Computers and Electronics in Agriculture 85 (2012) 1–7

Contents lists available at SciVerse ScienceDirect

Computers and Electronics in Agriculture

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

Intelligent systems for the assessment of crop disorders

Anyela Camargo [a], Juan P. Molina [b], Jorge Cadena-Torres [b], Nora Jimenez [b], Jan T. Kim [c,*]

[a] Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, Ceredidion, UK
[b] Corpoica, Km. 13 way Cerete, Colombia
[c] School of Computing, University of East Anglia, Norwich, Norfolk, UK

ARTICLE INFO

Article history:
Received 3 November 2011
Received in revised form 22 February 2012
Accepted 24 February 2012

Keywords:
Crop disorder diagnosis
Machine learning
Information systems
Food security

ABSTRACT

Crop disorders are a serious threat to food security of inhabitants of remote areas in developing countries. While farmers in developed countries have frequently access to various expert resources that help them to identify the onset of a disease, farmers in developing countries usually do not have such support. However, their access to the Internet and thus to the web has rapidly improved during the last few years. This provides a new opportunity to communicate crop pathology information to remote places.

We have developed the "Information system for the assessment of plant disorders" (Isacrodi) to support farmers in protecting their crop. Farmers are guided to use a controlled but extensible set of attributes to describe the state of their crop. On this basis, Isacrodi provides suggestions which disorders may affect the crop, and which measures would be effective against these disorders. Experts provide Isacrodi with descriptions of actual incidents where they have identified the disorder. Isacrodi uses a computational classifier to provide suggestions to users autonomously. The classifier is constructed based on expert's inputs. Suggestions of disorders and countermeasures are presented as ranked lists, leaving the final identification of the disorder and decisions of countermeasures to the user, as they may have additional information beyond the attributes used by Isacrodi.

The performance of the classifier was evaluated by generating data that reflects the envisaged usage of the Isacrodi system. Data on crop disorders provided by experts was used to train the classifier and data that simulated the growers wishing to find out which disorder affects their crop was used to test the classifier. The results show that with limited expert input and errors in data provided by users, the classifier is capable of identifying disorders with reasonable accuracy, particularly when the user considers the three top scoring disorders rather than just the top one. Human experts will attain a much better accuracy than the Isacrodi classifier, particularly when provided with samples from the affected crop. However, where such expertise is not available, Isacrodi can provide valuable support to farmers.

1. Introduction

Food security has become a major issue as it is predicted that the world's population will reach 9 billion by 2050. Most part of this increase will occur in developing countries (FAO) (UN, 2006). Measures to increase food production such as plant breeding are important but not sufficient as climate change represents a long-term risk for food security (Gilligan, 2008). Shortages of water, accelerated spread of plant pathogens, and emergence of new pathogens that are adapted to altered climatic conditions are substantial threats to crop yields. These adverse impacts are collectively referred to as disorders. Disorders cause loss in yield and thus huge economic losses. As an example, Thorne et al. (2004) assess a scenario of an outbreak of Karnal bunt in the UK and conclude that under current contingency plan controls, costs of €454 million would result over a period of 10 years. As another example, crown rot disease reduces yield of winter wheat by up to 35% (1550 kg/ha) (Smiley et al., 2005). Beyond economic losses, crop failures do are a serious threat to the livelihood and to food security, especially in developing countries (Strange and Scott, 2005).

Technological advances in plant management and disease diagnosis and prognosis help avert outbreaks and spread of crop diseases. However, such advances are mainly available and affordable to farmers in industrialised countries or to those who grow crops at large scale. However, most farmers in developing countries produce at a small scale, and often they are located in isolated areas where they have no access to technical assistance from crop experts. Such farmers primarily rely on local knowledge to detect presence of a disease and to decide on which treatment to apply to control or eradicate the disease. Such knowledge typically results from experience accumulated in a given environment over a long period of time, often spanning several generations. While the value of such local knowledge cannot be overestimated, farmers

* Corresponding author.
E-mail address: jttkim@googlemail.com (J.T. Kim).

0168-1699/$ - see front matter
http://dx.doi.org/10.1016/j.compag.2012.02.017

benefit from complementing it with networked and global information. This is especially important where the local environment is subject to change, e.g. due to adoption of new farming practices, climate change, or migration and displacement. In these conditions, traditional local knowledge may lead farmers to misdiagnose diseases and apply the an unsuitable treatment. Also, when new diseases emerge, farmers may fail to diagnose these at an early stage and apply treatment when it is too late to be effective. In these scenarios, diseases develop beyond a controllable stage, resulting in a severe negative impact on the crop's yield. Farmers in developing countries depend heavily on their yield and are not normally subsidised or protected against crop failures by the government. For these reasons, the economic and social impact of a disease outbreaks can extend to an entire community. According to Vurro et al. (2010), farmers in developing countries lost 31–42% of their crop production to plant diseases, weeds and other environmental factors.

The best way to minimise impacts of a disease outbreak is to regularly monitor the crop to detect the onset of potential diseases and to take timely action if a disease is present. Monitoring should be done on the basis of local knowledge and up-to-date and reliable global information. There now is a huge wealth of information assembled by global scientific communities commonly written in English. The challenge thus is to make this information available to farmers in a format and language that they can access, especially to those at remote sites in developing countries. In recent years, Internet access and availability of mobile phones has dramatically widened in many developing countries. This opens new opportunities to the development of systems that are accessible via these technologies. Generally, using the Internet as a way to communicate and to transfer information has already shown positive results. The e-Choupals community centres in rural India facilitates decision making by providing farmers with the technological means to access information on agriculture and marketing (Ali and Kumar, 2011). The Bahía Solano telecentre in rural Colombia provides Internet facilities so that women traders can buy and order the stock they later sell to their costumers (Rednatel, 2011).

Plant disease detection by means of computer aided systems is a topic that has received limited interest until now. A number of studies have focused on the use of digital images as a tool to disease detection. For example, Chaerle and Straeten (2000) showed how thermal, reflectance and fluorescence images can be use to assess stress in plants, Rumpf et al. (2010) used a data classifier based on spectral vegetation indices to detect sugar beet diseases, Camargo and Smith (2009a,b) reported an image-processing based algorithm to extract plant disease symptoms from coloured images. They used image features from extracted regions to feed a classifier and identify possible plant disorders. Casa and Jones (2005) reviewed the advantages and disadvantages of the application of remote temperature sensing of plants by infrared thermography and infrared thermometry to assess a plants level of stress. While image analysis has much potential for plant disorder diagnosis, it must be used in combination with other information as different disorders may have similar visual symptoms, and because visual symptoms may be absent in early stages of disorders. Experts have the most complete knowledge of visual symptoms and other attributes including environmental conditions, and it is their task to suggest these attributes to farmers who are scouting for diseases or other disorders. In developed countries, expert support is provided by organisations such as the Food and Environment Research Agency (FERA, 2011) and the National Institute of Agricultural Botany (NIAB, 2011) in the UK, the National Plant Diagnostic Network (NPDN, 2011) in the USA or the Commonwealth Agricultural Bureaux International (CABI, 2011). These organisations use the web to disseminate materials (e.g. Plant Pest and Disease Factsheets, 2011; Crop Protection Compendium, 2011) and for submitting requests for expert support (e.g. Plant Disease Clinic, 2011). Such requests are handled individually by an expert and are usually chargeable. Computational systems that support growers without directly involving human experts are not offered by the organisations mentioned above. Unfortunately, the demand for expert advice by far exceeds the number of experts that is available in many developing countries. This imbalance leaves farmers unable to perform appropriate crop monitoring, apply on-time disease control and prevent economic losses. Therefore, these farmers would benefit from a system that helps them monitor their crops and establish whether their crops are healthy, and if not, what treatment they should apply in order to control the crop disorder.

In this paper we report the "Intelligent system for the assessment of plant disorders" (Isacrodi), a web based system which we have designed to assist farmers in assessing disorders in their crops and in protecting their crops on this basis. Farmers query Isacrodi by entering a structured description of the crop growing in their field, and Isacrodi provides suggestions which disorders may affect the crop, and which measures would be effective against these disorders. Experts provide Isacrodi with descriptions of actual incidents where they have identified the disorder. Isacrodi constructs a computational classifier which is based on expert's inputs, and it uses this classifier to provide suggestions to users autonomously, i.e. without requiring an expert to review each user input individually.

The Isacrodi software architecture is designed to be generic, i.e. to be used with multiple kinds of crop in many regions of the world. Specifically, Isacrodi enables experts such as agronomists, entomologists and phytopathologists to specify attributes relevant to the disorders, crops and regions that their area of expertise covers. For the demonstration presented in this paper, we have populated Isacrodi with data relevant for the Sinú region of Colombia, based on interviews with local experts and farmers. The information gathered from these interviews and in field trips, provided the basis for a realistic test of Isacrodi.

2. System description

2.1. *Isacrodi usage*

Isacrodi is a typical web application that can be accessed with any standard browser. Users require a username and a password to log in. Once logged in, a user can complete a web form with information about their crop. This form, which is illustrated by Fig. 1, is highly structured and asks the user to enter numerous *attributes* describing the state of their crop. Attributes are typed, they are either numeric (e.g. the age of the crop or the altitude of the plot) or they are categorical (e.g. the type of soil or the irrigation system). Categorical attributes are either restricted to a single value (e.g. the soil type), or they allow multiple values (e.g. symptoms). For numeric attributes, a unit (e.g. degrees Celsius, metres above sea level) is specified as well. All attributes may be left unspecified, either because they do not apply, or because the user does not have the information. As an example, the "pest density" attribute would not apply to disorder caused by a microbial pathogen, and it may also be left unspecified by a farmer who is unsure. Once the user has completed the form to the best of their knowledge, they can save it. Isacrodi will then store it as a *crop disorder record (CDR)* in its database.

For a stored CDR, users may request a diagnosis. Isacrodi then uses its classifier, detailed in Section 2.3, to designate a score to each disorder it knows about. In the diagnosis section, shown in Table 1 disorders are reported in a ranked list, with the highest scoring disorder at the top. Users should use the scores as a guidance to help them decide which disorder their crop might be

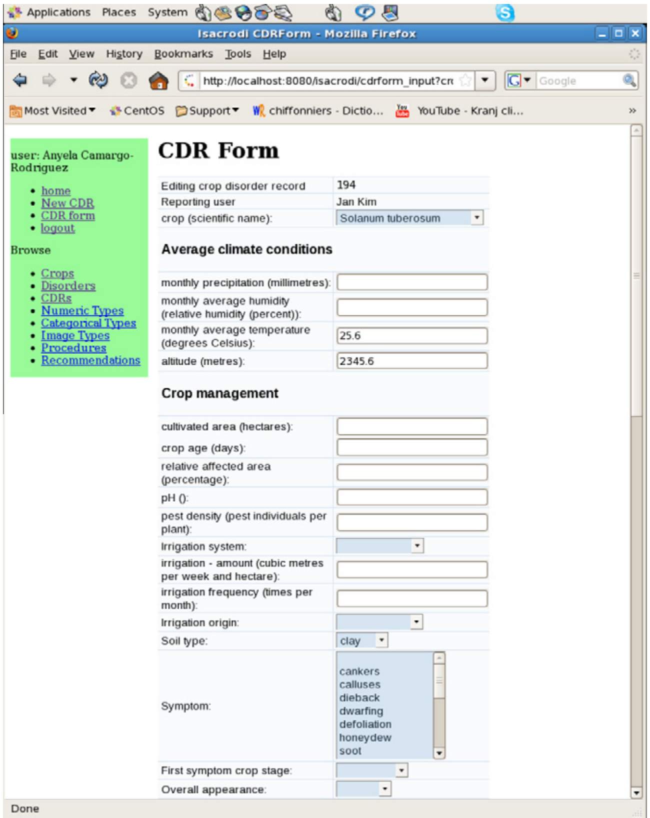

Fig. 1. CDR input form.

Table 1
Diagnosis score table.

Score	Disorder
0.8647998074877806	Tobacco budworm (*Heliothis virescens*)
0.0355477742777766765	Cabbage looper (*Trichoplusia ni*, *Pseudoplusia includens*)
0.01945471877587352	Cotton leafworm (*Alabama argillacea*)
0.0182511661141308128	Trips urticae (*Tetranychus urticae*)
0.0182142364436067	Fall armyworm (*Spodoptera frugiperda*)
0.014838698832329012	South American boll-worm of cotton (*Sacadodes pyralis*)

suffering from. Diagnoses are accompanied by suggestions of measures to take to control or eradicate the disorder.

2.2. *Architecture and design*

The architecture of Isacrodi follows the standard three-tier approach (Ramirez, 2000). It comprises various loosely coupled modules and is designed to be extensible. Fig. 2 gives an overview of the architecture and the main flows of data. Data of the Isacrodi system are held in a persistent database. The central entity of this database is the *Crop Disorder Record (CDR)*, which was introduced in Section 2.1 already. A CDR contains a reference to the *crop* that is affected, and it contains an arbitrary number of *attributes*, which are either *numeric*, *categorical* or *image attributes*. An *attribute type* is specified for each attribute, some examples for attribute types are given in Section 2.1. For each categorical type, the database details the possible *categorical values*, and also whether multiple values are permitted. The type of an image specifies what the image shows, e.g. the a field, a plant or a leaf (Fig. 2).

02
04/07/2014
Mdl. User

Types are held in the database, they are not hard-coded into the Isacrodi system. This allows experts to add new types, e.g. on the occasion of entering new disorders. Experts can thus continuously develop a controlled vocabulary by specifying and types as well as possible values for categorical types. Isacrodi also maintains tables of *crops* and *disorders*, which contain trivial and scientific names of the respective object, as well as further descriptive information. Crops and disorders are also part of Isacrodi's controlled vocabulary.

Where the disorder affecting a crop described by a CDR is known, this is recorded in the *expert diagnosis*. CDRs containing an expert diagnosis are called *labelled CDRs*. CDRs may further contain a *diagnosis* computed by Isacrodi's classifier. Such a computed diagnosis comprises a score for each disorder in Isacrodi's database. The highest scoring disorder is the one which, according to the classifier, most likely affects the crop described by the CDR.

Based on the diagnosis, Isacrodi computes a *recommendation*, consisting of a list of *procedures* suggested to control the disorder, or to prevent it in the future. Procedures are measures that a farmer might take, such as applying an insecticide or fungicide, draining the field, or removing and burning affected plant parts. For each disorder, the database details which procedures are effective to control it. A recommendation is comprised of the procedures that are effective against the highest scoring disorders.

2.3. *Computational diagnosis of disease disorders*

Computational *diagnosis providers* are modules that assign scores to disorders based on data contained in a CDR. Isacrodi uses the labelled CDRs to construct, or train, its classifier. The trained classifier is then used to compute diagnoses for unlabelled CDRs, and thereby to support users who submitted these CDRs in protecting their crop. Diagnoses comprise a score for each disorder, giving an indication of how likely the disorder affects the crop described by the CDR. Based on the diagnosis, Isacrodi computes a recommendation suggesting treatments to be considered to control the disorder.

Isacrodi's approach to diagnosis is inspired by the disease triangle concept (Scholthof, 2007) which states that disease outbreaks arise from interaction between the host crop, the pathogen, and the environmental conditions. If environmental conditions are permissive, the host is rendered susceptible, and if the pathogen is present, an outbreak of the disease results. The disease triangle concept is used by plant pathologists and entomologists e.g. to investigate crop disorders and track down its evolution over time (Parker and Gilbert, 2004) and as a method to diagnose potential crop disorders.

Realistically, it must be expected that users entering a CDR will not provide all attributes that are relevant for diagnosing a disorder. Also, errors in attribute entry must be expected. Therefore, a central goal in designing diagnosis providers is robustness to errors and incompleteness in user input. While such problems inevitably reduce accuracy of diagnosis, it is essential that this degradation is graceful, i.e. as the number of problems grows, accuracy should gradually decline rather than sharply drop off.

Isacrodi provides an extensible architecture that enables implementation of a wide range of diagnosis providers. The current diagnosis provider is based on multi-class Support Vector Machines (mc-SVM), as provided by the widely used libsvm implementation (Hsu and Lin, 2002). Each disorder corresponds to a class. For computing a diagnosis, the attribute data is mapped to a *feature vector*. The prediction method of the mc-SVM model is then applied to the feature vector. For each class, this method estimates a probability with which the feature vector belongs to that class. The diagnosis provider assigns these probabilities as scores to the respective disorders. Fig. 3 illustrates this process.

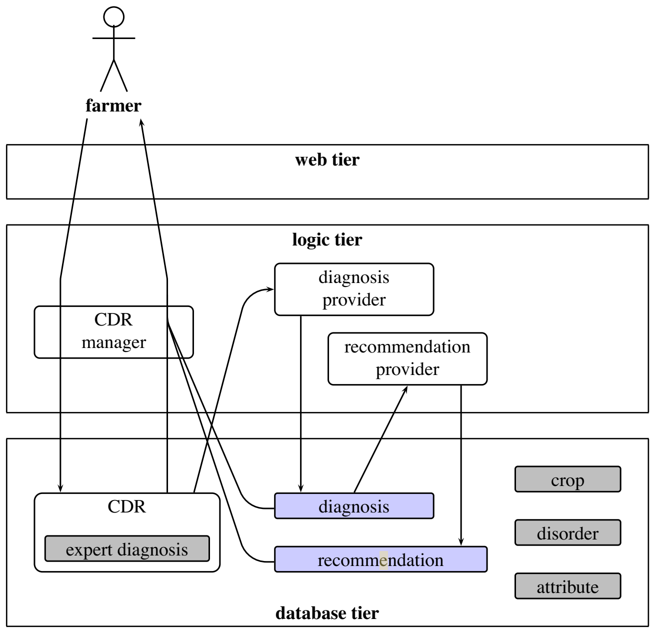

Fig. 2. Overview of the Isacrodi architecture and the main flows of data. Farmers provide CDRs using the web tier, which employs a CDR manager to validate the farmer's input and to store it in the database. The diagnosis provider is invoked to compute a diagnosis, and subsequently, a recommendation provider is invoked to recommend procedures. The farmer can then access the CDR along with the diagnosis and recommendation that were computed; the CDR manager retrieves these along with the CDR itself. Experts enter the items shown in grey (crops, disorders, attributes), as well as expert diagnoses, these items are shown in grey. Items computed by Isacrodi are shown in violet. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

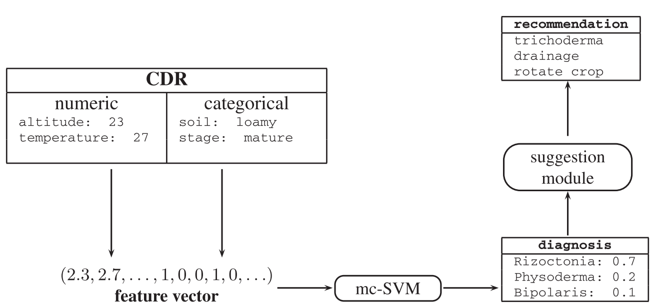

Fig. 3. Data flow schema of computation of diagnoses and suggested countermeasures.

The methods of feature vector mapping and training of the mc-SVM are adapted from Hsu et al. (2003). Each numeric attribute is mapped to one feature vector component. In preparation for training, the mean and the standard deviation of each attribute are determined. On this basis, the attribute data for each feature component are shifted and scaled so that the mean is 0 and the standard deviation is 1. The same scaling is applied when mapping unlabelled CDRs during diagnosis.

Categorical attributes are mapped to a set of feature vector components using orthogonal coding (Kim et al., 2004). Each component corresponds to one of the possible categorical values and is set to 1 if the attribute takes that value and to 0 otherwise. Where multiple values are permitted, multiple components may be set to 1.

Following the procedure suggested by Hsu et al. (2003), the diagnosis provider uses the RBF kernel and the hyper-parameters are selected using a grid search with $C = 2^{-5}, 2^{-5}, \ldots, 2^{16}$ and $\gamma = 2^{-15}, 2^{-14}, \ldots, 2^{4}$, and 10-fold cross-validation. The mc-SVM model for diagnosis is then trained using all labelled CDRs and the best hyper-parameter settings found by the grid search.

3. Testing the SVM diagnosis provider

3.1. Generating training and test data

We test and evaluate the performance of the SVM diagnosis provider using randomly generated CDR data. For this purpose we devised a method for generating data that reflects the envisaged usage of the Isacrodi system, where experts determine disorders and enter or curate CDRs in which attribute values are correct. On the other hand, CDRs entered by growers wishing to find out which disorder affects their crop will may be incomplete and contain errors. We use a simple model of ecological niches to randomly generate attribute values for a disorder.

For each disorder we designate a set of relevant environmental conditions, such as temperature, soil type, humidity. For a given disorder, all relevant conditions have to be within a permissive range to enable the disorder to affect the host crop. Relevant conditions vary with disorders. As an example, the soil type is relevant for a fungal pathogen if it infests crops on loamy or silty soils, but not on other soil types. On the other hand, for an insect pest that requires high temperatures but affects crops regardless of the type of soil type, temperature is relevant while the soil type is not.

Each relevant condition corresponds to an attribute, so we specify a set of relevant attributes for each disorder, and a range of permissive values for each relevant attribute. The ranges of all relevant attributes enclose a (high dimensional) hyper-cuboid which is our simple model of the disorder's niche and therefore called the *niche hyper-cuboid*. The value range for a numeric attribute is given by the minimal and maximal values. The range of a categorical attribute is the set of all permissive values.

Following our assumption that experts determine disorders and attribute values correctly, we generate labelled CDRs where the attribute values are drawn from a uniform distribution over the niche hyper-cuboid. This approach implies that all attribute values are independent (i.e. uncorrelated), which is a substantial simplification. Realistically, fairly strong correlations among some parameters are to be expected. As an example, for a contagious disorder the affected area will increase over time, and therefore, the affected area and the crop's age will be fairly strongly correlated. The volume of the hyper-cuboid can therefore be much larger than the volume of the actual niche. Additionally, hyper-cuboids of two disorders may substantially overlap (i.e. they may share a sizeable intersection) even though the actual niches are disjoint. However, as we do not have empirical data to determine niche structures with correlations among attributes, we use the uniform and uncorrelated distribution as the most unbiased one.

It should be noticed that CDRs that fall into a region where multiple niches overlap cannot be expected to result in accurate classification. The mc-SVM, like many other classifiers, is capable of fitting correlation structures that are expected in real data. Therefore, classifiers trained with real data can be expected to achieve higher accuracies than classifiers trained with data generated with the hyper-cuboid approach.

Differently from experts, we assume that growers who use Isacrodi may enter CDRs that are incomplete or that contain errors. The possible problems and corresponding details of simulating them in random CDR generation are:

- Users may fail to provide relevant attributes, e.g. because they do not have the corresponding information. Test data generation therefore comprises a *missing attribute probability* (map) that applies to both numeric and categorical attributes.

- Users may make mistakes when determining numeric attribute values, due to imprecise measurement or misunderstandings. We reflect this by drawing numeric values for testing from a Cauchy distribution rather than from a uniform distribution, so that values outside the niche hyper-cuboid can be generated. The scale of the Cauchy distribution is the difference between the maximal and the minimal attribute value, multiplied by the *numeric range multiplier* (nrm).
- Wrong values may also be provided for categorical attributes. We simulate such errors by drawing the categorical value randomly from all possible values, rather than from the permissive values only. The *categorical error probability* (cep) determines the probability of such a mistake.

We have chosen 14 disorders that affect cotton and maize, and that are important in the Sinú river valley in the Córdoba province of Colombia. For each disorder we have identified the relevant environmental conditions and their respective permissive range. These disorders are summarised in Table 2 and details on the permissive ranges for each disorder are provided in the electronic supplement. Since many attribute values (such as crop height) strongly depend on the crop, we define a separate hyper-cuboid for each host crop. Thus we have defined a total of 17 niche hyper-cuboids. Based on these, we constructed a training set comprised of 30 samples for each of the 17 disorders, drawn from a uniform distribution over the respective hyper-cuboid to simulate labelled CDRs provided by experts, as described above. We then constructed test sets by simulating CDRs provided by non-expert users, as described above. The control parameters for test set generation are given in Table 3.

Table 2
List of crop disorders and corresponding scientific names included in Isacrodi. Crop disorders commonly affect more than one host.

Disorder	Common name	Host crops
Agrotis ipsilon	Black cutworm	Cotton, maize
Alabama argillacea	Cotton leafworm	Cotton
Anthonomus grandis	Boll weevil	Cotton
Aphis gossyppi	Cotton aphid	Cotton
Bemisia tabaci	White fly	Cotton
Corythucha gossypii	Cotton lace bug	Cotton
Dysdercus sp.	Cotton stainer	Cotton
Helicoverpa zea	Corn earworm	Cotton, maize
Heliothis virescens	Tobacco budworm	Cotton
Pectinophora gossypiella	Pink bollworm	Cotton
Sacadodes pyralis	South American boll-worm of cotton	Cotton
Spodoptera frugiperda	Fall armyworm	Cotton, maize
Tetranychus urticae	Trips urticae	Cotton
Trichoplusia ni, Pseudoplusia includens	Cabbage Looper	Cotton

Table 3
Control parameter settings for test set generation. One test set was generated for each combination of control parameters.

Control parameter	Values	
Missing attribute probability	$0, 0.1, 0.2, \ldots, 1$	(11 values)
Numeric range multiplier	$0.1, 0.2, 0.3, 0.5, 0.7, 1.0, 1.3, 1.6, 2.0$	(9 values)
Categorical error probability	$0, 0.1, 0.2, \ldots, 1$	(11 values)

3.2. *Implementation*

The Isacrodi system is implemented using Java Enterprise Edition technologies (j2ee, 2011). The business tier, providing the logic of CDR management and disorder diagnosis is realised with Enterprise JavaBeans (ejb3, 2011). The web tier is implemented using the Apache Struts framework (Apache struts, 2011).

3.3. *Performance of the SVM classifier*

Using the control parameter settings given in Table 3 we produced $1189 = 11 \cdot 9 \cdot 11$ test sets. For each of these, we determined the accuracy, i.e. the frequency with which the classifier diagnoses the correct disorder. Fig. 4 shows accuracy as a function of missing parameter probability for control parameter settings that we consider representative of non-expert users. At a numeric range magnifier of 0.3, approximately 20% of numeric values are outside of the niche cuboid, and as we use the Cauchy distribution, such values are frequently far from the cuboid. This represents user errors that might be due to misunderstandings such as the use of a wrong unit of measurement (e.g. metres instead of centimetres). Similarly, a categorical error probability of 0.2 simulates that 20% of categorical values are misunderstood and as a result, the values entered amount to random noise.

As the missing attribute probability increases, accuracy deteriorates. At a missing attribute probability of 100%, CDRs contain no information and the scores for all disorders are identical as a consequence. The classifier then picks one of the disorders at random, resulting in an accuracy of $1/17 \approx 0.059$. This is the accuracy expected for random guessing.

Deterioration of accuracy is limited for missing attribute probabilities up to 0.4 and drops off more quickly as the missing attribute probability increases further. This means that users may submit CDRs in which up to 40% of the relevant attributes are missing and still obtain a quite accurate diagnosis.

The coloured lines in the plots shown in Fig. 4 depict results with increasing categorical error probability (left) and numerical range magnifier (right). As expected, increasing both of these reduces accuracy. Importantly, however, there are no sudden drop-offs in accuracy. This demonstrates that the SVM classifier's performance degrades gracefully: Loss of accuracy are proportionate to the respective loss of input quality but there is no sudden breakdown of accuracy at any point. The classifier is capable of using the information contained even in low quality CDRs.

Even in the most favourable conditions (no missing attributes, few categorical errors and a small numerical range magnifier) accuracy stays below 0.9, so there are still more than 10% of CDRs for which the SVM classifier yields a wrong diagnosis. As explained in Section 3.1, CDRs that fall into an area of overlap of multiple niches may result in incorrect diagnoses and accuracy is limited as a result. However, the score of the correct disorder should be elevated, and the second or third highest scoring disorder may be the correct diagnosis. Fig. 5 shows the frequency with which one of the three highest scoring disorders is the correct one. This means that the correct disorder will be displayed as one of the top results, giving the user a chance to identify it.

It should also be noted that drawing CDRs from a uniform distribution over the niche cuboid neglects the fact that there will be quite substantial correlation among some attributes. The support vector machine (Vapnik, 1998) used in the classifier can take advantage of such correlation structure. From this perspective, the method we used to generate training data represents a particularly difficult problem for the SVM classifier, and it is reasonable to expect that its performance will be better when applied to more realistic data.

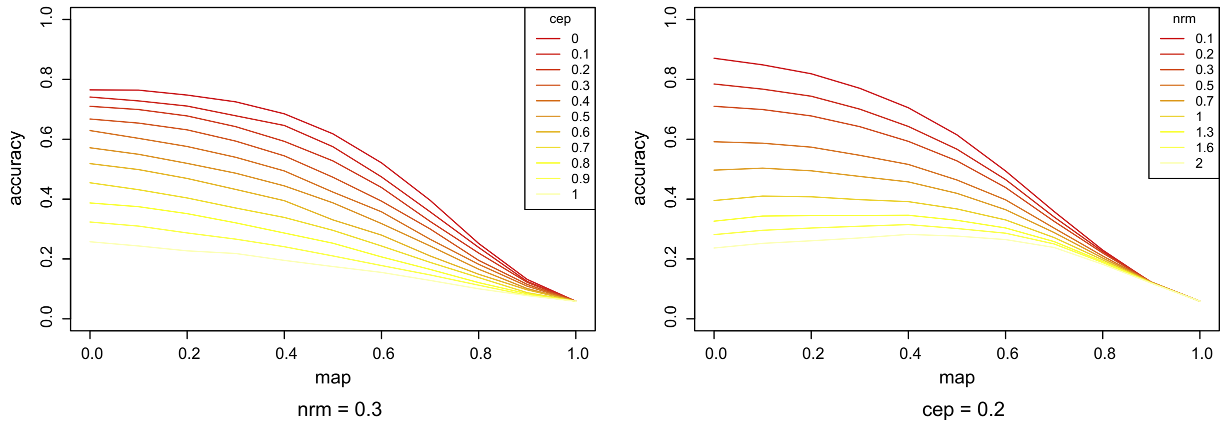

Fig. 4. Graphs showing accuracy of the SVM classifier as a function of missing attribute probability (map), numeric range magnifier (nrm) and categorical error probability (cep). The left panel shows accuracy values for a fixed numeric range magnifier of 0.3, the right panel shows accuracies for a categorical error probability fixed at 0.2.

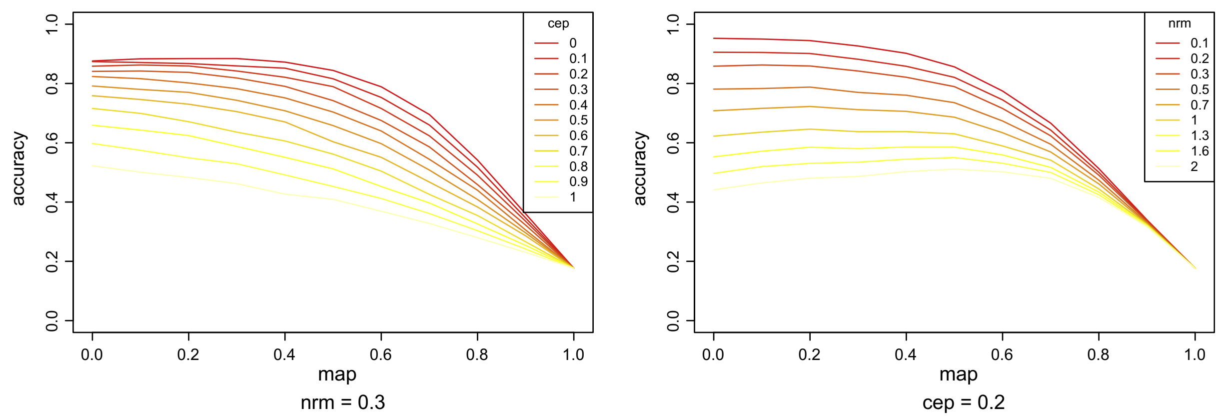

Fig. 5. Graphs showing the frequency with which one of the three highest scoring disorders is the correct diagnosis. The other control parameter settings are as in Fig. 4.

4. Conclusions

In this paper we have reported Isacrodi, a web based system designed to assist farmers in assessing disorders in their crops, and in protecting their crops on this basis. Isacrodi uses a controlled vocabulary to describe key components such as crops and crop disorders. This enables farmers to use the system without requiring training, and experts to continually expand and refine the system to include more crops, disorders and attributes. Crop disorders are described in the form of CDRs which can be labelled if the disorder has been determined by an expert, or unlabelled if the disorder is unidentified. Isacrodi uses diagnosis providers which are inspired by the disease triangle concept (Scholthof, 2007). Diagnosis providers use the expert knowledge contained in the labelled CDRs to support users in assessing unlabelled CDRs. Currently, Isacrodi uses a diagnosis provider which based on a multi-class support vector machine (Vapnik, 1998, 2002).

We evaluated the performance of the SVM diagnosis provider with CDR data generated by a random process that that captures key features of the envisaged usage of the Isacrodi system. A simple model of ecological niches is used to randomly generate attribute values for a disorder, and differences between experts and users with regard to observing and entering attribute data are reflected as well. The results of the test showed to that even in favourable conditions, diagnoses are not perfectly accurate. However, even where the diagnosed disorder, i.e. the disorder with the highest score, was not correct, the second or third highest scoring disorder could be the correct diagnosis. This means that the correct disorder will be displayed as one of the top results, giving the user a chance to identify it, e.g. by obtaining additional information or local expert knowledge.

The attributes in CDRs generated for testing the diagnosis provider were entirely uncorrelated However, in reality, there are significant correlations between attributes (e.g. crop age and affected area, see Section 3.1). The support vector machine used in the classifier is designed to be capable of capturing such correlation structures. Therefore, the diagnosis provider will likely perform better when trained with real data.

The current implementation of Isacrodi is sufficient for evaluation and basic demonstration, but functionality that is not essential for these basic purposes is missing or incompletely developed. However, the technical design of Isacrodi is based on architectural patterns, open standards and technologies for enterprise-grade software systems that enable scalability and extensibility. Thus, the Isacrodi software infrastructure supports development, ranging from improvements of diagnosis and recommendation providers to providing global accessibility by internationalisation and localisation, and thus it enables continuous development towards a fully productive system.

Isacrodi's main beneficiaries are farmers who have limited access to expertise in crop science. Typically, such farmers or farming communities do not have the time, expertise and funds to effectively use resources such as the CABI Crop Protection Compendium (Crop Protection Compendium, 2011). More accessible materials (e.g. FERA fact sheets Plant Pest and Disease Factsheets (2011)) are not available for their crops and in their language. Services such as the NIAB Plant Disease Clinic (Plant Disease Clinic, 2011) are often not affordable to farmers in developing countries. In a substantial part of the developing world, such resources are not available altogether, as the cost of setting up and operating them is prohibitive to societies and governments.

Isacrodi is designed to address this problem. The system computes diagnoses autonomously, i.e. without requiring a human expert. Therefore, the cost of obtaining diagnoses is small and they may well be provided free of charge to end users. Differently from e.g. downloading factsheets, Isacrodi users actively provide information about their crop. This enables Isacrodi to provide diagnoses and suggestions for treatment that are tailored to be relevant to the user. As a major benefit, Isacrodi may alert farmers to disorders that previously have not affected crops in their local environment. Disorders and countermeasures are presented as ranked lists, enabling and supporting farmers and farming communities to include their local knowledge in their decision making. Finally, the CDRs provided by farmers are also a potential resource for crop disorder experts, as they might be useful e.g. for monitoring geographical distributions of disorders.

Acknowledgment

This work was supported by funding from the Royal Society, Grant Number JP090307. The authors would like to thank the farmers and agronomists, entomologists and phytopathologists from the Sinú region in Monteria Colombia. They kindly provided us with vital information and feedback that enriched the Isacrodi system.

References

Ali, J., Kumar, S., 2011. Information and communication technologies (ICTs) and farmers' decision-making across the agricultural supply chain. International Journal of Information Management 31 (2), 149–159.

Apache struts, 2011. Apache struts. <http://struts.apache.org/>.

CABI, 2011. Centre for agricultural bioscience international. <http://www.npdn.com>.

Camargo, A., Smith, J.S., 2009a. Image pattern classification for the identification of disease causing agents in plants. Computers and Electronics in Agriculture 66, 121–125.

Camargo, A., Smith, J.S., 2009b. An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosystems Engineering 102, 9–21.

Casa, R., Jones, H., 2005. LAI retrieval from multiangular image classification and inversion of a ray tracing model. Remote Sensing of Environment 98 (4), 414–428.

Chaerle, L., Straeten, D.V.D., 2000. Imaging techniques and the early detection of plant stress. Trends in Plant Science 5 (11), 495–501.

Crop Protection Compendium, 2011. Crop protection compendium. <http://www.cabi.org/cpc/>.

ejb3, 2011. Enterprise javabeans. <http://www.jboss.org/ejb3>.

FERA, 2011. The food and environment research agency. <http://www.fera.defra.gov.uk>.

Gilligan, C.A., 2008. Sustainable agriculture and plant diseases: an epidemiological perspective. Philosophical Transactions of the Royal Society of London, Series B 363 (1492), 741–759.

Hsu, C., Lin, C., 2002. A comparison of methods for multiclass support vector machines. IEEE Transactions on Neural Networks 13, 415–425.

Hsu, C.-W., Chang, C.-C., and Lin, C.-J., 2003. A practical guide to support vector classification. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.

j2ee, 2011. Java 2 platform, enterprise edition (j2ee). <http://java.sun.com/j2ee/>.

Kim, J.T., Gewehr, J.E., Martinetz, T., 2004. Binding matrix: a novel approach for binding site recognition. Journal of Bioinformatics and Computational Biology 2, 289–307.

NIAB, 2011. National institute of agricultural botany (niab). <http://www.niab.com>.

NPDN, 2011. National plant diagnostic network. <http://www.npdn.com>.

Parker, I.M., Gilbert, G.S., 2004. The evolutionary ecology of novel plant–pathogen interactions. Annual Review of Ecology, Evolution, and Systematics 35 (1), 675–700.

Plant Disease Clinic, 2011. Plant disease clinic. <http://www.niab.com/pages/id/142/Plant_Disease_Clinic>.

Plant Pest and Disease Factsheets, 2011. Plant pest and disease factsheets. <http://fera.defra.gov.uk/plants/publications/plantPestDiseaseFactsheets.cfm/>.

Ramirez, A.O., 2000. Three-tier architecture. Linux Journal 75 (Article 7).

Rednatel, 2011. The `Red Nacional de Telecentros` home page. <http://www.telecentros.org.co/>.

Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., Plümer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Computers and Electronics in Agriculture 74, 91–99.

Scholthof, K.-B.G., 2007. The disease triangle: pathogens, the environment and society. Nature Reviews Microbiology 5, 1740-1526.

Smiley, R.W., Gourlie, J.A., Easley, S.A., Patterson, L.M., Whittaker, R., 2005. Crop damage estimates for crown rot of wheat and barley in the pacific northwest. Plant Disease 89, 595–604.

Strange, R.N., Scott, P.R., 2005. Plant disease: a threat to global food security. Annual Review of Phytopathology 43 (1), 83–116.

Thorne, F., Brennan, J., Kelly, P., Kinsella, A. 2004. Quantification of risks associated with plant disease: the case of karnal bunt of wheat. Technical Report RMIS 4919, Teagasc – Agriculture and Food Development Authority, <http://www.teagasc.ie/research/reports/crops/index.asp>.

UN, 2006. World population prospects. The 2006 revision. <http://www.un.org/>.

Vapnik, V., 1998. Statistical Learning Theory. John Wiley & Sons, New York.

Vurro, M., Bonciani, B., Vannacci, G., 2010. Emerging infectious diseases of crop plants in developing countries: impact on agriculture and socio-economic consequences. Food Security 2, 113–132. http://dx.doi.org/10.1007/s12571-010-0062-7.

Intelligent systems for the assessment of crop disorders

Camargo, Anyela; Molina, Juan P.; Cadena-Torres, Jorge; Jimenez, Nora; Kim, Jan T.

01 **Mendeley User** Page no. **2**

4/7/2014 2:28

Ver el enfoque, expertos hacen sus comentarios y de allí el sistema recomienda

02 **Mendeley User** Page no. **3**

4/7/2014 2:28

CDR: Crop Disorder Record