Original Research Paper

Early Detection and Classification Approach for Plant Diseases based on MultiScale Image Decomposition

[1]Assia Ennouni, [1]Noura Ouled Sihamman, [1]My Abdelouahed Sabri and [2]Abdellah Aarab

[1]*Department of Computer Science, Faculty of Sciences Dhar el Mahraz, University Sidi Mohamed Ben Abdellah, Fez, Morocco*
[2]*Department of Physique, Faculty of Sciences Dhar el Mahraz, University Sidi Mohamed Ben Abdellah, Fez, Morocco*

Corresponding author:
Assia Ennouni
Department of Computer
Science, Faculty of Sciences
Dhar el Mahraz, University Sidi
Mohamed Ben Abdellah, Fez,
Morocco
Email ennouni.assia@gmail.com

My Abdelouahed Sabri
Department of Computer
Science, Faculty of Sciences
Dhar el Mahraz, University Sidi
Mohamed Ben Abdellah, Fez,
Morocco
Email abdelouahed.sabri@gmail.com

Abstract: This paper presents a new and powerful approach for detecting and classifying leaf diseases for plant diagnosis with high accuracy. The main contribution of this paper is that a hybrid approach is proposed by using the combination of Partial Differential Equations (PDE) based image decomposition, segmentation, feature extraction, features selection and classification aiming to improve the classification accuracy and provide an excellent diagnosis. The TV-L^1 Total variation model is adopted to separate the original image into texture and object components. Segmentation will be done only on the object component. Then texture, color, vein and shape features are extracted and merged in a feature vector using the codebook method. Moreover, features are selected by the RelieF feature selection algorithm to keep only relevant ones. In the classification, only selected features will be used and passed to the Multiclass Support Vector Machine algorithm SVM. The proposed approach is implemented and tested on the PV Plant Village dataset and provided a good and greater classification accuracy compared with the existing approaches from the literature. The obtained results proved that the use of PDE influences on the segmentation, which in turn, allowed us to identify correctly the leaves and provide new and optimal features, those features improves the classification accuracy rate to 95.9%.

Keywords: Smart Agriculture, Image Processing, Machine Learning, Plants Disease Imaging, Early Diseases Detection, PDE, RelieF, Texture, Features Engineering, SVM

Introduction

Many countries like Morocco depend very much on agriculture. Thus, such a country like others must increase its production to meet the enormous demands. Recently, it has been observed that huge solutions have been put in place to improve the quality and quantity of production to increase agricultural production in general. On the other hand, diseases are the fearsome enemy of this agricultural progression and can have a direct impact on the quality and quantity of plant foods. These diseases often affect plants and are defined by professionals as anything that disturbs the natural behavior of plants and prevents sufficient production (UNL, 2019). A plant is classified as diseased when an external actor affects it and causes a change in its physiological and biochemical behavior leading to abnormal growth of these functions (EBI, 2019). Annually Morocco loses crop yield like any country in the world due to plant diseases. The protection of crops against plant diseases has a vital role and has to play in meeting the growing demand for food quality and quantity (Strange and Scott, 2005). So, disease identification in plants remains a difficult task for farmers due to lack of expertise which means that the beginners' farmers cannot identify the case of a diseased plant in the early stage because the visual identification is not precise (Wäldchen and Mäder, 2018). Recently we see many applications using Machine Learning (ML) and Deep Learning (DL) algorithms in this direction aiming to get the best analysis results (Ennouni *et al.*, 2017; Borra *et al.*, 2019). So, to help farmers, computer processing and machine learning can be utilized to develop a robust classification system that

Science Publications

can help in the detection and the classification of plant diseases using only images of leaves (Saleem *et al.*, 2019). Such a system to detect and classify plants diseases can involve five main steps:

1. Acquisition
2. Pre-processing
3. Segmentation
4. Features engineering
5. Classification and analysis

Figure 1 presents the whole process to detect diseased plants using image processing and ML.

The process of detection and classification of plant diseases consists of first identifying the plant to be studied and which is subject to a disease and then taking a snapshot of one of the infected leaves and transmitting it to the Computer-Aided Diagnosis (CAD) system.

The first step after image acquisition is preprocessing which aims to remove the artifacts and noises. The aim of the preprocessing is to enhance the input image and to prepare it for segmentation. Several preprocessing techniques have been proposed; multiscale decomposition, color space transformation, filtering, cropping, smoothing.

To extract the Region Of Interest (ROI) the segmentation is applied. The aim, in our context, is to identify correctly the leaves to subsequently identify whether it is healthy or diseased. In this step, we distinguish between healthy and unhealthy leaves. In this phase, many techniques have been used (Gonzalez and Woods, 2008; Cui *et al.*, 2010; Kwack *et al.*, 2005). For instance and since we used a good preprocessing decomposition, we can use K-means, edge detection, or any other algorithms in this direction, etc.

Feature engineering: The extracted features are the inputs for the classification algorithm. So, from existing features, we create the new ones to improve algorithm performance and it focuses on what's important. This step contains features normalization, feature selection (Haghighat *et al.*, 2016), dimensionality reduction (Filali *et al.*, 2020).
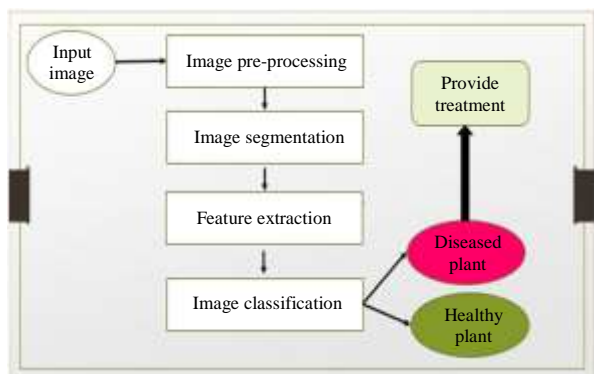

Fig. 1: Plant diseases detection and classification workflow

Finally, in the classification step, to classify and identify the healthy or infected leaf selected features will be used instead of all the extracted features (Pal and Mather, 2006). In this step, many classifiers can be used like Neural Network (ANN), Decision Tree, Random Forest and Support Vector Machine (SVM). Based on a comparative study we will use the SVM as the classification algorithm (Cui *et al.*, 2010; Khan and Ahmad, 2004).

Related Work

In recent times, a lot of methods are proposed in the classification of plant diseases (Warne and Ganorkar, 2015; Kadir *et al.*, 2013; Shergill *et al.*, 2015; Sumathi and Kumar, 2012; Khirade and Patil, 2015; Beghin *et al.*, 2010; Sladojevic *et al.*, 2016; Tulshan and Raul, 2019; Sibiya and Sumbwanyambe, 2019). Warne and Ganorkar (2015) proposed a machine vision approach for recognition of three types collected with one type characterized by the damages of a tormentor insect; green stink and two types visualized by symptoms of 2 pathogens; Bacteria angular and Ascochyta blight. Their approach is based on color, texture, shape, lacunarity, shape dimension and Fourier descriptors, which are necessary to achieve the classification by SVM algorithm. Their approach is tested only on 117 images, within which the recognition accuracy was ninety-three percent.

Kadir *et al.* (2013) introduced an approach that includes shape, vein and texture features. They have used Probabilistic Neural Networks (PNN) as an algorithm for the plant leaf classification. Usually, numerous approaches are there for plant leaf classification but none of them have taken color feature, because the color was not considered as an important aspect of the identification process. In this case, color also playing important role in the classification process. Their result indicates that the suggested method provides an average accuracy of 93.75% when it was tested on Flavia dataset which contains 32 kinds of plant leaves.

Shergill *et al.* (2015) proposed a recognition system of 5 diseases includes Early scorch, Cottony mould, Ashen mold, Late scorch and little achromatic color, which mainly attack cotton cultivation. Their method used Haralick texture features and neural networks for classification. The authors presented and tested their methodology on 192 images of six categories (5 types of diseases and one type portrayed by the normal leaves). The average accuracy was ninety-three percent.

Sumathi and Kumar (2012), they applied a feature fusion technique using the Gabor filter in the frequency range and merging the obtained features with edge-based feature extraction without considering color features. The resulted features were trained using 10 fold cross-validation and tested with CART and RBF classifiers to evaluate its accuracy.

RBF provides an accuracy of 85.93% with low comparative error for a nine class problem.

In this study (Khirade and Patil, 2015) they proposed an approach for detecting two diseases: Downy mildew and Powdery mildew, based on neural networks for classification, their approach is based on 4 categories of features including shape dimension, texture, color, and form. The identification rate was ninety-seven held on 85 images, 50 of Downy mildew and thirty-five of Powdery mildew.

Concerning (Beghin *et al.*, 2010) mentioned an approach that joins easy steps based on shape and texture features. The shape-based method extracts the contour descriptor from all leaves and then computes the dissimilarities between them. After they analyze the macro-texture of the leaf using the orientations of edge gradients. The obtained results are then joining with the aid of an incremental classification algorithm which provides 81.1% accuracy.

Besides that, (Sladojevic *et al.*, 2016) propose an approach using a deep learning method for classifying and detecting plant diseases from leaf images. The proposed approach is capable to classify just 13 different types of plant diseases out of healthy leaves, with the ability to distinguish plant leaves from their surroundings. In this case, the achieved results on the proposed approach provide an average accuracy of 94.60%.

Tulshan and Raul (2019) applied a plant leaf disease detection technique to detect a disease from the input images. This technique includes many steps as, image segmentation, feature extraction. Furthur K Nearest Neighbor (KNN) classification is applied to the results of these three stages. Obtained results have exposed 95.81% of accuracy in predicting plant leaf diseases.

The work was proposed in (Sibiya and Sumbwanyambe, 2019) used CNN method for classifying images of the maize leaf diseases that were collected by the use of a smartphone camera. The used method deals with three different types of maize leaf diseases out of healthy leaves. The northern corn leaf blight (Exserohilum), common rust (*Puccinia sorghi*) and gray leaf spot (Cercospora) diseases were chosen for this work as they affect most parts of Southern Africa's maize fields. Their results demonstrate that the suggested method provides an average accuracy of 92.85% when it was trained and tested using datasets from Plant Village's online website.

Although different methods have been stated and have been examined with almost all leaf features successfully extracted and classified, still those methods have their limitations. It is showed clear that someones are signaled to be inaccurate, mainly because the input image contains noise and texture and also no use of preprocessing step decrease classification performance.

The main advantage of our approach is considering the most important features to classify plants leaves for greater accuracy because good accuracy can be achieved by adding a more consistent number of features and also by increasing the dataset. So the novelty of our proposed method lies in its simplicity.

This paper focuses on plant disease diagnosis based on a multiscale preprocessing scheme to propose a new and powerful approach for identifying diseased plants in the early stage. The remainder of this paper is organized as follows: Section 2 presents related work, section 3 presents an overview of the proposed approach and describes it in detail. Section 4 depicts our results and related discussion. It also presents the dataset that is used and the evaluation metrics. In the end, section 5 summarizes this paper and gives some future works concerning this work.

Proposed Method

In this research paper, a powerful method is proposed for improving the performance of classification for plant disease. At first, our work contributes to the pretreatment step aiming to improve the quality of segmentation and then to get a greater classification accuracy. We suggest a pretreatment that uses an image decomposition model based on the PDE model to separate our input images were collected from Plant Village Dataset that covers 6215 images classified into 15 subsets. Then, color histogram, morphological features and GLCM features are used for extracting the features from the leaf region and textural region obtained by projection of the segmentation mask. After getting the hybrid features; the codebook is built using an easy concatenation based method, RelieF feature selection algorithm is used to limit the number of features for representing the data efficiently. Then, the output of feature selection is given as the input for SVM classifier for classifying leaf as healthy or diseased. At last, the proposed method performance is compared with the existing methods in light of specificity and sensitivity using PDE and accuracy using RelieF.

An overview of the proposed approach step by step is presented in Fig. 2.

The steps involved in the proposed approach are detailed as follow:

1. Image acquisition
2. Pre-processing: Apply PDE Image decomposition to obtain two components: Object component and texture component
3. Segmentation: Extract the leaf by segmenting only the object component
4. Features extraction:

   a. Extract texture features using only the textural part of the leaf after the projection of the segmentation mask on the texture component
   b. Extract color features from only the leaf obtained by projection of the segmentation mask on the input image

   c.   Extract shape features from the segmented image
5.   Features fusion using codebook method
6.   Features selection: Selecting the pertinent features using the Relief algorithm
7.   Classification: Apply SVM Algorithm for classification using only selected features

A. Pretreatment

For getting pertinent classification, attention to image pre-treatment is required so that one can improve image features by eliminating unwanted misrepresentation. This part emphasizes the image pre-processing method which is adopted. Partial Differential Equations (PDE) based decomposition method offer advantages and resolve misclassification that finally results in improved features. So our work contributes to the pretreatment stage aiming to improve the quality of classification. We propose a pretreatment that uses an image decomposition model based on PDE to separate our input image. Then our pretreatment based on a variational model using TV-L^1 model (Khan and Ahmad, 2004) is performed on it, aiming to reduce the effect of inner structure that makes the classification process not pertinent as result. We use TV-L^1 model to separate our images to both: The object part and not object part. In the last result, we have two components object and texture. In this case, the texture will add a second level of information which is necessary for our process. Besides that, we estimate the noise level by measuring the standard deviation of the grey-level histogram of continuous regions of the image.
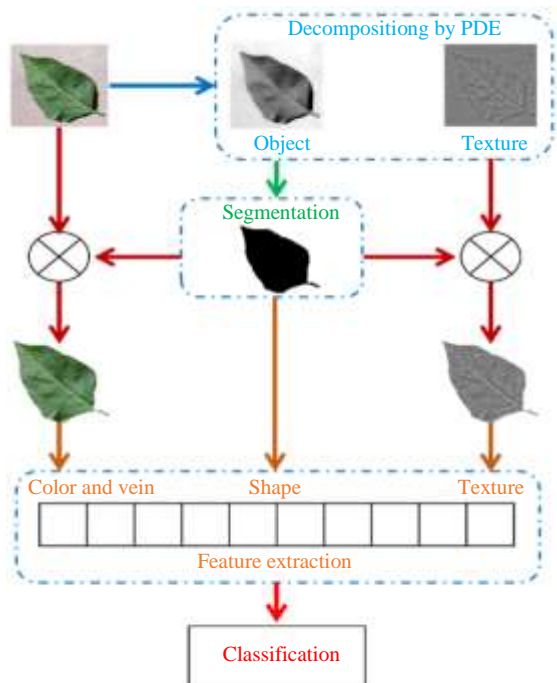

Fig. 2: An overview of the proposed approach

Several models have been proposed to decompose images. Meyer (2001) presented some limitations of the proposed model. Nikolova proposed in (Warne and Ganorkar, 2015) to use L^1 norm instead of the L^2 norm in the ROF model. The L^1 norm according to Nikolova is suitable to remove salt and pepper noises. Zhang (2002; Pham *et al.*, 2000), the authors proposed to use Gabor function and Hilbert norm to resolve the EDP equation. This model aims introduced orientation and frequency information of textures in the input image.

In this section, we present some models based on PDE. Many models have been proposed using Partial Differential Equations (PDE) to extract textures that are represented by oscillating patterns. For example, ROF model (Rudin *et al.*, 1992) and the TV-L^1 model (Khan and Ahmad, 2004). The rest of this section is devoted to the presentation of TV-L^1 that is adopted:

$$\min\left\{\alpha J(u) + \frac{1}{2}|u - I|^2\right\}$$

where, *J* represents the variation of total regularization:

$$J(u) = \int \nabla_u dx$$

∇:   Represents the spatial gradient,
α: Represents the compromise between the best fit to the noisy data and the regularization
I: Is the input image
u and *v = I-u*: The respectively the object component, texture and noise component obtained by the decomposition

In our proposed approach we will use the TV-L^1 proposed in (Khan and Ahmad, 2004) which can separate efficiently the texture from objects in our images. Table 1 presents two decomposition examples of images using the EDP.

Table 1: Images decomposition using EDP

Input image	Components obtained by EDP	
	Object	Texture

Table 2: Segmentation results with and without EDP decomposition

		Segmentation	
Input image	Ground of truth	On input image	On object component

B. Segmentation

Segmenting an image consists of partitioning it into several homogeneous parts with similar properties. These partitions will be used later on in the analysis and interpretation. In this phase, we will segment the object image using the K-means clustering technique (Chilvers, 2013). The choice of the K-means algorithm when knowing that there are several segmentation algorithms in literature is based on several studies (Borra *et al.*, 2019).

In our proposed approach and to improve the segmentation result we will apply the K-means algorithm on the object component obtained after decomposing the input image by the EDP. The object component contains only the shape of the objects without textures or noises.

A segmentation example using the K-means algorithm based on the original image and object component is presented in Table 2.

As you see, the images of the leaves contain a lot of textures which gave a bad segmentation. This is why segmenting the object component instead of the input image gives the best segmentation mask. This can help us in identifying correctly the leaf for the features extraction phase and will improve significantly the classification accuracy.

C. Features Extraction

To get the most relevant information from our segments; it is obligatory to apply the feature extraction task. In-plant analysis, generally they are four basic features that are commonly used: Color, Texture, shape and Vein.

a. Color Features

The color-based features are most widely used primitive to represent correctly the infected region in the leaves image. Color features can be acquired by diverse methods like color histogram (Bhagat and Atique, 2012), color structure descriptor, color moments (Albregtsen, 2008). Color histogram can describe well the leaves color characteristics because it can represent the distribution of the color in the leaves image. So we extract color features from only leaf region obtained by projection of the segmentation mask on the input image. We specified collections of color features. In the first collection, we calculated the standard deviation (σ) and mean (μ) of the intensity of each RGB canal, leaf kurtosis and leaf skewness. So, we will have 12 color features. The values of Mean (μ), kurtosis (σ), skewness (θ) and standard deviation (y) are calculated as follows:

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}P_{ij}$$

$$\sigma = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(P_{ij}-\mu\right)^{2}}$$

$$\theta = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(P_{ij}-\mu\right)^{3}}{MN\sigma^{3}}$$

$$y = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(P_{ij}-\mu\right)^{4}}{MN\sigma^{4}} - 3$$

where, *M* and *N* are the image size and p_{ij} is the intensity.

b. Textural Features

Textural features will be extracted from only the leaf texture obtained by projection of the segmentation mask on the texture component obtained by the EDP. Six measures are used in our proposed approach to characterize textural information. Those measures will be extracted from the famous Gray Level Co-occurrence Matrix (GLCM) method proposed by the (Syahputra, 2014; Zaletel *et al.*, 2016); Angular Second Moment (ASM) that measures the uniformity of texture in the leaf, leaf Contrast, Inverse Different Moment (IDM) that measures the homogeneity of texture, the Entropy to measure the non-homogeneity of leaf texture and the Correlation that measures the degree of linear dependency on the leaf. The six measures are given as follows:

$$ASM = \sum_{i,j=1}^{N} P_{i,j}^2$$

$$Dissimularity = \sum_{i,j=1}^{N-1} P_{i,j} |i - j|$$

$$Contrast = \sum_{i,j=1}^{N-1} P_{i,j} (i - j)^2$$

$$Homogeneity = \sum_{i,j=1}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2}$$

$$Correlation = \sum_{i,j=1}^{N-1} P_{i,j} \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$$

$$Entropy = \sum_{i,j=1}^{N-1} P_{i,j} (-\ln P_{i,j})$$

c. Shape Features

Shape features also considered as an important index in plant leaf and there are many techniques to extract the leaf shape features (Munisami *et al.*, 2015; Wu *et al.*, 2007), but in this study, we utilize the morphological features in (Aptoula and Yanikoglu, 2013), which is a common shape features used in the literature. To identify the global and local information of the leaf, ten shape features are extracted from the segmented leaf. These features are defined as; the leaf area, the leaf width, the leaf height, the leaf perimeter, the extent, the solidity, roundness, aspect ratio and major and minor axis which calculate respectively the longest and shortest distance between two points on the border of the leaf. All these shape features were used in classification are described as follows:

- Area: The value of leaf area is the actual number of pixels in the region:
  - The leaf width is calculated based on the sum of the number of pixels for the widest region of the leaf
  - The leaf height is calculated based on the sum of the number of pixels for the highest region of the leaf
- Perimeter: Perimeter of a leaf is the summation of the distances between each adjoining pair of pixels around the border of the leaf
- Aspect ratio: Is another feature sometime called slimness is defined as follow:

$$slimness = \frac{L1}{L2}$$

where, L_1 is the width of a leaf and L_2 is the length of a leaf.

- Major axis: The line segment connecting the base and the tip of the leaf is the major axis
- Minor axis: The maximum width, which is perpendicular to the major axis, is the minor axis of a leaf
- Extent: The extent of a leaf specifies the ratio of pixels in the region to pixels in the smallest rectangle containing the region

Solidity is measured as an area convex hull where the convex hull bounds the leaf shape as a polygon. Leaves with a large discrepancy between area and convex hull can be distinguished from leaves lacking such features using solidity.

Roundness is a ratio of area to the perimeter (true perimeter, excluding holes in the middle of the object) measured as:

$$Roundness = \frac{4\pi A}{P^2}$$

where, A is the area of leaf image and P is the perimeter of leaf contour.

d. Vein Features

Vein contains significant information despite its complex modality. Mishra and Pandey (2020) presents four features that are extracted from vein of the segmented leaf. The vein of leaf is constructed using the opening morphological operation. The structuring element used in the opening procedure is disk-shaped using respectively 1, 2, 3 and 4 radiuses which will give us 4 vein structures. The 4 extracted features are given as follows:

$$V_i = \frac{A_i}{A}$$

where, $i = 1$ to 4 and A_i is the number of pixels contained in the vein obtained by the opening morphological operation using a structuring element of radius equal to i.

D. Fusion of Features

In terms of feature fusion methods search, we find many feature fusion techniques have been developed like CCA (Sun *et al.*, 2005) and discriminant correlation analysis (Haghighat *et al.*, 2016) but here we adopt a simple concatenation method that merges all extracted features horizontally in only one vector per image. After the extraction of four types of features counting color, shape, texture and vein. The codebook is built by these features using an easy horizontal concatenation-based fusion method. So we should check the range of features if they were not from the same numerical range (varied),

they should be normalized by using the normalization technique to transform the feature vector into a common domain before concatenating them. In the absence of normalization features with large values have a stronger influence on the cost function in designing the classifier (Theodoridis *et al.*, 2010). Since we have a limited number of characteristics we have opted for merging these features using a direct concatenation by minimizing the correlation between them and omitting the interclass variations. Hence, the feature codebook is optimized by a new feature selection method.

E. Features Selection

Missing and noisy features in our process will degrade the performance of classification. Starting from the fact that all extracted features are not relevant and cannot correctly represent the leaves images, feature selection is a fundamental step in our process. The presence of features in our dataset increases the volume of our data. As a result, using this step we will describe well our features by removing redundant and irrelevant features to first achieve higher classification accuracy and second speed up the classification time. In the recent study (Duda *et al.*, 2001; Yin *et al.*, 2018). Many researchers have studied the approaches of feature selection and the improvement of the classification accuracy using Relief approach (Kira and Rendell, 1992). We start by a brief review of relief algorithm:

$$D = \left\{ \left(x_n, y_n \right) \right\}_{n=1}^N C\mathbb{R}^l \times \left\{ \pm 1 \right\}$$

Let *D* be a training dataset, where x_n is the *n*-th data sample and y_n is its equivalent class label. The principal idea of relief is to iteratively approximate feature weights according to their capability to discriminate between neighboring patterns. In each iteration, a pattern *x* is randomly selected and then two nearest neighbors of *x* are found, one from the same class (termed the nearest hit or NH) and the other from the opposite class (termed the nearest miss or NM). The weight of the *i*-th feature is then updated as:

$$w_i = w_i + \left| x^{(i)} - NM^{(i)}(x) \right| - \left| x^{(i)} - NH^{(i)}(x) \right|, for\ 1 \leq i \leq I.$$

They were found that the RelieF algorithm (Li *et al.*, 2011) is considered one of the most successful ones due to its easiness and effectiveness to detect the restrictive reliance between used features. This algorithm deals with multi-class and performed better. The feature weight of the instances is updated after the random selection of instances. Since this algorithm selects the instances in the random, the uncertainty in the feature weight is recorded iteration by iteration to get a suitable feature weight from the feature set.

Table 3: Selected features by RelieF

Selected features	Sub features
Shape	Aspect ratio
	Perimeter
	Extent
	Roundness
	Solidity
Color	variance of Red, Green and Blue, skewness
Texture	Contrast
	Correlation
	Homogeneity

So Table 3 presents the selected features obtained by the RelieF algorithm.

F. Classification

Finally, in the last phase, the selected features are stored in the feature dataset and are passed through SVM classifier. The SVM classification algorithm is chosen for our implementation even though there is a multitude of classification algorithms based on studies we have conducted that have shown that SVM is well suited for tree leaf classification (Filali *et al.*, 2018). The objective of classification in our case is to be able to design a very powerful classification model to correctly predict the class of a new image containing a tree leaf between the 18 present classes. Thus, in our case, it is a supervised classification where we have a labeled base that will be used for learning and another base for validation and evaluation of the classification model. Technically, we use multi-class classification algorithms to couple labeled input data with correct outputs (predictions).

Experimental Results

A. Dataset

In this study and to validate our proposed approach we will use the popular dataset called Plant Village (PV) Dataset. This dataset is one the most used for the evaluation of plant disease classification algorithms it contains 6215 images classified into 15 subsets (Shin *et al.*, 2016). It contains healthy and infected leaves with diseases. Some diseases such as Potato Late blight can only be found in Potato and tomato and respectively. The dataset is provided by the Ground of Truth (GT) that will help us to evaluate our segmentation proposed approach. Table 4 presents a detailed description of the used dataset and also the abbreviation of each class used in the next subsections.

Since we have only one labeled dataset, to design, validate and evaluate our classification model. Note that, the data comes with predefined training and validation subsets. In our work, we use the configuration as the one that gives the best performances in the research (Mohanty *et al.*, 2016), where 80% of the data is for the training and the remaining 20% for validation. Fig. 3 gives examples of each class.

Table 4: Description of PV dataset

Class name	Abbreviation	Number of images
Pepper bell healthy	PBH	745
Pepper bell Bacterial spot	PBB	802
Potato Early blight	PEB	431
Potato healthy	PH	152
Potato Late blight	PLB	224
Tomato Target Spot	TTS	336
Tomato Tomato mosaic virus	TTMV	373
Tomato Tomato YellowLeaf Curl Virus	TTYLCV	306
Tomato Bacterial spot	TBS	593
Tomato Early blight	TEB	283
Tomato healthy	TH	292
Tomato Late blight	TLB	233
Tomato Leaf Mold	TLM	504
Tomato Septoria leaf spot	TSLP	346
Tomato Spider mites Two spotted spider mite	TSMTSSm	303
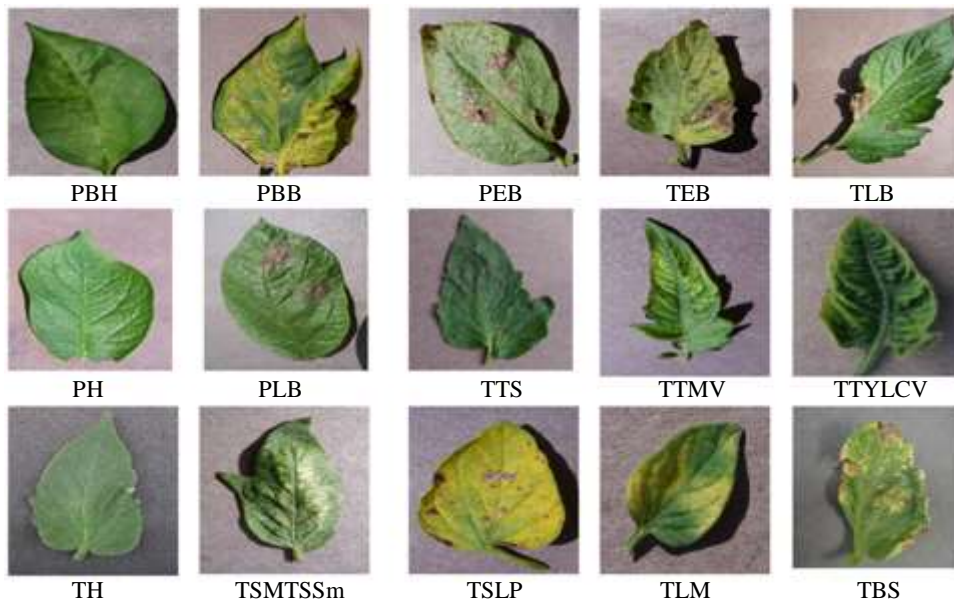
Fig. 3: Plant healthy and diseases of each class from the Plant Village dataset

B. Performance Measures

To evaluate the segmentation and classification phase, we will use three measures; Sensitivity or precision (the True Positive Rate), Specificity or recall (True Negative Rate) and accuracy. In our case, as we are working on a multiclass context, the Specificity and Specificity average over all classes is calculated by the macro-mean. It calculates first the Specificity and Sensitivity over each class and then calculates the average over all the classes as defined below:

$$precision = \frac{\sum_{i=1}^{n} precision_i}{n}$$

where:

$$precision_i = \frac{TP_i}{TP_i + FP_i}$$

$$recall = \frac{\sum_{i=1}^{n} recall_i}{n}$$

where:

$$recall_i = \frac{TP_i}{TP_i + FP_i}$$

$$accuracy = \frac{\sum_{i=1}^{n} accuracy_i}{n}$$

where:

$$accuracy_i = \frac{TP_i + FP_i}{TP_i + TN_i + FP_i + FN_i}$$

where, n is the number of classes, i the current class. *TP* is the True Positives, *FP* is the False Positives, *FP* is the False Positives, *FN* is the False Negatives.

C. Segmentation

Segmentation in our proposed approach is done on the object component obtained after decomposing the input image with the EDP. Table 5 presents the segmentation results of images that were selected randomly from the dataset.

Table 5 presents sensitivity and specificity values of the segmentation of all the images in the dataset: On images without preprocessing and on the object component after decomposition by EDP.

From values in Table 5 we can conclude that the preprocessing by the EDP decomposition allowed us to significantly improve the segmentation results. As you see we have considerable changes were applied decomposition-based PDE in Sensitivity/Specificity. So this step helps us to identify correctly the leaf we will extract relevant features and then have a good classification rate.

D. Classification

For the classification evaluation, the dataset has been divided into a training dataset and test dataset respectively as mentioned at the beginning of the results section. We note that 80% for training and 20% for validation. Both datasets are labeled. The classification model is created on the training dataset and serialized and used after that in a cross-validation process on the images in the test dataset. The accuracy and the confusion matrix will be used to evaluate our proposed classification model.

The Features Selection (FS) comparison study is first elaborated to keep only the most relevant extracted features.

From values in Table 6, the features selection process permits us to significantly improve the classification accuracy and to reduce the number of features used by selecting only 12 relevant features from 34 initial ones.

The confusion matrix is also used to evaluate our proposed approach.

The confusion matrix helps us to know exactly each of the classes have been wrongly classified and to improve more in the future the proposed classification model. From Table 7 we can see that the worst classifications have resulted for the Tomato Healthy (TH) class when it was misclassified especially as Potato Early Blight (PEB) and Tomato Bacterial Spot (TBS). This is due to that visually their leaves are very similar.

To situate the effectiveness of our proposed approach in comparison with the most relevant classification approach proposed in the literature, Table 8 presents a classification accuracy comparative study where our proposed approach is compared with Deep Learning-based approaches for plant diseases. Concerning these approaches (Sladojevic *et al.*, 2016; Tulshan and Raul, 2019; Sibiya and Sumbwanyambe, 2019) are presenting before in related work section.

Table 5: Segmentation sensitivity and specificity with and without PDE

	Segmentation	
	Without PDE	Using PDE
Sensitivity	75.25	98.20
Specificity	78.00	97.90

Table 6: Results of classification with and without features selection

	Without FS	RelieF
Number of features	34.0	12.0
Accuracy (%)	82.1	95.9

Table 7: The confusion matrix using the Relief algorithm with SVM

	PBH	PBB	PEB	PH	PLB	TTS	TTMV	TTYLCV	TBS	TEB	TH	TLB	TLM	TSLP	TSMTSSm
PBH	720	0	0	0	0	0	0	0	0	0	25	0	0	0	0
PBB	0	759	0	0	0	0	0	0	0	0	43	0	0	0	0
PEB	0	0	337	0	0	0	0	0	0	0	94	0	0	0	0
PH	0	0	0	152	0	0	0	0	0	0	0	0	0	0	0
PLB	0	0	0	0	215	0	0	0	0	0	9	0	0	0	0
TTS	0	1	0	0	0	320	0	0	0	0	15	0	0	0	0
TTMV	0	0	0	0	0	0	364	0	0	0	9	0	0	0	0
TTYLCV	0	0	0	0	0	0	0	287	0	0	19	0	0	0	0
TBS	0	0	0	0	0	0	0	0	569	0	24	0	0	0	0
TEB	0	0	0	0	0	0	0	0	0	268	15	0	0	0	0
TH	0	0	0	0	0	0	0	0	0	0	584	0	0	0	0
TLB	0	0	0	0	0	0	0	0	0	0	12	221	0	0	0
TLM	0	0	0	0	0	0	0	0	0	0	10	0	494	0	0
TSLP	0	0	0	0	0	0	0	0	0	0	3	0	0	343	0
TSMTSSm	0	0	0	0	0	0	0	0	0	0	4	0	0	0	299

Table 8: A comparison of the results of the classification of the approaches we propose with recent approaches

	Sladojevic *et al.* (2016)	Tulshan and Raul (2019)	Sibiya and Sumbwanyambe (2019)	Proposed approach
Accuracy (%)	94.60	95.81	93.45	95.90

Although our approach is based on machine learning using handcrafted features, the classification accuracy obtained exceeds recent approaches based on Deep Learning. This is due firstly to the use of PDE which allowed us to identify correctly the leaves and secondly to the use of the relevant extracted and selected features.

Conclusion

In this study, a powerful detection and classification approach for plant diseases is proposed and applied to plant village dataset. The main objective of this research work is to propose a suitable pretreatment, proper feature extraction and feature selection methods for improving the accuracy of classification for plant diseases. In this study, the use of PDE-based TV-L^1 model allows us to isolate the object from the texture which makes the segmentation step more reliable. As a result, we have two components, the first one contains the geometric part and the second is the texture. Then texture, color and geometric features are combined in a feature vector using the codebook method. Furthermore, the optimal features are selected by ReliefF feature selection method; the most dominant discriminative features are passed to Support Vector Machine for the last step to classify an input plant image into diseased or not. From the experimental results, the proposed method achieves 95.90% of accuracy, but the existing methods obtained a limited accuracy of 93.45% on the Plant Village dataset. In future research, we would like to reduce the time of the pretreatment from PDE based TV-L^1 into our approach to improve the accuracy of classification

Acknowledgment

Author's Contributions

Assia Ennouni: Developpement, experimentations testing, validation and writing the manuscript.

Noura Ouled Sihamman: Exprermimentations validation and proofreading.

My Abdelouahed Sabri: Concept developpement, experimentations testing, validation and proof reading.

Abdellah Aarab: Concept developpement, experimentations validation and proof reading.

All authors read and approved the final manuscript.

Ethics

All information provided in this study is confidential and unique. This paper has neither been published nor is under review elsewhere. There are no ethical issues associated with this research.

References

Albregtsen, F. (2008). Statistical texture measures computed from gray level coocurrence matrices. Image Processing Laboratory, Department of Informatics, University of Oslo, 5(01), 01-14. https://www.uio.no/studier/emner/matnat/ifi/INF4300/h08/undervisningsmateriale/glcm.pdf

Aptoula, E., & Yanikoglu, B. (2013, September). Morphological features for leaf based plant recognition. In 2013 IEEE International Conference on Image Processing (pp. 1496-1499). IEEE. https://doi.org/10.1109/ICIP.2013.6738307

Beghin, T., Cope, J. S., Remagnino, P., & Barman, S. (2010, December). Shape and texture based plant leaf classification. In International conference on advanced concepts for intelligent vision systems (pp. 345-353). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-17691-3_32

Bhagat, A. P., & Atique, M. (2012, March). Design and development of systems for image segmentation and content based image retrieval. In 2012 2nd National Conference on Computational Intelligence and Signal Processing (CISP) (pp. 109-113). IEEE. https://doi.org/10.1109/NCCISP.2012.6189688

Borra, S., Dey, N., Bhattacharyya, S., & Bouhlel, M. S. (Eds.). (2019). Intelligent Decision Support Systems: Applications in Signal Processing (Vol. 4). Walter de Gruyter GmbH & Co KG. https://doi.org/10.1515/9783110621105

Chilvers, M. (2013). Bacterial Blight and Septoria Brown Spot Appearing in Soybeans. Michigan State University Extension, Department of Plant, Soil and Microbial Sciences. https://www.canr.msu.edu/news/bacterial_blight_and_septoria_brown_spot_appearing_in_soybeans#:~:text=The%20disease%20is%20caused%20by,cause%20Septoria%20brown%20spot%20symptoms.

Cui, D., Zhang, Q., Li, M., Hartman, G. L., & Zhao, Y. (2010). Image processing methods for quantitatively detecting soybean rust from multispectral images. Biosystems Engineering, 107(3), 186-193. https://doi.org/10.1016/j.biosystemseng.2010.06.004

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). Pattern classification (2nd and.; New York; Chichester. ISBN: 978-0-471-05669-0.

EBI. (2019). Plant Disease. Encyclopædia Britannica, Inc. https://www.britannica.com/science/plant-disease/Epiphytotics

Ennouni, A., Filali, Y., Sabri, M. A., & Aarab, A. (2017, April). A review on image mining. In 2017 Intelligent Systems and Computer Vision (ISCV) (pp. 1-7). IEEE. https://doi.org/10.1109/ISACV.2017.8054968

Filali, Y., Ennouni, A., Sabri, M. A., & Aarab, A. (2018, April). A study of lesion skin segmentation, features selection and classification approaches. In 2018 International Conference on Intelligent Systems and Computer Vision (ISCV) (pp. 1-7). IEEE. https://doi.org/10.1109/ISACV.2018.8354069

Filali, Y., Sabri, M. A., & Aarab, A. (2020). Improving Skin Cancer Classification Based on Features Fusion and Selection. In Embedded Systems and Artificial Intelligence (pp. 379-387). Springer, Singapore. https://doi.org/10.1007/978-981-15-0947-6_36

Gonzalez, R. C., & Woods, R. E. (2008). Digital image processing. Upper Saddle River, N.J: Prentice Hall. ISBN: 013168728X.

Haghighat, M., Abdel-Mottaleb, M., & Alhalabi, W. (2016, March). Discriminant correlation analysis for feature level fusion with application to multimodal biometrics. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1866-1870). IEEE. https://doi.org/10.1109/ICASSP.2016.7472000

Kadir, A., Nugroho, L. E., Susanto, A., & Santosa, P. I. (2013). Leaf classification using shape, color and texture features. arXiv preprint arXiv:1401.4447. https://arxiv.org/abs/1401.4447

Khan, S. S., & Ahmad, A. (2004). Cluster center initialization algorithm for K-means clustering. Pattern Recognition Letters, 25(11), 1293-1302. https://doi.org/10.1016/j.patrec.2004.04.007

Khirade, S. D., & Patil, A. B. (2015, February). Plant disease detection using image processing. In 2015 International conference on computing communication control and automation (pp. 768-771). IEEE. https://doi.org/10.1109/ICCUBEA.2015.153

Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In Machine learning proceedings 1992 (pp. 249-256). Morgan Kaufmann. https://doi.org/10.1016/B978-1-55860-247-2.50037-1

Kwack, M. S., Kim, E. N., Lee, H., Kim, J. W., Chun, S. C., & Kim, K. D. (2005). Digital image analysis to measure lesion area of cucumber anthracnose by Colletotrichum orbiculare. Journal of General Plant Pathology, 71(6), 418-421. https://doi.org/10.1007/s10327-005-0233-0

Li, B., Wang, Q., & Hu, J. (2011). Feature subset selection: a correlation-based SVM filter approach. IEEJ Transactions on Electrical and Electronic Engineering, 6(2), 173-179. https://doi.org/10.1002/tee.20641

Meyer, Y. (2001). Oscillating patterns in image processing and nonlinear evolution equations, volume 22 of University Lecture Series. American Mathematical Society, Providence, RI, 364. https://doi.org/10.1090/ulect/022

Mishra, K. N., & Pandey, S. C. (Eds.). (2020). Cloud-Based M-Health Systems for Vein Image Enhancement and Feature Extraction: Emerging Research and Opportunities: Emerging Research and Opportunities. IGI Global. https://doi.org/10.4018/978-1-7998-4537-9

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419. https://doi.org/10.3389/fpls.2016.01419

Munisami, T., Ramsurn, M., Kishnah, S., & Pudaruth, S. (2015). Plant leaf recognition using shape features and colour histogram with K-nearest neighbour classifiers. Procedia Computer Science, 58, 740-747. https://doi.org/10.1016/j.procs.2015.08.095

Pal, M., & Mather, P. M. (2006). Some issues in the classification of DAIS hyperspectral data. International Journal of Remote Sensing, 27(14), 2895-2916. https://doi.org/10.1080/01431160500185227

Pham, D. L., Xu, C., & Prince, J. L. (2000). Current methods in medical image segmentation. Annual Review of Biomedical Engineering, 2(1), 315-337. https://doi.org/10.1146/annurev.bioeng.2.1.315

Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena, 60(1-4), 259-268. https://doi.org/10.1016/0167-2789(92)90242-F

Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant disease detection and classification by deep learning. Plants, 8(11), 468. https://doi.org/10.3390/plants8110468

Shergill, D., Rana, A., & Singh, H. (2015). Extraction of rice disease using image processing. International Journal of Engineering Sciences & Research Technology, 1(4), 135-143.

Shin, H. C., Roth, H. R., Gao, M., Lu, L.,... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging, 35(5), 1285-1298. https://doi.org/10.1109/TMI.2016.2528162

Sibiya, M., & Sumbwanyambe, M. (2019). A computational procedure for the recognition and classification of maize leaf diseases out of healthy leaves using convolutional neural networks. AgriEngineering, 1(1), 119-131. https://doi.org/10.3390/agriengineering1010009

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. Computational Intelligence and Neuroscience, 2016. https://doi.org/10.1155/2016/3289801

Strange, R. N., & Scott, P. R. (2005). Plant disease: a threat to global food security. Annual Review of Phytopathology, 43, 83-116. https://doi.org/10.1146/annurev.phyto.43.113004.13 3839

Sumathi, C. S., & Kumar, A. S. (2012). Edge and texture fusion for plant leaf classification. International Journal of Computer Science and Telecommunications, 3(6), 6-9. https://citeseerx.ist.psu.edu/viewdoc/download?doi= 10.1.1.680.8865&rep=rep1&type=pdf

Sun, Q. S., Zeng, S. G., Liu, Y., Heng, P. A., & Xia, D. S. (2005). A new method of feature fusion and its application in image recognition. Pattern Recognition, 38(12), 2437-2448. https://doi.org/10.1016/j.patcog.2004.12.013

Syahputra, H. (2014). Plant recognition using stereo leaf image using gray-level co-occurrence matrix. Journal of Computer Science, 10(04), 697-704. https://doi.org/10.3844/jcssp.2014.697.704

Theodoridis, S., Pikrakis, A., Koutroumbas, K., & Cavouras, D. (2010). Introduction to pattern recognition: a matlab approach. Academic Press. ISBN-10 : 0123744865.

Tulshan, A. S., & Raul, N. (2019, July). Plant Leaf Disease Detection using Machine Learning. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE. https://doi.org/10.1109/ICCCNT45670.2019.8944556

UNL. (2019). Plant Disease: Pathogens and Cycles. University of Nebraska–Lincoln. https://cropwatch.unl.edu/soybean-management/plant-disease

Wäldchen, J., & Mäder, P. (2018). Plant species identification using computer vision techniques: A systematic literature review. Archives of Computational Methods in Engineering, 25(2), 507-543. https://doi.org/10.1007/s11831-016-9206-z

Warne, P. P., & Ganorkar, S. R. (2015). Detection of diseases on cotton leaves using k-mean clustering method. International Research Journal of Engineering and Technology (IRJET), 2(4), 425-431. https://www.irjet.net/archives/V2/i4/Irjet-v2i474.pdf

Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y. X., Chang, Y. F., & Xiang, Q. L. (2007, December). A leaf recognition algorithm for plant classification using probabilistic neural network. In 2007 IEEE international symposium on signal processing and information technology (pp. 11-16). IEEE. https://doi.org/10.1109/ISSPIT.2007.4458016

Yin, C., Zhang, H., Zhang, R., Zeng, Z., Qi, X., & Feng, Y. (2018). Feature selection by computing mutual information based on partitions. IEICE TRANSACTIONS on Information and Systems, 101(2), 437-446. https://doi.org/10.1587/transinf.2017EDP7250

Zaletel, I., Milošević, N. T., Todorović, V., Kovačević-Filipović, M., & Puškaš, N. (2016). Fractal and gray level co-occurrence matrix texture analysis of senescent and non-senescent deciduous teeth stem cells: A pilot study. Fractal Geometry and Nonlinear Analysis in Medicine and Biology, 2(2), 1-6. https://doi.org/10.15761/FGNAMB.1000131

Zhang, Y. J. (2002). Image engineering and related publications. International Journal of Image and Graphics, 2(03), 441-452. https://doi.org/10.1142/S0219467802000755