Identification of Cotton Diseases Based on Cross Information Gain_Deep Forward Neural Network Classifier with PSO Feature Selection

P.Revathi[#1] and M.Hemalatha[*2]

[1, 2]Department of Computer Science
Karpagam University
Coimbatore-21.Tamil Nadu,India.
[1]kvrevathi1@gmail.com, [2]hema.bioinf@gmail.com

***Abstract*- This work exposes the automatic computation system to analyse the cotton leaf spot diseases. First to initialize the images from the database (Image features) that are highly related to the test image (new image), where test image is given by the user. Three features are used for matching the train image features in database images, namely color feature variance, shape and texture feature variance. These features are extracted by PSO. The feature selection method which helps to identify the injured leaf spot of cotton and at the same time improve the accuracy of the system and reduce the error rate also. These features are calculated by different techniques. The proposed Skew divergence color variance feature is calculated by color histogram and color descriptor. The shape Skew divergence feature is calculated by Sobel and Canny through the find out edge variance, edge location using Edge detection method. The skew divergence texture feature is calculated by Gober filter and texture descriptor. This investigation is based on six types of diseases like Bacterial Blight, Fusarium wilt, Leaf Blight, Root rot, Micro Nutrient, Verticillium wilt. This work utilizes these three features and combined the classifier of proposed Cross Information Gain Deep forward Neural Network (CIGDFNN) which helps to recognize and identify cotton leaf spot diseases. The forceful feature vector set is a combination of three features to obtain the higher accuracy rate and sensitivity, specificity when tested with the cotton leaf dataset.**

Keywords- Particle Swarm Optimization, Skew divergence Color Variance feature, Skew divergence Edge variance feature, Skew divergence texture variance feature, Cross Information Gain Deep forward Neural Network (CIGDFNN) classifier.

The paper is addressed with the following Sections: Section 1: Introduction,Section 2: Literature Review, Section 3: Material and Methods, Section 4: Result and Discussion, Section 5: Conclusion and Acknowledgement, Reference.

I. INTRODUCTION

The leaf spot of diseases plays a vital role in the plant growth environment. Also diseases can be easily identified with the help of the infected area of the crop [1]. Commonly the leaf will naturally show the infected part in a clear way which can be easily identified, So we can say the transformation in the crop colors are the significant aspect of the notification. When the health of the crop is in good stage then the color of the crop is different, but as soon as the crop is going to perish by some harming pathogens, the color changes automatically. Crop diseases have affected in a particular part. It may cause a reduction in productivity. The early days exposed eye observation of experts is the main approach adopted in practice for detection of plant diseases. In this investigation based on feature extraction the proposed skew divergence of edge, color, texture variance features has been used to analyze the affected part of the leaf. Feature extraction stage deals with the color, size and shape of the spot and finally classification is done by the proposed Cross information Gain Deep forward Neural Network. This technique has been used to predict the six types of diseases. Finally obtained performance evaluation results were obtained.

This work discussed is based on how we can reduce the computational complexity by enhancements of automatic detection of Crop diseases. As this can cause much damage in the agricultural fields of crops and detect the symptoms of diseases. Again we can see the potential prevention and treatment of the infected plant. In this way we need to look for the robust and less expensive and accurate method to detect and classify the crop diseases.

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. Several 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 the 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 feature level fusion carry out the joint feature vector (JFV) calculate, score level fusion techniques like sum rule and the product rule of edge, color, texture features variance calculate 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.

II. BACKGROUND STUDIES

The Proposed work describes the diagnosis of cotton leaf diseases using various methods suggesting the various implementation ways as illustrated below.

Hui Li et al., (2011) The work based on the Web-Based Intelligent Diagnosis System for Cotton Disease Control system used the proposed method in a BP neural network as a decision-making system. A research scheme was designed for the system test, in which 80 samples, including 8 main species of diseases, and 10 samples in each sort were included. The result showed the rate of correctness. The system could identify the symptom was 89.5% in average, and the average running time for a diagnosis was 900ms [2].

Syed A. Health et al., The work discussed the automated system that can identify the pest or disease affecting a cotton leaf, boll or flower by using image analysis. Here proposed method is CMYK based image cleaning technique to remove shadows, hands and other impurities from images. The outcomes are tested over a database consisting of 600 images. After that the images are cleaned and three techniques applied to classify the diseases [3].

Bernardes A. A. et al., (2011). This method 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. The image set of supposedly adulterated leaves was classified within one of the four other sub-classes, namely: MA, RA, AS, and NONE. Finally obtained results were: 96.2% accuracy for the SA class, 97.1% accuracy for the MA class, 80% accuracy for the RA class, and 71.4% accuracy for the AS class [4].

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 disease leaves. In classifying to get the best information for diagnosing and identifying, a subset of independent significant features is identified exploiting the fuzzy feature selection approach [5].

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. Finally out comes to segment image is filtered by Gabor wavelet which allows the system to analyze leaf disease color features more efficiently [6].

Gulhane. V.A et al., (2011) This work described the features that could be extracted intentionlly using the Self organizing feature map with a back-propagation neural network used to recognize the color of the images. Finally classification of the cotton diseases is done [7].

Ajay. A, et al., (2012) This work addresses the disease analysis 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 be able to detect three types of diseases of the cotton leaves by the methodology of Eigen feature regularization and extraction technique. In this method 90% of detection of Red spot is obtained, (I.e.) fungal disease is detected. It is most dangerous disease, it can highly affect the productivity of the cotton crop to a great extent. If it is detected in the early stage. We can say that, we are able to get better productivity [8].

Qinghai He et al., (2013) This work described the three different color models for extracting the injured image from cotton leaf. Images were developed, address the RGB color model, HIS color model, and YCbCr color model. The ratio of damage (γ) was chosen as feature to measure the degree of damage which is caused by diseases or pests. This work shows the comparison of the results obtained by implementing different color models. The comparison of outcomes shows good accuracy in both color models and YCbCr color space is considered as the best color model for extracting the injured leaf images [9].

III. MATERIAL AND METHODS

This proposed work is based on cotton leaf diseases. Classification has been used for the proposed CIG-DFNN classifier and proposed skew divergences. Edge, color, texture variance feature techniques are used to extract the affected spot of leaf features like the shape, color, texture features methods to investigate. Finally obtained the performance evaluation of six types of diseases accuracy, sensitivity, specificity of this system is obtained .

A. Proposed CIG-DFNN Algorithm

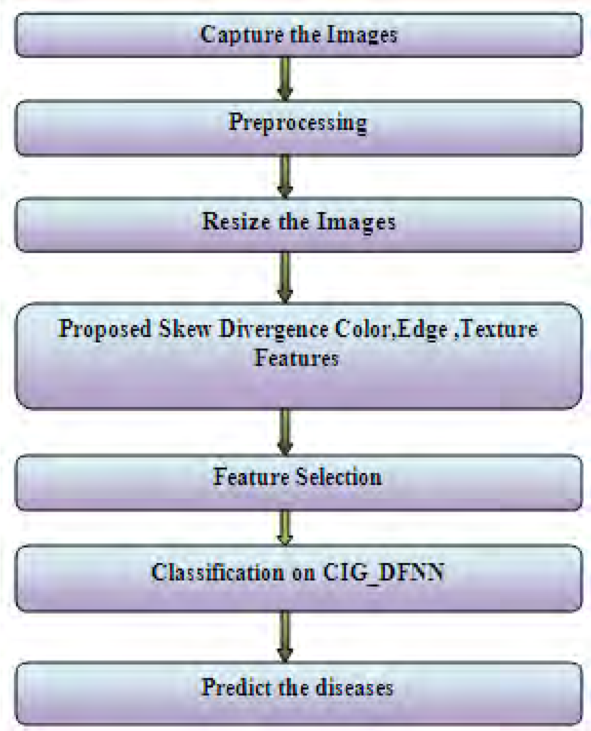

Fig:1 The Proposed System Architecture Diagram

Configuration

The architecture has followed us:

- Input : $1 < i < n$ where i= 1 to 150,
- One hidden layer : $1 < j < p$ where j= 1 to 20,
- Output : $1 < k < m$ where k= 1 to 11.

According to DFNN theory weight ranges from $\left[\frac{-1}{S_1} \text{ to } \frac{1}{S_1}\right]$, where S_1 is number of hidden layers in the network.

B. Proposed CIG- DFNN Algorithm

1. In the first phase, greedily train subsets of the parameters of the network using a layer-wise and unsupervised learning criterion, by repeating the following steps for each layer ($i \in \{1, \dots, l\}$)

Until a stopping criteria is met, iterate through training by

(a) Mapping input training sample $\mathbf{x}_t$ to representation $\widehat{\mathbf{h}^{i-1}}(\mathbf{x}_t)$ (if $i > 1$) and hidden representation $\widehat{\mathbf{h}^{i-1}}(\mathbf{x}_t)$.

(b) Updating parameters $\mathbf{b}^{i-1}$, $\mathbf{b}^{i}$ and $\mathbf{W}^{i}$ of layer is using some unsupervised learning algorithm.

Also, initialize (e.g., randomly) the output layer parameters $\mathbf{b}^{l+1}$, $\mathbf{W}^{l+1}$.

2. In the second and final phase, fine-tune all the parameters θ of the network using backpropagation and gradient descent on a global supervised cost function $C(\mathbf{x}_t, y_t, \theta)$, with input $\mathbf{x}_t$ and label y_t, that is trying to make steps in the direction $E\left[\frac{\partial C(x_t, y_t, \theta)}{\partial \theta}\right]$.

In general the deep forward neural network initialization of the parameters randomly to perform the classification becomes poor solution. Standard gradient descent from random initialization is done, so poor solution with deep neural networks. To overcome this problem in DFNN, first observe the influence of the nonlinear activation functions. Find that the logistic sigmoid activation is unsuited for deep networks with random initialization because of its mean value, which can drive especially the top hidden layer into saturation.

In this algorithm the first phase that initializes the parameters of the whole network will ultimately have an impact on the solution found in the second phase (the fine-tuning phase). In the first phase, initializations of the parameter are measured by the gradient descent error rate and also extend the work to which configuration of the neural network parameters can be different from the initial configuration. Hence, similarly to using a regularization term on the parameters of the model that constrains them to be close to a particular value (e.g., 0 for weight decay), the first phase here will ensure that the parameter solution for each layer found by fine-tuning will not be far from the solution found by the unsupervised learning algorithm. In addition, the non-convexity of the supervised training criterion means that the choice of initial parameter values can greatly influence the quality of the solution obtained by gradient descent with error gradient function.

In DFNN also we perform CIG rate to diminish the error rate of the hidden layer and input vector layer. Reduce the error rate calculate the CIG for the hidden layer and the input vector layer.

C. Pseudo code of CIG_DFNN

Procedure call train()

Read input feature vector w

Set weight within $\left[\frac{-1}{S_1} \text{ to } \frac{1}{S_1}\right]$

For($i = 1 ; i > l; i++$)

do

mapping x_t to representation $\widehat{h^{l-1}}(x_t)$

Fine tune (x_t)

Cross – information gain

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

Where

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

Compute activation function

$$f(x) = \frac{1}{1 + e^{-x}}$$

Calculate error (d_{ij}, a_{ij})

$$E_i = \sum_{i=1}^{l} (d_{ij} - a_{ij})^2$$

Update b^{l-1}, b^l and W^l

Compute cost function $C(x_t, y_t, \theta)$

$$E\left[\frac{\partial C(x_t, y_t, \theta)}{\partial \theta}\right]$$

If target is met

Then

End

Else goto step 2 until the target is met

Fig 2: Proposed CIG-DFNN Algorithm flow

D. Proposed CIG_DFNN Algorithm

First, simple edge detection is carried out and blocks with edge pixels inside are judged into the structural category. Then, color variance 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 in a 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 has been used for PSO feature selection of the best parameters for both global and local features results. In this algorithm of feature as input to extract the features. Next classify the diseases. Finally display the performance of outcomes. Initially select the feature vectors for training. The proposed CIG-DFNN classification algorithm the input file is read by the system. We define the decision function ( ) 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 ( ) 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.

IV. RESULTS AND DISCUSSION

Cotton leaf disease datasets 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 during upto one week. The disease images 270 data sets were collected from the field. Directly congregate the farmers and get the suggestion from them.

In this investigation dataset collection was done using camera mobile (Nokia) captured the infected cotton leaf images. In this dataset used has an input image and tested in advance version2010Matlab tool environments. In this investigation about six types of diseases like the Bacterial Blight, Fusarium wilt, Leaf Blight, Root Rot, Micronutrient, Verticillium wilt are analysed. This process has reduced the computational complexity and is 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 accuracy of the diseases is obtained.

TABLE : 1
The Actual Classes of six types of diseases

Actual Classes						
Predicted the Class	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	3.0	0.0	0.0	0.0
3 -Leaf blight	0.0	0.0	17.0	1.0	0.0	0.0
4 -Root rot	0.0	0.0	0.0	19.0	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

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

TABLE : 2
The Performance Evaluation of proposed CIG-DFNN Algorithm

Diseases Name	Precision	Sensitivity	Specificity
Bacterial blight	1.00	1.00	1.00
Fusarium wilt	0.87	1.00	0.97
Leaf blight	0.94	0.85	0.99
Root rot	1.00	0.95	1.00
Micronutrient	0.91	1.00	0.98
Verticillium wilt	1.00	0.90	1.00
Model Accuracy 0.95%			

Table 2: Described the performance Evaluation of proposed CIG-DFNN algorithm used to analyze the six types of diseases. Finally out comes is the true positive of diseases, likewise the precision, sensitivity, specificity the overall accuracy is 95% .

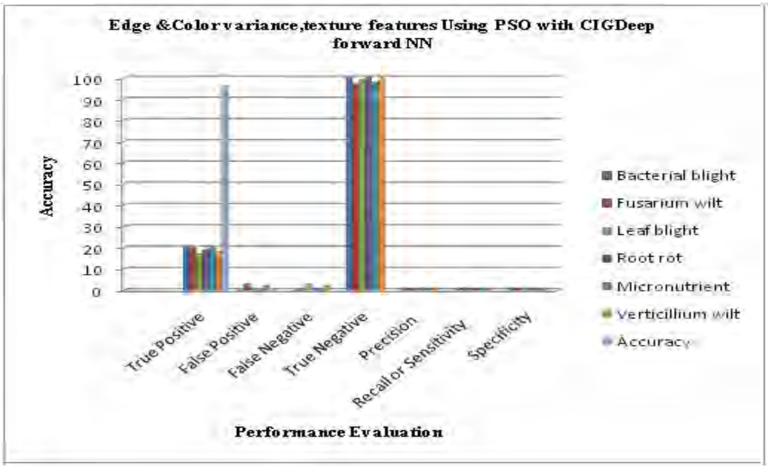

Fig 3: The Performance Evaluation of six types of diseases

Figure3:Represents the performance evaluation of six types of diseases used for proposed CIG-DFNN algorithm by correctly classifying the six types of diseases. Lastly outcome shows the overall accuracy of this system.

V. CONCLUSION

The proposed Skew divergence color variance feature is calculated by color histogram and color descriptor. The shape feature extracted used for Skew divergence feature is calculated by Sobel and Canny through the find out edge variance, edge location of Edge detection method. This has been used for the skew divergence texture feature calculated by Gober filter and texture descriptor. In this investigation the six types of diseases like Bacterial Blight , Fusariumwilt , Leaf Blight, Root rot, Micro Nutrient, Verticilium Wilt are classified. We utilizing these three features are combined with the classifier of proposed Cross Information Gain Deep forward Neural Network (CIG-DFNN) , which are the most appropriate features for training images, which one has been utilized for recognizing the cotton leaf spot diseases. Performance evaluation of proposed CIG-DFNN system based on the overall accuracy is 95%. This method helps to correctly classify the diseases. It has been used for accurate classification of the diseases by increasing the accuracy rate and diminishing the error rate.

ACKNOWLEDGMENT

I would like to express my special thanks of gratitude to Dr.S.Nakkeeran Savage Perumal, Associate Professor, (Pathology) Department of Plant Pathology of Tamil Nadu Agricultural University. I am thankful to Karpagam University for permitting me to do my Research work. My hearty thanks them to for their overall support.

REFERENCES

[1] Mrunalini R. Badnakhe and Prashant R. Deshmukh..”Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering”.2(3): 449-452. 2012.

[2] 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, 346: 483-490. 2011

[3] 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”.

[4] Bernardes.A.A, J.G.Rogeri, N.Marranghello, A. S. Pereira, A.F. Araujo and João Manuel R. S. Tavares. “Identification of Foliar Diseases in Cotton Crop”. SP, Brazil.

[5] 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, 2-4.2007.

[6] 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

[7] Gulhane.V. A & A. A. Gurjar, “Detection of Diseases on Cotton Leaves and Its Possible Diagnosis” (IJIP), 5 (5): 591-598. 2011

[8] Ajay A. Gurjar and Viraj A. Gulhane. “ Disease Detection On Cotton Leaves by Eigenfeature Regularization and Extraction Technique”. IJECSCSE.1 (1): 1-4.2012.

[9] 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): 3445- 3450.2013.