Automatic annotation of plant diseases symptoms from digital images

Perla A Troncoso Rey [1], Ammar Halabi [2], Flavio M Santana [3] Anyela V Camargo [4,*]

1

[4] *National Plant Phenomics Centre, IBERS, Aberyswyth University, Gogerddan, Aberystwyth, SY23 3EB, UK*
[1] *Earlham Institute, Norwich Research Park, Norwich, NR4 7UG, UK*
[2] *Department of Informatic, University of Fribourg, Bd de Prolles 90, Switzerlands*
[3] *Embrapa, Brazil*

Correspondence*:
x
x, avc1@aber.ac.uk

ABSTRACT

Keywords: computer vision; plant diseases; machine learning; ontologies

1 INTRODUCTION

The future of global agriculture and its impact on food security is one of the most urgent issues in todays world. Farmers must prepare for changes in the climate that is likely to feature more erratic weather patterns that will necessarily have an effect in the emergence and re-emergence of plant diseases. Early and accurate diagnosis systems on local, regional, and global scales are necessary to predict pest and disease outbreaks and allow valuable time to formulate and develop mitigation strategies. Forecasting the appearance and development of a disease is difficult, as many environmental and other factors influence the complex interactions between pathogen, host, and vector.

Fortunately, Internet access and mobile phone technologies have much improved during the last few years and are becoming increasingly accessible. This provides a new opportunity to communicate crop pathology information more widely. Containing the spread of plant diseases in a profoundly interconnected world requires active vigilance for signs of an outbreak, rapid recognition of its presence, and diagnosis of its cause, in addition to strategies and resources for an appropriate and efficient response. Due to the rapid spread of plant diseases across the world, disease surveillance and monitoring systems based on multi-country, multi-institution partnerships are necessary to predict pest and disease outbreaks and allowing a valuable time to formulate and develop mitigation strategies.

Early detection is essential for the control of emerging, re-emerging, and novel infectious diseases, whether naturally occurring or manually introduced as a result of human mobility. Containing the spread of such diseases in a profoundly interconnected world requires active vigilance for signs of an outbreak, rapid recognition of its presence, and diagnosis of its cause, in addition to strategies and resources for an appropriate and efficient response. Considerable time often elapses between the introduction of an agricultural pathogen and its detection. Given sufficient warning prior to the introduction of a new plant disease threat, researchers can reduce the impact of disease by identifying chemical control measures or by breeding resistant crop varieties [11].

2 MATERIAL & METHODS

2.1 Images

2.2 Image processing

Digital images were processed using Matlab MATLAB (2016) in the following sequence: 1) Nearest neighbour foreground/background colour separation was used to classify pixels. Two sets of colour intensities, corresponding to foreground (target) and to background (non target) are selected. RGB pixels matching selected intensities are mapped into the image. A search around mapped pixels is performed to identify pixels with similar intensities that might be part of the foreground/background regions. Once the search is performed, the image is converted to binary, where 1 is the target (disease symptom) and 0 the background (e.g. leaf, soil and stones). 2) Morphological techniques were then applied to deal with pixels incorrectly classified. First, morphological erosion was applied to remove small and isolated pixel regions incorrectly classified as plant. Second, morphological dilation was applied to correct for those pixels located in the border of the images that were incorrectly classified as background. Once the images were segmented, [**todo:** n?] image features (Table [**todo:** n] shows the list of features - descriptors) describing the geometry, shape and size of the lesions were extracted. Some of these features are illustrated in Figure [**todo:** ?] (an example set is shown in Figure S1).

2.3 Ontology

The Plant Protection Ontology (PPOntology) was developed in 2008 by Ammar Halabi at the International Center for Agricultural Research in the Dry Areas (ICARDA), near Aleppo, Syria. This project aimed to model the knowledge of agricultural experts in diagnosing barley diseases as an initial stage to building a system for automatic detection of barley disorders. As a result, PPOntology establishes a vocabulary to describe symptoms, environmental conditions and agricultural practices, which are the main factors taken into account when experts carry out their diagnosis.

The ontology was developed by reviewing various relevant systems, agricultural manuals, and through interviewing agricultural experts in ICARDA. The researcher reviewed existing attempts to implement systems for detecting plant disorders, including the Regional Wheat Expert System, the tomato expert system developed by Al-Shamaa , and the Knowledge Acquisition Tool (KAT) Boose (1990) developed at ICARDA and CLAES El-Sebae Ahmed et al. (2002). This helped establish an initial understanding of the vocabulary and categories used to describe and model factors related to plant disorders, and to conceptualize the problem domain as consisting of three main concept categories: disorders, symptom groups, and control plans.

After reviewing the use of Intelligent Systems for Plant Protection project (ISPP) developed by several organizations, in addition to the Barley Disease Handbook Neate and McMullen (2010), PPOntology was refined to account for various types of disorders (e.g. biotic, including micropests like fungi or viruses; or abiotic, including mineral deficiencies or environmental stresses). Another main concept category was added as well to account for environmental conditions. The researcher validated the resulting knowledge model with two experts in entomology and plant pathology in ICARDA. This brought up the importance for accounting for the growth sage of the plant, which strongly influences the development of symptoms of a certain pathogen. Subsequently, the ontology was updated to split cases (a renaming of the previous category of ”sympom groups”) into various growth stages. With this scheme, the same disorder (or pathogen) can be linked to various cases belonging to different growth stages. Another concept category named ”observations” was added to group all possible observable symptoms, and this was split to various categories to correspond to observations on different plant parts.

The final refinement included further nuances in concept categories to account for plant attributes (e.g. leaf color), growth stages, and to link them with observations. This was done through building a partial knowledge base with the ontology. The researcher filled-in most barley disorders common in Syria, including fungal, bacterial, viral, and insect disorders along with phytotoxicities and mineral deficiencies [**todo:** Appendix A]. He also entered information about biological organisms and materials causing disorders for barley plants. This was accumulated based on Berkey (1992), Neate and McMullen (2010), [**todo:** CANNOT FIND: UISPP 2006], as well as [**todo:** CANNOT FIND: Yahyaoui et al 2003] for verification and validation.

2.3.1 Ontology structure

Here we summarize the structure of basic concepts in PPOntology. In the following list, a concept that is nested under another concept represent a subclass of that higher class concept.

- Organism: represents the base class of all classes and instances of biological organisms that are related to barley disorders.
  - Animal: represents the base class for all classes and instances of animals (e.g. Rat).
  - Microorganism: represents the base class for all classes and instances of microorganisms (e.g. Cochliobolus Sativus).
- Material: represents the base class of all classes and instances of materials that are related to barley disorders.
  - Mineral: represents the class of nutrient minerals of the barley plant (e.g. Nitrogen).
  - Pesticide: represents the class of pesticides used for the control of barley disorders (e.g. Dicamba).
- Disorder: represents the base class of all classes and instances of barley disorders (e.g. barley stripe). Every Disorder instance is linked to a set of Organism or Material instances that are considered the agents of the corresponding disorder. Also, each Disorder instance is also linked to a set of Case instances that represents a group of possible manifestations.
  - Abiotic Disorder: represents the base class for all classes and instances of barley disorders caused due to non-biotic reasons (e.g. Boron Phytotoxicity).
  - Biotic Disorder: represents the base class for all classes and instances of barley disorders caused by biotic pathogens (e.g. Powdery Mildew).
- Plant Attribute: represents the base class of all classes and instances of possible attributes of plant parts (e.g. Leaf Color).
- Growth Stage: represents the class of different growth stages of the barley plant (e.g. Heading).
- Observation: represents the observations of different plant parts, where an Observation instance links between a set of Plant Attribute instances and a set of Growth Stage instances to mark the stage of growth at which these attributes were observed.
- Environmental Conditions: represents the base class of environmental conditions surrounding barley plants.
- Control Plan: represents groups of control measures and directions associated for the treatment of specific cases of disorders.

- Case: represents the class of barley disorder cases, which are associated to their causing disorders. Each Case instance is associated with a set of Observation instances that represents a group of consistent observations which can be made in the course of corresponding disorder, a set of Environmental Condition instances, and a Control Plan instance.

2.4 Plant infection and phenotyping

2.5 Automatic annotation

3 RESULTS

4 DISCUSSION

DISCLOSURE/CONFLICT-OF-INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

We are grateful to *Funding*: for finantially suportting the project.

REFERENCES

Berkey, J. P. (1992). Storytelling, preaching, and power in mamluk cairo. *Mamluk Studies Review* 4, 53–73

Boose, J. H. (1990). Knowledge acquisition tools, methods and mediating representations. *Knowledge Acquisition for Knowledge-Based Systems (Motoda, H. et al., Eds), IOS Press, Ohinsha Ltd., Tokyo* , 123–168

El-Sebae Ahmed, S., Rafae, A., and Shaalan, K. (2002). Expert systems: Useful tools for enhancing agricultural research and production. *ICARDA Caravan (ICARDA)*

MATLAB (2016). *version 7.10.0 (R2010a)* (Natick, Massachusetts: The MathWorks Inc.)

Neate, S. and McMullen, M. (2010). Barley disease handbook

FIGURES