Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 26 (2023), No. 1, pp. 147–158
DOI : 10.47974/JSMS-954

Screening based sparse classification method for the detection of corn leaf diseases

Anuja Kumar Acharya †
Minakhi Rout *
School of Computer & Engineering
KIIT (Deemed to be Univesity)
Odisha
India

Srinibasa Padhy §
School of Electronics Engineering
KIIT (Deemed to be Univesity)
Odisha
India

Satyaranjan Jena @
Pradeep Kumar Sahu #
School of Electrical Engineering
KIIT (Deemed to be Univesity)
Odisha
India

Abstract

Detection of diseases in corn plant at early stage reduces the chance of productivity loss. Recognizing the different kind of disease in corn plant requires the identification of lesion part of the plant along with the extraction of proper features from the region of interest. Although several imaging tools and classifier were developed in the recent past towards identifying different kinds of disease in corn plant but most of the classification models do not meet the specific requirement due to improper selection of feature and classifier. In this paper, we proposed a two-step based sparse recognition model for the identification of corn disease. First step emphasizes with the construction of image dictionary by extracting multiple features from the lesion part of training image and followed it in the second phase,

† *E-mail:* **anuja.signin@gmail.com**
E-mail:* **minakhi.routfcs@kiit.ac.in (Corresponding Author)
§ *E-mail:* **srinibas.padhyfet@kiit.ac.in**
@ *E-mail:* **srj.kiit@gmail.com**
# *E-mail:* **pksahu.nitrkl@gmail.com**

uses the screening and angle rule based sparse classifier to recognize the different corn diseases. Utilizing the data from the kaggle dataset and comparing with the other state of art, our method yielded an average classification accuracy of 88.55%. This makes the framework feasible and effective for the detection of corn plant disease.

Subject Classification: *68T07.*

Keywords: *Angle rule, Feature extraction, Lesion segment, SLIC, SRC.*

1. Introduction

CORN plant is one of the most popular and demanding crops having wider adaptability to grow under varied agro-climatic conditions. In India, maize or the corn is the third most important food crops after rice and wheat. Corn production puts a larger impact on the economy of many countries but the target efficacy is not achieved sometimes due to certain disease infection in the plant. Diseases, like gray leaf spot, corn leaf blight, tar spot and common and southern rust are included in this category. It is difficult to diagnose by planting personnel based on experience. However, the misdiagnosis reduces production efficiency. With the development of image based technique, nowadays it is possible to identify the kind of disease infected to the plant and measure can be taken at early stage to avoid the damage to these plant. Automated Image based tools are quite capable in detecting the diseases in the early stage of the plant infection. Automated tools plays significant role in identifying the crop disease at its early stage and hence reduces the agriculture productive loss [1-4]. Moreover, farmers of isolated and poor areas will also get benefits in their farming by adopting these methodology [5]. In past various methods based on machine learning and image processing tool were developed for the disease detection. These algorithms generally use the pipeline stages [6-7] such as preprocessing the image, separating the lesion part, extracting the feature and then recognizing the specific diseases. Sun et. al. [8] developed an improved threshold based image segmentation algorithm along with the multiple linear regression for the recognition of diseased plant. With the development of regression analysis and the L1 minimization, SRC becomes a popular choice for many researchers to use the technique for the classification and recognition task[9]. SR based classification (SRC) rely on the linear combination of a small number of elementary samples called atoms, which are usually selected from an over-complete dictionary[10]. SRC has been used extensively in applications like tumor classification, visual tracking, video surveillance and plant species identification [11-15]. Zhang et. al. developed a three stage method that uses sparse coding for recognizing the disease in the cucumber plant. This method utilize the *K*-mean clustering for identifying the lesion segment of the plant and then extract the local feature from lesion region and then using those feature, recognises the diseased leaf. Although the SRC methods are useful tools for the classification task but the L1–minimization step of SRC involves high computational cost. Previously many classification model of SRC were developed like L1 homotopic method[16-17], Orthogonal Matching Pursuit(OMP)[18] but the computational complexity is almost same. Towards concurring the remarks of L1 minimization step of SRC and to classify the plant disease into minimal computation, we have used screening and angle based SRC method in this paper for the identification and recognition of corn disease. Proposed screening method reduces the number of attribute participation for the classification.

The rest of this paper is organized as follows. Section 2 introduces the use of screening and angle rule for the sparse based classification. Section 3 describes details about the proposed classification model. Experimental results and analyses are given in Section 4. Finally the last section focus on the concluding remarks and the future works to improving the disease recognition algorithm.

2. Sparse Representation Based Classification Using Screening And Angle Rule

The general idea behind the Sparse Representation Based Classification (SRC) deals with the mechanism to represent and express the test element as the linear combination of all input training data accumulated in a dictionary and it also aims at yielding the lowest reconstruction error in classifying the test data to the class. SRC mainly composed of three stags that is subset selection for selecting the samples from the dictionary and followed to this are least square regression and classification via regression residual. In the proposed paper subset selection and the least square computation is handled by the screening method [19] and for identifying the disease class of test sample, angle rule estimation procedure is used.

In order to visualize the concept, we collect all training sample as $X = \{x_1, x_2, \dots\dots .x_n\} \in R^{d\,X\,n}$ with the known class label $Y = \{y_1, y_2 \dots\dots ., y_n\} \in [K]^n$, where n is the number of samples with each sample defined with numbers of features and K is the number of classes represented as [K] = {1, 2, ..K}. Mathematically the test sample $y \in R_d$ in

SRC is represented as the linear combination of selected sample $\hat{X} = \{x_1, x_2, \ldots\ldots.x_s\} \in R^{d\ X s}$ as

$$y = \beta_1 x_1 + \beta_2\ x_2 + \cdots\ldots\ldots\ldots\ldots\ldots + \beta_s x_s \qquad ...(1)$$

where β denotes the $s\ X\ 1$ least square regression coefficient between class specific dictionary $\hat{X}$ and y. In the context of the work, we have used the screening method for the sample selection $\hat{X}$ of samples. This method of the SRC begins with computing the dot product between each sample x_i with the testing sample y and then arrange these products on the basis of decreasing order of their value. Mathematically, this is expressed as $\alpha = \{x_1^T\ .y,\ x_2^T.y, \ldots\ldots,\ x_n^T.y\}$. From the above set α, we select s number of elements and represented as $\hat{X} = \{x_1, x_2, \ldots..x_s\}$. Here s is representing the number of selected elements also known as maximal sparsity level and this value chosen as $S = \min\{n/(\log n), m\}$.

Next the regression coefficient β for the above is computed using the moore-penrose inverse. i.e

$$\beta = \hat{X}^{-1} y \qquad ...(2)$$

where $\hat{X}^{-1}$ is the moore–penrose inverse.

Once the β and $\hat{X}$ is estimated, then computing the angle between the actual testing vector y and the predicted one i.e. $\hat{X}_k \beta_k$ will determine the class of the sample. Mathematically, it can be expressed as:

$$f_n(y) = \arg\min \theta(y\ , \hat{X}_k \beta_k) \qquad ...(3)$$

The angle between the testing vector y and the product $x_k \beta_k$ will estimate the class level of the testing sample. Less the angle between these products, more is the class similarity.

3. Materials and methods

A. Data Preprocessing and Augmentation

In the proposed model we have **c**ollected four different kinds of crop disease from the plant leaf disease library (https://www.kaggle.com/smaranjitghose/corn-or-maize-leaf-disease-dataset.) and these images are classified as corn blight, common rust, corn gray spot and corn healthy as shown in Figure 1. Approximately 4000 images are present in the repository. Each of these leaf images of different classes were preprocessed and resized to 224 X 224 before it is submitted to our model for the recognition of class of new test sample.

Figure 1

Sampling the image of 4 different class of Corn disease (a) Corn blight (b) Corn common rust (c) Gray spot and (d) Healthy

B. System Model

Our proposed model consists of two phase i.e construction of dictionary and classification step. Construction of dictionary consists of many sub steps such as preprocessing of leaf image, identification of lesion segmentation, extraction and construction of feature vector followed by the classification step used for the recognition of type or class of leaf disease by using the screening and angle rule method of SRC. Flow diagram of the proposed model is shown in the Figure 2. Details of each steps of the proposed model have been discussed in the subsequent section of the articles.

C. Preprocessing stage

In this stage, each of the raw image of the dataset is preprocessed to improve the quality of the image. Quality of the image is enhanced by

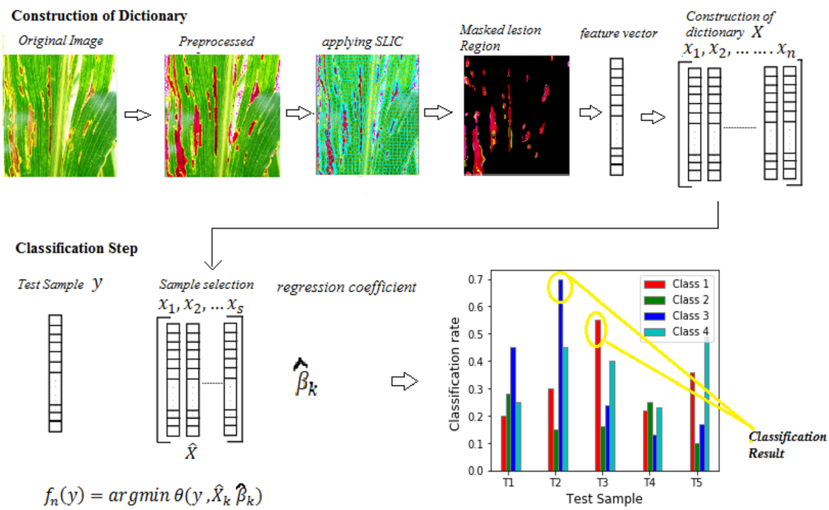


Figure 2

Flow chart of Corn leaf disease classification

adjusting the intensity and also to track the lesion segment, green channel of the image is extracted and then with the thresholding technique, lesion region is identified. In the current context of the paper we have used, Simple Linear Iterative Clustering popularly known as SLIC method[20]. This method is used to extract the super pixel from the image and this produces the high quality segmentation. Using the masking, these lesion regions are extracted and forwarded to the next stage for the construction of feature vector.

D. Extraction of feature vector and construction of dictionary

Extracting the valuable feature from the lesion part of the image and then constructing the feature vector is the key step for disease recognition process. Attribute such as color, shape and texture plays major role for this purpose. In this paper, we have considered multiple attribute such as color, texture, shape etc for detection of disease in order to enhance the accuracy of classification rate. In the current context of the paper, we have used a total of 21 attribute including 4 color attribute, 6 feature attribute

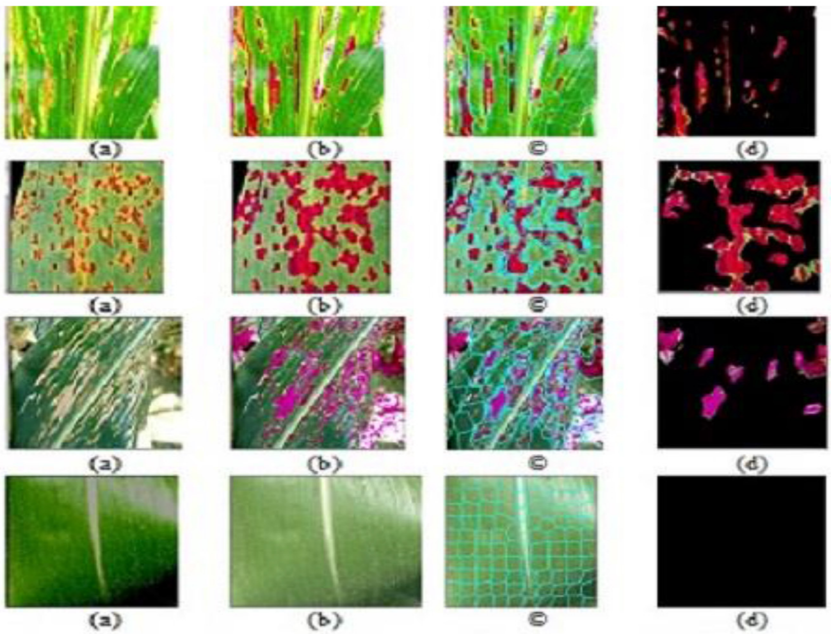


Figure 3

Identification of healthy and lesion region (a) Disease leaf images; (b) Enhancement leaf image; (c) Lesion region identification with SLIC algorithm (d) Image masking to extract the lesion region

and 11 texture attribute. These 21 attributes are extracted for each lesion segment of the image found from the previous section. These attribute or the feature extracted from the image are normalized and then connected with each other to construct the feature vector. As shown in the Figure 2, all the training samples of different classes undergo the feature extraction process and then the corresponding feature vector is constructed and inserted into dictionary.

E. Classification of corn leaf disease with screening and angle rule

This step is mainly concerned with discriminating the class of the test sample from the other classes on the basis of the regression residual. As described in the earlier section, screening process is used for the selection of sample from the dictionary and the angle rule helps in minimizing the class wise residual of the sample. This sample selection process is judged by estimating the dot product between the test observation y with the training sample X_i of the dictionary and then arranges these dot product in non-increasing order of their values. In particular, number of sample selection depends on the maximum sparsity level $S = \min\{n \div \log(n), m\}$ and this works well for our proposed model. Subsequently after selection of sample, regression coefficient β is estimated by Moore-penrose inverse and then the test sample is assigned to the class by minimizing the class wise residual in angle θ between the two vector x_i and y. Summarized steps for the proposed classification method is illustrated below:

Algorithm:

Step 1: Input: Training data matrix X and the corresponding class label y.

Step 2: Calculate the dot product between each element of the dictionary X with the testing observation y i.e $\alpha = \{x_1^T .y,\ x_2^T.y, \ldots\ldots,\ x_n^T.y\}$ and then arrange the element on the basis of non-increasing order of α. Select s number of elements from the set and assign it to $\hat{X} = \{x_1, x_2, \ldots..x_s\}$.

Step 3: Compute the regression coefficient $\beta = \hat{X}^{-1}y$ where $\hat{X}^{-1}$ is the moore–penrose inverse and then assign the testing observation y by evaluating $f_n(y) = \arg\min\theta\ (y, \hat{X}_k\beta_k)$ where $\in [K]$. Here θ denotes the angle between the vectors and this determine the class level of the testing observation.

4. Results and discussion

The experiments were performed under the windows platform with a 2.20 GHz system supported with 6 GB RAM. All the simulation

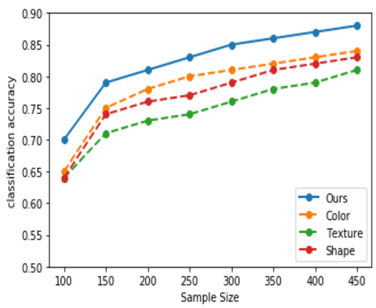


Figure 4

Accuracy testing of the model using the multiple features

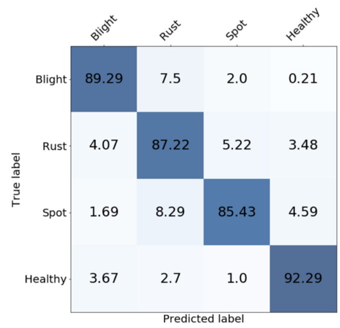


Figure 5

Performance evaluation of the model with confusion matrix

work performed with the Python software packages. At the time of data acquisition, the kaggle data set contains approximately 4000 images and after passed through the image quality selection, 500 images including different diseased class were extracted and used for the construction of the dictionary.

Figure 3 illustrated the images of different type of diseased corn leaf that is passed through the preprocessing, segmentation and masking stage for the extraction of the feature vector from the lesion region. After passing these stages, corresponding feature vector is constructed and added to the dictionary. Finally, in order to measure the performance of the model, different experiments were conducted by applying the screening and angle rule of SRC method on the test sample.

Case 1: Accuracy of the model is being tested with the multiple features in comparison to the single feature. Extraction of multiple features such as color, space and texture enhance the accuracy of the classification rate. Figure 4 shown the experimental outcomes of accuracy testing of the model using the multiple features and from the result, it is observed that that multiple feature for classification produce better accuracy rate as compared to the single feature.

Case 2: In this case the performance parameter is expressed in terms of confusion matrix precision recall plot and ROC plot. Confusion matrix provide the performance of actual target values of test samples measured with predicted one, on the other hand precision and recall parameter provide the true positive and true negative that were present in the prediction mix.

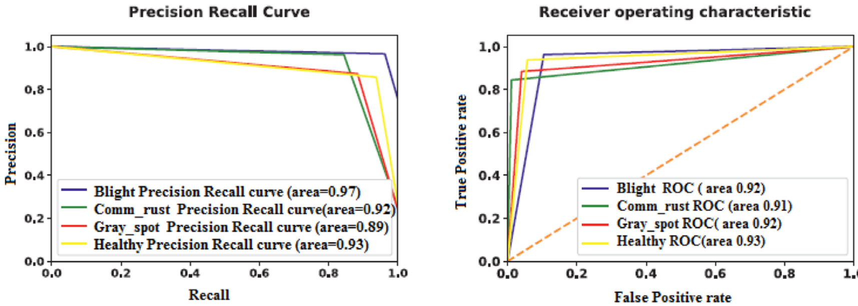


Figure 6

Precision recall and ROC curve

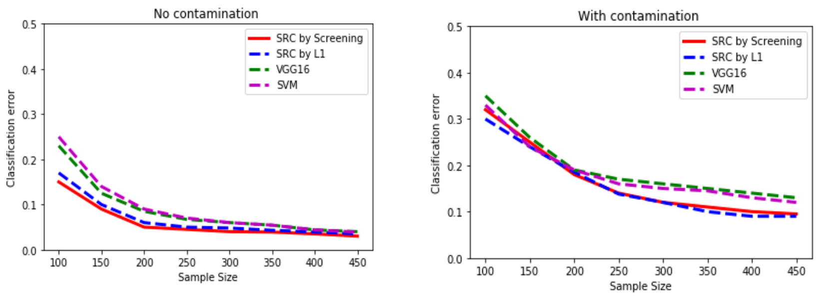
Experimental output for the confusion matrix, precision recall and ROC curve is shown in Figure 5. From the confusion matrix, we have observed that the predictive classification accuracy for the true positive rate for the different class is quite satisfactory. It has also found from the Figure 6 of precision recall and ROC curve that the performance of true positive rate under the different predictive probability for the recognition of different disease class is above 90% and this makes the model suitable for using the plant disease classification.

Case 3: This experiment is conducted to compare the performance of the proposed model against the other classifier. We compared with four existing corn leaf disease recognition method, namely, Support Vector Machines (SVM), Sparse representation classification using L1 norm and Convolutional Neural Networks (CNNs) namely VGG16 and the KNN. The effectiveness of the proposed method is measured with three performance parameter i.e accuracy and classification error. Table-1 shows the classification accuracy of our method with other state of art and the classification error rate with and without addition of random noise

Table 1

Comparison of accuracy rate between ours and other state of the art

Disease name	SVM	VGG16	KNN	SRC using L1	Ours
Blight	82.95	86.28	79.48	87.34	89.29
Common rust	73.81	84.25	77.28	85.21	87.22
Gray leaf Spot	78.25	83.22	75.29	84.34	85.43
Healthy	85.62	87.21	80.21	83.91	92.29
Average	80.15	85.24	78.06	85.20	88.55



(a) Classification error without contamination

(b) Classification error with contamination

Figure 7

Illustrating classification Error

is shown in the Figure 7. In both the case, it was observed that with the growing size of samples, classification error for the corn leaf disease is decreased.

5. Conclusions

The ease of using the screening and angle method makes the sparse representation based classification algorithm more convenient to use for classification task. Screening process is also useful in minimizing the number of attribute participation for the classification task. With the added support of multi-feature in construction of feature vector enhance the accuracy of our classification model and also enables our classifier in recognizing the disease class even under the contamination of noise in the sample space. Experimental results shows that the current model is superior in the disease identification as compared to the SRC using L1 minimization and other state of art with recognition accuracy of 88.55%. Compared with the human experts, such type of machine learning model helps agriculturist in identifying the corn plant disease in real time. Although this sparsity approach in this paper applied to corn plants, it can also be applied to other plant for the disease identification.

References

[1] KR Gavhale and G Ujwalla, "An overview of the research on crop leaves disease detection using image processing techniques," *IOSR J. Comput. Eng*, 2nd ed., vol. 16, 2014, pp. 10–16.

[2] T Rumpf, AK Mahlein and U Steiner, “Early detection and classification of crop diseases with support vector machines based on hyperspectral reflectance,” *Comput. Electron. Agric.*, 2010, vol 74, pp. 91-99.

[3] S Zhang, XWu, Zhu-Hong You and L Zhang, “Leaf image based cucumber disease recognition using sparse representation classification,” *Computers and Electronics in Agriculture,* 2017, vol. 134 , pp. 135-141.

[4] G Dhingra, V Kumar and HD Joshi, “A novel computer vision based neutrosophic approach for leaf disease identification and classification,” *Measurement 2019,* vol. 135, pp. 782-794.

[5] JGA Barbedo , LV Koenigkan and TT Santos, “Identifying multiple plant diseases using digital image processing,” *Biosyst. Eng,* 2016, vol. 147, pp. 104–116, https://doi.org/10.1016/j.biosystemseng.2016.03.012.

[6] Z Iqbal, MA Khan, M Sharif, JH shah, MH Rehman and K Javed, “An automated detection and classification of citrus plant diseases using image processing techniques: a review,” *Computers and Electronics in Agriculture,* 2018, vol. 153,pp. 12-32.

[7] S Arjunagi and NB Patil, “Texture based Leaf Disease classification using Machine Learning Techniques.,” *International Journal of Engineering and Advanced Technology (IJEAT),* 2019.

[8] G.L. Sun, X.L. Jia, T.Y Geng, “Plant diseases recognition based on image processing technology,” *Journal of Electrical and Computer Engineering,* 2018, pp. 1-7.

[9] M Osborne, B Presnell and B. Turlach, “ A new approach to variable selection in least squares problems,” *IMA Journal of Numerical Analysis*, 2000, vol. 20, pp. 389–404.

[10] J Wright, AY Yang, A Ganesh, SS Sastry and Y Ma, “Robust face recognition via sparse representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, vol 31(2), pp. 210-227.

[11] L Zhang, M Yang and X. Feng, “Sparse representation or collaborative representation: which helps face recognition?,” *International Conference on Computer Vision (ICCV),* 2011.

[12] R Rigamonti, M Brown and V. Lepetit, “ Are sparse representations really relevant for image classification?,” *In Computer Vision and Pattern Recognition (CVPR),* 2011.

[13] L Yong, S Tuo and J Shi, "Sparse solution of some special optimization problems", *Journal of Interdisciplinary Mathematics*, 2017 Vol 20, 595-602.

[14] N Akhtar, M. M. S Beg and M Hussain, "Sparse two level topic model for extraction of general summary words", *Journal of Interdisciplinary Mathematics*, 2020, Vol 23, pp 303-310.

[15] Y Zhang, Q Lv, H Lin and Ke-Xin Qi, "An adaptive pose tracking method based on sparse point clouds matching" , *Journal of Interdisciplinary Mathematics,* Volume 21, 2018 - Issue 5, pp 1115-1120.

[16] M Osborne, B Presnell and B Turlach, "A new approach to variable selection in least squares problems," *IMA Journal of Numerical Analysis,* 2000, vol. 20, pp. 389–404.

[17] S Shekhar, V Patel, N Nasrabadi and R Chellappa , " Joint sparse representation for robust multimodal biometrics recognition", *IEEE Trans.Pattern Anal. Mach. Intell.,* 2014, vol. 36, pp. 113-126.

[18] J Fan and J Lv, " Sure independence screening for ultrahigh dimensional feature space," *Journal of the Royal Statistical Society: Series B.* 2008, vol. 70, no. 5, pp. 849–911.

[19] C Shen, Li Chen, Y Dong and CE Priebe, "Sparse Representation Classification Beyond L1 Minimization and the Subspace Assumption," *IEEE Transaction on Information Theory,* 2020, vol. 6, pp. 5061–71.

[20] R Achanta, A Shaji, K Smith, A Lucchi, P Fua and S Susstrunk, " SLIC Superpixels.", *EPFL Technical Report.,* June 2010.