IDENTIFICATION OF POSITIVE GRAM BIOPESTICIDE BACTERIA USING FUZZY CLUSTERING LEVEL SET and RANDOM FOREST

Budi Dwi Satoto¹, Imam Utoyo², Riries Rulaningtyas³ and Muhammad Yusuf⁴

¹,⁴ Information system, Engineering Faculty, Trunojoyo University of Madura
²,³ Faculty of Science and Technology, Airlangga University of Surabaya

Abstract. The quality of agricultural land is an important factor for corn farmers in Madura. Control of maize plants is affected by the use of pesticides when viewed from a positive perspective can help humans in terms of eradicating pests that damage agricultural land. But on the other hand, pesticides also have a negative impact on humans and the surrounding environment, namely the breakdown of the food chain. To overcome this problem, biopesticides can be used in the form of bacteria that can kill plant pest organisms (OPT). Bacillus thuringiensis is one of the bacteria that can produce protein crystals that are insect killers (insecticides) when undergoing a sporulation process. By studying this identification, it is expected that farmers can analyze the use of pesticides and replace them with bio-pesticides that are more environmentally friendly. In this study, an image processing approach was used to identify the presence of biopesticide bacteria. At the pre-processing stage, the stages of a culture of bacterial colonies were carried out on Blood Agar Plate media, followed by repairs to the results of image morphology. Fuzzy Clustering level set is one of the methods used in the image segmentation process. The results of form extraction are then used in the training process to determine the type of bacteria in the sample. This method makes identification of agricultural land easier, faster and costs less. The result is that in processing 100 training data and 25 data testing using Bacilli bacteria with 9 bacterial morphological attributes and 2 identification classes the accuracy value of the Random Forest decision tree was 91%

Keyword: Bacillus thuringiensis, Form extraction, Fuzzy Clustering level set, Random Forest

1. Introduction

The role of bacteria in human life is divided into two parts: beneficial bacteria and harmful bacteria. The beneficial bacteria include decomposing bacteria. These bacteria break down proteins, carbohydrates, and other organic compounds into CO₂, ammonia gas, and other simpler compounds. Therefore the presence of these bacteria plays an important role in mineralization in nature and in this way bacteria cleanse the world of organic waste [1, 2].
With the increasing population growth, the fulfilment of food needs is a necessity. In order to meet the needs of the community as consumers, it can be improved by maximizing the yield of agricultural processes so as to increase crop yields as food products. Agricultural processes that are not running optimally are caused by pest attacks on agricultural land. Some pests that commonly interfere with agricultural crops are from insects such as caterpillars, beetle larvae, and fruit flies. The insect can eat the leaves, gnaw at the stem and roots, and decompose the fruit. The solution made by farmers in controlling the attack of insect pests is to use pesticides. [3, 4].

Spraying pesticides effectively can kill insects because of the effects of the toxic chemicals they contain. But the effects of chemicals contained in pesticides are not only effective in killing insects but also are toxic to humans and other animals when they accumulate in the body. To reduce the impact of these pesticides, an insecticide is needed that is effective and does not endanger other organisms involved, namely biopesticides or biological pesticides. One finding that is effective enough to eradicate disruptive insects but is safe for other organisms especially humans is the use of Bacillus thuringiensis bacteria in agriculture [5, 6].

This paper describes how to identify thuringiensis bacteria as an indicator of the use of biopesticide ingredients to reduce pests. This research is needed considering there are pest-prone agricultural lands that are difficult to control while excessive use of insecticides can cause damage to the environment and ecosystems. The research gap with previous research is that previous studies identified Bacillus thuringiensis bacteria using visual and biological identification, in this study using a morphology and machine learning approach to determine the percentage of land that must be carried out by biopesticides. The purpose of this research is primarily to find out the condition of land that is free of pests and ready for use by farmers based on indicators of the presence of Bacillus thuringiensis bacteria. Farmers do not require the use of insecticides in the area in their care so as not to damage the land and agricultural products.

2. Literature review

In the literature discussed about the role of bacillus thuringiensis bacteria as biopesticides, bacterial isolation, previous research on identification of thuringiensis bacteria, image processing, Fuzzy Clustering Level set and Random Forest algorithms.

2.1. The role of Bacillus thuringiensis bacteria

Bacillus thuringiensis (Bt) is a rod-positive, aerobic and spore-forming gram-positive bacterium. Many strains of this bacterium produce proteins that are toxic to insects. Since the potential of Crystal or cry Bt proteins as insect control agents is known, various Bt isolates contain various types of crystal proteins. Therefore Bacillus thuringiensis (Bt) bacteria are widely used as an alternative to plants that are resistant to pests [7]. Bacillus thuringiensis isolates can be isolated from soil, plant parts, animal feces, insects and carcasses and other sources. [8]. Morphological characteristics of the bacterial colony of Bacillus thuringiensis on Blood Agar Plate show a solid colony growth shape and glossy appearance. Whereas the morphological characteristics of the bacterium Bacillus thuringiensis after Gram staining showed that the bacterial cells were blue-purple, rod-shaped and Gram-blue[9, 10].

2.2. Previous research

According to Temel Gokturk (2016) to determine the efficiency of biopesticides that will be used in the struggle against Ricania simulans, pests that eat hundreds of products by absorbing juice extracts and causing danger on the Black Eastern coastline, the Turkish Sea Region. Vegetable pesticides used are Pyrethrum and Bacillus thuringiensis which have been proven effective in eradicating nymphs and juvenile Ricania simulans during the period of pest attack [7]. In the Subbanna research (2019), the pesticide properties of Bacillus thuringiensis and related toxic proteins are a science that continues to evolve with potential implications in the management of biological pests[8]. The distinguishing element in this study was that the isolates of thuringiensis bacteria were identified in the corn plant area which
was suspected of having a rat attack using an image processing approach. Some samples are taken and then checked whether there is an isolate. If not, then thuringiensis bacteria can be cultured during the period of pest attack and do not need to spray insecticides

2.3. Image Processing

Long (2017) defines Digital Image Processing (Digital Image Processing) as a field of study about how an image is formed, processed, and analysed so as to produce information that can be understood by humans. Stages in image processing include Pre-processing, Segmentation, Feature Extraction and Identification [11, 12].

2.3.1. Fuzzy Clustering Level Set

Balla Arabe et al (2013), conducted research by paying attention to the development of parallel programming and the Lattice Boltzmann (LBM) method by offering an alternative approach to solving partial differential equations. Using the gradient descent method, an appropriate level set equation was obtained, namely fuzzy power for LBM solvers based on Zhao’s model [13, 14]. FCM assigns pixels to each class using the fuzzy membership function. Think of \( X = (x_1, x_2, x_3 \ldots x_N) \) shows images with N pixels categorized into Clusters[15]. FCM is the repeated minimization of the following object functions:

\[
J = \sum_{j=1}^{N} \sum_{i=1}^{C} \mu_{ij}^{m} \left\| x_j - u_i \right\|^{2}
\]

with \( \mu_{ij} \) is membership of pixels \( x_j \) in cluster ke-i, \( u_i \) is the center of the i-cluster, m is the fuzzifier control parameter that controls obscurity resulting in partitions and is located between 1 < m ≤ ∞ and \( ||.|| \) is a matrix norm. Usually Euclidean distance between pixels \( x_j \) and center of cluster \( u_i \), norm matrix are used. The membership function and the cluster center are updated to be

\[
\mu_{ij} = \left( \sum_{k=1}^{C} \frac{||x_j - u_k||^{2/(m-1)}}{||x_j - u_k||} \right)^{-1/m} \quad \text{and} \quad u_i = \frac{\sum_{j=1}^{N} \mu_{ij}^{m} x_j}{\sum_{j=1}^{N} \mu_{ij}^{m}}
\]

Cluster centres can be initialized randomly or with estimates. In images that are audio, FCM does not include spatial information that makes it sensitive to noise and other images. A level set is a method for detecting a moving curve. Opening performs a curve search outside the object of an image. While Closing searches curves in the object of an image[16].

In this study, the shape morphology obtained through Perimeter values (comparison between circumference with length and width), Metric (form factor/roundness circle) and Eccentricity (the comparison of distances from minor ellipses with a major ellipse function of a circular object) for feature selection

2.3.2. Random forest Classification

Random Forest algorithm is a continuation of the Classification and Regression Tree (CART) method in which the bootstrap aggregating (bagging) method is implemented and there is a feature selection that is not according to standard rules. The RF method combines many trees, unlike a single tree which only consists of one tree to make classification and prediction classes. The selection of variables to be used for the split process is taken irregularly [17, 18]. The advantage of the Random Forest application is that this algorithm can overcome data problems that have incomplete attributes, can be implemented for classification or regression processes but are not very good for regression, are more suitable for classifying data and can be implemented to overcome large amounts of data samples. The classification process in the Random Forest begins by randomly dividing existing sample data into a Decision tree. After the tree is formed, voting will be carried out in each class from the sample data. Then, a combination of votes from each class is then performed with the most votes taken. By using Random Forest in the data classification, it will produce a better vote [19, 20].
3. Research methodology

The methodology used for bacterial identification is the image processing approach. This method was chosen to replace visual observation macroscopy because it would save time in the identification process. The research step is shown in the flowchart of Figure 1.

![Flowchart of Research Methodology](image)

From Figure 1, it can be explained that the research method begins with the pre-processing specimen, which is taking a bacterial sample and breeding it in the Blood Agar Plate media. Results can be obtained after the bacteria have been isolated for a maximum of 12 hours. Then Gram staining was carried out. To observe this bacterial image, using a fluorescence microscope with a 10x magnification lens and 100x an objective lens so that a 1000x magnification is obtained. The training process is based on the histogram area and the ovality level of the object area using random forest as shown in figure 2.

4. Result

The results of this study are in the form of a process of identifying bacterial images under a microscope that shows the presence of bacillus thuringiensis bacteria from objects observed after going through the image processing process. The results of the segmentation process with Fuzzy Clustering Level Set as shown in Figure 3.

![Segmentation object](image)

![Labelling object](image)
After the segmentation process is completed, the next step is the process of extracting morphological features as shown in Figure 4. The results from feature extraction or feature selection are used to obtain ground truth after being validated by microbiologists. The approved object is used as a reference object on observation. The attributes used for feature selection are using the Estimated, Perimeter, Metric and Eccentricity area values as shown in Table 1.

| No. of Object | Area   | Perimeter | Metric | Eccentricity | No. of Object | Area   | Perimeter | Metric | Eccentricity |
|---------------|--------|-----------|--------|--------------|---------------|--------|-----------|--------|--------------|
|  3            | 98     | 106.      | 0.10   | 0.98         |  32           | 148    | 155.      | 0.39   | 0.98         |
|  8            | 691    | 179.      | 0.27   | 0.98         |  33           | 94     | 34.9      | 0.96   | 0.76         |
|  9            | 75     | 30.7      | 0.99   | 0.77         |  35           | 758    | 155.      | 0.39   | 0.98         |
| 10            | 139    | 43.2      | 0.93   | 0.68         |  36           | 80     | 32.7      | 0.93   | 0.81         |
| 11            | 79     | 33.5      | 0.88   | 0.85         |  39           | 157    | 127.      | 0.12   | 0.81         |
| 16            | 422    | 276.      | 0.06   | 0.95         |  42           | 114    | 39.5      | 0.91   | 0.76         |
| 17            | 96     | 107.      | 0.10   | 0.97         |  44           | 89     | 34.1      | 0.95   | 0.74         |
| 18            | 161    | 47.7      | 0.88   | 0.78         |  47           | 204    | 64.7      | 0.61   | 0.79         |
| 19            | 71     | 34.7      | 0.73   | 0.91         |  48           | 74     | 30.7      | 0.98   | 0.76         |
| 20            | 70     | 29.8      | 0.98   | 0.78         |  49           | 161    | 46.6      | 0.93   | 0.72         |

In machine learning using ID3 J.48 and Random Forest, data from feature extraction is tested to obtain the accuracy value of Precision, Recall and F-Measure as shown in Table 2.

| TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class       |
|---------|---------|-----------|--------|-----------|----------|-------------|
| 0.932   | 0.122   | 0.917     | 0.932  | 0.924     | 0.965    | b.thuringiensis |
| 0.878   | 0.068   | 0.9       | 0.878  | 0.889     | 0.965    | fungi or other |

Table 2 shows the accuracy of the training process using 100 data. The result is an accuracy rate of 91.7%. This research hopes to get an indicator of land that has not been carried out by bioinsecticides so that farmers can make initial anticipation if there are crop pests.

Conclusion
The results of this study are that the process of identifying Gram-positive bacillus thuringiensis bacteria can be carried out, where samples can be obtained from land that has not been done bio-insecticide. Land becomes pest-resistant by paying attention to the use of bioinsecticides which are carried out by considering the duration of pest attacks on plants. If compared to previous studies, there has never been an identification stage with machine learning. Previous research discusses the enzymes and toxicity of bacillus thuringiensis bacteria as pesticide ingredients. This research has contributed to using an image processing approach to identify bacteria. The method used is the Fuzzy Clustering Level Set algorithm in the segmentation process and Random Forest to identify 100 samples of the Region of interest in Bacilli bacteria. The result is the level of identification accuracy of 91%

Acknowledgment
Thank you to the Faculty of Engineering, University of Trunojoyo, who helped publish this paper.

References
[1] Subbanna, A.R.N.S., H. Rajasekhara, et al., 2018, Pesticidal prospectives of chitinolytic bacteria in agricultural pest management, Soil Biology and Biochemistry, 116, 52-66.
[2] Kumar, B.L. and D.V.R.S. Gopal, 2015, Effective role of indigenous microorganisms for sustainable environment, 3 Biotech, 5, 867-876.
[3] War, A.R., M.G. Paulraj, et al., 2012, Mechanisms of plant defense against insect herbivores, Plant signaling & behavior, 7, 1306-1320.

[4] Bravo, A., D.L. Martínez de Castro, et al., 30 - Mechanism of action of Bacillus thuringiensis insecticidal toxins and their use in the control of insect pests, in The Comprehensive Sourcebook of Bacterial Protein Toxins (Fourth Edition), J. Alouf, D. Ladant, and M.R. Popoff, Editors. 2015, Academic Press: Boston. p. 858-873.

[5] Akhtar, M.W., D. Sengupta, and A. Chowdhury, 2009, Impact of pesticides use in agriculture: their benefits and hazards, Interdisciplinary toxicology, 2, 1-12.

[6] Damalas, C.A. and I.G. Eleftherohorinos, 2011, Pesticide exposure, safety issues, and risk assessment indicators, International journal of environmental research and public health, 8, 1402-1419.

[7] Palma, L., D. Muñoz, et al., 2014, Bacillus thuringiensis toxins: an overview of their biocidal activity, Toxins, 6, 3296-3325.

[8] Schuch, R. and V.A. Fischetti, 2009, The Secret Life of the Anthrax Agent Bacillus anthracis: Bacteriophage-Mediated Ecological Adaptations, PLOS ONE, 4, e6532.

[9] Monnerat, R.G., C.M. Soares, et al., 2009, Translocation and insecticidal activity of Bacillus thuringiensis living inside of plants, Microbial biotechnology, 2, 512-520.

[10] Rampersad, J. and D. Ammons, 2005, A Bacillus thuringiensis isolation method utilizing a novel stain, low selection and high throughput produced atypical results, BMC microbiology, 5, 52-52.

[11] Long, B., D. Long, and A. Koyfman, 2017, Emergency Medicine Evaluation of Community-Acquired Pneumonia: History, Examination, Imaging and Laboratory Assessment, and Risk Scores, The Journal of Emergency Medicine, 53, 642-652.

[12] Li, C., X. Jia, et al., 2011, Digital image processing technology applied in level measurement and control system, Procedia Engineering, 24, 226-231.

[13] Balla-Arabé, S., X. Gao, and B. Wang, 2013, A Fast and Robust Level Set Method for Image Segmentation Using Fuzzy Clustering and Lattice Boltzmann Method, IEEE Transactions on Cybernetics, 43, 910-920.

[14] Benamour, M., E. Liberge, and C. Beghein, 2017, A new approach using lattice Boltzmann method to simulate fluid structure interaction, Energy Procedia, 139, 481-486.

[15] Chaghari, A., M.-R. Feizi-Derakhshi, and M.-A. Balafar, 2018, Fuzzy clustering based on Forest optimization algorithm, Journal of King Saud University - Computer and Information Sciences, 30, 25-32.

[16] Picelli, R., S. Townsend, et al., 2018, Stress-based shape and topology optimization with the level set method, Computer Methods in Applied Mechanics and Engineering, 329, 1-23.

[17] Kursa, M.B., 2014, Robustness of Random Forest-based gene selection methods, BMC Bioinformatics, 15, 8-8.

[18] Shaihina, T., D. Lowe, et al., 2017, Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation, Biomedical Signal Processing and Control.

[19] Chaudhary, A., S. Kolhe, and R. Kamal, 2016, An improved random forest classifier for multi-class classification, Information Processing in Agriculture, 3, 215-222.

[20] Subudhi, A., M. Dash, and S. Sabut, 2019, Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier, Biocybernetics and Biomedical Engineering.