Computers and Electronics in Agriculture 107 (2014) 104–114

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Knowledge discovery and Leaf Spot dynamics of groundnut crop through wireless sensor network and data mining techniques

A.K. Tripathy [a,*], J. Adinarayana [a], K. Vijayalakshmi [b], S.N. Merchant [c], U.B. Desai [d], S. Ninomiya [e], M. Hirafuji [f], T. Kiura [g]

[a] Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra 400076, India
[b] Department of Entomology, ANGR Agricultural University, Rajendranagar, Hyderabad 500030, India
[c] Department of Electrical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, Maharastra 400076, India
[d] Indian Institute of Technology Hyderabad, Yeddumailaram, Andhra Pradesh 502205, India
[e] Institute for Sustainable Agro-ecosystem Services (ISAS), The University of Tokyo, Nishitokyo, Tokyo 188-0002, Japan
[f] Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization & University of Tsukuba Memuro-cho, Hokkaido 082-0071, Japan
[g] Agroinformatics Division, Agricultural Research Center, National Agriculture and Food Research Organization (NARO), 3-1-1, Kannondai, Tsukuba, Ibaraki 305-866, Japan

A R T I C L E I N F O

Article history:
Received 16 November 2012
Received in revised form 5 October 2013
Accepted 12 May 2014

Keywords:
Data mining
Knowledge discovery
Wireless sensor network
Precision farming
Pest/disease management at micro-scale level

A B S T R A C T

Data driven precision agriculture aspects, particularly the dynamic disease management, require dynamic crop-weather-environment data at micro level. An experiment was conducted during four consecutive seasons (2009 Kharif, 2009–10 Rabi, 2010 Kharif and 2010–11 Rabi) in a semi-arid tropic region of India to understand the crop-weather-environment-disease relations using wireless sensory and field-level surveillance data on the groundnut crop for Leaf Spot (LS) disease, which is economically important yet more prone in the semi-arid tropic. Tailor-made various data mining techniques (Naïve Bayes classification with Gaussian distribution, rapid association rule mining and multivariate regression mining) were developed and applied to turn the data into useful information/knowledge/relations/trends and correlation to understand crop-weather-environment-disease continuum. These dynamics obtained from the data mining techniques and trained through proposed multivariate regression (MVR) mathematical models were validated with ground level surveillance data as well as ARI model (derived from ongoing long term weather-based experiment with diversely pooled data experimented from 10 seasons in semi-arid and arid zones). It was found that LS disease infection is strongly influenced by minimum temperature (18–20 °C), prolonged duration of leaf wetness (7–10 h), high humidity (75% or more) and age of the crop. These findings have been used for development of prediction models (One week and cumulative predictions), which can assist the user community to take respective ameliorative measures and it has been found that cumulative prediction model has performed better than ARI model with respect to ground level observations in all 16 diverse dates of sowing experiments spanning for two model-years.

1. Introduction

Crop losses, particularly oilseeds, due to pests and diseases are quite substantial, particularly in the semi-arid conditions (Wang et al., 2007). Among the oilseed crops, groundnut extent its large space despite its prone to pests and diseases. Significant crop losses by pests/diseases have been reported from many countries (AgroMet-Cell, 2009; Izge et al., 2007; Singh et al., 1992). Disease is a product of the interaction between host and pathogens in favorable environment overtime. Even though susceptible host and aggressive pathogen are presence, the absence of congenial weather at canopy level, generally referred to as microclimate, restricts the disease pathogen development.

The major worldwide foliar diseases of groundnut, particularly in Semi-arid tropic region, caused by fungi are Rust (*Puccinia arachidia*), late Leaf Spot (LS) (*Cercosporidium personatum*), and early LS (*Cercospora arachidicola*), are the most important foliar diseases of groundnut worldwide, and can cause yield losses as high as 50% (Vijayalakshmi et al., 2009; Wu et al., 1999). Early LS develops small necrotic flecks that usually have light to dark-brown centers, and a yellow halo. The spots may range from 1–10 mm in

* Corresponding author. Tel.: +91 2225040508, deputed from Don Bosco Institute of Technology, Mumbai, India.

E-mail addresses: aktripathy@iitb.ac.in, tripathy.a@gmail.com (A.K. Tripathy), adi@iitb.ac.in (J. Adinarayana), vijayalkorlipara@yahoo.com (K. Vijayalakshmi), merchant@ee.iitb.ac.in (S.N. Merchant), ubdesai@iith.ac.in (U.B. Desai), snino@isas.a.u-tokyo.ac.jp (S. Ninomiya), hirafuji@affrc.go.jp (M. Hirafuji), kiura@affrc.go.jp (T. Kiura).

http://dx.doi.org/10.1016/j.compag.2014.05.009
0168-1699/

diameter and sporulation appears on the abaxial (upper) surface of leaflets. Late LS develops small necrotic flecks that enlarge and become light to dark brown. The yellow halo is either absent or less conspicuous in late LS with sporulation on the abaxial (lower) surface of leaves. LS cause premature defoliation and are commonly a major constraint to both seed yield and good quality haulm production (Butler et al., 2000).

Weather and climate are very important in determining the precise epidemiology of outbreak of the disease. Critical threshold of the meteorological elements for the incidence, spread and intensification of disease determined in the laboratory condition have little relevance to the field condition. Therefore, they have to be determined and monitored under field conditions through simultaneous observation of micrometeorological parameters and the pertinent data (Prabhakar, 2010). Temperature, humidity, rainfall, leaf wetness are the foremost weather parameters influencing the LS disease incidence (Bailey et al., 1993; Butler et al., 2000; Davis et al., 1993; Jhorar et al., 1987; Wu et al., 1999). It is rather difficult to establish a direct cause and effect relationship between any single climatic factor and disease activity as the effect of these weather elements on the disease is usually confounded (Vijayalakshmi et al., 2009). The combined effects of temperature and wetness duration on a variety of diseases have been described by polynomial equations comprising numerous parameters that lacked a clear epidemiologic significance (Monroe et al., 1997). Duthie (1997) and Wu et al. (1999) described a nonlinear model for evaluating the combined effects of temperature and wetness duration. While parameters of this model can be interpreted to provide information on the mechanisms involved in the disease response, reports on fitting empirical data and interpreting the response have not been established.

With the ever-increasing amount of information about the farms, farmers are not only harvesting in terms of agriculture output but also a large amount of data. In order to extract only the useful information, these data need to be analyzed and used for optimization (Georg et al., 2008; Tripathy et al., 2011). Such data can be used in productive decision making if appropriate data mining (DM) techniques are applied. DM is a process of extracting the most important and useful information from large set of data and to uncover previously unknown patterns and hidden relationships within the data (Agrawal et al., 1993; Monroe et al., 1997; Tripathy et al., 2011) that may be relevant to current dynamic agricultural problems. In fact, applications of DM in agriculture emerged just recently, and therefore, there are not yet any specific DM techniques specifically tailored to solve agricultural-related problems (Monroe et al., 1997). However, different data mining techniques/algorithms can be developed and tailor-made to turn the data into useful information/knowledge/relations/trends and obtain correlation to understand crop-weather-environment-disease continuum.

In the practices of agricultural information, the agricultural intelligence monitoring system is an important and indispensable link (Yuanguai et al., 2009). The advanced sensor technology and the intelligence information processing technology are important methods to guarantee correctly and quantitatively to gather real time data on different parameters pertaining to weather, crop, soil and environment, which in turn help in developing open solutions for majority of the agricultural processes. With the advent of Wireless Sensor Network (WSN), information acquisition, dissemination and the processing technology has seeped gradually into agricultural domain by the characteristics such as low-cost, low-power consumption, small devices equipped with limited sensing, data processing and wireless communication capabilities, which perfectly suites for precision agriculture where decisions are to be made at micro-climatic level at right time/place/input (Liu et al., 2008; SPANN Lab, 2009).

Weather based pest/disease forewarning models have been developed to certain extent (Bailey et al., 1993; Butler et al., 2000; Davis et al., 1993; Jacobi et al., 1995; Jhorar et al., 1987; Vijayalakshmi et al., 2009; Wu et al., 1999). However, a functionally viable model for disease forecast considered to be one of the important components in the integrated pest/disease management strategy. With the current scenario of disease resistance to fungicides/pesticides and their high costs, it is still more important to develop an early warning model to provide caution to the farmers regarding the occurrence of disease, their peak activity and migration to develop effective and efficient management system. Hence, the present investigation has been carried out to discover the correlation (direct/hidden) and dynamics of disease with micro-weather parameters obtained from WSN in field conditions, which lead to developing a forewarning model with.

2. Materials and methods

In order to study the crop-weather-disease interactions, a test bed for WSN experiment was chosen at Agriculture Research Institute (ARI) of Acharya N G Ranga Agricultural University, Hyderabad, which falls in semi-arid tropic region. The test bed, where long term weather-based experiments are being carried out on groundnut crop, had provided a platform for validation (including comparisons) of the proposed model with existing pest/disease model. This work is a part of Indo-Japan initiative to develop a real time decision support system, called GeoSense (Sudharsan et al., 2011; AgroMet-Cell, 2009), integrating Geographical Information Communication Technology (Geo-ICT) and WSN for Precision Agriculture.

2.1. Standard experimental setup

Studies on crop-weather-disease interaction were carried out consecutively in two Kharif (monsoon) 2009 and 2010, two Rabi (post-monsoon) (2009–10 and 2010–11) agriculture seasons. A standard field experiment design was laid out in the test bed (Figs. 1 and 2). Where P_1 is unprotected plot (a normal situation in farmers field) and P_2 is weather based protection plots. Four different dates of sowing (D_1, D_2, D_3 and D_4) were taken into consideration. These different dates determine the impact of disease incidence in order to observe dynamics in pre (D_1), normal (D2 and D3) and post normal weeks of sowing (D_4).

Apart from this, to have uniform and unbiased observation, surveillance data has been collected from each plot in randomly selected one square meter area locations of the plot (S_1D_1, S_1D_2, S_1D_3, S_1D_4, S_2D_1, S_2D_2, S_2D_3, S_2D_4, etc.) during flowering to harvest crop phenological stages (Fig. 2) in three replicated plots (R_1, R_2 and R_3).

2.2. Surveillance data collection

LS population dynamics (surveillance data) were obtained at every week from flowering to reproductive phenological stages, where majority of disease incidences occur, at various locations in the experimental site. The surveillance data has been collected weekly instead of daily as there will not be any significant visible changes in disease incidences. A total of 48 observations (12 weeks $\times$ 4 different dates of sowing) were made with respect to different dates of sowing in each season. Along with this, Groundnut crop age (that is at which stage of the crop the disease attack takes place and their dynamics trends) also recorded week-wise to understand the infection dynamics of LS. The weeks in a year are mapped into integer values by considering the first week of January as first standard week.