Mobile Application for Tomatoes and Potatoes Diseases Identification using Leaf Analysis

Iwan Syarif, Achmad Basuki, Nana Ramadijanti, Tessy Badriyah, Edi Satriyanto, Wahyu Abid Arfiyanto
Politeknik Elektronika Negeri Surabaya, Indonesia
{iwanarif, basuki, nana, tessy, kangedi}@pens.ac.id, wahyu.abied@gmail.com

Abstract— **Several types of diseases affect the tomato and potato plants, which could potentially damage the plants and even lead to crop failure. Among many diseases, some attack specifically tomato and potato plants' leaves such as the early and late blight diseases. These are the most dangerous plant diseases for plants growing in subtropical and tropical regions. These diseases cause crop failure. Early detection of the existence of this disease is crucial to ensure maximum yield. In this study, we develop mobile applications to detect diseases affecting tomato and potato plants. This application detects early blight or late blight on leaves by capturing leaf images through the smartphone's camera directly. We use the Gabor filter for image feature extraction and naïve Bayes algorithm for identifying the leaf condition. Our application could detect diseases on potato plants with a 92% accuracy rate and 78% on tomato plants.**

Keywords— Data Mining, Image Processing, Gabor Filter, Naïve Bayes

I. Introduction

Early and late blight are diseases that affect many plants grown in sub-tropical and tropical regions. They are considered threatening to plants as they can cause death, which, in effect, create huge losses for farmers due to crop failure [1].

Athanikar et al. [2] carried out a study related to the detection of diseases on potato plants by examining the shape and color of leaves. This study uses three labels, namely "Healthy", "Early Blight" and "Insect Damage". They use the k-Means algorithm and Backpropagation Neural Network. It is concluded that BPNN could effectively show critical areas showing signs of disease with an accuracy rate of 92%. Hence, this is considered a suitable system to detect and clarify diseases on potato plants, contributing to the reduction in crop losses for farmers [3].

Another similar study is conducted by Debabrata et al. [3]. They use a color histogram to distinguish between healthy and diseased potatoes. However, rather than the leaves, the potato itself is examined for signs of disease. The goal of the study is to detect scab disease on potatoes using a color histogram and k-means method, which are shown to be considered accurate. Melik Sarogan et al. [4] identifies four different types of diseases affecting the leaves of tomato plants such as "Late Bright", "Septoria", "Yellow Leaf Curl" and "Bacterial Spot". To increase the accuracy of the classification, they use several filters and different measures of convolution.

A similar study is done by Chun-Hua et al. [1] by analyzing leaves and fuzzy set algorithm. This experiment is conducted on several pictures of plants, which then show the effectiveness of methods are suggested. They used several edge detection algorithms to analyze plant leaves. The other study proposed Local Gabor Phase Quantization for classifying leaf [5]. This study provided a Gabor Phase to determine feature values for finding the shape and texture of the leaf. Madiwalar presented the comparative study for identifying plant disease with some methods [6].

In this study, we propose a new method to develop a mobile application that can identify Early and Late Blight diseases on tomato and potato plants. This application uses the smartphone's camera, with the possibility of further application on other types of plants. We use the Gabor filter for image extraction and Naïve Bayes algorithm for the classification process.

II. System Design

The system design of the application is shown in Figure 1, including the Training and Testing Phases.

2.1 Training data

The labelled data is obtained from a repository, open-sourced project owned by P. Mohanty [7]. In total, 120 sample pictures of potatoes leave and 120 pictures of tomatoes leave are used as data training. Of the 120 pictures of leaves, 40 are infected by Late Blight, 40 Early Blight and 40 are considered healthy. The composition for photos of potatoes have the same distribution.

2.2 Preprocessing

The pre-processing part of the data includes three steps, namely resizing, background removal and grayscaling.

2.2.1 Resizing and GrayScaling

Resizing is a method that alters the size of an image. In this study, we specifically alter each image to have the dimensions of 256x256 pixel. In this case, the purpose of gray scale is to simplify an image, thus reducing the processing time needed to examine the image. The equation used for this method is shown below:

2.2.2 Remove Background

In this process, the researcher uses GrabCut algorithm, which will be used to extract foreground by minimizing user interaction [7]. Gaussian Mixture Model is a probability distribution that can be used background removal techniques. The GMM formula is explained in the equation (1).

$$D(N) = -log \sum_{i=1}^{K} \pi(a_n, i) \frac{1}{\sqrt{det \sum(a_n, i)}} \tag{1}$$

Where:

$$\pi(a_n, i) = e^{-\frac{1}{2}[Z_n - \mu(a_n, i)]^T \sum(a_n, i)^{-1} [Z_n - \mu(a_n, i)]}$$

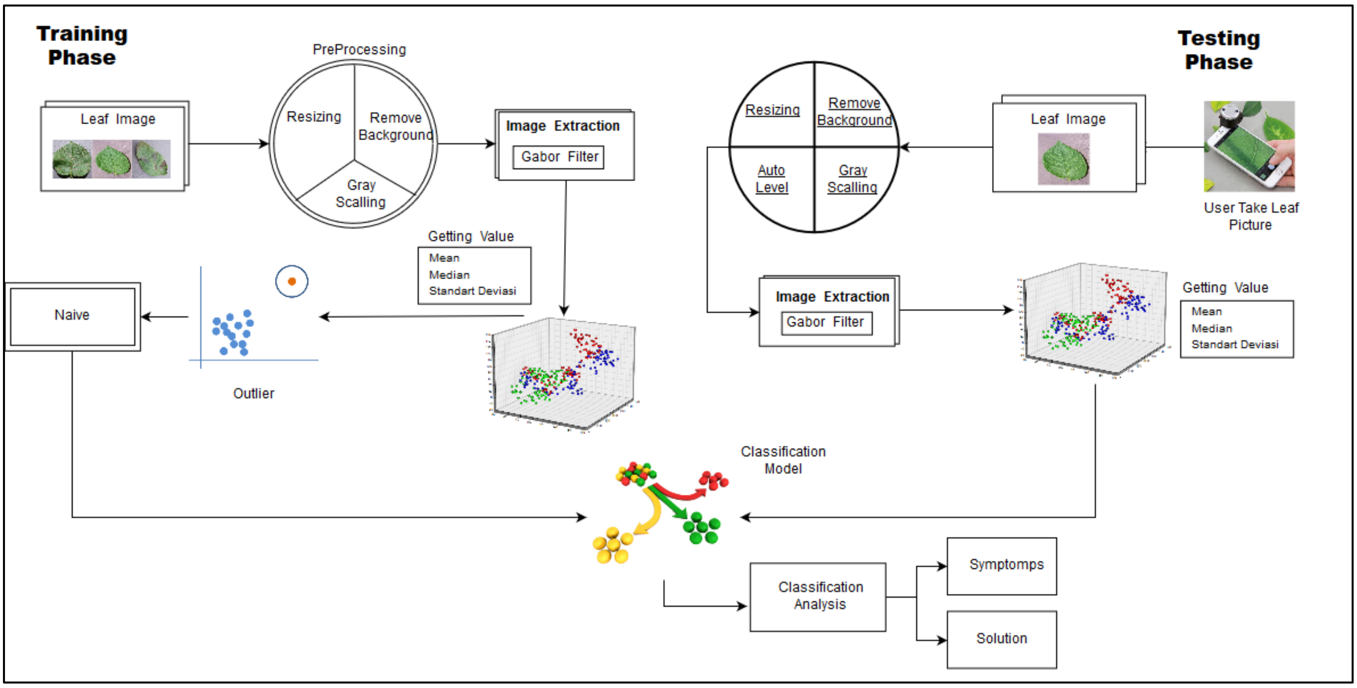


Figure 1. System Design

2.2.4 AutoLevel

Auto level is one of image enhancement methods using histogram equalization. In this research, auto level is used to enhance the color level and contrast the leaf images which taken by smartphone's camera. There are two methods for auto-level; histogram equalization and auto stretching. This research use auto-stretching using balance distribution [8]. The equation for this method is shown below:

$$x' = 255\left(\frac{x - x_{min}}{x_{max} - x_{min}}\right) \quad (2)$$

where x is original image and x' is result image

$x_{min} = \min(x) + \partial_L$

$x_{max} = \max(x) + \partial_R$

∂_L and ∂_R is half distance between mean(x) and median(x).

2.3 Image Extraction using Gabor Filter

We use Gabor filter for image extraction. This Filter, founded by Dennis Gabor, is a linier filter that can be used in various image processing application to detect edges, texture feature extraction, etc. There are several parameters that control how Gabor filter is shaped and which features are responded. Filter 2D Gabor can be seen as a sinusoidal signal from certain frequency and orientation, which are modulated by Gaussian waves [9]. The filter has real and imaginary components that represent orthogonal direction. Two components can be shaped into complex numbers or used separately.

The similarities are shown in Equation 4 below [10].

λ - The length of sinusoidal component wave
ϴ - Length of Gabor orientation line, normal to parallel
Ψ - Offset sinusoidal function phase
σ - Sigma or standard deviation from Gaussian….
ɣ - Spatial aspect ratio and determining ellipticity supported by Gabor function.

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = \exp\left(-\frac{x^2+\gamma^2 y^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x^l}{\lambda}+\psi\right) \quad (3)$$

where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$

2.4 Determining Feature Vector

This process has the aim to find feature vector or image characteristics taken from the data obtained from the previous process. Three features are obtained, namely mean, median, standard deviation on each filter. Next step we calculate mean, median and deviation of each result images and find the value on each filter kernel. The set of feature values of one image yield the feature vector. The number of bins of feature vector is the number of filter kernels.

2.5 Naïve Bayes

Naïve Bayes is a simple technique that builds clarification model and set class labels. Problems are represented by feature value vector, where class label is taken from several set. No single algorithm is used to train the clarification, but a family of algorithm based on general principal. All Naïve Bayes clarification assumes that feature value is not dependent on others, but rather the class variable. Naïve Bayes is an interesting technique because it has an explicit basic theory that is strong, ensuring optimal induction. However, one of the weaknesses is the independent feature assumption may be violated in several cases. Nonetheless it has been shown that the technique remains very strong in tackling this problem [11].

Naïve Bayes algorithm is quick, easy to implement with a simple feature and effective. This can also be used for higher dimension data as the probability of each features are predicted independently. Naïve Bayes clarification, by using the assumption that Features X1, X2, …, X are not conditionally dependent on one another in terms of class, we obtain:

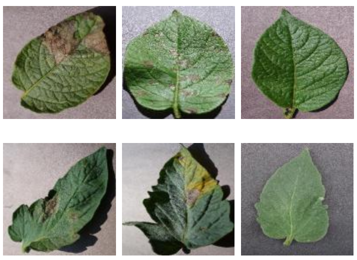
$$P(C|X) = \frac{P(C)\Pi_{i=1}^{\eta} P(X_i|C)}{P(X)} \qquad (4)$$

2.7 Classification Analysis

Classification analysis is the result obtained by the application after checking and counting input image as data testing. The result generates diagnose of a type of disease found on a leaf. The app offers symptoms and solutions related information to the user. This includes the signs found on leaves of a tomato or potato plant that indicate a certain illness and advices on how to care for and treat the diseased plants in a natural way, including suggestion on the suitable pesticide to use.

III. Experiments

. In this study, 120 trainings and 30 tests of the chosen potato and tomato plants' leaves are reduced into 256x256 dimensions. Three different labels are used, of which two indicate diseases and one healthy plant. The images below are some examples. Figure 6 shows the example of leaf images. There are three states of tomato and potato leaf; healthy leaf, leaf with early blight disease and leaf with late blight disease. We have 30 images for each state for training and 10 images for each state for testing. The examples of images are shown in figure 2.



(d) (e) (f)

Figure 2 (a) leaf of a potato plant affected by late blight, (b) early blight, (c) healthy, (d) leaf of a tomato plant affected by late blight, (e) early blight, (f) healthy

Leaf data that are inputted will be pre-processed first to obtain the desired data matrix. The next step is generating a kernel using four frequency parameters (0.06, 0.13, 0.18, 0.25), and six rotations for potatoes (0, 30, 60, 90, 120, 150). We use 24 Gabor filter that shown in Figure 3.

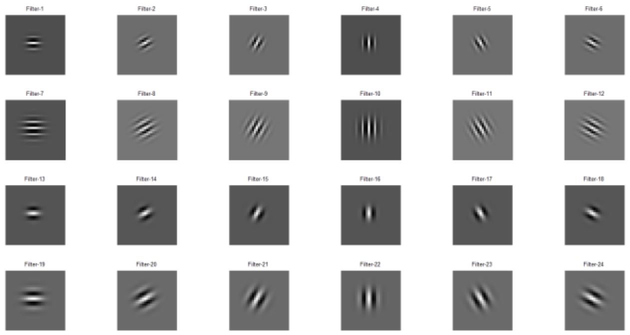


Figure 3. Kernel generation results

The formed kernel will be convoluted with images that have been pre-processed to create an energy/power in the form of matrix. We have 24 kernel of Gabor filter. We determine the feature vector uses mean, median and deviation of filter result [5]. Then, we select a half of 24values with biggest value. So, we have 3x12 or 36 bins feature vector. Figure 4 shows the segmentation of feature vector using Gabor filter. Generally, healthy leaf, leaf with early blight and leaf with late blight has different feature vector. This character applies to both potato leaves and tomato leaves.

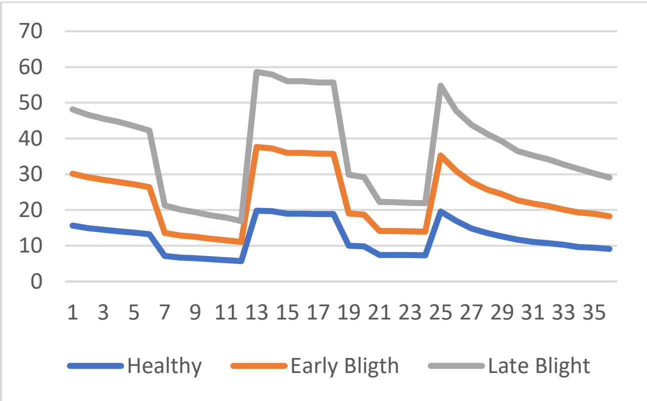


Figure 4. Segmentation of feature vector

The data produced from data training is then grouped into three separate groups, namely high, middle and low. This grouping is done using Equation 11.

$$Threshold_i = Min + i\frac{Max - Min}{3} \qquad (5)$$

The calculation results are shown in Table 2 and Table 3. Potatoes and tomatoes have different range values.

Table 2 Training results from potatoes images

Group	Standard Deviation	Median	Mean
High	>60	>84	>81
Medium	56-60	73-84	74-81
Low	<56	<73	<74

Table 3 Training results from tomatoes images

Group	Standard Deviation	Median	Mean
High	>70	>154	>119
Medium	64-70	133-154	108-119
Low	<64	<133	<108

After feature extraction using Gabor filter, Naive Bayes algorithm is applied to the training data using 10-fold cross validation. It yields different results for tomato and potato plants. Naïve Bayes algorithm and 10-fold cross validation gives an accuracy rate of 92% for images of potato plants' leaves and 78% for tomato plants.

The next step is testing application that has been developed by using smartphone camera. As a data tester, we use potato and tomato plants' leaves consisting of healthy, early blight and late blight plants. From several experiments, the accuracy rate is shown in Table 6 below.

Table 4 Classification Accuracy on Testing Phase

Accuracy	Early Blight	Late Blight	Normal	Total Accuracy
Potatoes	80%	70%	80%	76.6%
Tomatoes	60%	60%	70%	63.3%

Table 4 shows that the accuracy rates for tomato plants are generally lower than the potato plants. This is because the texture shown on tomato plants with early and late blight are very similar. The two diseases are detected almost interchangeably because the browning texture is blurred by the fur of a tomato plant's leaf.

Meanwhile there is a significant difference between the learning and testing phases. This is because images obtained from smartphone cameras cannot be directly used, because brightness, contrast and pixel size are different than ones in data training. The solution offered by this study is by processing these images using autolevel to correct for the contrast and brightness, hence minimizing differences of images between different cameras.

IV. CONCLUSIONS

The mobile application developed in this study is quite successful in detecting early and late blight diseases on tomato and potato plants using smartphone cameras. In the training step, using 120 pictures of potato plants' leaves, we obtain an accuracy rate of 92%, whereas for tomato the figure is lower at 78%. In the testing step using smartphone camera, the ability to identify disease in potato plants reaches an accuracy of 76.6% whereas for tomato only 63.3%. Diseases on tomato plant are harder to detect, because its leaves have fur that covers the signs of a disease, hence impairing the image extraction process.

REFERENCES

[1] Chun-hua Hu,Ping-pingLI "Edge Detection for Hardwood Seedings Leaves Based on Intuitionistic Fuzzy Set" *International Conference on Information Technology and Applications* ,2013.

[2] G. Anthanikar ,P. Badar "Potato Leaf Disease Detection and Classification System". *International Journal of Computer Science and Mobile ComputingI.* (2016).

[3] D. Samanta, P.P.Chaundhury, A Ghos. "Scab Diseases Detection of Potato using Image Processing". *International Journal of Computer Trends and Technology***.** (2012)

[4] M. Sardagon, A. Tuncer, Y. Ozen. "Plant Leaf Disease Detection and Classification based on CNN with LVQ Algorithm". *International Conference on Computer Science and Engineering***.** (2018)

[5] S. K. Venkatesh and R. Raghavendra, Local Gabor Phase Quantization Scheme for Robust Leaf Classification, Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, Hubli, Karnataka, 2011, pp. 211-214. doi: 10.1109/NCVPRIPG.2011.52

[6] S. C. Madiwalar and M. V. Wyawahare, "Plant disease identification: A comparative study," International Conference on Data Management, Analytics and Innovation (ICDMAI), Pune, 2017, pp. 13-18.

[6] P. Mohanty, "Plant Village Dataset", https://github.com/spMohanty/PlantVillage-Dataset

[7] Carsten Rother, Valdimir Kolmogorrov, Andre Blake, Grabcut: Interactive Foreground Extraction using Iterated Graph Cut, ACM Transaction on Graphics, Vol 23 Issue 3, 2004, pp. 309-314.

[8] Achmad Basuki, Nana Ramadijanti, 2016, Improving auto level method for enhancement of underwater images, International Conference on Knowledge Creation and Intelligent Computing (KCIC),

[9] Rosniza Roslan, Nursuriati Jamil, Texture feature extraction using 2-D Gabor Filters, Computer Applications and Industrial Electronics (ISCAIE), 2012 IEEE Symposium, DOI: 10.1109/ISCAIE.2012.6482091

[10] Richard Szeliski, Computer Vision: Algorithms and Applications, 2010 Springer.

[11] Sona Taheri, "Learning the Naïve Bayes Classifier with Optimization Model". *International Journal Application Mathematics and Computer Science***.** (2018)