Implementation of Case-Based Reasoning and Nearest Neighbor Similarity for Peanut Disease Diagnosis

Minarnii*, Indra Warman1, Yuhendra1

1Teknik Informatika, Institut Teknologi Padang, Indonesia

Email: minarni1706@gmail.com

Abstract. This research discusses the development of an expert system to diagnose peanut disease using Case-Based Reasoning (CBR) and Nearest Neighbor Similarity. CBR is a computer reasoning system using old knowledge to overcome new problems. It provides solutions by looking at the closest old case to new case. The diagnosis process is carried out by entering a new case containing the symptoms to be diagnosed into the system, then calculating similarity values between new cases on a case base using the nearest neighbor method. The average test results of the system to make an initial diagnosis of peanut disease indicate that the system is able to correctly recognize 100% peanut disease. Accuracy calculation uses the nearest neighbor similarity method with a threshold of 0.5, 0.6 and 0.7 respectively 97.22, 88.89%, and 80.55%.

1. Introduction

Peanuts (Arachis Hypagaea L) have an economic value which is quite important in Indonesia because it produces oil that can be used for various purposes [1]. Therefore, it is necessary to keep the productivity of peanuts. In West Sumatra in 2016 peanut production decreased by 0.13% [2]. One factor that influences the level of productivity is an attack of disease [3]. To diagnose peanut disease is carried out by experts, namely field agricultural extension workers where there are limited numbers, time and place [4]. For this reason, the farmers need a system that can help them in diagnosing peanut disease.

Expert systems have the ability to facilitate practical problems when experts are not present. One of the systems implementation experts in agriculture to diagnose plant diseases. There are many types of plant diseases and various names and their consequences. The characteristics of plants are affected by diseases which have similarities with other diseases, so people are confused about diagnosing them.

Development of expert systems in agriculture has been carried out. Expert systems use two methods of plant diagnosis, namely descriptive methods based on visible symptoms, and graphical methods based on images. The results of the study showed the tendency of users to choose the graphical method compared to the descriptive method [5]. An expert system built by knowledge base, inference mechanism or control system and user interface used to diagnose rice plant disease [6]. An expert
The system uses the Multi-Criteria Decision technique namely Analytic Hierarchy Process and Sensitive Simple Additive Weighting. This system uses two types of weights: relative weights and scales. The diagnostic process involves calculating the characteristic utility values. The test results show an accuracy value exceeding 90% [7]. Expert system for diagnosing rule-based Oyster Mushroom disease and forward chaining inference engine. This system helps users to recognize diseases caused by mold, bacteria, viruses, insects and other pests [8]. Research on methodological trends is used in developing expert systems for managing plant pests and diseases by comparing rule-based expert systems with case-based systems. The results obtained by rule-based expert systems are more interesting than case-based systems [9].

This study aims to develop an expert system to identify peanut plant diseases using case-based reasoning method with the nearest neighbor similarity measurement method.

Some of the theories used in this study are as follows:

1.1. The Expert System

The Expert systems are computer-based applications that are used to solve problems as the expert thinks. Experts are people who have special skills that can solve problems that cannot be solved by common people [10].

In the structure of the expert system, there are two parts, namely the development environment and the consultation environment. The expert system makers build a knowledge base in development environments. The user consults in a consulting environment [11] (See Figure 1)

Figure 1. The Expert System Architecture [11]

1.2. Case-Based Reasoning

Case-Based Reasoning (CBR) is a process of remembering a case in the past, then reusing it and adapting it in a new case [12]. The stages of CBR are as follows:
Retrieve to recover the most similar/similar cases with the new case. This section refers to terms of identification, initial similarity, search and recovery, and execution. Then Reuse to use information and knowledge of the case to solve the problem. The process of reuse of case solutions obtained in the new context is focused on two aspects: the difference between the previous and the current case, what part of the case has been obtained that can be transferred into a new case. Revise to review/ refine proposed solutions. Retain is parts of the experience that may be useful for solving next problems.

1.3. Nearest Neighbor Similarity

The Nearest Neighbor algorithm is one technique for finding the closest distance of each case (cases) in the database, and how similar the size (similarity) of each source case in the database with the target case. Similarities usually are at a value of 0 to 1. A value of 0 means that the two cases are not similar absolute, contrary to the value of 1 case is similar to the absolute. The similarity function is formulated as follows [13]:

\[
\text{Sim}(T,S_i) = \frac{\sum_{i=1}^{n} f(T,S_i) \cdot w_i}{\sum_{i=1}^{n} w_i}
\]  

Where: T=New Case; S=Existing Case in Storage; n=Number of Attribute; i=number of attribute each case; f=similarity function of T and S Attribute; w_i=Attribute Weighting to i

1.4. Peanut Disease

The peanut disease classification in this study consisted of 6 (six) diseases where several diseases had the same symptoms. The intended diseases are leaf spot, stem rot, bacterial wilt, peanut mottle, peanut stunt, and peanut green mosaic [1].

2. Method

2.1. Description of System

The system is designed to diagnose peanut diseases that can be used by three groups of users, namely admin, farmer, and expert. The process of diagnosis by entering data on visible symptoms. Then it calculates similarity with the case in the base case using the nearest neighbor similarity method. Furthermore, the reuse process, the solution is the highest similarity value. The output module will display a diagnosis result. However, if the similarity value is less than the threshold, it is considered that the new case does not have a solution and will be stored as a new case which will be evaluated by an expert (revise). Then the results are saved into the system as a new case with the solution provided (retain)(See Figure 2).

2.2. Case Representation

A case must be represented in a particular form for storage purposes in the case base and retrieval process. Representation of a case must include problems that explain the situation and the solution. In this study the case representation using the frame model. Symptom features are used to look for similar cases, while diseases and solutions are the output of the system (See Table 1)
The retrieval process used in this study is by comparing the symptoms of new cases with existing case symptoms in the case base, then the comparison results will be calculated the level of similarity. In this study, calculating similarity values using a nearest neighbor.
Table 2. The Sample Case

| Case         | New Case (T) (wi) | Old Case (S1)               | f(T, Si) |
|--------------|-------------------|-----------------------------|---------|
| K1           | Brown spots (3)   | Brown spots (3)             | 1       |
|              | Brown Spot on leaves (3) | Brown Spot on leaves (3) | 1       |
|              | Dark Reddish-Brown Top Surface of The Leaf (3) | Dark Reddish-Brown Top Surface of The Leaf (3) | 1       |
|              | Leaves are Yellow (1) | Leaves are Yellow (1)       | 1       |
|              | Dry Leaves (3)    | Dry Leaves (3)              | 1       |
|              | Fall Out Leaves (3) | Fall Out Leaves (3)         | 1       |
|              | Irregular Yellows Spot (3) | Smaller and Rounder Spots (3) | 0       |
|              | Yellow Leaves Edges (3) |                           | 0       |

Calculation of similarity using equation (1)

\[
Sim(T, S_i) = \frac{(1 \times 3) + (1 \times 3) + (1 \times 3) + (1 \times 1) + (1 \times 3) + (1 \times 3) + (0 \times 3) + (0 \times 3)}{3 + 3 + 3 + 1 + 3 + 3 + 3 + 3} = 0.7273
\]

With the same calculation process, the similarity between new cases of K1 and some cases that have been stored in the database is obtained as shown in Table 3.

Table 3. Similarity Value

| Case   | The similarity of CBR (%) |
|--------|---------------------------|
| K1-S1  | 72.73                     |
| K1-S2  | 13.64                     |
| K1-S3  | 0                         |
| K1-S4  | 4.55                      |
| K1-S5  | 0                         |
| K1-S6  | 27.27                     |
| MAX    | 72.73                     |

The reuse processes the identification solution provided is a solution that has the highest similarity value. For the above case, the highest similarity results were found in S1, namely leaf spot disease with a similarity value of 0.7273 or 72.73%, so the recommended solution was a solution for leaf spot disease.

2.4. System Testing

The test results need to be evaluated to find out whether the system built is feasible to be applied in diagnosing peanut diseases. Evaluation is done by calculating sensitivity and accuracy using equations (2) and (3) [14]

\[
Sensitivity = \frac{T_P}{T_P + T_N}
\]

\[
Accuracy = \frac{T_P + T_N}{T_P + F_P + T_N + F_N}
\]
3. Result and Discussion

3.1. Test Result

Testing is done by the user directly with the system that has been designed. The test was carried out in two stages, namely the first stage of testing in accordance with cases in the case base, while the second stage of testing was carried out using 36 data from 120 test data. The results of the first phase of the test showed that the system was able to correctly identify 100% peanut disease. Recapitulation of the results of the second stage testing with a similarity threshold $\geq 0.5$, $T \geq 0.6$, and $T \geq 0.7$ are shown in Table 4.

### Table 4. Test Result Recapitulation

| No. | Type of Disease      | No of Test Data | Correct Diagnoses |
|-----|----------------------|-----------------|-------------------|
|     |                      |                 | $T \geq 0.5$ | $T \geq 0.6$ | $T \geq 0.7$ |
| 1   | Leaf Spot            | 6               | 6 | 5 | 5 |
| 2   | Stem Rot             | 6               | 6 | 5 | 5 |
| 3   | Bacterial Wilth      | 6               | 6 | 6 | 5 |
| 4   | Peanut Stunt         | 6               | 5 | 5 | 4 |
| 5   | Peanut Green Mosaik  | 6               | 6 | 6 | 5 |
| 6   | Peanut Mottle        | 6               | 6 | 5 | 5 |
|     | Amount               | 36              | 35 | 32 | 29 |

3.2. Discussion

The test results in Table 4 will be used as an evaluation system. This is important to determine the feasibility of the system to diagnose peanut diseases. Evaluation of test results is done by calculating the value of sensitivity and accuracy. Table 5 shows the confusion matrix based on the results of system testing.

### Table 5. Confusion Matrix of Test Result

| No. | Type of Disease      | $T \geq 0.5$ | $T \geq 0.6$ | $T \geq 0.7$ |
|-----|----------------------|--------------|--------------|--------------|
|     |                      | Peanut Diseases | Non-Peanut Diseases | Peanut Diseases | Non-Peanut Diseases | Peanut Diseases | Non-Peanut Diseases |
| 1   | Peanut Diseases      | 35 (TP) | 1 (FP) | 32 (TP) | 4 (FP) | 29 (TP) | 7 (FP) |
| 2   | Non-Peanut Diseases  | 0 (FN) | 0 (TN) | 0 (FN) | 0 (TN) | 0 (FN) | 0 (TN) |

### Table 6. The result of Sensitivity and Accuracy

| Threshold | Sensitivity (%) | Accuracy (%) |
|-----------|-----------------|--------------|
| $\geq 0.5$ | 100             | 97.22        |
| $\geq 0.6$ | 100             | 88.89        |
| $\geq 0.7$ | 100             | 80.55        |

The results of the calculation of the sensitivity value of 100% with the system accuracy value for peanut with a threshold of similarity $T \geq 0.5$, 0.6 and 0.7 respectively 97.22%, 88.89%, and 80.55% (See Table 6). This shows that an expert system using CBR can diagnose peanut disease well.
4. Conclusion

Based on the research and test results it was concluded that:

- This study produces an expert system diagnoses the peanut diseases using case-based reasoning inference and nearest neighbor similarity.
- The results of the system testing for the identification of peanut diseases using the nearest neighbor similarity method indicate that the system is able to identify the disease according to the cases in the case base of 100%.
- The test results of 36 test data for peanuts with a similarity threshold of 0.5, 0.6 and 0.7 indicate that the system has good performance with a sensitivity of 100% and an accuracy level for each threshold respectively 97.22%, 88.89%, and 80.55%.

Acknowledgment

The research was funded by the Directorate of Research and Community Service of Ministry of Research, Technology and Higher Education of the Republic of Indonesia for the funding of the National Strategic Research Institute (PSNI) research grant for 2018 under contract number: 324/27.O10.5/PN/ II / 2018.

References

[1] Semangun H 2008 Penyakit-penyakit Tanaman Pangan di Indonesia (Gadjah Mada University Press, Yogyakarta)
[2] Tim Penyusun 2017 Sumatera Barat dalam Angka 2017 (Badan Pusat Statistik Provinsi Sumatera Barat)
[3] E Paturohman and S Sumarno 2014 Iptek Tanaman Pangan 9(2) p 97
[4] Tim Penyusun 2015 Laporan akuntabilitas Kinerja Instansi Pemerintahan (LAKIP) 2015, (Dinas Pertanian Tanaman Pangan Sumatera Barat, Padang)
[5] S.S. Abu-Naser K.A. Kashkash and M. Fayyad 2008 Journal of Artificial Intelligence 1(2) p 78
[6] Kalita, H. Sarma S. K., and Choudhury, R. D. (2016, September). International Conference Automatic Control and Dynamic Optimization Techniques (ICACDOT), IEEE p 723-730
[7] Goodridge, W., Bernard, M., Jordan, R., & Rampersad, R. (2017). Computers and Electronics in Agriculture (133) p 80-87.
[8] Munirah, M. Y., Rozlini, M., & Mariam, Y. S. (2013, October). 13th International Conference Control, Automation and Systems (ICCAS) IEEE p. 329-332
[9] Shafinah K Sahari N Sulaiman, R Yusoff MSM and Ikram M M 2013 Journal of Theoretical & Applied Information Technology 58(1) p 182
[10] Kusrini 2008 Aplikasi Sistem Pakar (Andi Offset, Yogyakarta)
[11] Sutojo T Mulyanto E and Suhartono V 2011 Kecerdasan buatan (Penerbit Andi, Yogyakarta)
[12] Pal SK and Shiu Simon K 2004 Foundation of Soft Case-Based Reasoning (John Wiley &Sons Inc Publication, USA)
[13] Kusrini and Luthfi ET 2009 Algoritma Data Mining (Andi Offset, Yogyakarta)
[14] Witten IH and Frank E 2005 Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann, San Francisco)