Detection of diseases on tomato leaves based on Sub-Classifiers Fuzzy Combination

F.Jakjoud, A.Hatim, A.Bouaddi

***Abstract:** Improving agricultural productivity is the subject of several research field, and in order to prevent huge losses automatic identification of plant diseases seems a good solution. In this paper we present an approach based on the K-nearest neighbours (KNN) algorithm and Support vector machine (SVM) to classify tomato leaves images into two classes (normal and sick). We developed two schemes based on a combination of sub-classifications based on KNN and SVM combined with a fuzzy decision maker. The features extractor used in this work is the Haralick approach based on the co-occurrence Matrix. The result of each classifier can reach over than 80% and the best accuracy is given by the Fuzzy Combination of KNN Sub-classifier more than 98%.*

***Index Terms:** Plant disease, image classification, machine learning, SVM, KNN.*

I. INTRODUCTION

Crops monitoring is a significant issue for detecting plant disease, the naked eye observation of expert is the most popular approach in the past, with the technological development in image processing[18], one can use automatic monitoring for detecting plant disease. These techniques will help the farmers to enhance the quality and the quantity of agricultural productions [1]. Image processing can be used for multiple purpose in agricultural applications such as: detecting plant disease[17], finding the affected region using its color and/or the shape, sorting plants using either their color and their shape. In this paper, we are interested in the application of the image classification algorithm for the detection of plant disease and especially on Tomato leaves. Indeed in most of the cases diseases are seen on the leaves of the plant [2]. Machine Learning for image processing has shown excellent result in the detection and the recognition of plant disease and can successfully be applied as an efficacious recognition mechanism. Weizheng et al. [3] propose a method based on image processing for gradient of plant diseases, they use Otsu method [4] for segmentation and Sobel operator to detect the disease spot edges, for the classification they calculate the quotient of disease spot and leaf areas. A prediction approach was also presented [5] the approach is based on support vector machine SVM, which led to a better result comparing with several type of machine learning. Camargo et al. [6] have developed a machine learning for the identification of the visual symptoms of plant diseases from colored images using SVM. Rumpf et al. [7] propose a method for the diagnosis and classification of grapevine leaves using the K-nearest neighbours (KNN) method for segmentation on limited data set which had proved a good performance and accurate result.

Revised Manuscript Received on March 10, 2019.

Fatimazahra Jakjoud, Laboratory of Energy Engineering, Materials and Systems, National School of Applied Sciences, Ibn Zohr university, Agadir, Morocco.

Anas Hatim, National School of Applied Sciences, Cadi Ayad university, Marrakech, Morocco.

Abella Bouaddi, Laboratory of Energy Engineering, Materials and Systems, National School of Applied Sciences, Ibn Zohr university, Agadir, Morocco.

II. RELATED WORK

Revathi et al [19] propose a method to identify the affected part of the disease using Homogeneous Pixel Counting Technique for Cotton Diseases Detection (HPCCDD) algorithm, which analyse input image by the RGB pixel Counting values features to extract the features then for edge detection they used Homogenization techniques Sobel and Canny and for the classification they tested different methods the best accuracy was reached by the proposed HPCCDD (98.1%). [20] in this paper, after masking green pixels using specific threshold value the authors used Color Co-occurrence method for feature extraction, finally the extracted features are passed through the SVM Classifier. ArunPriya et al [21] propose an efficient machine learning approach for the plants diseases recognition using Digital Morphological feature and SVM algorithm for classification, Finally the accuracy of the proposed method is compared with the KNN classification approach, the accuracy obtained by the SVM in real dataset is 96.8% whereas the accuracy obtained by the KNN is 81.3%. In [22] a new approach was proposed to diagnosing plant disease using image processing and artificial intelligence techniques, they began by thresholding images to mask green pixels, then the segmentation step is done using K-means clustering and for the classification the best result was obtained by Feed forward back propagation neural network. For Paddy plant, Jagan Mohan et al [23] propose an approach to identify and classify the diseases affecting these plants by the use of Scale Invariant Feature Transform (SIFT) feature, KNN classifier and SVM, the accuracy rate is 91.1% with SVM and 93.33% using KNN. A.Gulhane [24], has proposed a method to find the sicknesses on cotton leaf using Principal Component Analysis PCA and KNN, this approach has reached an accuracy of 95%

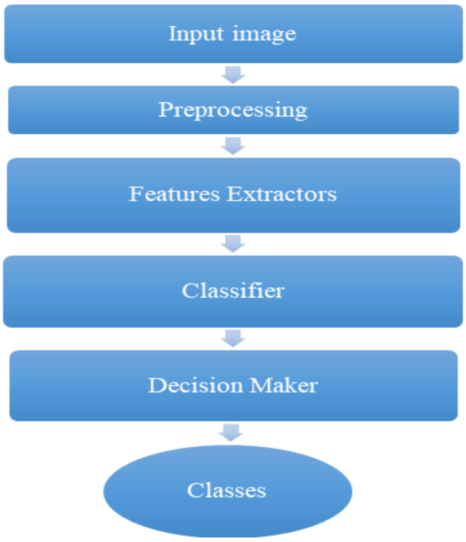

Fig.1.The proposed scheme

III. PROPOSED METHODS

We proposed, here, a scheme based on two methods of classification and diseases detection on tomato leaves. The general structure of the classification system consists of three steps. It consists on the Pre-processing, Features extraction, Classification stage and Decision maker. The decision maker added is a fuzzy inference system. The most important part during the development of a machine learning is the training of the data to generate the classification model, in this work we collect the image-dataset from the internet and others are taken by mobile phones, the size of the dataset is 200 images of which 80% are used for training and 20% for the test and validation of the model.

A. Pre-processing Stage

The images used are taken in nature so the presence of different objects in the background is intolerable in this work, it is for this reason that we used a color filter to extract the leaves from their background. The first step is to apply a color thresholding on the images to convert them from the RGB color space to the HSV [16], the second step, consists on eliminating all the pixels that are not connected to each other. Finally, we apply a color filter [12] with an interval corresponding to the green shade on the Tomato leaves, any pixel value that does not belong to this threshold would be replaced by a white pixel (Fig2).

Fig.2.Pre-processing Result

B. Features extraction

We choosed Co-occurrence matrix for the features extraction which is a statistical method applied to a NxN size matrix , where N is the number of levels of each color component. To evaluate all color components we had calculate this matrix for each one (gray scale, R,G,B,H,S,V).

From the Co-occurrence matrix, Haralick et al. [15] have extracted 14 features: Contrast, Angular Second Moment, Correlation, Variance, Inverse Difference moment, Sum Average, Sum variance, Sum entropy, Entropy, Difference variance, Difference entropy, information Measures of correlation1, Information measures of Correlation 2 and maximal correlation coefficient.

Main equations to calculating Haralick features:

- Entropy:

$$\mathrm{ENT} = -\sum_i \sum_j p(i,j) \log p(i,j)$$

- Contrast

$$\mathrm{CON} = \sum_i \sum_j (i-j)^2 \, p(i,j)$$

- Energy (Angular Second Moment)

$$\mathrm{ASM} = \sum_i \sum_j \{p(i,j)\}^2$$

- Inverse Difference Moment:

$$\mathrm{IDM} = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$$

C. Machine Learning models

Machine learning methods are divided into two categories, supervised and unsupervised. The supervised learning is based on the learning by result, while the unsupervised is based on the learning by example [5]. In this work we have elaborated two classification methods based on the following algorithms: SVM, KNN. To test the performance of each of them. The classification database consists of the 14 Haralick parameters extracted from the Co-occurrence matrix for each color level (Gray-scale R,G,B,H,S,V), each image is represented by 14 parameters.

1) K-nearest neighbours

The KNN algorithm [7] is a very efficient approach and does not need too much time for learning, indeed it relies on the distance between features. KNN is used to classify unlabelled features by assigning them to the class of the most similar labelled features by comparing them to all the stored data to choose the majority class among the K nearest neighbours based on the Euclidian distance in our case.To use this method, we have two parameters to be tuned, the value of K and the distance metric. The Euclidean distance metric (1) is the most popular used to define the closeness between the input data in KNN classifier

$$d = \sqrt{\sum_{i=1}^{N}\left(x_{input} - P_i\right)^2} \quad (1)$$

With: N size of data base

x_{input} the feature of classification

P_i Features stored in the classifier

In this work, we have used the Euclidian distance since we use statistical parameters, and for the value of K we have run the classification process for different value from 1 to 10 and finally we chose the optimal value of K

KNN algorithm:

- Training data $[P_i, y_i]$:
  - Pi: Feature of the image I
  - Yi: Class corresponding to the image
  - X: Parameters to classify
- Algorithm:
  - Compute distance d(x, Pi) and store the result in the database
  - Select K closet data points to be analysed to determine which class label is most common among the selection
  - Output the most common class to be assigned to the input data point

Even if the KNN is an efficient classifier but it does not reacha high accuracy when used for each Haralick features We hence propose to combine many KNN classifiers based on the most significant Haralick features (comparing the accuracy of the 14 Haralick features) and for the final verdict we used a fuzzy decision maker based on Sugeno fuzzy inference system that generates the final decision from the results of each classification stage.

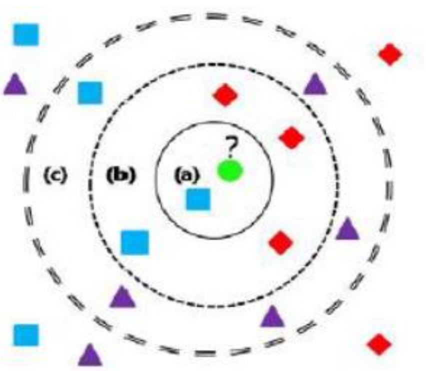

Fig 3. KNN Classification principle

2) Support Vector Machine

Support Vector machine is a supervised Learning technique that analyse and classify data [13], by representing each data point in n-dimensional space then we preform classification by finding the Hyper-plane that maximizes the margin between different classes.

The aim of SVM Classification is to produce a model, based on the training data in order to predict class of the test data, training SVM classifier is the problem of finding the optimal hyperplane that gives the largest minimum distance to the training data points that turns SVM to an optimization problem.

To find the optimal Hyperplane [13] we use the function (2)

$$min_{\omega,b,\xi}\ {}^{1}/_{2}\, w^T w + C \sum^{i=1} \zeta i \qquad (2)$$

Initially, the hyperplane can be a linear (3) classifier, but we can apply several kernel to create nonlinear classifiers such as, Gaussian radial basis Function (RBF)(4), polynomial kernel.

$$K(x_i, x_i) = x_i^T x_j \ (3)$$

$$K(x_i, x_i) = \exp\left(-\gamma \|x_i - x_i\|^2\right) \quad \gamma > 0 \quad (4)$$

As a result of the optimization dilemma, we will obtain two parameters: C and γ, C parameter represents the trades off between misclassification of training example and simplicity of the decision surface and γ is a kernel parameter.

SVM performs well on datasets that have many attributes [14] however, this kind of classifiers include limitations in speed and size during the training and the testing stage.

Following the same process as the KNN to improve the classification efficiency we have combining different sub-SVM classifiers and used a fuzzy inference system to generate the final decision.

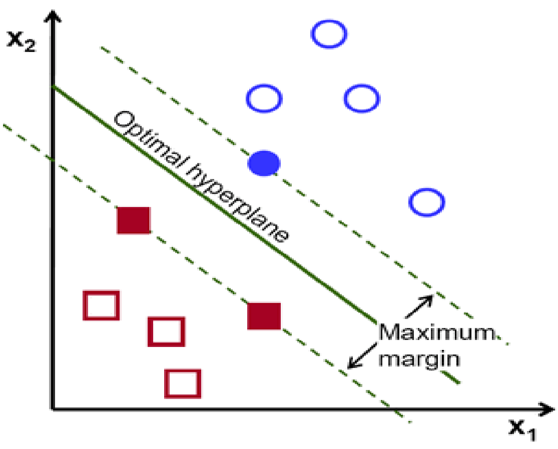

Fig 4. The Optimal Hyperplane

IV. Results and Discussion

The KNN and SVM learning algorithms are combined with a fuzzy decision maker. For each one we developed the same structure in Figure 1, the input database contains Haralick features classed into two classes (Normal and Sick leaves).

1) KNN Combined with fuzzy decision maker classification results

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

- **Global KNN**

We have developed a KNN classifier with the 14 Haralick features as inputs and used them as a global representation of images.
We tested different values of K from 0 to 10, the values that gave the best results are presented in Table1.

Table 1. KNN Results

K value	Training Accuracy	Validation Accuracy
1	100%	92.11%
2	83.89%	89.47%
3	83.89%	86.84%
4	83.89%	92.10%
5	83.90%	86.75%
6	81.88%	86.85%
7	80.54%	89.47%
8	89.47%	80.54%
9	81.88%	92.12%
10	79.87%	89.47%

We can notice that for K=1, the classifiers gave an accuracy of 100% during the training phase, and 92.11% in the validation phase. For other values, we notice a huge difference between the validation accuracy and the training accuracy.

- **KNN Combined with fuzzy decision maker**

In this approach we tested each Haralick feature by the KNN classifiers, the features that gave the best results are presented in Table 2

Table 2. Sub-classifiers Results

Feature	K value	Training Accuracy	Validation Accuracy
Energy (grayscale)	4	80%	72.41%
Entropy (Hue)	10	81.12%	70.96%

To improve the accuracy of each sub classifier, we propose to combine the two sub classifiers (Table 2) and we developed a fuzzy inference system (Sugeno) to generate the final decision from the results of each classification stage. The fuzzy inference system is used to generate fuzzy rules from a given input-output dataset, these rules are represented in the form of IF-THEN, where the output is in the form of a constant or linear equations (5). When the function f(x,y) is a constant (in our case), we have a Zero-order Sugeno Fuzzy model. The decision is divided into two states (Normal and Sick) that is why we used trapezoidal membership functions.

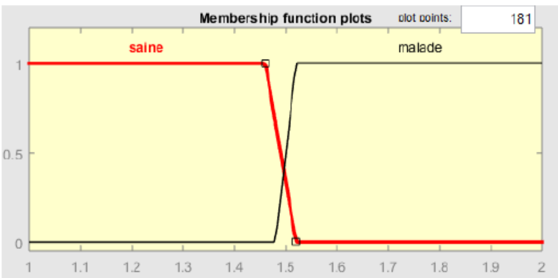

Fig 5. Shape of membership Function

$$if\ x\ is\ Normal\ and\ y\ is\ Sick\ then\ z = f(x,y) \quad (5)$$

The accuracy of the proposed scheme has reached 98.38% in validation step, and 99% during the training phase.

2) SVM Combined with fuzzy decision maker classification results

- **Global SVM**

The SVM Classifier that we have developed gave the best result when using RBF kernel, we trained the classifier on 1500 iterations and using the 14 Features as a global representation of image. The overall accuracy rate was 85.03% during the training stage and 71.87% for the validation set.

Table 3. Global SVM result

Classifier	Training Accuracy	Validation Accuracy
Global SVM	85.03%	71.87%

- *Fuzzy Combination of SVM Sub-classifiers*

We tested SVM classification on each features to evaluate the performance of each one, the best results are given by Correlation, entropy, homogeneity and Difference entropy (Table4).

Table 4. SVM Sub-classifier results

Feature	Training Accuracy	Validation Accuracy
Correlation (gray scale)	71%	70%
Entropy(Green)	69%	71%
Homogeniety(Green)	74%	78%
Difference entropy(green)	71%	75%

Next step is the combination between these sub classifiers, and for generate the final decision we have used the same fuzzy inference system. The accuracy of the proposed approach has reached 82.14% during the training phase and 82.01% on the validation set.
We Summarize the major results in Table 5:

Table 5. Result Synthesis

Classifiers	Results on validation set
Global KNN	92.11%
Global SVM	71.78%
FC-KNN sub-classifier	98.38%
FC-SVM sub-classifier	82.01%

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 6. Classification results comparison of state of art's method and the proposed approaches.

Method of classification	Results
HPCCDD [19]	98.10%
SVM with Morphological feature [21]	96.80%
KNN with Morphological feature [21]	81.30%
SIFT detector with KNN [23]	93.33%
SIFT detector with SVM [23]	91.10%
PCA with KNN [24]	95.00%
Global KNN with Haralick Features	92.11%
Global SVM with Haralick Features	71.78%
FC-KNN subclassifier with Haralick Features	98.38%
FC-SVM subclassifier with Haralick Features	82.01%

Table.6 shows that the FC-KNN sub-classifier gave the best result compared to the different method presented in the state of the art and to the FC-SVM sub-classifier. KNN classifier still the optimal method for the classification of statistical data as the Haralik feature [15] and PCA descriptors [24].

V. CONCLUSION

The application of machine learning in the field of plant disease detection is very important and considered as an optimal solution to increase the quality of cultivation. In this work we presented two different approaches of classification: Global classification, and Fuzzy Combination Sub-classification, based on KNN and SVM. The aim of classification task is to build a system who has the ability to detect anomalies on the Tomato leaf using the Co-occurrence matrix for the feature extraction.Combining classifiers gave good result compared to global classification and the addition of fuzzy inference system as a decision maker has increase the efficiency of sub-classifiers. The best result was obtained by FC-KNN sub-classifier with 98.38% accuracy on the validation set. The advantage of KNN is the storage of data without going through a training step which requires a lot of time, also we do not have many parameters to tuning in contrary of SVM. The disadvantage of SVM is the size and the type of data, with the statistical features, the dependence between parameters makes hyperplane tuning very difficult.

REFERENCES

1. Al bashish.d, Braik.m, bani-ahmad.s,2011,detection and classification of leaf diseases using k-means-based Segmentation and Neural networks based classification, Information Technology Journal, ISSN1812-5683/DOI:10.3923/itj.2011.267.275
2. Jayamala K. Patil, Raj Kumar, 2011, Advances in image processing for detection of plant diseases, Journal of advanced Bioinformatics Applications and Research ISSN 0976-2604 Vol2, Issue 2, June-2011, pp 135-141
3. Weizheng.S, Yachun.W, Zhanliang.C,Hongda.W,2008, Grading Method of Leaf Spot Disease Based on Image Processing, International Conference on Computer Science and Software Engineering, DOI 10.1109/CSS
4. Otsu.N,1979, A Threshold Selection Method from Gray-Level Histograms, IEEE, Transactions on systems man and cybernetics, Vol. SMC-9,NO1
5. Kaundal.R, Kpoor S A, Raghava G, 2006, Machine learning techniques in disease forecasting: a case study on rice blast prediction, BMC Bioinformatics 2006, 7:485 doi:10.1186/1471-2105-7-485
6. Camargo A, Smith J.S, 2009, An image-processing based algorithm to automatically identify plant disease visual symptoms, biosystems engineering 102.
7. Dhiriti,Kaur.M,2012,K Nearest Neighbor Classification Approach for Face and Fingerprint at Feature Level Fusion, International Journal of Computer Applications (0975-8887) Volume 60 No.14, December 2012
8. H.Al-Hiary, S.Bani-ahmad, M.Reyalat, M.Braik& Z.AlRahamneh , 2011, Fast & accurate de-tection & classification of plant diseases, international journal of computer applications,0975-8887,volume17-no1,pp-31-38
9. Stephen Gang Wu,Forrest Sheng Bao, Eric You Xu, Yu - Xuan Wang Yi - Fan Chang 2007 A Leaf Recognition Algorithm for plant Classification Using Probabilistic Neural Network, IEEE 7th international Symposium on Signal Processing and information Technology.
10. Baldomero Manuel Sanchez Rangel, Marco Antonio Aceves Fernandez, Jesus Cardoso Murillo [2016] KNN-based image segmentation for grapevine potassium deficiency diagnosis, International Conference on Electronics, Communications and Computers.
11. Al Maghrabi, K. (2014). The pattern of skin diseases among patients attending primary health care centers in Jeddah city, Saudi Arabia, 2012. International Journal of Research in Applied, Natural and Social Sciences.
12. Ms. M.S Priya, G. M. Kadhar Nawaz,Matlab Based Feature Extraction and Clustring Images using K-Nearest Neighbour Algorithm, IJAICT Volume2, Issue 11 2016.
13. Gajanand Gupta, Algorithm for Image Processing Using Improved Median Filter and Comparison of Mean, Median and Improved Median Filter, International Journal of Soft Computing and Engineering, ISSN 2231-2307,Volume1, Issue-5, 2011
14. Jain S, 2013, A Machine Learning Approach: SVM for Image Classification in CBIR, International Journal of application or innovation in engineering and management, volume2.issue4, ISSN 2319-4847
15. Jinho K, Byung-soo K, Silvio Savarese, Comparing Image Classification Methods: K-Nearest-Neighbor and Support Vector Machines, Applied Mathematics in Electrical and Computer Engineering, ISBN: 978-1-61804-064-0
16. Robert M. Haralick, K.Shanmugam, Its'Hak Dinstein, Textural Features for Image Classi-fication,IEEE Transactions on systems, Man and Cybernetics, Vol. SMC-3 NO.6,1973
17. Ms. M.S Priya, G. M. Kadhar Nawaz,Matlab Based Feature Extraction and Clustring Im-ages using K-Nearest Neighbour Algorithm, IJAICT Volume2, Issue 11 2016.
18. Jayme Garcia ArnalBarbedo, Digital image processing techniques for detecting, quantify-ing and classifying plant diseases. Barbedo Springer Plus 2013,2:660
19. Mbark IGGANE,Driss Mammass, A.KHENCHAF, Segmentation d'images sonar par ma-trice de co-occurrence, SETIT2009, 5th International conference: Sciences of Electronics, Technologies of Information and Telecommunication.
20. Kumar, S., Walia, S. S., & Kalanidhi, A. Fuzzy model identification: a new parallel BB-BC optimization based approach. International Journal of Electronics and Communication Engineering (IJECE) ISSN (P), 2278-9901.
21. P.Revathi, M.Hemalatha, Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques" ISBN, pp 169-173,2012 IEEE.
22. S.Arivashagan, R.Newlin Shebiah, S.Ananthi, S.Vishnu Varthini, Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features, CIGR,vol. 15(1), pp.211-217, March 2013.
23. Arunpriya.C, Balasaravanan.T, Antony selvadoss.T, An efficient leaf recognition algorithm for plant classification using Support Vector Machine, Proceeding of the international Conference on Pattern Recognition, informatics and Medical Engineering, 978-1-4673-1039-0/12, March 21-23,2012
24. S.Sannakki, V.Rajpurohit, V.Nargund, P.Kulkarni, Diagnosis and classification of grape leaf diseaces using neural networks, 4th ICCCNT, July 2013
25. K.Mohan, M.Balasubramanian, S.Palanivel, Detection and recognition of diseases from Paddy plant leaf images, international Journal of Computer application,0975-8887,vol(144)-No.12,June 2016
26. Chakraborty, S., Bandopadhya, S., Debnath, A., & Dutta, S. Evaluation of Turmeric Germplams for Tolerance to Foliar Diseases in Terai Region of West Bengal.
27. A.Gulhane, H, Kolekar, Diagnosis of diseases on Cotton leaves using Principal Component Analysis Classifier, Annual IEEE india conference,2014.