IDENTIFICATION OF DISEASE TYPES ON TEA – PLANT VARIETIES BASED IMAGE PROCESSING WITH K-NEAREST NEIGHBOR METHOD

Adi Kurniawan Saputro [1], Kunto Aji Wibisono [2], Faiddhatus Puji Pratiwi [3]
Universitas Trunojoyo Madura
Jl. Raya Telang, Perumahan Telang Indah, Kamal, Kabupaten Bangkalan, Jawa Timur 69162
Adi.kurniawan@trunojoyo.ac.id [1], Kunto.ajiw@trunojoyo.ac.id [2], faiddhatuspujipratiwi@gmail.com [3]

Abstract. Tea is a plant typed Camellia sinesis, frequently consumed by many people. It can cure various diseases and can be enjoyed to relax. Within the traded tea leaves, the quality of tea leaves determine the tea beverage. Many factors influence quality of tea plants. One of them is biotic disturbance caused by macro-organism and micro-organism. Disease problems could be noticed when the farmers plant them on the field. Generally, the detections of disease and pest on the plants can be done in various ways but the method used in this case was HSV (Hue, Saturation, Value) to detect the colors. Classification method in this system used KNN (K-Nearest Neighbor) by classifying the system based on the closes data with the subject. The disease classification on this observation consisted of 3: tea leaf chicken pox, caterpillar roller, and tea mites. To compare which one is diseased plant to the healthy one, this research had classified the leaf diseases by using microscope with accuracy level of normal eye 100%, dealing with the chicken pox disease of the leaf 71.4%, and 28.5% for empoasca sp with successful level 85.7%, and error level 14.2%. Dealing with caterpillar roller pest, the successful level was 92.8% with error level 7.2%.

Keywords: Tea Plants, HSV (Hue, Saturation, Value), KNN (K-Nearest Neighbor).

1. Introduction

Tea (Camellia sinesis (L) is a plant with various functions. It has function as medication and can be enjoyed while relaxing. This plant was firstly planted in China. In Indonesia, there are many tea plants, especially in Java Island. Tea plant can raise National expenditure because of its benefits and is liked by many people. Consuming tea regularly could recover body system and improve immune system [1]. To consume tea, it only needs the leaf. In processing the qualified tea leaf needs freshness of tea leaf [2].

In growing up tea, there are many organisms disturbing the growth and lowering the production value of the tea. Pest and disease of tea plant can attack on three points: the trunk, leaf, and root. Farmers can minimize this attack if they handle it correctly and understand the influential factors of the disease. The disease attacking tea are tea leaf chicken pox, caterpillar rollers, and tea mites. Dealing with tea – growth problems attacked by various diseases will decrease quality of the tea
during production. This digression of tea production amount at each farm will hinder general need of tea consumption.

In this research, there was an observation dealing with morphological features of tea leaves, such as colors and shape of the cells to identify the types of diseases. In the identification process, it was done by feature extraction method on image processing. The classification method used observational method using Microscope by observing an unseen object by bare eye.

2. The Underlying Theory

Tea is a plant from *Theaceae* family. It was initially planted in RRC and now spreads into Indonesia, especially Java Island. The suitable soil to plant due to its high rainfall level and sunray exposure. Tea is frequently planted because it has many benefits and can be enjoyed to relax. The chemical compound within tea is located in its leaf. The changes of its chemical compounds into nutritious beverage needs to keep its vitamin contents [6]. In producing tea, there are many processing ways and methods to use. However, different tea types will have different nutritious tea beverages.

2.1. KNN Method (*K*-Nearest Neighbor)

This method is a classification method upon an object based on nearest – space data with the object. It aims to classify new objects based on their attributes and training samples. It is done by giving a query point, then the number of *K* – object of the closest training point is determined by the closest point to the query point. Then, a calculation process is done to find out the space by using *Euclidean*. Then, calculation stage with KNN method will be done.

\[
d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2}
\]

Notes

\( (x_i, x_j) \): Euclidean distance
\( (x), (x_j) \): record of \(i^{th}\) and record of \(j^{th}\)
\( (a) \): the \(r\) – data
\( I, j \): 1, 2, 3, . . . \(n\)
\( n \): The object dimension

3. Designing the System

The designed system in this research is an identification system of disease and pest types on tea – plant varieties based on image processing. The detection uses KNN (*K* – Nearest Neighbor) method and GUI to show the detection result. To differ between pictures and graphics of the inputted in this designing research methodology, it used software and hardware.
3.1. The Flowchart

4. Discussion

On the second test, it was done by observing the tissue of tea – leaf cell under microscope. It aimed to detect the disease within the leaf by using the cell.

Table 4.1 Results of Microscopic Figures

| No | Normal Leaf (1) | Chicken Pox (0) | Daun *Empoasca* sp (2) | Caterpillar Roller Leaf (3) |
|----|-----------------|-----------------|------------------------|---------------------------|
| 1  | ![Image](1)      | ![Image](1)     | ![Image](1)            | ![Image](1)               |
| 2  | ![Image](1)      | ![Image](1)     | ![Image](1)            | ![Image](1)               |
| 3  | ![Image](1)      | ![Image](1)     | ![Image](1)            | ![Image](1)               |
| 4  | ![Image](1)      | ![Image](1)     | ![Image](1)            | ![Image](1)               |
The method classified the disease by using KNN method, by analyzing the distance or k. When distance of similarity is closer, then the accuracy gets higher. In this research, the distance was \( k = 1 \) in which to read each characteristic of the disease would use labelling of each infected leaf figure. In the parameter figure, the normal leaf is labeled 1, chicken pox leaf labeled 0, Empoasca sp infected leaf labeled 2, then caterpillar roller – leaf labeled 3.

| No | Figure | Label | Label KNN of Feature vektor | Detection Mark | Label KNN of the Histogram | Detection Mark | Remark |
|----|--------|-------|-----------------------------|----------------|-----------------------------|----------------|--------|
| 1  | ![Image](image1.png) | 1     | 1                           | Normal         | 1                           | Normal         | Appropriate |
| 2  | ![Image](image2.png) | 1     | 1                           | Normal         | 1                           | Normal         | Appropriate |
| 3  | ![Image](image3.png) | 0     | 1                           | Normal         | 1                           | Normal         | Inappropriate |
| 4  | ![Image](image4.png) | 0     | 0                           | Chicken Pox    | 0                           | Chicken Pox    | Appropriate |
| 5  | ![Image](image5.png) | 2     | 2                           | Empoasca sp    | 2                           | Empoasca sp    | Appropriate |
| 6  | ![Image](image6.png) | 2     | 2                           | Empoasca sp    | 2                           | Empoasca sp    | Appropriate |
| 7  | ![Image](image7.png) | 3     | 3                           | Caterpillar Roller Leaf | 3                   | Caterpillar Roller Leaf | Appropriate |
| 8  | ![Image](image8.png) | 3     | 3                           | Caterpillar Roller Leaf | 3                   | Caterpillar Roller Leaf | Appropriate |

The investigation on normal cellular leaf for labeled rows means each leaf sample with normal types, labeled 1 after being checked by using the program software one by one. It showed the vector result 1 or normal. It meant the detection was normal and on the histogram since it histogram reading showed 1, then the result was normal leaf. It happens when the score of feature vector and the histogram have equal output score. Then, the result would be appropriate. It is concluded that from 14 sample test used for training and testing data, the accuracy of match levels in normal category were 100%. It meant that normal tea leaf did not have any similarity to other tea leaves. The evidence by using KNN method dealing with chicken pox leaf, labelled 0, had accuracy 71.42% with error level 28.58%. From the error or the failures, it meant there was similarity of photograph realizations or other form realizations which fitted on the others. From 14 tested samples, only three were different from the demanded output results. The evidence by using KNN method dealing with Empoasca sp, labelled 2, had accuracy level 85.8%, with error level 14.2%. From the failure, it meant there was similarity from the photograph realization or other form
realization which were in line with the other. From 14 tested samples, only 1 was different to the demanded output. The evidence by using KNN dealing with caterpillar rollers, labeled 3, had accuracy 92.8% and error level 7.2%. From the failures, there was similarity from photograph realization or other form realizations which were in line with the others. From 14 samples, only 1 was different with the demanded output results. The pixel score and histogram score would be used as reference to calculate by using KNN method. The RGB [68 41 94] score referred to RGB histogram from a photograph which was using 1x1 matrix to find out the score. The scores of RGB from a normal leaf sample was used on pixel calculation of normal leaf. To calculate it, manual formula of KNN Eucliden distance with equation 2.1

\[
E = \sqrt{(R - R)^2 + (G - G)^2 + (B - B)^2} = \sqrt{(68 - 68)^2 + (41 - 41)^2 + (94 - 94)^2}
\]

5. CLOSING
5.1. Conclusion

Based on the identification trial of the disease types of tea plants, it is concluded:
1. Under microscopic observation, there was a very thin slicing to make the photograph of the cell could be clear and not to thick.
2. The test by using KNN method showed that normal leaf 100%, chicken pox – leaf 71.4% with error level 28.5%, Empoasca sp 85.7% with error level 14.2%, caterpillar rollers 92.8% with error level 7.2%. The tested samples consisting of 14 samples where in 1 leaf was not completely infected on its middle point. The score of KNN method process was gained by finding out the feature vector and histogram scores.

5.2. Suggestion

Based on the trial of types of disease identification and aroma of the leaf, a clearer photograph could be taken with more than 1 method as the comparison of successful accuracy level. Future investigation may also detect tea aroma.

References

[1] H. S. YAP, “Penyakit Cacar Daun Teh (Exobasidium vexans) dan Tingkat Infeksinya di Perkebunan Gunung Mas (PTPN VIII) dan Bah Butong (PTPN IV),” 2017.
[2] S. Palgunadi dan N. Pratiwi, “Prediksi Umur Dan Kandungan Klorofil Daun Teh Berdasarkan Image Daun Dengan Menggunakan Vektor Ciri Warna Hijau,” 2015.
[3] Nurmahaludin dan r. G. cahyono, “Metode Penalaran Sistem Pakar Menggunakan Metode Hibrid Fuzzy Dempster Shafer untuk Identifikasi Hama dan Penyakit Tanaman Jagung,” vol. 9, 2017.
[4] A. I. Puji, I. Hermandi, A. Buono dan M. K. H, “Pengembangan Sistem Pakar Identifikasi Awal Penyakit Kendali Dengan Pendekatan Naive Bayes Berbasis Android,” vol. 14.
[5] Y. G. Hari, R. Magdalene dan H. Fauzi T, “Identifikasi Jenis Penyakit Kakao Dengan Pengolahan Citra Digital Dan K-Nearest Neighbor,” vol. 3, 2016.
[6] J. Towaha dan B., “Kandungan senyawa Kimia Pada Daun Teh (Camellia sinensis),” Warta penelitian dan Perkembangan Tanaman Industri, vol. 19, 2013.
[7] Y. H. Sartika, “Penyakit Cacar Daun Teh (Exobasidium vexans) Dan Tingkat Infeksiyang Di perkebunan Gunung Mas (PTPN VIII) Dan Bah Butong (PTPN IV).”

[8] “www.pdfactory.com,” [Online].

[9] “Potensi Daun Teh (Camellia sinensis) dan Daun Antinganting Acalypha indica L. dalam Menghambat Pertumbuhan Salmonella typhi,” Jurnal AL-Azhar Indonesia Seri Sains Dan Teknologi, vol. 2, p. 7, 2013.

[10] N. P. Sarngadi Palgunadi, “Prediksi Umur Dan Kandungan Klorofil Daun Teh Berdasarkan Image Daun Dengan Menggunakan Vektor Ciri Warna Hijau,” dalam Universitas Wahid Hasyim Semarang, Semarang, 2015.

[11] N. Vina, “Keanekaragaman Serangga pada Tanaman Teh (Camellia sinensis L) pada Perkebunan Teh Bantaran Blitar,” p. 7, 2015.

[12] A. F. Wulandarai, “Inventarisasi Dan Identifikasi Penyakit-Penyakit Pada Beberapa Klon Teh (Camellia sinensis L.) Di PT Rumpun Sari Kemuning,” Surakarta, 2016.

[13] A. W. Joko Ragiliyanta, “Sistem Pakar Diagnosta Hama Dan Penyakit Tanaman Teh Menggunakan Metode Forward Chaining”.

[14] R. H. G. Ratih Kartika Dewi, “Identifikasi Penyakit Pada Daun Tebu Dengan Gray Level Cooccurrence Matrix Dan Color Moments,” vol. 1, 2014.

[15] B. D. S. P. P. A. Yerry Anggoro, “Implementasi Metode Fuzzy K-Nearest Neighbor Untuk Klasifikasi Penyakit Tanaman Kedelai Pada Citra Daun,” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 2, p. 9, 2018.

[16] H. Alustia Dewi, “Sistem Perbaikan Citra Berkabut dengan Park Channel Prior dan Contrast Limited Adaptive Histogram Equalization,” 2018.

[17] Vademicum Bagian Tanaman teh, Surabaya: PT. Perkebunan Nusantara XII.

[18] Standart Operating Procedure, Surabaya: PT. Perkebunan Nusantara XII, 2014.