Computers and Electronics in Agriculture 90 (2013) 76–85

Contents lists available at SciVerse ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Rice diseases classification using feature selection and rule generation techniques

Santanu Phadikar [a,*], Jaya Sil [b], Asit Kumar Das [b]

[a] Department of Computer Science and Engineering, West Bengal University of Technology, Salt Lake, Kolkata 700 064, India
[b] Department of Computer Science and Technology, Bengal Engineering and Science University, Shibpur, Howrah 711 103, India

ARTICLE INFO

Article history:
Received 20 June 2012
Received in revised form 21 October 2012
Accepted 7 November 2012

Keywords:
Rice diseases
Fermi energy
Feature extraction
Rough set theory
Rule base classifier
Genetic algorithm

ABSTRACT

Development of an automation system for classifying diseases of the infected plants is a growing research area in precision agriculture. The paper aims at classifying different types of rice diseases by extracting features from the infected regions of the rice plant images. Fermi energy based segmentation method has been proposed in the paper to isolate the infected region of the image from its background. Based on the field experts' opinions, symptoms of the diseases are characterized using features like colour, shape and position of the infected portion and extracted by developing novel algorithms. To reduce complexity of the classifier, important features are selected using rough set theory (RST) to minimize the loss of information. Finally using selected features, a rule base classifier has been built that cover all the diseased rice plant images and provides superior result compare to traditional classifiers.

1. Introduction

Precision farming is defined as an information technology based farm management method aims at inferring more accurate decision in detecting diseases automatically. Accurate and timely diagnosis of rice plant diseases saves the products from quantitative and qualitative loss and plays significant role in country's economic growth. The paper aims at developing an automated system to classify the rice diseases by performing the steps: (i) Identification of the Infected Region, (ii) Extraction of Features (iii) Selection of Features, (iv) Rule Generation and (v) Classification of Diseases.

Performance of the system depends on the result of segmentation of the infected region of the diseased plant images. Thresholding (Gonzalez and Woods, 2007; Savakis, 1998) is a widely used segmentation technique (Cai and Liu, 1998; Guo and Pandit, 1998; Jawahar et al., 1997; Li and Tam, 1998), that determines threshold value experimentally and so time consuming and depends on the image quality. The energy function based segmentation methods (Ray and Saha, 2007) are designed to capture an object that exhibits high image gradients and shape compatible with the statistical model (Xavier et al., 2003). In the paper, Fermi energy (Rolf, 2001; Wei and Nigel, 2000) based segmentation method has been proposed to identify the infected regions using colour components of the images.

According to the experts' opinions, different types of rice diseases create distinct visual symptoms in the infected and its surrounding regions. Change of colour of the infected region in comparison with the background is considered as one of the important features for disease classification. In the paper, colour features are measured by calculating mean and standard deviation of the infected and background pixels as well as change of colour of the infected region in comparison with the background in three different colour planes, Red (R), Green (G) and Blue (B).

Shape of the infected region is a major symptom to identify the diseases. Oval, elliptical, circular, rectangular and irregular spot are created by the pathogens when the plant is infected. Genetic algorithm (GA) is a population based search algorithm applied earlier for shape matching (Kim and Kim, 2001). In the paper, GA is applied in two steps to detect the shape of the infected region more accurately. First, the center of the infected region is approximated and in the next step, several primitive shapes are positioned at the centre and the one that closely matches with the infected region is selected as the desired shape.

Position of infection with respect to the boundary of the leaf varies from disease to disease. A novel approach for shape detection has been proposed where the infected region is divided into number of blocks. The process continues until either a single pixel block or at most one sub-block with non-uniform pixels has been achieved representing contour of the infected portion. The smallest such block is selected globally from the image and considered as the unit block, based on which the average block distance is calculated. A coding mechanism is applied to compute the average distance unlike the conventional distance measures, which are computationally expensive.

* Corresponding author. Tel.: +91 9434185367.
E-mail address: sphadikar@yahoo.com (S. Phadikar).

0168-1699/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.compag.2012.11.001

To reduce complexity of the classifier, most important features are selected based on the essential (core) attributes and completeness of the reduced attribute set using discernibility matrix (Pawlak, 1998) concept of rough set theory (RST). The noncore attributes satisfying certain criterion are added to the core attribute set and thus the reduced set of attributes is obtained, called *reduct* (Huang et al., 2004; Raymer et al., 2000).

Rule learning techniques have largely focused on finding possible set of rules that satisfy a set of constraints (typically based on support and confidence). Classification Based on Associations (CBA) (Liu et al., 1998) and Classification based on Multiple *Association Rules* (CMAR) (Li et al., 2001) typically discover all the association rules in the first phase and then post-process the resulting rule sets in order to retain only a small number of suitable rules. In the paper to form the classification rules, a conditional attribute and its each value in the dataset is considered as an antecedent of the rule. Next all the objects in the data set satisfying the antecedent are checked for their decision values. If it is same then the respective class label determines consequence of the rule. Each rule is included in the final rule set provided it covers at least one additional object that has not been classified yet. The process continues for single antecedents and multiple antecedents until all objects are classified.

The paper is organized as follows: In Section 2, different rice diseases are described in brief. Fermi energy based region segmentation method has been discussed in Section 3. Section 4 explains feature extraction procedures. Rough set based feature selection and rule generation for disease classification have been presented in Section 5. Experimental results are given in Section 6 and 7 finally concludes the paper.

2. Description of rice diseases

Different types of rice diseases which are classified by the proposed system, described below.

2.1. Leaf brown spot

Brown spot is caused by fungus Bipolarisoryza resulting both quantitative and qualitative losses (Padwick, 1950; Rice Doctor, 2003). Typical classical *brown spot* symptoms are observed at tillering stage and beyond. Small foliar lesions are found and the shapes of the lesions vary from circular to oval with light brown colour to gray at the center and reddish brown colour at margin, shown in Fig. 1a.

2.2. Rice blast

Rice blast (Webster, 2000; Zeigler et al., 1994) is caused by the fungus Magnaportheoryzae (Couch and Kohn, 2002) and observed in both lowland and upland. Initially white to grayish green circular lesions or spots with dark green borders are found on the leaves. Older lesions on the leaves are elliptical or spindle-shaped with more or less pointed ends. A leaf infected by *blast* is shown in Fig. 1b.

2.3. Sheath rot

Sheath rot caused by the pathogen Sarocladiumoryzae. Rotting occurs on the leaf sheath that encloses the young panicles. The lesions start as oblong or some irregular spots with variation in colour from gray to light brown at centers, surrounded by distinct dark reddish brown margins (OU, 1985; Sreenivasaprasad and Johnson, 2001) are observed, shown in Fig. 1c.

2.4. Bacterial blight

Bacterial blight is caused by bacteria Xanthomonasoryzae creates water-soaked lesions usually start at leaf margins, a few cm from the tip and spreading towards the leaf base (Rice Doctor, 2003; OU, 1985). Length and width of the affected areas increase while colour changes from yellowish to light brown due to drying, shown in Fig. 1d.

3. Fermi energy based region detection

A concept based on Pauli Exclusion Principle states that Fermi energy is attained when temperature of a material is lowered to absolute zero (Rolf, 2001). The Fermi energy or Fermi level (Wei and Nigel, 2000) is expressed using Eq. (1).

$$E_F = \frac{\hbar^2 \pi^2}{2ML^2} \left(\frac{3N}{\pi} \right)^{\frac{2}{3}} \tag{1}$$

where N is the number of particles, M is the mass of the particles, L is the length of the cube and h is the Planck constant. When an image is acquired using a physical source, the information content in the image is proportional to the energy radiated by the source. Fermi energy (E_F) of an image is computed using Eq. (1), where N is mapped as the number of pixels having distinct colour values in the image, number of grey levels (256) is equivalent to length L

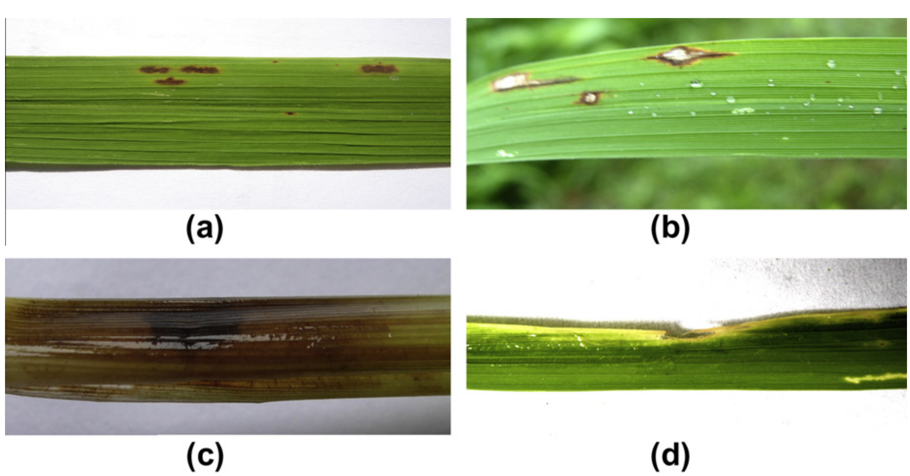

Fig. 1. (a) Leaf infected by *brown spot* disease, (b) rice leaf infected by *blast* disease, (c) rice stem infected by *sheath* rot, and (d) rice leaf infected by bacterial *blight* disease.