Integrating Diagnostic Expert System with Image Processing via Loosely Coupled Technique

Mohammed El-Helly
Central Lab. For Agricultural Expert System (CLAES)
Agriculture Research Center (ARC)
helly@mail.claes.sci.eg

Ahmed Rafea
Computer Science Dept.
American University in Cairo (AUC)

Salwa El-Gamal
Computer Science Dept.
Faculty of Computers &Information
Cairo University

Reda Abd El Whab
I.T. Department.
Faculty of Computers &Information
Cairo University

Abstract

Conventional Expert systems especially those used in diagnosing diseases in agricultural domain depends only on textual input. Usually abnormalities for a given crop are manifested as symptoms on various plant parts. To enable an expert system to produce correct results, end users must be capable of mapping what they see in a form of abnormal symptoms to answers to questions asked by that expert system. This mapping may be inconsistent if a full understanding of the abnormalities on any plant or in questions being asked does not exist.
This paper presents a novel approach for integrating image analysis techniques into diagnostic expert systems. The result of applying this approach is presented through the use of cucumber fungal diseases as a case study.

Keywords: Expert System – Diagnosis - Image Processing - Feature Extraction - Classification

1. Introduction

Expert systems are intelligent computer programs that are capable of offering solutions or advices related to specific problems in a given domain, both in a way and at a level comparable to that of human experts in a field. One of the advantages of employing expert system is its ability to reduce the information that human users need to process, reduce personnel costs, and increase throughput. Another advantage of expert systems is it performs tasks more consistently than human experts [1]. “Knowledge-Based expert system technology has been successfully applied to a variety of agricultural problems since the early 1980s” [2]. At CLAES (Central Lab. Of Agricultural Expert System in Egypt) several expert systems have been developed for different agriculture activities. Many of these expert systems are used to diagnose disorders from observation symptoms. These diagnosing expert systems depend on the ability of an end user to understand abnormal symptoms of the plant and to convey these symptoms through a textual dialogue. Depending on the user’s level of understanding of the abnormal observations, the expert system can reach the correct diagnosis. If, however, the end user interprets the abnormal observations in a wrong way and chooses a wrong textual answer to a presented question, then the expert system will reach a wrong conclusion. Devising a method whereby abnormalities are automatically detected, would greatly reduce the risk of human error, and would accordingly lead to a more accurate diagnosis. This could be achieved through the integration of an image-processing component with a diagnostic problem solver. Image processing is a powerful tool that has been applied in many domains such as intelligent remote sensing via satellite, medical image analysis, radar, sonar, robotics, and automated inspection. Image information can play a crucial role, in the diagnosis of different diseases in the agricultural domain where the understanding of image symptoms is often essential to problem solving. Consider for example the information stored in the following fungal images:

The expert system can reach a correct and accurate diagnosis through extracting symptoms from those defected images (e.g.: yellow spots and angular shape in Figuer1), (white spots and circular shape in Figure2), and (white spots and linear shape in Figure3), apply the reasoning process while taking into account the extracted symptoms.


Fig. 1


Fig. 2


Fig. 3

This work aims to facilitate the process of integrating expert system with image processing. The diagnosis of cucumber diseases is used as an example. The motivation for automating image feature extraction in a diagnostic expert system can be summarized as follows:

- -The expert interpretation/analysis of defected image content is subjective,
- -The normal behavior of a real expert is to detect the symptoms from an image before employing his expertise. So automating symptom detection via feature extraction simulates real experts.
- -Such integration alleviates the need for relying heavily on user inputs.
- -Such integration assures that the final decision of the expert system is accurate, because an accurate decision requires accurate inputs.

This paper is divided into four sections in addition to its introduction section. Second section presents CLAES diagnostic model, Third section describes the architecture of the proposed diagnostic model. Fourth section present experiments and discussions. Fifth section displays conclusions.

2 The CLAES Diagnostic model [3]

This section presents the CLAES diagnostic model. The description of this model is based on the notation provided in CommonKADS methodology. The inference knowledge of this model is depicted in Figure 4. An inference structure diagram is used to describe this type of knowledge. Three types of components are used in this diagram, namely: inference steps, dynamic roles, and static roles. An inference step (represented by an ellipse) is a declarative definition of the directional relationship between the input and output roles. A dynamic role (represented by a rectangle) refers to the data used as input or output of an inference step. A static role (represented by a bold-line rectangle) refers to the domain knowledge (domain model) on which an inference step operates. Figure 4 represents the inference structure, which consists of three-inference steps, namely: expand, generate hypothesis, and differentiate. The goal of the expand inference step is to derive parameters used in the system according to the available data involved in the case description role. To perform its function, it uses the expansion model, which contains knowledge that derives and/or abstracts new parameters based on the available known parameters.

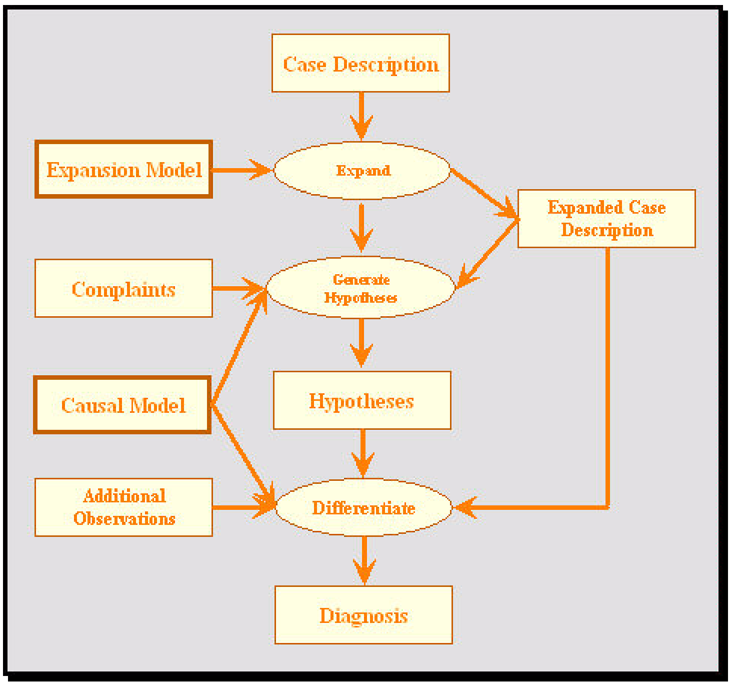

Fig. 4: Diagnosis Inference Structure

The output of this inference step is new case description that has been expanded. The objective of the generate hypotheses inference step is to use the set of observations, involved in the case description role to generate a set of suspected disorders (hypotheses). The step uses the knowledge included in the causal model, which contains a collection of causality relationships between the observations and the disorders to generate the hypotheses. The function of the differentiate inference step is to confirm/disconfirm each hypothesis within the generated hypotheses set based on the acquisition of some additional observations related to the current hypotheses set producing a set of confirmed disorders.

2.1 Cucumber Diagnostic Expert System

Cucumber diagnostic expert system is one of the five expert systems available in CUPTEX [22]. CUPTEX is the crop management expert systems, which is developed at CLAES to manage cucumber crop. The diagnosis expert system finds out the causes (Diseases) of the user complain (Observation). The system interacts with the end user the through textual dialogue to get the user complaint and to ask additional observations that is required to confirm the user assumptions.

The diseases can be:

- Infectious diseases cased by Fungi, Bacteria, Viruses, etc.
- Non-infectious diseases or disorder caused by mineral toxicities, soil acidity, nutrient deficiencies, or environmental factors.

Disorder identification is obtained by considering symptoms observed by eye, which denotes a malfunction in the physiology of plant.

One or more possible diseases may cause the presence of these symptoms. So the system continues to ask the user additional symptom that may be existed on the plant or in the environment. The diagnosis system uses these additional observations to reject some of the possible diseases and proceed with the selection of the most probable ones. Figure 5 shows the hierarchical disorders and observation, which is included in CUPTEX knowledge base.

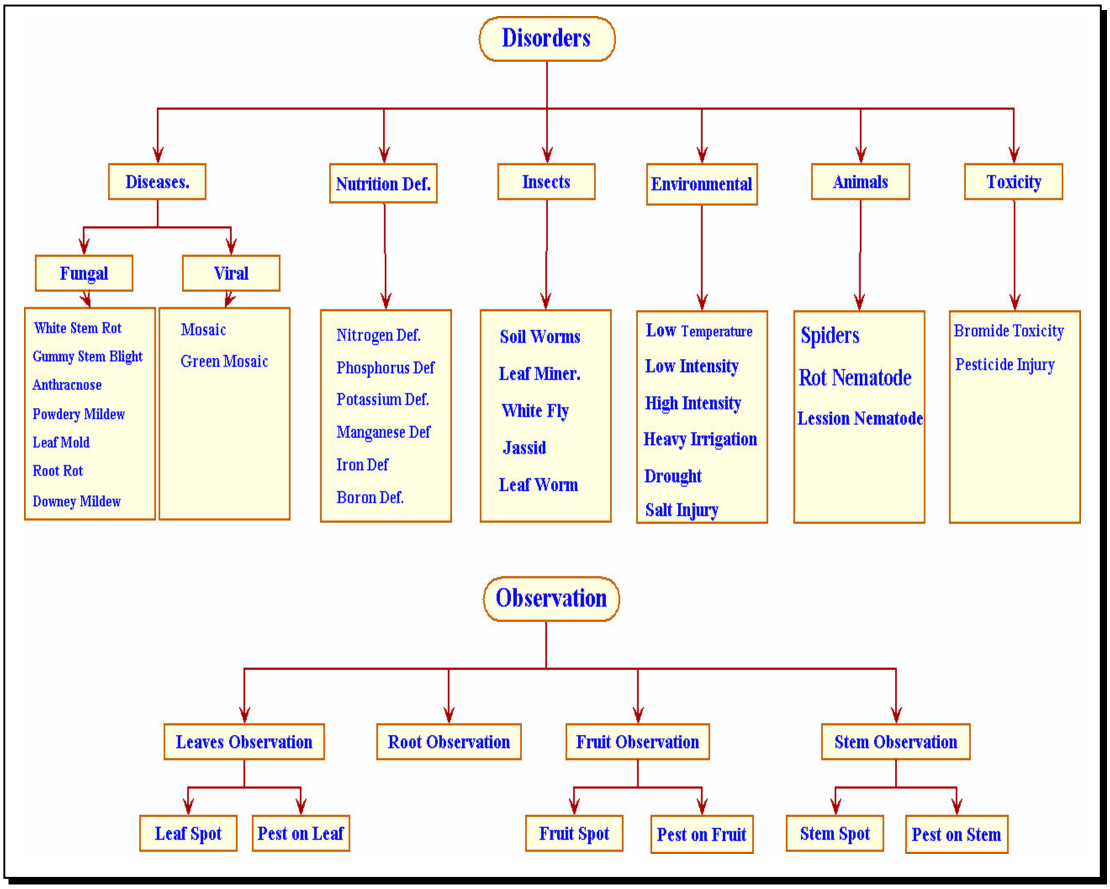

Fig. 5: Disorders and Observations Hierarchy

3 The Proposed Diagnostic model

Figure 6 presents the architecture of the new proposed diagnostic model integrated with the image analyzer.

In this model, the defected image of the defected plant is used as an input to the model. The image analyzer component detect the abnormal symptom in the defected image then extract their features, and classify those features to specific

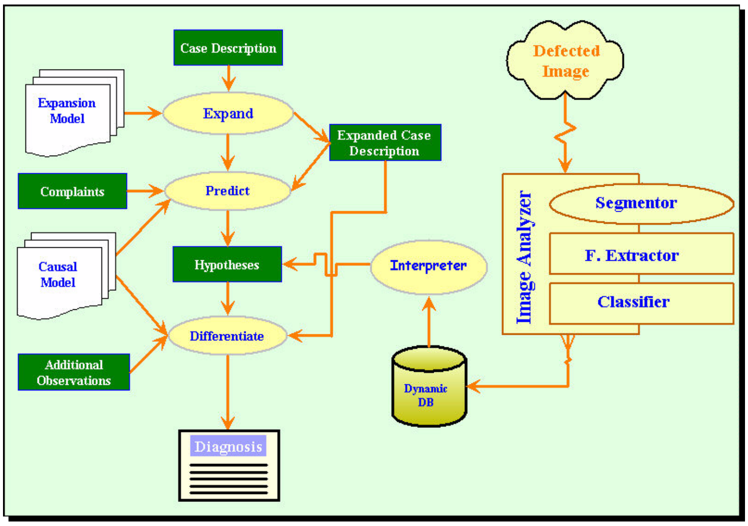

Fig. 6: Proposed Diagnostic model

class(s). Those classes are stored in a dynamic database to be interpreted by the interpreter. The interpreter converts each record in the dynamic database into a disorder(s) name and sets this disorder name into working memory as a hypothesized disorder. Figure 7 presents the algorithm of our interpreter.

```
Function DisorderInterpreter ()
Begin
      While Not End of Dynamic DB Do
      Begin
      ClassVal = Get First Record ()
      If(ClassVal <0 OR ClassVal >4) Then
                  DisName = "Unknown"
                              ElseIf(ClassVal >=0 & ClassVal <= 1) Then
                              DisName = "Downy Mildew"
                              ElseIf (ClassVal >1 AND ClassVal <= 2) Then
                                        DisName = "LeafMiner"
                                        ElseIf (ClassVal >2 And ClassVal <= 3)  Then
                                              DisName = "Powdery Mildew"
                                              ElseIf (ClassVal >3 And ClassVal <= 4) Then
                                              DisName = "Normal"
                                              Else
                                              DisName = "Unknown"
      End If
      If(DisName <> "Unknown" OR DisName <> "Normal") Then
            SetInWM(DisName)
      End If
      ClassVal = GetNextRecord ()
      End
End
```

Fig. 7: Pseudo code of the Interpreter

Our model provides two different paths to generate the hypothesized disorder(s). The first one is based on textual interaction with the user through predict inference step which uses the knowledge base in causal model. The second one is based on analyzing the defected image through the image analyzer component, which is more accurate than the textual input.

Differentiate inference step uses also causal model and additional textual input through additional observation input role in addition to another input role which is in expanded case description to differentiate between the hypothesized disorder(s). In the next section we will discuss the image analyzer components.

3-The Architecture of the Image Analyzer

The image analyzer was developed in order to automate the process of determining cause(s) of the abnormal symptoms. An image-processing component is employed to identify and classify leaf batches into a hypothesized disorder. The proposed architecture of the image analyzer is depicted in Figure 8.

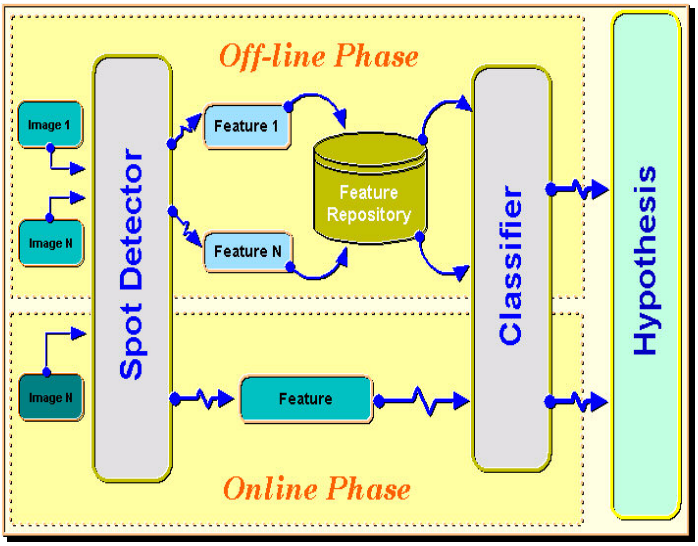

Fig. 8: Image Analyzer Components

This system consists of three main components: a spot detector, a feature repository, and a classifier. The processing done using those components is divided into two phases. The first processing phase is the offline training phase in which a large set of defected cases caused by disorders which are known before hand are processed by a spot detector for extracting spot features. Those features are then stored in a feature repository for later usage by the classifier. The second processing phase is an online phase by which the abnormal features of any given case are extracted by the spot detector and then classified resulting in a specific hypothesis. In the following section we will describe each of the image analyzer's components in more details.

4-The Spot Detector

The main purpose of the spot detector is to extract spot features from a defected color image represented by spot color and spot shape. As depicted in Figure 9, the input of the spot detector is the acquired defected color image and the output of the spot detector is the extracted features of the defected image.

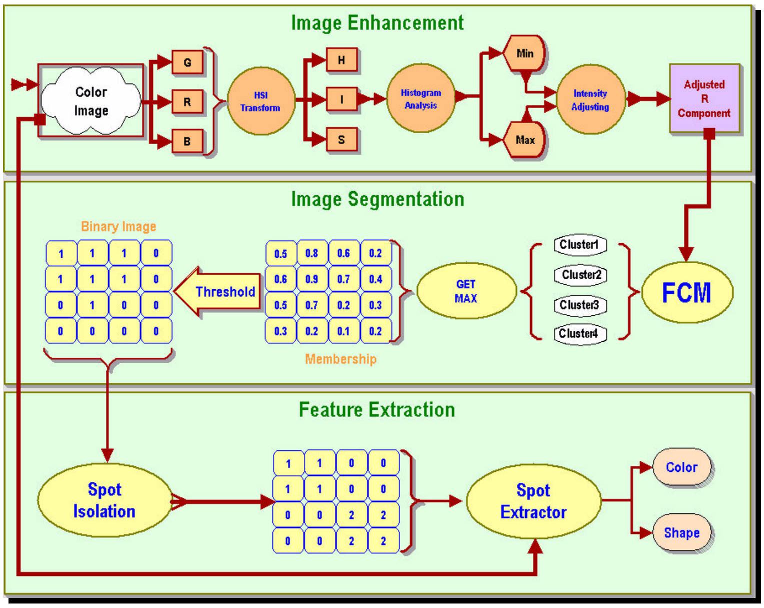

Fig.9: Spot Detector Phases

There are number of phases involved in the process of spot detection. These are: The image acquisition phase, the enhancement phase, the segmentation phase, and the feature extraction phase.

4.1The Image Acquisition Phase

In this phase, images were captured using a high-resolution 3-CCD color camera (DSC-P1 Cyber-shot, Sony) with 3.3 million-pixel sensitivity, 3X optical zoom lens, auto focus illuminator light, and Focal Length 8 - 24mm. The camera was placed at about 60mm from top of the leaves. The image from the camera was digitized into a 24-bit image with resolution 720 x 540 pixels. The storage format of the images is the bitmap format. The used data set consisted of three categories of spotted images, based on three disorders, which are: powdery mildew, leafminer, and downy mildew. This data set was taken in the cucumber green house at CLAES. In addition some images are taken from literature for testing purpose

4.2 The Enhancement Phase

From the inspection of the infected leaves, it was found that spots have a higher intensity than other normal tissue. So to extract abnormal spot tissue features easily, we concentrated on enhancement preprocessing to clarify those spots so as to facilitate segmentation processing. Figure 9 describes the three steps of the enhancement phase. The first step is the transformation of the defected image into HSI color space. The second step is to analyze the histogram of the intenisty channel so as to get the threshold by which we can increase the contrast of the image, and the final step is to adjust the intensity of the image by applying the thresholds.

4.3The Segmentation Phase

Image segmentation is the first step in image analysis and pattern recognition. It is a critical and essential step and is one of the most difficult tasks in image processing, as it determines the quality of the final result of analysis. The problem of segmentation has been broadly investigated by scientists using both classical [4,5] and fuzzy based techniques [6-9]. Classical segmentation approaches take crisp decisions about the regions. However, regions in an image are not always crisply defined, and uncertainty can arise within each level of image processing as in our addreses. Most plant images are represented by overlapping gray-scale intensities for different tissues. In addition, borders between tissues are not clearly defined and memberships in the boundary regions are intrinsically fuzzy. Fuzzy set theory provides a mechanism to represent and manipulate uncertainty and ambiguity. Therefore fuzzy clustering turns out to be particularly suitable for the segmentation of plant images. One widely used algorithm is the fuzzy c-means (FCM) algorithm which was first presented by Dunn [10], further developed by Bezdek [11], and subsequently revised by Rouben[12], Gu [13], and Xie [14]. However, Bezdek's FCM remains the most commonly used algorithm.

The segmentation of defected plant images involves partitioning the image space into different cluster regions with similar intensity image values. The success of applying FCM to fit the segmentation problem depends mainly on adapting the input parameter values [15,16]. As a consequence, if any of the parameter is assigned an improper value, the clustering results in a partitioning scheme that is not optimal for the specific data set and that leads to a wrong decision. These parameters include, the feature of the data set, the optimal number of clusters, and the degree of fuzziness. Based on experiments with these parameters, we've shown that the optimal cluster number for leaf spots is 4, and the degree of fuzziness is 2 [17]. We've applied those parameters to our data set and the results are presented in Figure 10,11,12,13 and 14. In those Figures, the top row represents the original defected input images, while the bottom row represents abnormalities, detected via segmentation.

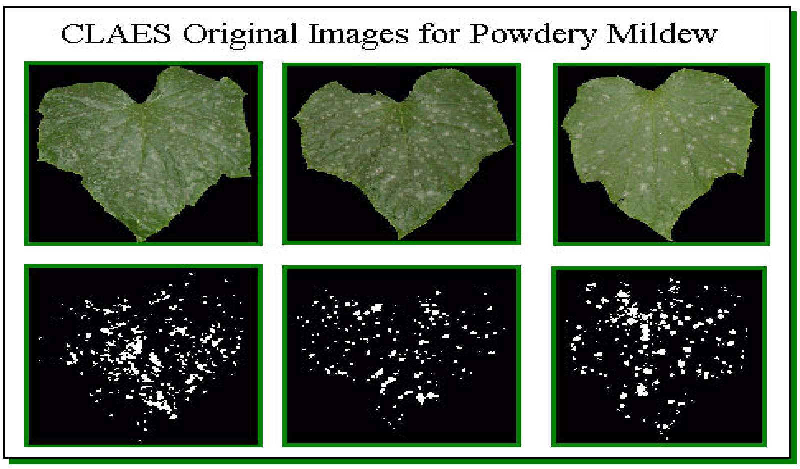

Fig. 10: Original and Segmented Images (Powdery Mildew)

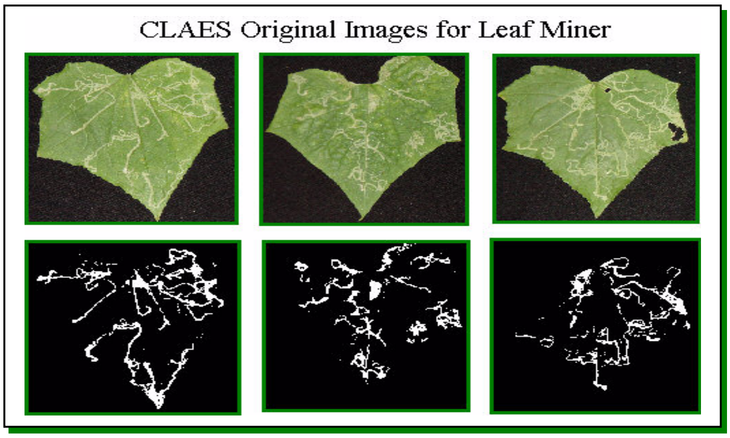

Fig. 11: Original and Segmented Images (LeafMiner)

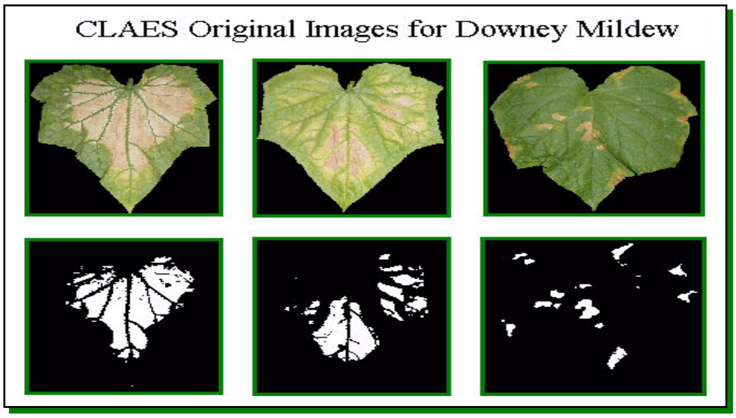

Fig. 12: Original and Segmented Images (Downey Mildew)

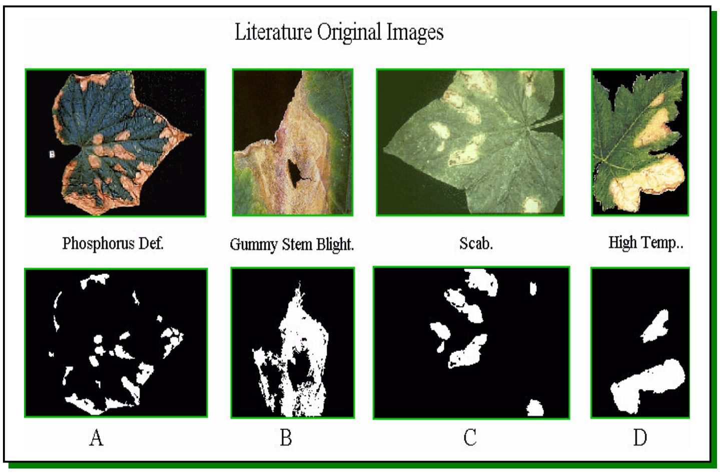

Fig.13: Segmentation results on some literature images

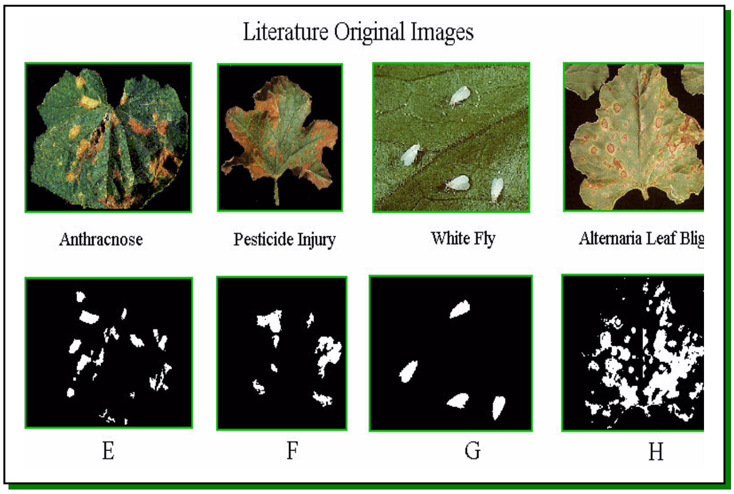

Fig.14: Segmentation results on some literature images

4.3 The Feature Extraction phase

The third phase is the feature extraction phase. The purpose of feature extraction is to reduce image data by measuring certain features or properties of each segmented region such as color, shape, or texture. This phase consists of two steps, mainly spot isolation, and spot extraction. In the spot isolation step we label each spots with a unique integer number where the largest integer label denotes the number of spot in the image. In the spot extraction step we measure several numbers of features from the segmented image to be later used for classification purposes. These features correspond to color characteristics of the spots such as the mean of the gray level of the red, green, and blue channel of the spots. Other features correspond to morphological characteristics [18] of the spots such as: the length of the principal axes, the diameter, eccentricity, compactness, extent, euler's number, and the orientation of the spots.

5 The Features Database

The features database is the component used to store the outputs of the feature extraction phase for later usage by the classifier. The database is a relational one, which consists of two tables namely a disorder table, which is used to keep track of disorders that have been processed and a feature table, which is used to store the spot features for each disorder. The created database contains 1500 records with 300 records per each class.

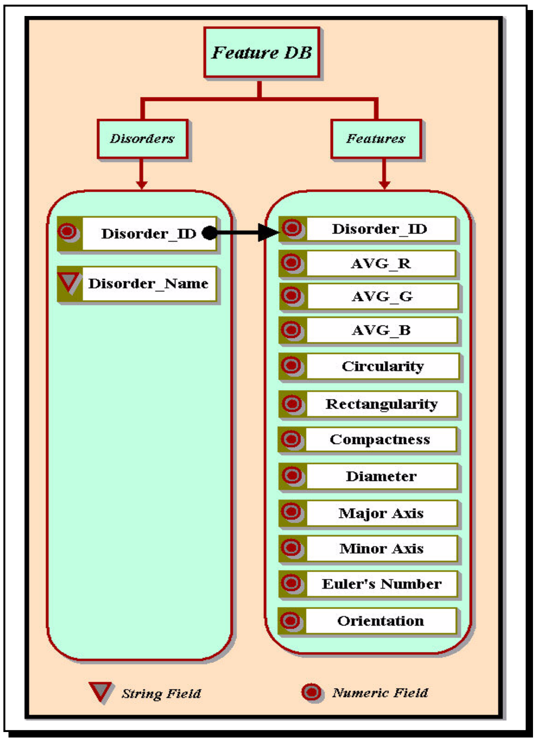

Fig. 15: Feature Database

6 The Classifier

Before on-line processing is done, the system needs to be manually trained using a set of training examples. An artificial neural network ANN was used to perform our classification task. There are many different types of ANNs. The most widely used is the Back Propagation ANN. This type of ANN is excellent for performing classification tasks [19,20]. We have used a feed-forward ANN with two hidden layers, five neurons per each hidden layer, and standard back propagation as the training algorithm. The following table summarizes the results of the training phase [21].

Class Name	Training Sample	Testing Sample	Classifier Accuracy
Downy	250	50	86%
Leafminer	250	50	74%
Powdery	250	50	94%
Normal	250	50	98%
Negative	250	50	92%

Table 1: Classifier accuracy for five classes

As shown in Table 1, the capability of the classifier to identify normal leaves is 98%, which is the highest recognition percentage. This can be attributed to the fact that normal leaves exhibit no abnormal features, which makes it easy for the classifier to identify them. The third highest recognition percentage was that for unknown diseases. This high recognition percentage is due to the fact that the classifier has been trained extensively to recognize 3 specific diseases. As a result the classifier is capable of identifying the features that best point to them as well as features that do not indicate their presence. The second highest identification percentage was that for the powdery mildew disease. This can be explained by the ease by which features related to that disease can be detected. The fourth, and fifth percentages are those for downy mildew and leafminer diseases. These percentages are acceptable for those diseases because there was an overlap in the appearance of some symptoms related to those disorders.

7 Implemented System

Our system was built using MATLAB 6.5, Visual Basic 6, and the knowledge base was built using KSR shell. The development process was done through three phases. In the first phase, the MATLAB was used to build the core of the image processing. In the second phase, the core that was developed in the first phase was transferred into a Component Object Model (COM) using a COM builder tool, which is delivered with MATLAB. Component Object Model is a technology by which the component can communicate with the outer-world. Such component can run in the same process, in different processes on the same machine, or even on different machines. The user interface was developed using Visual Basic in the third phase, this user interface uses the Image Analyzer component and knowledge base component.

The user interface is depicted in Figure 16. It consists of three list boxes for navigating symptoms, which is *Initial Symptom* List box, *Properties* List box, and *Values* List box. The other List box is for displaying suspected and confirmed disorders as well as some controls to browse the defected image. The working memory is displayed as a grid, which contains the user input symptoms. In addition to some push buttons to provide diagnosis functionality.

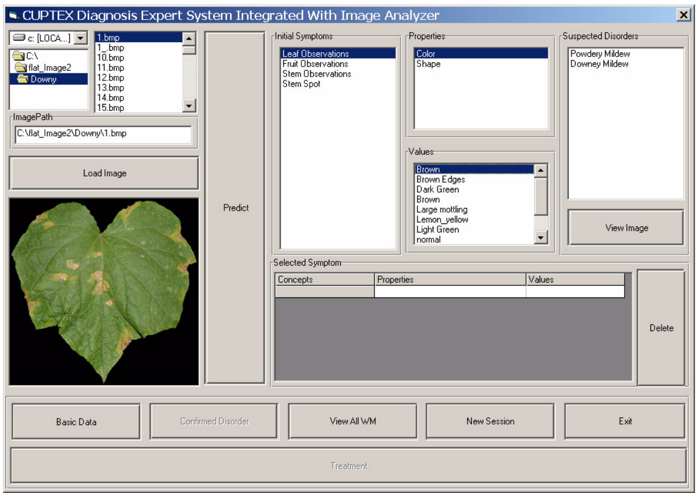

Fig. 16: Interface for Diagnostic Expert System

The model provides different two paths to generate the hypothesized disorder(s). The first one is based on textual interaction with the user through predict inference step which uses the knowledge base in causal model. The second one is based on analyzing the defected image through the image analyzer component, which is more accurate than the textual input.
Differentiate inference step uses also causal model and additional textual input through additional observation input role in addition to another input role which is in expanded case description to differentiate between the hypothesized disorder(s). as presented in Figure14.

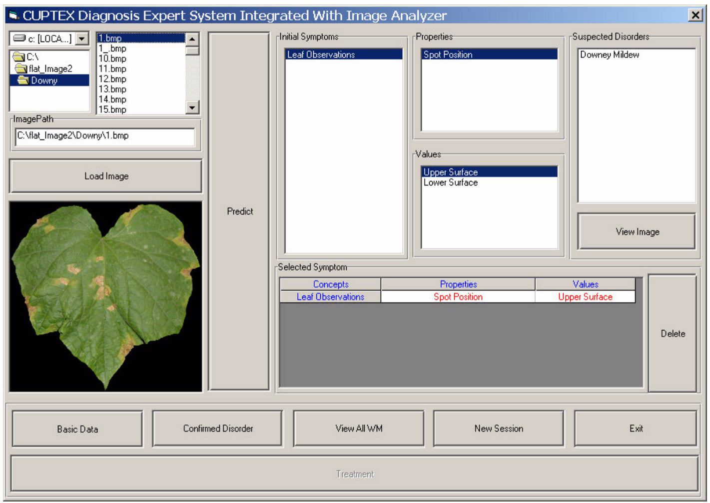

Fig. 14: Confirmation Phase of the Diagnostic Expert System

8 Conclusions

This paper has demonstrated the usefulness of integrating an image analyzer within a diagnostic expert system model through a real life example.

In order to diagnose a disorder from a leaf image, four image-processing phases were applied: enhancement, segmentation, feature extraction, and classification.

In order to employ this system we first have to train it using a set of images of disorders. We have tested the system on 3 cucumber disorders. The results of this test indicated that this system could indeed identify disorders with a high level of accuracy.

Applying this model to any other crop disorders requires only special care to be taken in order to acquire a sufficient set of images representative of these disorders for use in the training step.

Integrating this model within a diagnostic expert system then will greatly reduce any error prone dialogue between the system and the user while resulting in increased accuracy in the system's diagnosis.

9 Future Work

The work presented in this paper opens up lots of avenues for future research. The following are some of the points that can investigate.

Currently image processing system focus on three disorders identification, it would be helpful to extend the system in order to include other disorders.

Also it would be helpful to extend the system to be able to detect and identify abnormalities on the other parts of the plant such as fruits, stem, and root.

The system could be extended in order to cover all crops.

Building an expert system robotic capable to see abnormalities of a plant and understand it and give a treatment operation directly.

10 Acronyms & abbreviations

CLAES	Central Lab. for Agricultural Expert Systems
ES	Expert Systems
KBS	Knowledge Base System
CUPTEX	Cucumber Expert Systems
FCM	Fuzzy C-Mean
COM	Component Object Model

Reference

[1] Liebowitz, Rada, R., "*The Handbook of Applied Expert Systems*", Chapter 6, CRC Press LLC, 1998.

[2] Liebowitz, Rafea A., "*The Handbook of Applied Expert Sys*tems", Chapter 35, CRC Press LLC, 1998.

[3] El-Sayed E., Hesham A., Rafea A., London, "*Pest Control Expert System for Tomato (PCEST)*", Knowledge and Information Systems, Vol. 2, pp. 242-257, Spring-Verlag Ltd, 2000.

[4] Gonzaliz RC., Woods RE., "*Digital image processing*", Pearson Education Inc., 2002.

[5] Spirkovska. L., "*A Summary of Image Segmentation Techniques*", NASA Technical Memorandum 104022, June 1993.

[6] Bezdek J.C., R. Ehrlich, W. Full., "*FCM: The Fuzzy c-mean Clustering Algorithm*", Computers and Geosciences, Vol. 10, pp.191-203, 1984.

[7] Chi Z., Yan H., Pahm T., "*Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition*", World Scientific, 1996.

[8] Dzung L., Chenyang X., Jerry L., "*A Survey of Current Methods In Medical Image Segmentation*", Technical Report JHU/ECE 99-01, The Johns Hopkins University, 2001.

[9] S. K. Pal., "*Image Segmentation Using Fuzzy Correlation*", Information Sciences, Vol. 62, pp. 223-250, 1992.

[10] Dunn, J. C., "*Well-separated clusters and the optimal fuzzy partitions*", Journal of Cybernetics, Vol. 4, pp. 95-104, 1974.

[11] Bezdek, J. C., New York, "*Pattern Recognition with Fuzzy Objective Function Algorithms*", Plenum Press, 1981.

[12] Rouben, M., "*fuzzy clustering algorithms and their clustering validity*", European Journal of Operational Research, Vol. 10, pp. 294-301, 1982.

[13] Gu. T., Dubuission B., "*Similarity of classes and fuzzy clustering*", Fuzzy Sets and Systems, Vol. 34, pp 213-221, 1990.

[14] X. L. Xie, G. Beni., "*A Validity Measure for Fuzzy Clustering*", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 13, No. 8, PP. 841-847, 1991.

[15] M. Halkidi, Y. Batistakis, M., "*Vazirgiannis., Cluster Validity Methods: Part I*", SIGMOD Record, June, 2002.

[16] M. Halkidi, Y. Batistakis, M., "*Vazirgiannis., Cluster Validity Methods: Part II*", SIGMOD Record, September 2002.

[17] El-Helly M., Onsi H., Rafea A., El-Gamal S., "*Segmentation Technique for Detecting Leaf Spots in Cucumber Crop Using Fuzzy Clustering Algorithm*", 11th International Conference on AI (ICAIA) , 2003.

[18] Sonka M., Hlavac V., BoyLE R., USA, "*Image Processing Analysis and Machine Vision*", Brooks/Cole Publishing Company, 1999.

[19] Hecht-Nielsen, R., "*Neurocomputing*", Addison Wesley, 1990.

[20] Hertz, J., Krogh, A., and Palmer, R., Redwood City, California, "*Introduction to the Theory of Neural Computation",* Addison-Wesley, 1991.

[21] El-Helly M., Rafea A., El-Gammal S., "*An Integrated Image Processing System for Leaf Diseases Detection and Diagnosis*". 1st Indian International Conference on AI Dec-18-20 (IICAI-03), 2003.

[22] El-Dessouki, A., S. Edrees, and S. El-Azhari, Cairo-Egypt, "*CUPTEX: An integrated Expert System for Crop Management of Cucumber*". Proceedings of the Second National Expert Systems and Development Workshop (ESADW-93) MOALR, 1993.