificial intelligence to classify microscopic images of bacteria present in sputum samples.

Several studies were conducted with the objective of identifying ARI bacteria through the implementation of computer vision techniques. Mithra was able to identify *Mycobacterium tuberculosis* bacteria with ZN staining, while image processing is used such as median filtering, intensity modification on grayscale images, and CAT segmentation [13]. K-Means clustering has been utilized in the detection of *Mycobacterium tuberculosis* bacteria; however, the lack of explicit selection criteria and inherent challenges in estimation have the potential to influence the accuracy of the resulting segmentation [14].

ARI diseases are not only tuberculosis, but also other diseases such as pneumonia, diphtheria and pharyngitis. A study was made, which was able to classify 5 types of ARI bacteria. The aim of this research is to classify the types of bacteria that cause ARI (tuberculosis, pneumonia, diphtheria, and pharyngitis) using the KNN method with the best accuracy of 90.63% for K values of 9 [15]. Moreover, researchers employ the LVQ method to classify bacillus-shaped bacteria that cause ARI (tuberculosis and diphtheria), with an accuracy rate of 97% at a learning rate (α) of 0.01 and a reduced learning rate (dec α) of 0.25 [8]. Finally, we analyzed the classification of ARI bacteria using the Naïve Bayes method with an accuracy of 86.84% [16].

Based on those references, the researcher developed a bacterial classification system for ARI by comparing several methods, including the Naïve Bayes method, backpropagation, radial basic function, and support vector machine. The chosen method had the highest accuracy level.

3 Materials method

Previous studies involved several research stages, including collecting datasets from sputum smear bacterial images. Then, several stages of digital image processing were carried out, such as converting color images from RGB to HSV color images.

Next step splits the components in HSV color space and segmentation process based on grey level thresholding and Channel Area Thresholding (CAT). The final stage of the image processing process involves feature extraction. This study employs morphological features, including area, perimeter, number of bacteria, and shape factor [15], as inputs for the intelligent system algorithm used to classify the bacteria dataset (Fig. 1).

3.1 ARI bacteria dataset

The bacteria dataset is unpublished primary data obtained from the Surabaya Health Laboratory Centre (BBLK) and has been validated by medical analysts, as well as by literature related to the field of medical microbiology [8, 15, 16]. Bacteria are classified into three shapes: cocci (spherical), bacilli (rod-shaped), and spirochetes (spiral) [7]. Our data include chain-shaped cocci (Streptococci), grape-like clustered cocci (Staphylococci), club-shaped and pleomorphic rods (Corynebacteria), diplococci (Neisseria) and aerobic acid-fast rod group (mycobacterium). Bacteria are cultured and Gram stained, except for *Mycobacterium tuberculosis* bacteria which are ZN (Ziehl-Neelson) stained [17] as shown in Fig. 2. The bacteria image used is 151 × 151 pixels to reduce the computational load [15].

3.2 Preprocessing

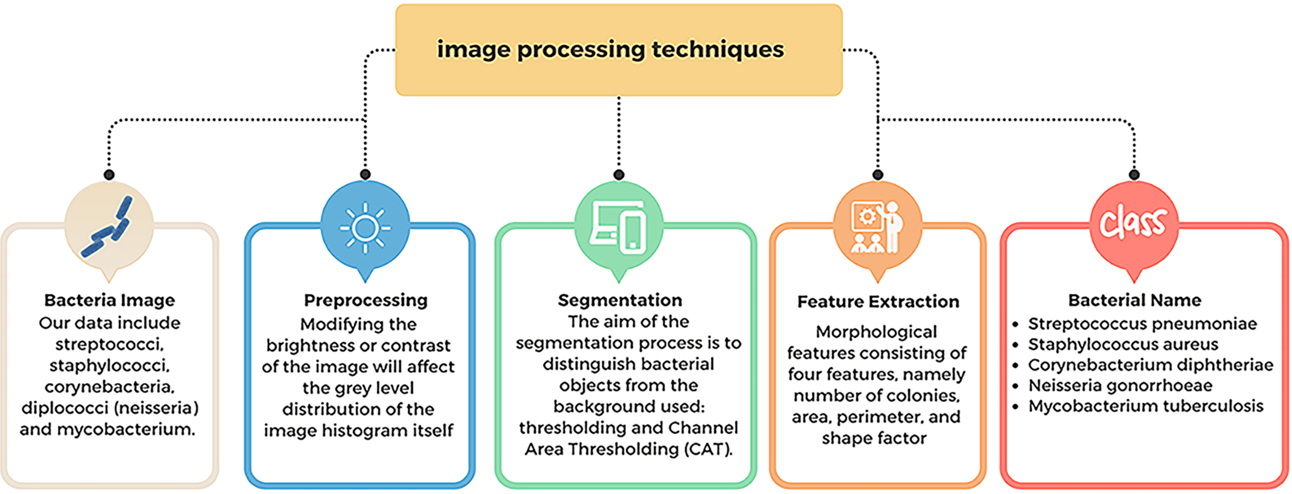
The pre-processing phase represents the initial stage of the digital processing technique. Its objective is to normalize the data collected, such that the processed data is homogeneous. The resulting homogeneous data set allows for more efficient and effective further image processing. Imron and Fitri [18] explained that the RGB color space has a large size of 24 bits, so splitting components helps the object segmentation process in RGB color space.

The objective is to identify the RGB color space component that most effectively represents the shape of the bacteria. This approach will be supplemented by the conversion to alternative color spaces, including HSV, CIELab, etc., should this be necessary [15]. Moreover, image enhancement is crucial for enhancing visual perception in computer vision. Techniques such as brightness and contrast enhancement can be employed to increase the intensity of the RGB color space [19].

Modifying the brightness or contrast of the image will affect the grey level distribution of the image histogram itself using the Eq. (1) [20].

$$\text{Brightness}(x,y) = B(x,y) + b \tag{1}$$

Where $B(x,y)$ is the input image, Brightness(x,y) is the output image resulting from the addition of brightness and b is the variable brightness value. Contrast enhancement or contrast stretching in an image applied on a linear scale.



Bacteria Image
Our data include streptococci, staphylococci, corynebacteria, diplococci (neisseria) and mycobacterium.

Preprocessing
Modifying the brightness or contrast of the image will affect the grey level distribution of the image histogram itself

Segmentation
The aim of the segmentation process is to distinguish bacterial objects from the background used: thresholding and Channel Area Thresholding (CAT).

Feature Extraction
Morphological features consisting of four features, namely number of colonies, area, perimeter, and shape factor

Bacterial Name
- Streptococcus pneumoniae
- Staphylococcus aureus
- Corynebacterium diphtheriae
- Neisseria gonorrhoeae
- Mycobacterium tuberculosis

Fig. 1 Steps in the research of the ARI bacterial classification

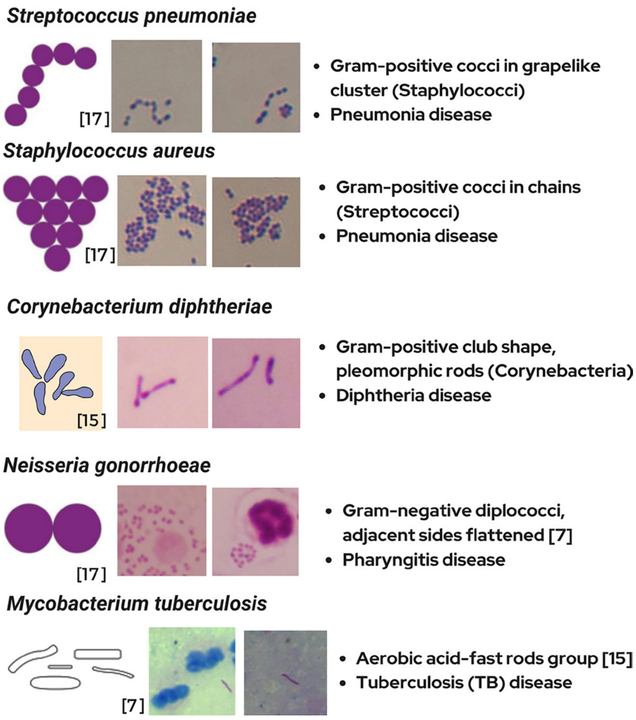


Fig. 2 Five types of bacteria in acute respiratory infection

Erwin and Ningsih [21] found that to normalize the contrast stretching process in the image, we must determine the minimum and maximum values of the image that will serve as the contrast boundary in the image, often referred to as clamping [20].

RGB color space has a range of grey intensity values from 0 to 255, so the process of adding contrast uses the Eq. (2) and then the process of clamping uses the Eq. (3).

$$\text{Contrast}(x,y) = C(x,y) \times c \tag{2}$$

$$b = \min\left(\max\left(f(a), a_{\min}\right), a_{\max}\right) = \begin{cases} a_{\min} & \text{for } f(a) < a_{\min} \\ a_{\max} & \text{for } f(a) > a_{\max} \\ f(a) & \text{otherwise} \end{cases} \tag{3}$$

Furthermore, a clamping process is required to limit the range of contrast values in the new image, so that the histogram of the image appears stretched according to the minimum and maximum values that have been defined. To improve the image quality, we also use an image subtraction operation to obtain an image that clearly shows the shape of the bacteria using the Eq. (4).

$$C(x,y) = A(x,y) - B(x,y) \tag{4}$$

$C(x,y)$ is the new image resulting from the image subtraction operation, $A(x,y)$ is the brightness image and $B(x,y)$ is the contrast image.

3.3 Segmentation

Segmentation aims to separate objects from the background, which can be considered as noise in the image. Segmentation of images is divided into three types, namely threshold-based segmentation, edge-based segmentation, watershed segmentation and region-based segmentation [22]. Segmentation based on threshold means segmentation based on one or two predetermined grey levels. This study uses a threshold value (T) using the Eq. (5). $S(x,y)$ is the input image matrix value and $S'(x,y)$ is the segmentation result image matrix value based on the thresholding value (T).

$$S'(x,y) = \begin{cases} 1, & \text{for } S(x,y) < T \\ 0, & \text{for } S(x,y) > T \end{cases} \tag{5}$$

Researchers segmented the image based on area size using the Channel Area Thresholding (CAT) method as well as implementing those methods [13] using the Eq. (6). Area_{old} is the input image, Area_{new} is the output image from the CAT segmentation process, T_{area1} and Tarea2 are the minimum and maximum threshold values of the predefined areas.

$$\text{Area}_{\text{New}} = T_{\text{area1}} \leq \text{Area}_{\text{old}} \leq T_{\text{area2}} \tag{6}$$

3.4 Feature extraction

The classification of the five types of bacterial images that cause ARI, based on the results of previous research, requires the extraction of features such as bacterial count, area, perimeter and shape factor [15] using pseudocode (Algorithm 1).

3.5 Machine learning for classification

Machine learning is a computerized model that learns patterns from data (data mining) to classify, predict and make automated decisions [23]. The selection of machine learning algorithms is a critical task that should be based on the type and amount of available data, as well as the specific

Algorithm 1 Feature extraction

```
% B_seg is the input image
CC = bwlabeln(B_seg,8);
label = max(max(CC));
S=regionprops(CC,'Area','Perimeter');
I_seg=ismember(CC, find([S.Area]>=10 & [S.Area]<=100));

J. Bacteria =label;
Area. Bacteria =S.Area;
Perimeter.Bacteria=S.Perimeter;
Shape= Perimeter.Bacteria ^2/ Area. Bacteria;
```

problem being addressed [24]. Machine learning methods are divided into two categories: unsupervised learning and supervised learning. Unsupervised learning involves grouping data into clusters using automatic methods or algorithms on unclassified data. Supervised learning involves labelled data that is categorized into the target, allowing the algorithm to predict new unclassified data [23].

This research employs traditional machine learning methods, such as K-Nearest Neighbors (KNN), Naive Bayes, Multilayer Perceptron, and Support Vector Machine (SVM), which use supervised learning algorithms. These methods require manual selection of features taken from image objects, such as area, perimeter, and object eccentricity [24]. The KNN algorithm classifies objects based on their features, which are calculated to determine neighborhood distance. The K-value is calculated and classified based on the majority vote of its neighbors [25]. Naïve Bayes classifies objects based on the probability of each independent variable. This approach considers the presence or absence of events in a particular group that is independent of the environment (other events) [26].

Multilayer Perceptron (MLP) is a feed-forward neural network algorithm consisting of an input layer, one or more hidden layers and an output layer, where each layer consists of a few neurons with an activation function [27]. Imron and Fitri [18] state that the MLP algorithm updates the weights in each layer to improve accuracy when the classification results do not match the desired target, which is an advantage of the method. Support Vector Machine (SVM) is a statistical learning algorithm that uses support vectors to create decision boundaries for classifying objects. It requires an optimal hyperplane to separate classes using these support vectors [28].

4 Experimental result

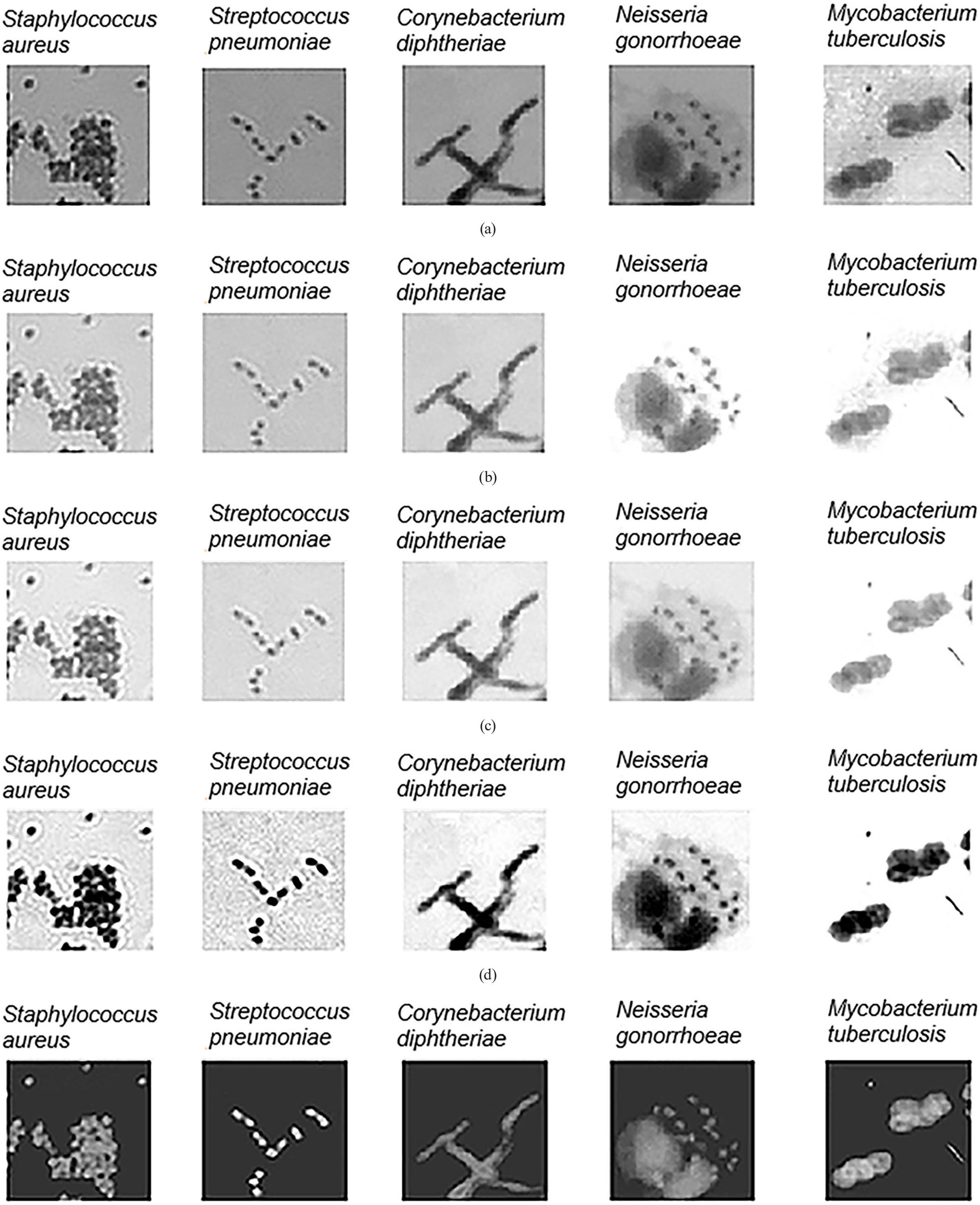
The research is initiated with a channel separation process in the RGB color space, which aims to identify the color space channel that most accurately represents the shape of the bacteria. This step is necessary given the differing color backgrounds of the bacteria. *Staphylococcus aureus* and *Streptococcus pneumoniae* have a grey background color, *Corynebacterium diphtheriae* and *Neisseria gonorrhoeae* have an orange background mixed with pink, while *Mycobacterium tuberculosis* has a blue or yellow background (Fig. 2).

The five bacteria are clearly visible in the green color channel (image with gray level in green channel in RGB color space in Fig. 3 (a)); however, the image quality could be improved by enhancing the brightness and contrast (Fig. 3). The green image is given the addition of brightness to the variable $b = 50$ in Eq. (1) (Fig. 3 (b). Then adding contrast with two values, such as $c_1 = 0.05$ and $c_2 = 0.9$ (Fig. 3 (c)). Next processed clamping by determining the value of $a_{\min} = 0$ and $a_{\max} = 255$, so that there is a contrast stretching value and produces an image with high contrast whose purpose is to clarify the shape of the bacteria.

The green channel images of the bacteria are provided in Fig. 3 (b) with the additional brightness $b = 50$. These results of the bacterial images of *Neisseria gonorrhoeae* and *Mycobacterium tuberculosis* are brighter, and there is a significant difference between the object and the background, which is certainly different from the results in the images of *Streptococcus pneumoniae*, *Staphylococcus aureus* and *Corynebacterium diphtheriae*. The images are observed to exhibit increased brightness; however, the background and object demonstrate comparable grey values. Following contrast enhancement (Fig. 3 (c)), the background of the *Mycobacterium tuberculosis* bacteria image appears brighter compared to the *Neisseria gonorrhoeae* and *Streptococcus pneumoniae* bacteria images. During the clamping process (Fig. 3 (d)), it is observed that the three bacterial images become darker than the background, thus allowing for enhanced visualization of the bacterial shape. Ultimately, the process of reducing the contrasting image matrix with the brightness image resulted in a different image (Fig. 3 (e)), which clearly revealed the bacterial shape with the background becoming black.

The image is advantageous as the background is black, allowing for enhanced segmentation of the *Streptococcus pneumoniae* and *Staphylococcus aureus* bacteria. The objective of the segmentation process is to distinguish bacterial objects from the background. The present research employs two segmentation techniques: thresholding values and area threshold values, also referred to as Channel Area Thresholding (CAT). The segmentation process uses a threshold (T) value such as $T > 50$, $T > 60$, $T > 75$, $T > 100$, $T > 150$ and $T > 200$ as shown in Fig. 4.

Fig. 4 illustrates that a value of $T > 150$ results in the elimination of numerous bacteria, accompanied by alterations in cell morphology. The optimal threshold value for the accurate representation of the shape of *Neisseria gonorrhoeae* bacteria is $T > 100$. Similarly, a value of $T > 75$ is an effective method for separating two adjacent *Corynebacterium diphtheria* bacteria. Contrary to the use of thresholds in the image of *Mycobacterium tuberculosis* bacteria, the original background image of the bacteria has several colors, namely greenish-yellow, turquoise blue, violet-yellow, violet-turquoise blue and green.



(a)

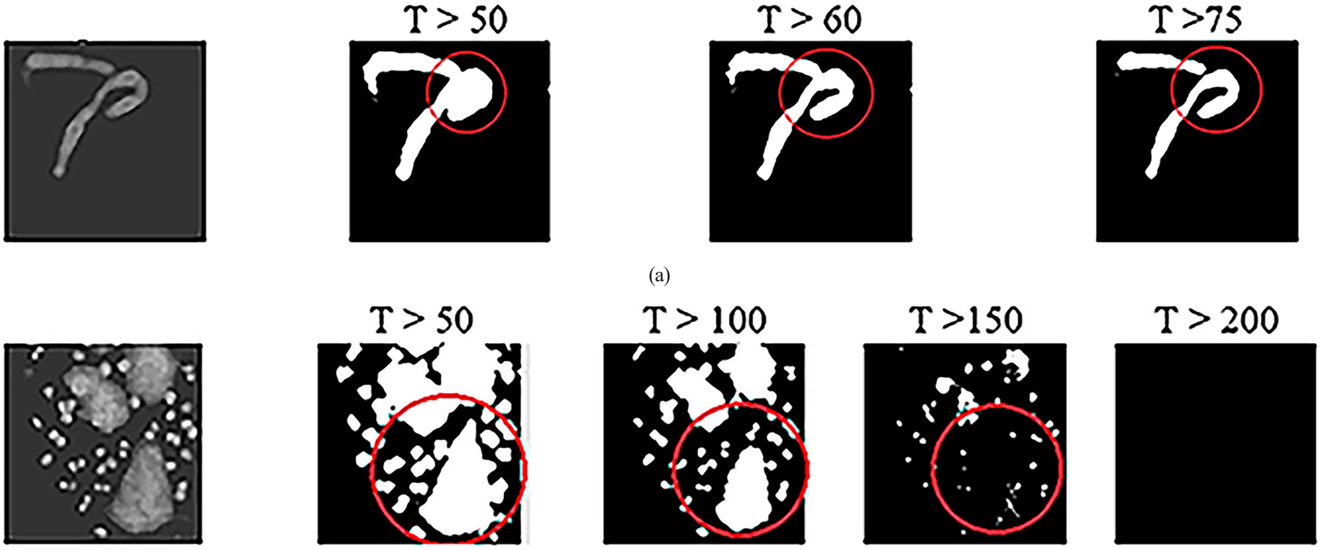
(b)

(c)

(d)

(e)

Fig. 3 Image enhancement techniques include (a) channel green image, (b) brightness enhancement, (c) contrast stretching, (d) clamping, and (e) subtraction between the contrast image and the clamping matrix



(a)

(b)

Fig. 4 Binary image segmentation results with varying *T*-values on images of bacteria, specifically (a) *Corynebacterium diphterae* and (b) *Neisseria gonorrhoeae*

This affects the determination of the threshold value in the segmentation process based on the matrix reduction value.

The optimal threshold values for representing the shape of *Mycobacterium tuberculosis* bacteria with background colors are $T > 70$ or $T > 100$ for yellow-green and purplish yellow, and $T > 230$ for turquoise blue and green. However, some bacterial binary images may require Channel Area Thresholding (CAT) segmentation to remove noise caused by paint coloration as shown in Figs. 5–8.

Fig. 5 illustrates the shape of *Neisseria gonorrhoeae* bacteria obtained using threshold values of $T_{\text{area1}} = 200$ and $T_{\text{area2}} = 10$, which eliminates the shape of blood cells as noise. The shape of *Mycobacterium tuberculosis* bacteria is obtained under two conditions:

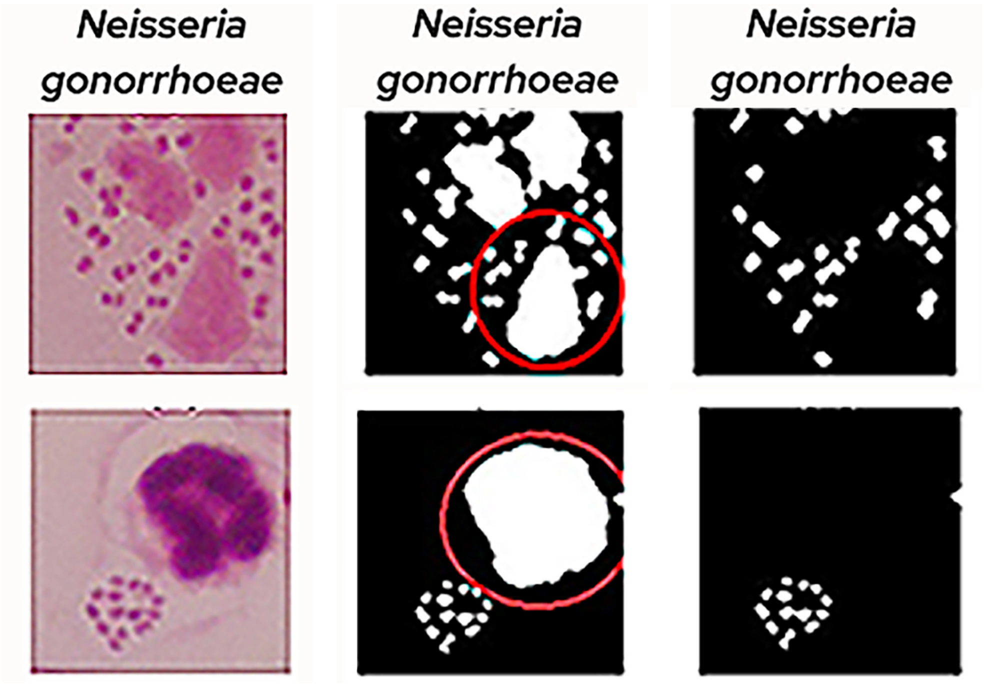
1. using a threshold value of $T_{\text{area1}} = 200$ and $T_{\text{area2}} = 50$ and
2. using a threshold value of $T_{\text{area1}} = 150$ and $T_{\text{area2}} = 100$,

in the bacterial image (Fig. 6), where the blue color indicates cells, and the stain color is noise. The process of clustering these bacteria is contrary to the case of *Staphylococcus aureu*s bacteria, as illustrated in Table 1. These bacteria are clustered like grapes to obtain the largest bacterial form, eliminating the small cocci form with a threshold of $T_{\text{area}} > 300$ unit (see Fig. 8). Thus, the resulting bacterial form is unique and differs from other bacterial forms.

To achieve a chain-like shape in *Streptococcus pneumoniae* bacteria (as illustrated in Fig. 7), a T_{area} value of greater than 50 was employed to eliminate the cocci shape (Fig. 4 (a)) and noise resulting from image blur (Fig. 3 (a)). The morphological features, consisting of four parameters (number of colonies, area, perimeter, and shape factor), were extracted and are presented in Table 1.

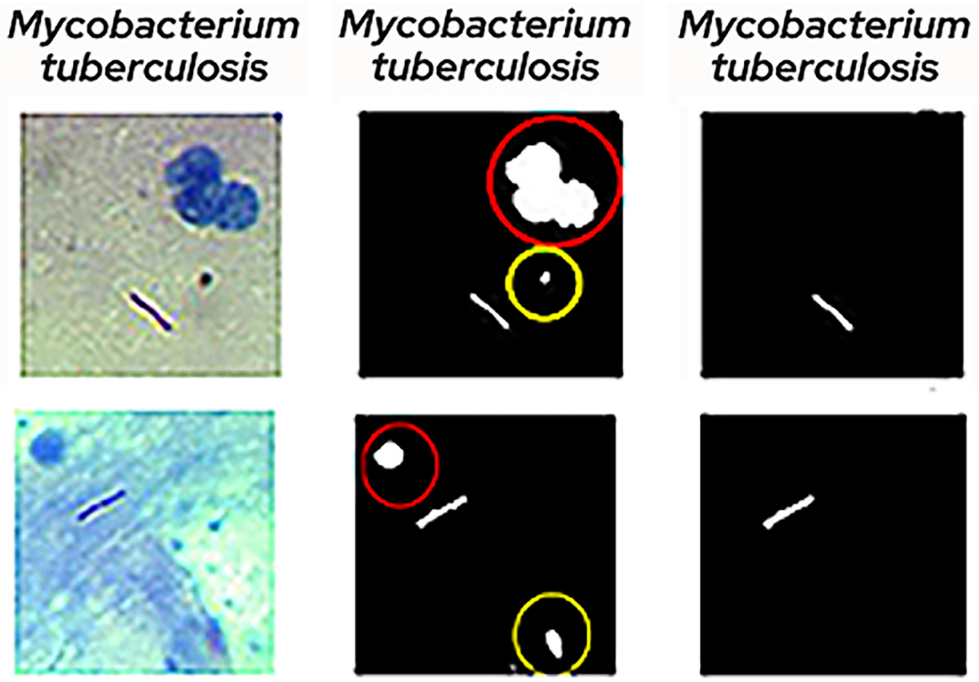
Table 1 illustrates that the greatest number of colonies is observed in *Neisseria gonorrhoeae*, with a total of 41 colonies. Conversely, the largest area is observed in *Staphylococcus aureus*, with a total of 10012 units (pixels). Area and perimeter values are directly proportional, and the calculated shape factor results are undoubtedly influenced by this proportionality. The feature values are input to the intelligent system and then compared with the accuracy of the system itself. Here, we have applied several algorithms, including K-Nearest Neighbor (KNN), Naive Bayes, Multilayer Perceptron, Radial Basic Function (RBF) and Support Vector Machine (SVM), to compare the accuracy of the system using the WEKA machine learning application. The results of this comparison are presented in Table 2.

Following the application of image processing techniques, it was found that the comparison of data set 80:20 showed the highest level of accuracy in the implementation of all five intelligent systems, with an average accuracy of 91.63%. Consequently, the system's accuracy rate has been enhanced. Fitri et al. [15] observed that the KNN method exhibited an accuracy rate of 90.63% when utilizing a value of $k = 9$. Following the development of image processing techniques and training using KNN with a value of $K = 9$, there was found to be an increase in system accuracy, reaching 93.07%.



(a) (b) (c)

Fig. 5 The (a) original image and the binary image results of (b) thresholding segmentation and (c) CAT segmentation are presented for *Neisseria gonorrhoeae* bacteria



(a) (b) (c)

Fig. 6 The (a) original image and the binary image results of (b) thresholding segmentation and (c) CAT segmentation are presented for *Mycobacterium tuberculosis* bacteria

The improvement in system accuracy was observed not only in the KNN method but also in the other four intelligent system methods. The SVM method was the most accurate, with an accuracy of 94.06%. SVM is a classification algorithm that learns using a dimensional space with a linear function, thereby dividing the data into two

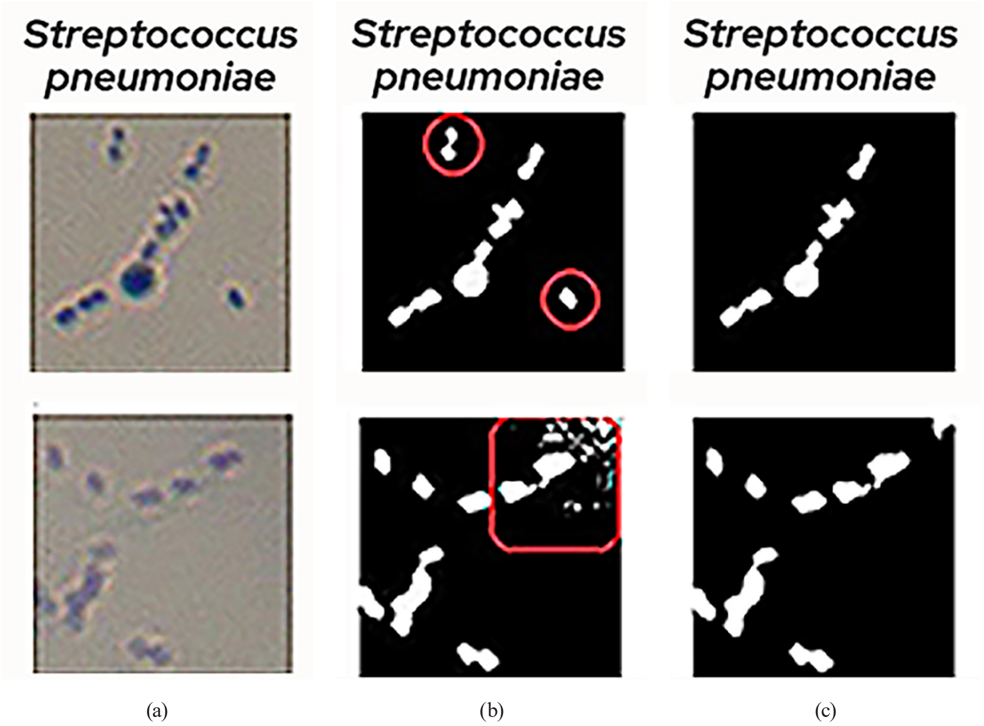


Fig. 7 The (a) original image and the binary image results of (b) thresholding segmentation and (c) CAT segmentation are presented for *Streptococcus pneumoniae* bacteria

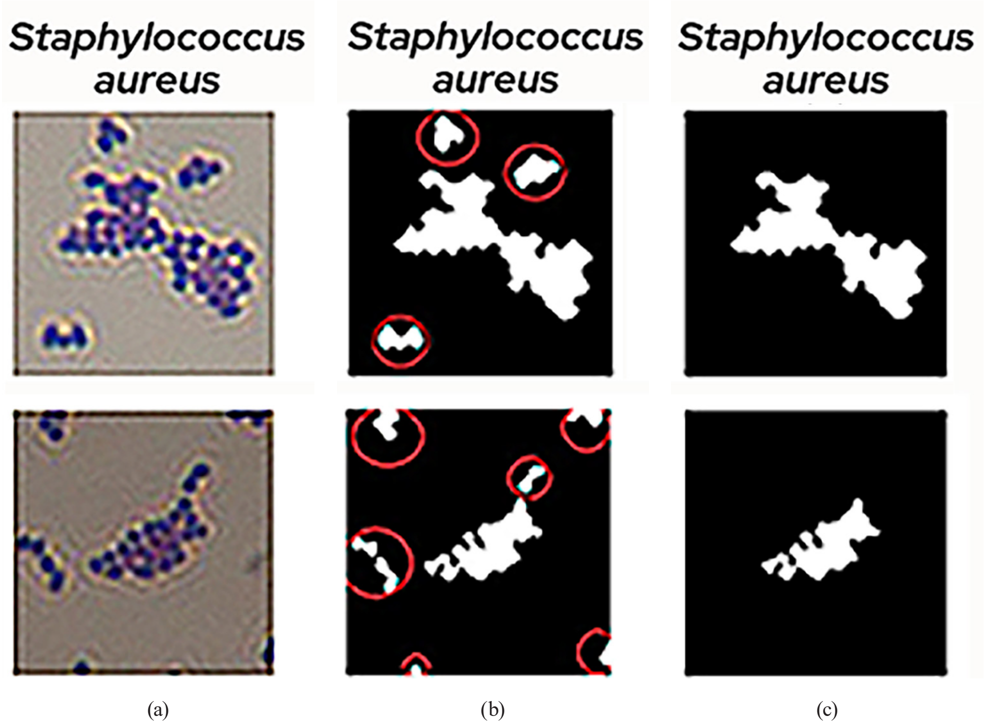


Fig. 8 The (a) original image and the binary image results of (b) thresholding segmentation and (c) CAT segmentation are presented for *Staphylococcus aureus* bacteria

classes [29]. The SVM algorithm is continually evolving, with the objective of enabling the classification of multiple classes (*Corynebacterium diphtheriae*, *Neisseria gonorrhoeae*, *Mycobacterium tuberculosis*, *Streptococcus pneumoniae* and *Staphylococcus aureus*).

Table 1 Morphological feature values for each ARI bacteria

Feature		Bacterial name				
		A	B	C	D	E
Colony count	Min	1	2	1	1	1
	Max	4	41	1	6	14
Area	Min	159	183	30	1219	494
	Max	3845	2780	311	10012	2091
Perimeter	Min	46.19	83	24.83	194	152
	Max	659.21	885	122.65	1195	609
Shape factor	Min	13.42	20.38	15.62	21.72	40.58
	Max	135.91	349.29	68.69	223.48	190.29

Description: A: *C. diphtheriae*, B: *N. gonorrhoeae*, C: *M. tuberculosis*, D: *S. aureus*, E: *S. pneumoniae*

Table 2 The accuracy of each intelligent algorithm within the system

Algorithms	System accuracy percentage (%)			
	60:40	70:30	80:20	90:10
KNN $k = 9$	91.09	91.39	93.07	88
Naïve Bayes	86.76	84.96	86.84	86.84
Multilayer perceptron	87.62	86.09	93.07	88
RBF	89.60	90.07	91.09	86
SVM	92.08	92.72	94.06	92
Average	89.43	89.05	91.63	88.17

5 Conclusion and future work

The implementation of image quality improvement processes, including the addition of brightness, contrast, clamping and matrix reduction to the image, has been demonstrated to enhance the performance of the system. This can be demonstrated by implementing the ARI bacteria dataset in the K-Nearest Neighbor method. There was an improvement in the system's accuracy, rising from 90.63% to 93.07%. The SVM method was identified as the most effective approach for classifying the types of bacteria that cause ARI, with an accuracy rate of 94.06%.

The limitations of our existing methodology require the input of multiple parameters [30] for the classification of bacteria based on their shape. The future research focus is an augmentation of data variations to overcome the limitations of the current dataset and the implementation of a Convolutional Neural Network (CNN) classification method to enhance the accuracy of the system.

Acknowledgement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Islam, S. U. "Infectious Diseases: Smart Study Guide for Medical Students, Residents, and Clinical Providers", Academic Press, 2023. ISBN 978-0-443-18742-1 https://doi.org/10.1016/C2022-0-00956-9

[2] Ciotti, M., Ciccozzi, M., Terrinoni, A., Jiang, W.-C., Wang, C.-B., Bernardini, S. "The COVID-19 pandemic", Critical Reviews in Clinical Laboratory Sciences, 57(6), pp. 365–388, 2020. https://doi.org/10.1080/10408363.2020.1783198

[3] Ministry of Health of the Republic of Indonesia "Profil Kesehatan Indonesia Tahun 2020" (Indonesia's Health Profile 2020), Ministry of Health of the Republic of Indonesia, Jakarta, Indonesia, 2021. [online] Available at: https://www.kemkes.go.id/id/profil-kesehatan-indonesia-2020 [Accessed: 24 January 2024] (in Indonesian)

[4] Struthers, J. K. "Clinical Microbiology", CRC Press, 2017. ISBN 9781315113708 https://doi.org/10.1201/9781315113708

[5] Ministry of Health of the Republic of Indonesia "Profil Kesehatan Indonesia Tahun 2022" (Indonesia's Health Profile 2022), Ministry of Health of the Republic of Indonesia, Jakarta, Indonesia, 2023. [online] Available at: https://www.kemkes.go.id/id/profil-kesehatan-indonesia-2022 [Accessed: 24 January 2024] (in Indonesian)

[6] East Java Provincial Health Office "Profil Kesehatan Provinsi Jawa Timur Tahun 2022" (East Java Province Health Profile 2022), [pdf] East Java Provincial Health Office, Surabaya, Indonesia, 2023. Available at: https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://dinkes.jatimprov.go.id/userfile/dokumen/PROFIL%2520KESEHATAN%2520JATIM%25202022.pdf&ved=2ahUKEwi23J-B1vyJAxX8ywIHHTGhB0kQFnoECBQQAQ&usg=AOvVaw18gySW3BejSzgpe_QQ0ezc [Accessed: 24 January 2024] (in Indonesian)

[7] Mahon, C. R., Lehman, D. C. "Textbook of Diagnostic Microbiology", Elsevier, 2023. ISBN 9780323829977

[8] Fitri, Z. E., Sahenda, L. N., Puspitasari, P. S. D., Imron, A. M. N. "Identification of Bacilli Bacteria in Acute Respiratory Infection (ARI) using Learning Vector Quantization", In: Proceedings of the 2nd International Conference on Social Science, Humanity and Public Health (ICOSHIP 2021), Atlantis Press, 2022, pp. 26–32. ISBN 978-94-6239-539-8 https://doi.org/10.2991/assehr.k.220207.005

[9] Fitri, Z. E., Yunita, F. P., Imron, A. M. N., Sahenda, L. N., Raharjo, A. M. "Combination of forward chaining and certainty factor methods for the early detection of Acute Respiratory Infections (ARI)", Engineering and Applied Science Research, 50(4), pp. 316–323, 2023. https://doi.org/10.14456/EASR.2023.34

[10] Falzone, L., Gattuso, G., Tsatsakis, A., Spandidos, D. A., Libra, M. "Current and innovative methods for the diagnosis of COVID-19 infection (Review)", International Journal of Molecular Medicine, 47(6), 100, 2021.
https://doi.org/10.3892/ijmm.2021.4933

[11] Fenn, D., Abdel-Aziz, M. I., Brinkman, P., Kos, R., Neerincx, A. H., Altenburg, J., …, Bos, L. D. J. "Comparison of microbial composition of cough swabs and sputum for pathogen detection in patients with cystic fibrosis", Journal of Cystic Fibrosis, 21(1), pp. 52–60, 2022.
https://doi.org/10.1016/j.jcf.2021.08.031

[12] Cho, M.-C., Kim, H., An, D., Lee, M., Noh, S.-A., Kim, M.-N., Chong, Y. P., Woo, J. H. "Comparison of Sputum and Nasopharyngeal Swab Specimens for Molecular Diagnosis of Mycoplasma pneumoniae, Chlamydophila pneumoniae, and Legionella pneumophila", Annals of Laboratory Medicine, 32(2), pp. 133–138, 2012.
https://doi.org/10.3343/alm.2012.32.2.133

[13] Mithra, K. S., Emmanuel, W. R. S. "Segmentation of Mycobacterium Tuberculosis Bacterium From ZN Stained Microscopic Sputum Images", In: 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2018, pp. 150–154. ISBN 978-1-5386-5874-1
https://doi.org/10.1109/ICSSIT.2018.8748294

[14] Sadikin, N. D. H., Sadikin, N. D. H., Marco, N., Wati, S., Agung, I. W. P. "Deteksi Penyakit Tuberkulosis Melalui Segmentasi Citra Menggunakan Algoritma K-Means" (Tuberculosis Disease Detection Through Image Segmentation Using K-Means Algorithm), JSI: Jurnal Sistem Informasi (E-Journal), 15(2), pp. 3290–3300, 2023. (in Indonesian)
https://doi.org/10.18495/jsi.v15i2.21902

[15] Fitri, Z. E., Sahenda, L. N., Puspitasari, P. S. D., Destarianto, P., Rukmi, D. L., Imron, A. M. N. "The Classification of Acute Respiratory Infection (ARI) Bacteria Based on K-Nearest Neighbor", Lontar Komputer: Jurnal Ilmiah Teknologi Informasi, 12(2), pp. 91–101, 2021.
https://doi.org/10.24843/LKJITI.2021.v12.i02.p03

[16] 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), pp. 14–22, 2024.
https://doi.org/10.20473/iapl.v5i1.55626

[17] Gnarpe, J. "Medical Microbiology for Health Professionals", Kendall Hunt Publishing, 2023. ISBN 9798765753743

[18] Imron, A. M. N., Fitri, Z. E. "A Classification of Platelets in Peripheral Blood Smear Image as an Early Detection of Myeloproliferative Syndrome Using Gray Level Co-Occurence Matrix", Journal of Physics: Conference Series, 1201(1), 012049, 2019.
https://doi.org/10.1088/1742-6596/1201/1/012049

[19] Veluchamy, M., Subramani, B. "Image contrast and color enhancement using adaptive gamma correction and histogram equalization", Optik, 183, pp. 329–337, 2019.
https://doi.org/10.1016/j.ijleo.2019.02.054

[20] Burger, W., Burge, M. J. "Digital Image Processing: An Algorithmic Introduction", Springer, 2022. ISBN 978-3-031-05743-4
https://doi.org/10.1007/978-3-031-05744-1

[21] Erwin, Ningsih, D. R. "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), pp. 30–41, 2020.
https://doi.org/10.5815/ijigsp.2020.02.04

[22] Abdulrahman, A., Varol, S. "A Review of Image Segmentation Using MATLAB Environment", In: 2020 8th International Symposium on Digital Forensics and Security (ISDFS), Beirut, Lebanon, 2020, pp. 1–5. ISBN 978-1-7281-6940-8
https://doi.org/10.1109/ISDFS49300.2020.9116191

[23] Berry, M. W., Mohamed, A., Yap, B. W. "Supervised and Unsupervised Learning for Data Science", Springer, 2020. ISBN 978-3-030-22477-6
https://doi.org/10.1007/978-3-030-22475-2

[24] Sajedi, H., Mohammadipanah, F., Pashaei, A. "Image-processing based taxonomy analysis of bacterial macromorphology using machine-learning models", Multimedia Tools and Applications, 79(43–44), pp. 32711–32730, 2020.
https://doi.org/10.1007/s11042-020-09284-9

[25] Balakrishna, K., Rao, M. "Tomato Plant Leaves Disease Classification Using KNN and PNN", International Journal of Computer Vision and Image Processing (IJCVIP)), 9(1), pp. 51–63, 2019.
https://doi.org/10.4018/IJCVIP.2019010104

[26] Arfah, J., Purnawansyah, Darwis, H., Sastra, R. "Klasifikasi Penyakit Bawang Merah Menggunakan Naive Bayes dan CNN dengan Fitur GLCM" (Classification of Shallot Diseases Using Naive Bayes and CNN with GLCM Features), The Indonesian Journal of Computer Science, 12(3), pp. 1231–1240, 2023. (in Indonesian)
https://doi.org/10.33022/ijcs.v12i3.3236

[27] Bikku, T. "Multi-layered deep learning perceptron approach for health risk prediction", Journal of Big Data, 7(1), 50, 2020.
https://doi.org/10.1186/s40537-020-00316-7

[28] Manikandan, J., Devakadacham, S. R., Shanthalakshmi, M., Arockia Raj, Y., Vijay, K. "An Efficient Technique for the Better Recognition of Oral Cancer using Support Vector Machine", In: 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 1252–1257. ISBN 979-8-3503-9726-0
https://doi.org/10.1109/ICICCS56967.2023.10142687

[29] Alita, D., Fernando, Y., Sulistiani, H. "Implementasi Algoritma Multiclass SVM Pada Opini Publik Berbahasa Indonesia Di Twitter" (Implementation of Multiclass SVM Algorithm on Indonesian Language Public Opinion on Twitter), Jurnal Tekno Kompak, 14(2), pp. 86–91, 2020. (in Indonesian)
https://doi.org/10.33365/jtk.v14i2.792

[30] Imron, A. M. N., Fitri, Z. E., Hasan, B., Widiastuti, R. Y. "Web-based Learning Media for Alphabet Recognition in Early Childhood using LeNet-5", Journal of Information Systems Research and Practice, 2(3), pp. 32–43, 2024.