

# Measurement of Similarity and Design Interface for Soybean Disease Diagnosis

Adriana Sari Aryani, Dian Kartika Utami, Hermawan

**Abstract:** The similarity of the target case is determined by measuring how close each attribute of the target is similar to the stored case in the case base. Similarities are usually normalized to fall within a range of 0 to 1. The soybean is one of the most important bean in the world, providing vegetable protein for millions of people and ingredients for hundreds of chemical products. Several diseases, including root and stem rot, pod and stem blight, frog eye leaf spot, brown spot, downy mildew, leaf blight and purple seed stain, and stem rot (white mold). In a case-based reasoning system for the identification of diseases of soybean plants provide solutions recommended by experts in diseases of soybean in accordance with a similar case or a similar matches within the database storage plant disease cases. Similarity value where 0 is totally dissimilar and 1 is an exact match. if similarity value equal zero then system will keep a set of data will be save temporary and need validation as a new case from the expert. The system give a recommended solution from similarity formula with the threshold which given by the expert.

**Index Terms:** Similarity, Design, Soybean disease

## I. INTRODUCTION

Plant disease can be defined as the condition when a plant is not able to carry out vital functions properly by losing form and integrity due to pathogenic attack (infectious disease) and other factors (noninfectious disease). Plant diseases are influenced by environmental factors. It is recommended that in an effort to control this new fungicide used after other efforts did not give results. In a case-based reasoning system for the identification of diseases of soybean plants provide solutions recommended by experts in diseases of soybean plants in accordance with a similar case or a similar matches within the database storage plant disease cases.

Plant disease epidemiology is the study of disease in plant populations. Much like diseases of humans and other animals, plant diseases occur due to pathogen such as oomycetes, nematodes and parasitic plants. Plant disease epidemiologists strive for an understanding of the cause and effects of disease and develop strategies to intervene in situations where crop losses may occur.

Commonly the elements of an epidemic are referred to as the "disease triangle": a susceptible host, pathogen, and conducive environment. For a disease to occur all three of these must be present. For some diseases it is important to

**Revised Manuscript Received on April 25, 2019.**

**Adriana Sari Aryani**, Department of Computer Science, Pakuan Bogor University, 16143, Indonesia.

**Dian Kartika Utami**, Department of Information System, Pakuan Bogor University, 16143, Indonesia.

**Hermawan**, Department of Computer Science, Pakuan Bogor University, 16143, Indonesia.

consider the disease occurrence over several growing seasons, especially if growing the crops in monoculture year after year or growing perennial plants.

The soybean (*Glycine max*) soya bean is a species of legume native to East Asia, widely grown for its edible bean, which has numerous uses. The soybean disease are many kind, there are : anthracnose, Brown Leaf Spot, Downy Mildew, Purple Seed Stain, etc. Brown Leaf Spot. In severely infected fields, the lower half of the plant may lose all its leaves. The extent of defoliation depends on weather conditions following initial infection

Seed Decay and Seedling Diseases Poor seed quality, whether due to physical, physiological or pathological causes, predisposes seedlings to disease organisms. Poor seed quality is a major problem confronting soybean producers. Seed quality can be lowered by improper storage conditions, which include too high or too low moisture or temperatures. Both seed and soil-borne organisms can cause seed decay and seedling diseases. Anthracnose (fungus—*Colletotrichum dematium* var. *truncatum*): The fungus infects stems, petioles and pods of plants nearing maturity. Dark brown or reddish-brown areas may cover the surface of infected stems and pods. Lower branches die. Downy Mildew, First symptoms appear as indefinite yellowish-green areas on the upper leaf surface. Later, these areas become light to dark brown spots with yellow-green margins. Grayish downy tufts of mold growth appear on the lower surface. Purple Seed Stain, A pink or light to dark purple discoloration of the mature seed coat provides easy identification. Size of the discoloration may vary from a small spot to the entire seed surface. Affected seed may be cracked, rough and dull. Seed quality is lowered.

Usually, one of the first diseases to appear on young plants. Small, angular spots varying from yellow to dark brown develop on leaves of infected plants. The brown area is often by a water-soaked margin. Spots later dry and portions of the leaf may drop out. Leaves may have a torn, ragged appearance. Charcoal rot is usually found in mid-summer, mostly on sandy soils. This fungus is a weak pathogen and generally attacks young plants when their growth is retarded by drought. The fungus attacks the roots and lower stem. Rhizobium – Induced Chlorosis: A chlorosis or yellowing ranging from light green to nearly white may occur about 6-8 weeks after planting when the plants are making rapid growth. Target Spot (Infection primarily on leaves, but may also occur on pods and stems. Spots on leaves are reddish-brown, circular in shape and vary in size from a pinpoint to more than 1/2 inch in diameter.

In computer science AI research is defined as the study of "intelligent agents": any device that perceives its



environment and takes actions that maximize its chance of successfully achieving its goals.

Artificial intelligence can be applied in farming for identification disease plant. Knowledge engineering is a core part of AI research. Machines can often act and react like humans only if they have abundant information relating to the world.

The cases in the CBR system are often represented by many attributes. In the research field about the big data technology, the number of attributes has increased a lot and it takes much time in the computation for case retrieval [1]. A large number of invalid or unrelated attributes could be considered to reduce the solution efficiency greatly. Therefore, attributes reduction is essential to improve the solution efficiency of CBR, especially in the fast modern industrial production.

CBR have four process or step, there are:

- a. Retrieve the most similar to the new problem past case, or cases.
- b. Reuse the information and knowledge in that case to solve the problem.
- c. Revise the proposed solution.
- d. Retain the parts of this experience likely to be useful for future problem solving [2].

In more detail, an initial description of a problem defines a *new case*. This new case is used to RETRIEVE the most similar case, or cases, from the library of *previous cases*. The subtasks of the retrieving procedure involve:

- a. Identifying a set of relevant problem descriptors.
- b. Matching the case and returning a set of sufficiently similar cases, given a similarity threshold of some kind.
- c. Selecting the best case from the set of cases returned [3].

New problem are solved by remembering and adapting the solution found for similar case that happen in the past[4].

## II. SIMILARITY

Identification or diagnosis disease plant can be using similarity technique in artificial intelligence. Similarity is define as the measure of how much two or more objects are similar [5]. Similarity can also be seen as the numerical distance between multiple data objects that are typically represented as value between the range of 0 (not similar at all) and 1 (completely similar). In approaches to retrieval based on surface features, the similarity of each case to the target problem.

There are many ways of measuring similarity and different approaches are appropriate for different case representations. For example, it is common in practice for each case to be represented as a simple feature vector (or set of attribute-value pairs) [6]. With this representation, it is usual to define a *local* similarity measure for each attribute. A *global* similarity measure aggregates the degrees of similarity that are computed by the local measure [7]. Then, global similarity can be computed as a weighted average of the local similarity [8]. Basic assumption: similar problems

have similar solutions. Similarity global formula is shown in Fig. 1 and similarity local formula is shown in Fig. 2.

$$sim(A, B) = \frac{1}{\sum w_i} \cdot \sum_{i=1}^p w_i \cdot sim_i(a, b)$$

Where,

- A* is new case,
- B* is previous cases,
- a* is new feature from local similarity,
- b* is previous features from local similarity,
- p* is the number of attributes,
- i* is the iteration
- w<sub>i</sub>* is weight of attributes  $i \sum_{i=1}^p w_i = 1$ , and
- sim<sub>i</sub>* is local similarity calculate for attribute *i*.

Fig. 1 Similarity Global Formula

$$sim(a, b) = 1 - \frac{|a - b|}{range}$$

Where,

- a* is new feature,
- b* is previous features, and
- range* is the value of difference between the upper and lower boundary of the set.

Fig. 2 Similarity Local Formula

A flowchart is a diagram that depicts a process, system or computer algorithm. Flowcharts, sometimes spelled as flow charts, use rectangles, ovals, diamonds and potentially numerous other shapes to define the type of step, along with connecting arrows to define flow and sequence. Flowchart for diagnosis soybean disease is shown in Figure 3.

The user must choice the few symptoms and then the system will retrieve to database case for searching the match disease based on the similar symptoms. The output of the system is the diseases name with the value similarity.

## III. RESULTS AND DISCUSSION

There are many Soybean Disease. Every disease have different symptoms. The detail of few disease is shown in single table. In each table will be detailed symptoms, control & prevention, and image of the disease.

Soybean Anthracnose detail can be seen in Fig. 4. Soybean Downy Mildew detail can be seen in Fig. 5. Soybean Purple Seed Stain detail can be seen in Fig. 6.



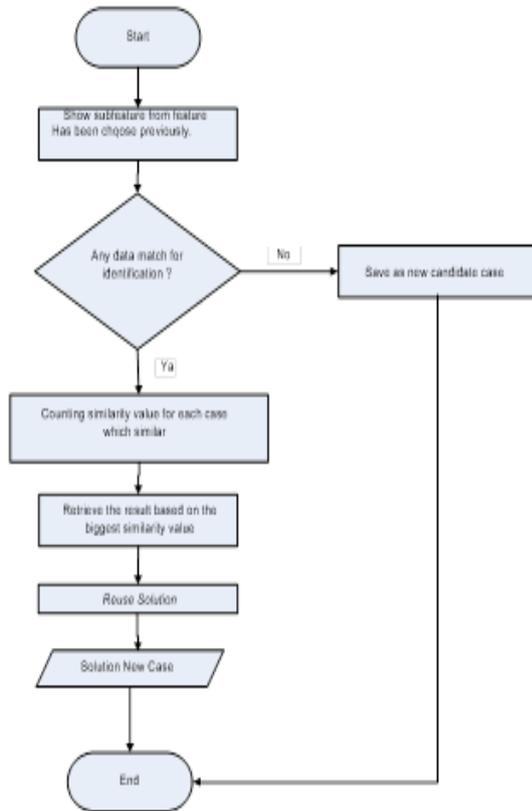


Fig. 3 Flowchart

**Disease name : Anthracnose**

**Symptoms :**

- reddish veins,
- rolled leaves,
- dark blotches on stems, pods and leaf petioles.

**Control & Prevention :**  
The disease is seedborne, so do not use seed from infected plants. Seeds from plants infected by anthracnose may or may not show discoloration.

**Image**

Fig. 4 Detail Soybean Anthracnose

**Disease name : Downy Mildew**

**Symptoms :**

- small, light green spots (not water-soaked) on upper leaf surfaces.
- the spots enlarge and turn pale to bright yellow,
- coalesce into large irregular brown areas

**Control & Prevention :**  
Select resistant soybean varieties and use appropriate seed treatments. Bury infested residue where feasible and where disease has been severe. Rotate soybeans with a non-bean crop for at least one year.

**Image**

Fig. 5 Detail Soybean Downy Mildew

**Disease name : Purple Seed Stain**

**Symptoms :**

- Infected seeds may appear healthy or have pink to purple spots that range in size from specks to large
- Reddish-purple, angular to irregular spots occur on both upper and lower surfaces of infected primary and secondary leaves.

**Control & Prevention :**  
Seed infection by either *Cercospora* fungus may cause poor seed vigor and reduced germination. Beans with substantial amounts of discoloration should not be saved for seed. Rapid decay of infested residue prevents a build-up of the pathogen and potentially greater infection in the next soybean crop.

**Image**

Fig. 6 Detail Soybean Purple Seed Stain

User interface (UI) design is the design of user interfaces for software or machines, such as the look of a mobile app, with a focus on *ease of use*. UI design usually refers to the design of graphical user interfaces. An interface is a point of interaction between the user and the hardware and/or software they are using. UI design is the skill employed to visual the interface used to complete the task it is designed for. Good UI design facilitates making the completion of tasks as frictionless as possible and increasing usability. The design of the user interface system consists of several designs, among others: the design of the home page, diagnostic page, results page, login page.

User interface design for the home page system for soybean disease diagnosis is shown in Fig. 7. In the menu will be some features: diagnosis, tips for soybean plant, information soybean disease, and login page.

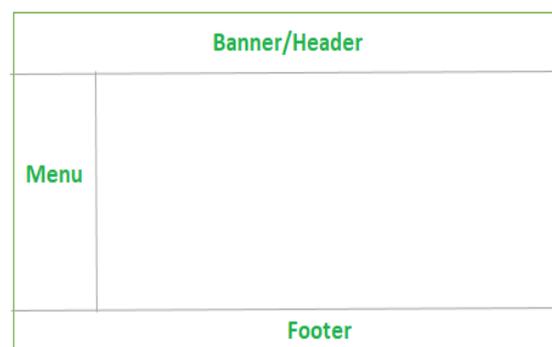


Fig. 7 Home Page User interface design

Login page page is reserved for administrator and the experts. After administrator or the expert login successfully then the page will appear to be able to edit or add data symptoms of diagnosis, disease data, and user profile data. User interface design for the login page is shown in Fig. 8.



Fig. 8 Login Page User interface design

The user data profile page displays the user data itself and can change the profile data as well as password data. Administrator and expert can be edit or add data symptoms of diagnosis. Data collection of symptoms contained in a disease. Data collection of symptoms can be add or delete. Symptom data sets contained within a disease should only be altered by plant disease experts to ensure validity of the data.

#### IV. CONCLUSION

Plant disease epidemiology is the study of disease in plant populations. Plant disease epidemiologists strive for an understanding of the cause and effects of disease and develop strategies to intervene in situations where crop losses may occur.

Diagnosis of soybean plant disease can be done with similarity technique. Similarity are typically represented as value between the range of 0 (not similar at all) and 1 (completely similar). New problem are solved by remembering and adapting the solution found for similar case that happen in the past. The soybean disease are many kind, there are : anthracnose, Brown Leaf Spot, Downy Mildew, Purple Seed Stain. The design of program flow is explained that the user must choice the few symptoms and then the system will retrieve to database case for searching the match disease based on the similar symptoms.

#### REFERENCES

1. A. Aamodt and E. Plaza. "Case-based reasoning: Foundational issues, methodological variations, and system approaches." *AI communications*, vol. 7, no. 1, pp. 39-59, 1994.
2. M. G. Voskoglou. "Case-Based Reasoning: A recent theory for problem-solving and learning in computers and people." In *World Summit on Knowledge Society*, pp. 314-319, 2008.
3. C. Jian, T. Zhe and L. Zhenxing. "A review and analysis of case-based reasoning research." In *2015 International Conference on Intelligent Transportation, Big Data and Smart City*, pp. 51-55, 2015.
4. Y. Liang, H. Wang, A. Xu and N. Tian. "A Two-step Case-based Reasoning Method Based on Attributes Reduction for Predicting the Endpoint Phosphorus Content." *ISIJ International*, vol. 55, no. 5, pp. 1035-1043, 2015.
5. S. Nasiri and M. Fathi. "Case Representation and Similarity Assessment in a Recommender System to Support Dementia Caregivers in Geriatric and Palliative Care." In *ICCBR (Workshops)*, pp. 157-166, 2017.
6. H. Núñez, M. Sanchez-Marre, U. Cortés, J. Comas, M. Martínez, I. Rodríguez-Roda and M. Poch. "A comparative study on the use of similarity measures in case-based reasoning to improve the classification of environmental system situations." *Environmental Modelling & Software*, vol. 19, no. 9, pp. 809-819, 2004.
7. M. K. Kiragu and P. W. Waiganjo. "Case based Reasoning for Treatment and Management of Diabetes." *International Journal of Computer Applications*, vol. 145, no. 4, pp. 20-29, 2016.
8. I. Watson. *Applying case-based reasoning: techniques for enterprise systems*. Burlington: Morgan Kaufmann Publishers Inc, 1998.