ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# CLASSIFICATION OF COTTON DISEASES USING CROSS INFORMATION GAIN_MINIMAL RESOURCE ALLOCATION NETWORK CLASSIFIER WITH PARTICLE SWARM OPTIMIZATION

[1] P.REVATHI , [2] M.HEMALATHA
[1]Ph.D.,Research Scholar, Department of Computer Science,
Karpagam University,Coimbatore,TamilNadu,India
[2] Assistant Professor, Department of Computer Science,
Karpagam University,Coimbatore,TamilNadu,India
E-mail: [1]kvrevathi1@gmail.com ,[2] hema.bioinf@gmail.com

## ABSTRACT

This paper is developed based on machine vision system and data mining techniques to identify the cotton leaf spot diseases. The leaves are most probably affected by the fungi, viral and bacterial diseases in the leaf spot areas which plays a vital role of crop situation. This paper clarifies six types of diseases in the cotton plant. The significance of this research work design is based on advanced computational techniques to reduce the complexity, cost and time. The proposed techniques correctly identify the diseases. In preprocessing, the image resolution value is resized to the 150* 150 pixels. The paper uses Enhanced Particle Swarm Optimization [EPSO] for feature selection to identify the affected region of a leaf. The Proposed Skew divergence (statistical method) is based on calculating the edge, color, texture variance features for analysis of the affected part of a cotton leaf. The Proposed *Cross Information Gain of Minimal Resource Allocation* (CIG-MRAN) Classifier has been used to classify the six types of diseases and increases the accuracy of the classification system.

**Keywords:** *CIG-MRAN Classifier, Skew divergence Edge, Color, Texture Variance Features, Cotton leaf, Feature Selection EPSO, Classification.*

## 1. INTRODUCTION

Classification is the most vital role in real world problem with the target of finding the fundamental patterns of the data and making use of the found patterns [1]. It is obvious that we are often flooded by data but lack of the information and clear data cannot tell us anything without processing [2]. The essential idea of classification is to learn from the given dataset in which patterns with their classes are provided; with the output of the classifier is a model or hypothesis that provides the relationship between the attributes and the class [3]. Composite classification problems are likely to present large numbers of features, many of which will be redundant for the job of classification. Hence if the number of features is very large, the classifier will take more time to classify the data set. Classification requires careful consideration when it comes to dataset before giving the data to be classified [4]. It is better to consider only necessary features rather than adding many irrelevant features since it will make the classification process much harder. It is very adequate to use methods of competence for selecting the most appropriate and informative features needed to come out with accurate and consistent result for classification problems.

Feature selection can be considered as one of the main preprocessing steps of machine learning [5]. Feature selection is different from feature extraction (or feature transformation), which creates new features by combining the original features [6]. The advantages of feature selection are manyfold. First, feature selection significantly saves the operating time of a learning procedure by eliminating irrelevant and redundant features. Second, without the intervention of irrelevant, redundant, and noisy features, learning algorithms can centrally point on most essential features of data and build simpler but This is of more precise data models. This work is based on use of particle swarm optimization(feature selection). This is because we have gained an optimal solution of the result, due to main best fitness function from matching the features of affected leaf features.

**Journal of Theoretical and Applied Information Technology**
10th February 2014. Vol. 60 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

## 2. LITERATURE REVIEW

The proposed work described the cotton leaf spot diseases using various techniques suggesting the various development ways as illustrated in this part.

Hui Li et al., (2011) The proposed work based on the Web-Based Intelligent Diagnosis System for Cotton Diseases Control system used the proposed method in a BP neural network as a decision-making system [7].

Syed A. Health et al., The proposed work discussed the automated system that can identify the pest or disease affecting a cotton leaf, boll or flower by using image analysis. In this proposed method CMYK based image cleaning technique is used to remove shadows, hands and other impurities from images [8].

Bernardes A. A. et al., (2011). This method proposed for automatic classification of cotton diseases through feature extraction of leaf symptoms from digital images. Wavelet transform energy has been used for feature extraction while SVM has been used for classification [9].

Yan Cheng Zhang et al., (2007) The proposed paper discussed the fuzzy feature selection approach - fuzzy curves and Fuzzy surfaces are to select features of cotton leaf disease [10].

Meunkaewjinda. A, et al., (2008) The proposed work organized the cotton leaf disease segmentation process using a self organizing feature map with a genetic algorithm for optimization and support vector machines for classification [11].

Gulhane. V.A et al., (2011) This proposed work described in the features that could be extracted using the Self organizing feature map with a back-propagation neural network which is used to recognize the color of the images [12].

Ajay. A, et al., (2012) This work addresses that disease analysis is possible for the cotton leaf disease. The analysis of the various diseases present on the cotton leaves can be effectively detected in the early stage before it will injure the whole crop. Initially we are able to detect three types of diseases of the cotton leaves by the methodology of Eigen feature regularization and extraction technique [13].

Qinghai He et al., (2013) This work described the three different color models for extracting the injured image from cotton leaf images developed, addressing the RGB color model, HIS color model, and YCbCr color model [14]

Rothe.P.R and R.V. Kshirsagar, (2012). The work discussed the select bacterial leaf blight, myrothecium and grey mildew as study objects. In this work, based on image enhancement techniques include histogram equalization, decorelation stretching, linear contrast stretch, intensity adjustment and filtering [15].

## 3. MATERIALS AND METHOD

This proposed work is based on cotton leaf disease classification. Initially the images are captured and preprocessed such as the resizing and removal of noise. Proposed Skew divergences statistical method is calculated by the edge, color, texture variance features. It has been used to identify the affected part of the leaf. Then, the proposed EPSO feature selection technique has been used to match the best features (appropriate features) owing the optimal solution of the affected part of the leaf. Feature extraction such as affected spot of leaf based on features such as shape, color, texture features are tested. Fusion method is applied and one of feature level fusion completes the joint feature vector (JFV) calculating, score level fusion techniques like sum rule and the product rule of edge, color, texture features variance calculating the affected part in that block. Testing image feature for the leaves are extracted and compared with the training part features, if pattern matching based on the affected region of cotton leaf disease is correct. Finally it recognizes the six types of diseases used by the proposed CIG-MRAN classifier. Then the performance evaluation of six types of diseases is obtained.

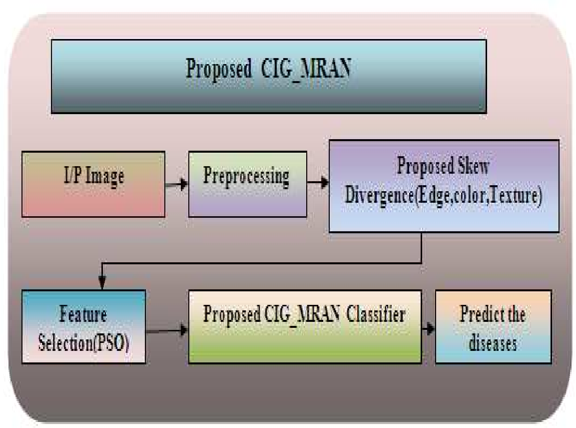


*Fig 1: The Proposed Cotton disease detection System flow Diagram*

Journal of Theoretical and Applied Information Technology
10th February 2014. Vol. 60 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*A. .PSEUDOCODE– Proposed CIG_MRAN*

1. Procedure **MRAN**
2. Begin **feature vector W**
3. Read W for training stage
4. Train Network $(X_n, Y_r)$
5. Fine tune $(X_n)$
6. Cross information gain

   $$CIG(S,A) = H(S,q) - \sum_{v \in values(A)} \frac{|\{s \in S \mid value(s,A) = v\}|}{|S|} \cdot H(\{s \in S \mid value(s,A) = v\})$$

   Where

   $$H(S,q) = -\sum_s p(s) \log q(s)$$

7. Compute $f(x)$
8. **MRAN**

   $$f(x) = v_0 + \sum_{i=1}^{m} \sum_{j=1}^{p} v_{ji} \varphi_i(X)$$

   Where

   $$\varphi_i(X) = \exp\left(-\frac{1}{\sigma_j^2} \|X - \mu_j\|^2\right)$$

   $\varphi_k(X)$ : Response of the $k^{th}$ hidden unit

   $v_{ji}$ : Connecting weight to the output unit

   $\mu_j$ : Center of the $j^{th}$ hidden neuron

   $\sigma_j$ : Width of the $j^{th}$ hidden neuron

   $v_0$ : Bias term.

9. Calculate error rate(f(x), actual value)

   9.1 Set the hidden neuron criterion $(\mu_j, \sigma_j)$

   $$\|X_n - \mu_{nr}\| > e_n \quad \text{where} \quad |e_n| = |y_n - f(x_n)| > e_{min}$$

   $X_n$ : Input data

   $y_n$ : Output data

   $\mu_{nr}$ : Center which is nearest to $x_n$

   $e_n$ : Thresholds to be selected appropriately

   $e_{min}$ : Thresholds to be selected appropriately

10. if $(|X_n - \mu_{nr}| > e_n)$

    then add a new hidden neuron

11. else adjust $(\mu_j, \sigma_j)$
12. Set pruning criteria $(V_k, \mu_{k+1}, \sigma_{k+1})$
13. $V_k + 1 = e_n$

    $\mu_{k+1} = X_n$

    $\sigma_{k+1} = k\|X_n - \mu_{nr}\|$

14. If prune criteria satisfied

    then

    prune the hidden neurons

15. Else if prune criteria not satisfied

    then

    goto step 14

16. get the next training data until the stopping criteria is met
17. End

*Fig 2: Proposed CIG_MRAN Algorithm System flow*

In general, neural network algorithms the error rate values are minimized with the help of the BPN and Feedforward neural network. But still the classification accuracy is less and the computationally expensive.

In normal MRAN based algorithm the neural network values are applied selecting only for assigning the weight values randomly and the hidden layer by combining the different input layers. It is not obvious to all hidden layer output which becomes less error value improving the accuracy of the classification result. To overcome the disadvantages of the BPN system a EMRAN Neural network based on the error rate is proposed. Some percentage only decreases. In order to reduce the error rate, here include the *cross information gain ratio* between the number of input feature vectors and the hidden layer based on the IG values adjusting the weight values. Finally minimize the error rate. It differs from normal MRAN algorithm. Cross information gain ratio (CIG (S, A)) is calculated for the input layer to another hidden layer. In this work IG values by adjusting the weight values of the input and hidden layer, the information gain ratio is satisfied.Then the pruning criteria then else the prune hidden neurons. If it doesn't satisfy then recalculate the CIG.

Journal of Theoretical and Applied Information Technology
10th February 2014. Vol. 60 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**Proposed Cig_Mran Algorithm**

Initially, simple edge detection is carried out and blocks with edge pixels inside are judged into the structural category. Then, color variance $C_v$ is calculated in the remaining blocks. Find the variance across the edge with canny edge detector and color splitting methods. Variance in the gray level region in the neighborhood of a pixel is a measure of the texture. Here we calculate the feature level fusion to combine the color and texture feature sets after normalization in order to yield a joint feature vector (JFV). Instead of using GA for selection of the best features in the feature vector. This work uses for PSO feature selection of the best parameters for both global and local features results. In this algorithm of feature $W$ as input to extract the features. Next classify the diseases. Finally display the performance of outcomes. Initially select the feature vectors $W_b$ for training. The Proposed CIG-MRAN means Cross information gain of Minimal Resources Allocation based on classification algorithm. The input file is read by the system. We define the decision function $(W)$ to extract the feature vectors W and based on decision function only, we classify the training samples into -1 or +1. If the objective or decision function f $(W)$ is greater than zero, we define the class label as +1 otherwise, it is considered as class label of -1. Finally the performance evaluation based on six types of diseases is obtained.

## 4. RESULTS AND DISCUSSION

Cotton leaf disease dataset collected from south zone Tamil Nadu at Andhiyaur district from 2012 in June month. On cotton the incidence on hybrids and Surabi varities from maximum incidence was recorded upto one week. The disease images 270 data set was collected from the field. Directly congregate the farmers and get the suggestion from them.

Cotton diseases dataset collection was done using camera mobile captured that infected cotton leaf images. In this dataset used has an input image tested in advance version 2010a Matlab tool environment. In this investigation analysis six types of diseases, such as the Bacterial Blight, Fusarium wilt, Leaf Blight, Root Rot, Micronutrient, Verticillium wilt are analysed. This process has reduced the computational complexity and was less expensive. Identification process has been used for proposed image processing with data mining techniques which are combined and analyzed classifying the diseases image of the output. Finally performance evaluation of six types of diseases is obtained.

First, test images of various cotton leaves are captured using a digital mobile phone. Then image processing techniques are applied to acquire the images to extract useful features that are necessary for further analysis. More than a few statistical techniques are applied to classify the images according to the specific problem at affected leaf spot area. Feature selection PSO has been used to analyze the best matching of affected leaf feature results owing to optimal solution. In the classification phase, Edge, color, texture feature variance corresponding feature values are stored in the image domain. Fusion method has been applied and one of the feature level fusion carryout the joint feature vector (JFV) calculation, score level fusion techniques like sum rule and the product rule of edge, color, texture features variance are calculated in the affected part of that block. Testing image feature for the leaves are extracted and compared with the training part features, if pattern matching based on the affected region of cotton leaf disease is correct.

*TABLE 1:The Actual Classes Of Six Types Of Diseases Classified CIG_MRAN Algorithm*

| Diseases Name | Precision | Sensitivity | Specificity |
|---|---|---|---|
| Bacterial blight | 1.00 | 1.00 | 1.00 |
| Fusarium wilt | 0.95 | 1.00 | 0.99 |
| Leaf blight | 1.00 | 0.95 | 1.00 |
| Root rot | 1.00 | 1.00 | 1.00 |
| Micronutrient | 0.91 | 1.00 | 0.98 |
| Verticillium wilt | 1.00 | 0.90 | 1.00 |
| Model Accuracy 0.98% | | | |

Table1:Shows the Actual Classes of six types of diseases used for CIG-MRAN classifier to classify the diseases.

*Table 2:The Performance Evaluation of Proposed CIG_MRAN Algorithm*

| Actual Classes | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1-Bacterial blight | 20.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 2-Fusarium wilt | 0.0 | 20.0 | 1.0 | 0.0 | 0.0 | 0.0 |
| 3-Leaf blight | 0.0 | 0.0 | 19.0 | 0.0 | 0.0 | 0.0 |
| 4-Root rot | 0.0 | 0.0 | 0.0 | 20.0 | 0.0 | 0.0 |
| 5-Micronutrient | 0.0 | 0.0 | 0.0 | 0.0 | 20.0 | 2.0 |
| 6-Verticillium wilt | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 18.0 |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 2: Described the performance Evaluation of proposed CIG-MRAN algorithm has been used to analyze the six types of diseases, finally out comes the true positive of diseases, likewise the precision, sensitivity, specificity the overall accuracy is 98% .

## 5. CONCLUSION

In this work, the proposed CIG_MRAN classifier uses EPSO with Skew divergence techniques (statistical analysis). The experimental results have been obtained by testing the proposed method with the cotton leaf diseases dataset. The outcomes obtained shows higher accuracy and better performance when our proposed EPSO algorithm is combined with a proposed CIG_MRAN classifier. The accuracy of the obtained system is 98%. This research work classifies six types of diseases in the cotton leaves correctly. The feature selection proves to work efficiently and hence the accuracy of the classifier improves significantly.

## REFERENCES

[1] J. Hand, H. Mannila , P. Smyth, "Principles of data mining," MIT Press, 2001.
[2] H.Hasan, N.M. Tahir, “Feature selection of breast cancer based on principal component analysis,” 6th International Colloqium on Signal Processing and Its Application, 2010.
[3] A.Ahmad, “Data transformation for decision tree ensembles,” PHD Degree Thesis, University of Manchester, 2009.
[4] Norshafarina Binti Omar ,Fatimatufaridah Binti Jusoh , Mohd Shahizan Bin Othmannd and Roliana Binti Ibrahim, “Review of Feature Selection for Solving Classification Problem,”Journal Of Research And Innovation In Information Systems.pp64-70.
[5] L. Yu And H. Liu, “Feature Selection For High-Dimensional Data: A Fast Correlation Based Filter Solution,” In Proceedings of the 20th International Conference On Machine Learning, August 2003, pp. 856– 863.
[6] F. Tan, “Improving Feature Selection Techniques For Machine Learning [Ph.D. Thesis], 2007, Computer Science Dissertations: Paper 27, Http://Digitalarchive.Gsu.Edu/Cs Diss/27.
[7] Hui Li, Ronghua Ji, Jianhua Zhang, Xue Yuan, Kaiqun Hu and Lijun Qi, “WEB-Based Intelligent Diagnosis System for Cotton Diseases Control,” IFIP Advances in Information and Communication Technology, 2011. 346.pp 483-490.
[8] Syed A. Hayat , Ahsan Abdullah, Muhammad, A.Chaudary , Yasir J. Malik and Waseem Gillani. “Automatic Cleansing and classification on Cotton leaves, bolls and flowers using CMYK color splitting”.
[9] Alexandre A. Bernardes, Jonathan G. Rogeri, Roberta B. Oliveira, Norian Marranghello, Aledir S. Pereira, AlexF.Araujo, JoãoManuelR.S. “Identification of Foliar Diseases in Cotton Crop Image Processing and Computational Vision,” Lecture Notes in Computational Vision and Biomechanics . Volume 8, 2013,pp: 67-85.
[10] Yan Cheng Zhang, Han Ping Mao, Bo Hu, Ming Xili,“Features selection of Cotton disease leaves image based on fuzzy feature selection techniques,” IEEE Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2007,pp:2-4 .
[11]Meunkaewjinda.A,P.Kumsawat,K.Attakitmongcol and A.Sirikaew,” Grape leaf disease Detection n from color imaginary using Hybrid intelligent system,” Proceedings of ECTI-CON. 2008.
[12] Gulhane.V. A & A. A. Gurjar,”Detection of Diseases on Cotton Leaves and Its Possible Diagnosis (IJIP),” 5 (5), 2011, pp: 591-598.
[13] Ajay A. Gurjar and Viraj A. Gulhane. “ Disease Detection On Cotton Leaves by Eigenfeature Regularization and Extraction Technique”, (IJECSCSE 1) 2012, pp 1-4.
[14] Qinghai He, Benxue Ma, Duanyang Qu, Qiang Zhang,Xinmin Hou, and Jing Zhao.” Cotton Pests and Diseases Detection based on Image Processing,” Telkomnika.11 (6), 2013, pp: 3445- 3450.
[15] Rothe.P.R and R.V. Kshirsagar,”A Study on the Method of Image Preprocessing for Recognition of Crop Diseases,” ICBEST and IJCA, 2012,pp:8-10.