Classification of features shape of Gram-negative bacterial using an extreme learning machine

B D Satoto¹, M I Utoyo², R Rulaningtyas³ and E B Koendhori⁴

¹ Doctoral Student of Applied Mathematics, Universitas Airlangga, Surabaya, Indonesia
² Mathematics Department, Universitas Airlangga, Surabaya, Indonesia
³ Physic Department, Universitas Airlangga, Surabaya, Indonesia
⁴ Microbiology Department, Medical Faculty, Universitas Airlangga, Surabaya, Indonesia

E-mail: budids@gmail.com, m.i.utoyo@fst.unair.ac.id

Abstract. Gram-negative bacteria are one of microorganism responsible for nosocomial infections in Indonesia. Nosocomial bacteria can cause nosocomial diseases, which are difficult to cure with antibiotic treatment. This bacterial observation was carried out using image processing to replace visual inspection. The process of this research consists of four stages, namely Pre-processing preparing image data, dividing objects by segmentation, obtaining and selecting features, ending with classification. At the segmentation stage, the bacterial image object was chosen that best suits the expert representation, in this case, a medical analyst. Feature extraction is done to get the pixel object information to be processed. At the classification stage, the use of extreme learning machines was chosen due to its shorter training process time. Two different bacteria were used, namely nosocomial bacteria and Gram-negative bacteria. In this research, selected bacterial Klebsiella pneumonia was obtained from 50 patients until a total of 2520 images were obtained. At the classification stage, the results of the bacterial object feature extraction are used 120 photographs in the training process and 40 image data in the testing process, with the total amount of data used, is 160 images with 512x512 pixel size and 24-bit depth. Extreme learning accuracy results are obtained in 96.71% of the process testing.

1. Introduction

Research on the bacteria related to pneumonia caused by Gram-negative bacteria has not been done in many studies. According to Grosso et al. [1], Gram-negative bacteria are bacteria that respond to the cause of pneumonia in tropical climates such as Indonesia. Gram-negative bacteria, in most cases, tend to be pathogenic. Nosocomial is Gram-negative bacteria that causing the main bacteria such as Klebsiella bacteria to become resistant [1].

In medical research, nursing, pharmacy there are studies related to Gram-negative bacteria including those conducted by Kumade et al. [2], regarding the examination of sputum pneumonia patients who showed that Klebsiella bacteria caused damage to the right upper lobe of the lung. Research by Dehghan et al. [3] discussed the awareness of patterns of Gram-negative and
antimicrobial resistance. To improve infection control systems in hospitals, selecting appropriate antibiotics are to reduce the duration of receiving Intensive Care Unit (ICU) and lowering the cost of hospitalisation, and to implement effective methods to prevent and control bacteria so that it can ultimately reduce mortality.

Gram-negative bacteria is the bacteria causing nosocomial infections, pneumonia, and meningitis [4]. It because these bacteria, in general, can cause severe illness. *Escherichia coli* is the most commonly found Gram-negative pathogenic bacterium, but other Gram-negative organisms responsible for Acquired Health Care Infections (HAI) are Klebsiella, Pseudomonas, Enterobacter, and Acinetobacter. With the illustration above, research on Gram-negative bacteria as a cause of disease and as pathogenic bacteria is necessary [5]. Shamir et al. [6] stated that a software device is needed to be able to recognise images of bacteria produced by a microscope device known as image processing software. They further added that this software consists of 4 stages, namely repairing the vision captured through the camera, segmentation, feature extraction and finally classification or identification.

This research is essential to be done to help visual observation in the laboratory. It is needed along with the development of information systems to recognise bacterial patterns. This research is one of the needs of clinical services in regional hospitals. Many previous studies have reported the identification of Gram-negative bacteria, but have not highlighted the use of machine learning for the bacterial classification process. New knowledge offered is the use of extreme education in image processing and data processing time-consuming processes that do not take long to get a model. Future research can be developed the method of optimisation research.

2. Materials and Method

According to Sousa et al. [7], the colony's characteristics were carried out based on observations on agar plates or Mac Conkey Plate. Colony shape, size, margins, elevation, colour, surface, consistency.

2.1. Gram-negative bacteria

There are few important things that can be used to recognise patterns of gram-positive and gram-negative bacteria. According to Bruner [8], is based on differences in the structure of cell walls. The Gram staining procedure was discovered in 1884 by a Danish scientist named Hans Christian Gram [9].

2.2. Shape feature

Indicators of bacterial macroscopy, according to Young [10] are in the form of dots, small, moderate or medium, large. Pigmentation (colony colour): white, yellow, red, purple, etc..

2.2.1. Area

Area (A) is the number of pixels that make up objects in an image. Unit of the area is a pixel, as shown in Figure 1.

![Figure 1. Image of bacteria (a) and shape feature approach (b)](image-url)
2.2.2. Perimeter
The length of the object's border was measured. Perimeter is the number of matching pixels found as the edge of the object \( C \), Area \( A \) and Perimeter \( P \) \[^{[11]}\], calculated using the equation below:

\[
C = \frac{P^2}{A}
\]  

(1)

2.2.3. Metric
Calculation of form factor/circle roundness was based on the following equation:

\[
\text{Slinness/Metric} = \frac{4\pi A}{P^2}
\]  

(2)

2.2.4. Eccentricity.
Eccentricity is a value that considers the distance of the minor ellipse to the focus of the central circle of an object. Eccentricity value is in the range of benefits are in the field of 0 to 1. The long-form of an object, eccentricity values is close to 1. The objects that are round/circular, eccentricity values close to 0.

\[
e = \sqrt{1 - \frac{b^2}{a^2}}
\]  

(3)

With \( e \) = eccentricity, \( a \) = mayor axis, \( b \) = minor axis \[^{[12]}\].

2.3. Extreme Learning Machine
This method works using the concept of artificial neural network algorithms with a feedforward design consisting of hidden layers. It is known as a neural layer feedforward neural networks (SLFNs). The ELM method is designed to correct one of the weaknesses that feedforward artificial neural networks have. The weakness is in terms of speed of learning. The ELM algorithm does not perform the training process of input weights and biases used. ELM trains to obtain output weights using solving the least-squares norm problem and the More Penrose inversion in a linear system. The principle is to find the node that provides the maximum output value. Parameters such as weights and biases at the input are chosen randomly. It shows how ELM can have a fast learning speed but still can produce overall performance well \(^{[13]}\).

2.3.1. Single Hidden Feedforward Neural Network (SLFNs).
Figure 2 shows the structure of extreme learning machine.

[Figure 2. The structure of Extreme learning machine \(^{[13]}\).]
The SLFN standard with hidden units of m and the activation function $g(x)$ can be described mathematically by Huang [13], as follows:

$$\sum_{j=1}^{m} \beta_j g(z_{netj}) = t_i$$

(4)

$$z_{netj} = w_{ji} \cdot x_i + b_j$$

(5)

with value:

- $i = 1, 2, ..., n, j = 1, 2, ..., m$
- $x_i$ input layer which consists of i-number of inputs
- $b_j$ is a bias vector in the $j^{th}$ hidden layer
- $w_{ji}$ is a vector weight that connects the $i$ - $t_h$ input layer to the $j^{th}$ hidden layer [14].

The value of $w_{ji}$: artificial neural network output is a vector weight that connects the $i$ - $t_h$ input layer to the $j^{th}$ hidden layer [15].

$$\sum_{j=1}^{m}||y_i - t_i|| = 0 \text{ so that } y_i = t_i$$

(6)

$$MSE = \frac{1}{n} \sum_{i=1}^{n}||y_i - t_i||^2$$

(7)

with value:

- $y_i$: Artificial neural network output value
- $t_i$: The desired target value or output for each output

For the number of input units as many as n and the number of hidden layer units as many as m, a matrix H can be composed containing the output of the hidden layer that is $m \times n$ in size [16].

2.3.2. Activation function

The activation function in the activation of neurons aims to reduce computational time. There are several functions including sigmoid, Hard limit transfer function (HARDLIM), radial basis function (RADBAS), sinusoidal (SINE) function and triangular base function (TRIBAS) [17].

2.4. Method

The research method carried out in this research is illustrated in Figure 3.

![Figure 3. Research methodology](image-url)
3. Results and Discussion

From the training process using data from Gram-negative bacteria, identification models are stored for validation and testing. The software used to compile is MATLAB version 2016a with GPU. The hardware uses an Intel core i7 laptop, 8GB RAM and Nvidia Ge force GTX1050 4GB. This research uses primary samples of Gram-negative bacterial pneumonia patients for training, testing, and validation of each class totalling 840 images so that the total images used by 2,520 images differ with a resolution of 256x256 pixels. The stages of image processing used were pre-processing to reshape image, segmentation to classify an object, feature extraction, and classification stages. In the segmentation process, fuzzy clustering level was used. At the feature extraction stage, morphological features were obtained by finding the Area, Metric, perimeter and eccentricity values. The steps of the process segmentation, feature extraction and labelling are shown in Figure 4.

![Figure 4](image)

**Figure 4.** Gram-negative bacteria: (a) original image, (b) Segmentation, and (c) feature extraction

Machine learning algorithm analysis was carried out using primary data from 50 patients who were indicated to be exposed to Gram-negative bacteria. The research revealed that 5 Gram-negative bacteria are often identified as pathogens including *Klebsiella pneumonia*, *Pseudomonas aeruginosa*, *Acinetobacter baumanii*, *Enterobacter cloacae* and *Escherichia coli* through the phoenix system test. In this research, data has been isolated and cultured for 12 hours, and then Gram staining was carried out. One image selected from the bacterium *Klebsiella pneumonia* with a resolution of 512x512 pixel and has a bit depth of 24. Table 1 shown the feature extraction result.

| No | Area  | Perimeter | Metric | Eccentricity | No | Area  | Perimeter | Metric | Eccentricity |
|----|-------|-----------|--------|--------------|----|-------|-----------|--------|--------------|
| 1  | 1,812 | 3,612     | -      | 0.50         | 14 | 363   | 76        | 0.78   | 0.81         |
| 2  | 638   | 110       | 0.66   | 0.92         | 15 | 462   | 97        | 0.61   | 0.94         |
| 3  | 361   | 80        | 0.70   | 0.89         | 16 | 62    | 34        | 0.68   | 0.93         |
| 4  | 194   | 56        | 0.76   | 0.76         | 17 | 218   | 54        | 0.92   | 0.72         |
| 5  | 723   | 114       | 0.68   | 0.92         | 18 | 28    | 18        | 1.12   | 0.74         |
| 6  | 711   | 127       | 0.54   | 0.95         | 19 | 443   | 96        | 0.60   | 0.93         |
| 7  | 251   | 64        | 0.75   | 0.82         | 20 | 72    | 33        | 0.81   | 0.76         |
| 8  | 580   | 118       | 0.51   | 0.94         | 21 | 37    | 22        | 0.94   | 0.80         |
| 9  | 540   | 111       | 0.54   | 0.95         | 22 | 117   | 228       | 0.02   | 0.99         |
| 10 | 659   | 116       | 0.61   | 0.94         | 23 | 4     | 6         | 1.39   | 0.96         |
| 11 | 98    | 42        | 0.68   | 0.79         | 24 | 15    | 28        | 0.24   | 0.99         |
| 12 | 310   | 87        | 0.50   | 0.93         | 25 | 4     | 6         | 1.39   | 0.96         |
| 13 | 599   | 131       | 0.43   | 0.96         | 26 | 81    | 160       | 0.03   | 0.99         |
Accuracy is calculated using training and testing stages. The amount of data used consisted of 140 images of Gram-negative bacteria in the training process and 50 images for the testing process. The number of hidden nodes used is between 10 and 100. There were three activation functions used, namely sigmoid, hard limit transfer learning and radians basis function. The use of the three activation functions indirectly influenced the accuracy produced. In this research, the target composed of 2 classes bacteria were used, namely Klebsiella bacteria or not Klebsiella bacteria. The accuracy results are shown in Figure 5.

Figure 5. (a) Training accuracy and testing accuracy and (b) Comparison of MSE with activation functions

Figure 5 shows the accuracy value of the training process produced is 95.92%, while in the testing process, the value obtained was 96.71%. While the error rate in Figure 5b shows the MSE using the sigmoid estimation function was 0.0328, the hard limit indicates the number of 0.880 and the radial basis function accuracy of 0.0903. Thus, the best results were using the sigmoid estimation function.

From Table 2, Average training time consuming is 0.0103 second and average testing time 0.0033 second. Average Deviation time between Training and Testing time on average is 0.0219 second.

| Hidden layer | Trainingtime | Testingtime | Train_Acc | TestAcc | devTestAcc |
|--------------|--------------|-------------|-----------|---------|------------|
| 10           | -            | -           | 0.8950    | 0.9180  | 0.0602     |
| 20           | -            | 0.0125      | 0.9171    | 0.9359  | 0.0181     |
| 30           | -            | 0.0094      | 0.9379    | 0.9487  | 0.0171     |
| 40           | 0.0297       | -           | 0.9571    | 0.9641  | 0.0216     |
| 50           | -            | -           | 0.9714    | 0.9718  | 0.0189     |
| 60           | 0.0203       | 0.0109      | 0.9779    | 0.9897  | 0.0179     |
| 70           | 0.0188       | -           | 0.9800    | 0.9821  | 0.0124     |
| 80           | -            | -           | 0.9850    | 0.9897  | 0.0179     |
| 90           | 0.0344       | -           | 0.9857    | 0.9897  | 0.0179     |
| 100          | -            | -           | 0.9857    | 0.9821  | 0.0173     |
| Average      | 0.0103       | 0.0033      | 0.9593    | 0.9672  | 0.0219     |
4. Conclusions
The ELM method can classify Gram-negative bacteria through phlegm images as well. The results of the bacterial classification process have an average accuracy rate of 96.71%. Based on system testing, the number of hidden neurons and the type of activation function in ELM dramatically affects the accuracy of the classification. The higher the number of hidden neurons used increases the level of classification accuracy. These findings confirm that the best accuracy results are obtained from the sigmoid activation function, with 50 to 100 hidden neurons. The average time needed for the testing process is 0.0033 seconds. Future research that can be done in this research is the use of the filtering layer and batch normalisation to improve accuracy and optimisation so that the number of neurons used is optimal.

References
[1] Grosso A, Famiglietti A, Luna C M 2015 Community-acquired pneumonia due to gram-negative bacteria
Community Acquire Infect. 2 117-122.
[2] Kumade E, Furusyo N, Takeshima N, Kishihara Y, Mitsumoto-Kaseida F, Etoh Y, Murata M, Hayashi J 2016 A case of lobar pneumonia and sepsis with death caused by invasive Klebsiella rhinoscleromatis infection J. Infect. Chem. 22 707-711.
[3] Dehghan F, Zolghadri N, Boostani V, Shafii A, Eftekhaari T E 2016 Resistance of gram-negative bacteria in hospital-acquired pneumonia: a prospective research Brazil. J. Infect. Diseases 20 113-114.
[4] Bassetti M, Welte T, Wunderink R G 2016 Treatment of Gram-negative pneumonia in the critical care setting: is the beta-lactam antibiotic backbone broken beyond repair? Critical Care 20 19-9.
[5] Munson D A, Evans J R 2012 Chapter 40 - Health Care–Acquired Infections in the Nursery, in Gleason C A, Devaskar S U (Eds.) Avery's Diseases of the Newborn, Ninth Edition, Elsevier Saunders, Philadelphia, pp. 551-564.
[6] Shamir L, Delaney J D, Orlov N, Eckley DM, Goldberg I G. 2010 Pattern recognition software and techniques for biological-image analysisPLoS Comp. Biol. 6 11 1-10.
[7] Sousa A M, Pereira M O, Lourenço A 2015 MorphoCol: An ontology-based knowledgebase for the characterisation of clinically significant bacterial colony morphologies J. Biomed. Inform. 55 55-63.
[8] Bruner D W 1933 Differentiation between Gram-positive and Gram-negative microorganisms by the use of enzymes J. Bacteriol. 26 4 361-371.
[9] Coico R 2005 Gram staining Curr. Protoc. Microbiol. Appendix 3, A.3.C.1-A.3.C.2.
[10] Young K D 2007 Bacterial morphology: why have different shapes? Current Opinion Microbiol. 10 596-600.
[11] Bribiesca E 2010 Computation of the Euler number using the contact perimeter Comp. Math. Appl. 60 1364-1373.
[12] Caccamo P D, Brun Y V 2018 The molecular basis of noncanonical bacterial morphology Trends Microbiol. 26 191-208.
[13] Huang G B, Zhu Q Y, Siew C T S 2004 Extreme learning machine: a new learning scheme of feedforward neural networks 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541) 2 985-990.
[14] Huang J, Yu Z L, Gu Z 2018 A clustering method based on extreme learning machine Neurocomp. 277 108-119.
[15] Li L, Wang C, Li W, Chen J 2018 Hyperspectral image classification by AdaBoost weighted composite kernel extreme learning machines Neurocomp. 275 1725-1733.
[16] Park Y, Yang H S 2019 Convolutional neural network based on an extreme learning machine for image classification Neurocomp. 339 66-76.
[17] Peng J, Li L, Tang Y Y 2013 Combination of activation functions in extreme learning machines for multivariate calibration Chemometrics Int. Lab. Sys. 120 53-58.

Acknowledgement
The author thanks the Dr Microbiology Lab team at Dr Hospital. Soetomo in Surabaya who has provided primary data support. And the promoters and co-promoters who have permitted this research publication