Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 44 (2023), No. 6, pp. 1127–1138
DOI : 10.47974/JIOS-1451

Automatic feature selection using gray-level co-occurrence matrix and detection of mustard plant disease using feed forward neural networks

Anita Sharma *
Chirag Sharma †
Department of Computer Science & Engineering
Lovely Professional University
Phagwara
Punjab
India

Vikrant Sharma §
Satvik Vats ‡
Department of Computer Science & Engineering
Graphic Era Hill University
Dehradun
Uttarakhand
India

Abstract

The Mustard crop is one of the important oil seed crops, but the crop suffers from various diseases causing loss in its yield. Diseases can be detected by the naked eyes by experts, but this method is time consuming when applied to large farms. The recent trends in computer vision with the help of various techniques show considerable results in detecting diseases in various crops. As a result, this paper presents a system for feature selection and classification of leaves for detecting diseases at an early stage to prevent the plants from further damage. For image segmentation OTSU technique is applied, GLCM feature selection is used for extracting the features and at last the feed forward ANN classifier is used to classify the infected mustard leaves. The performance of ANN is tested with various other classifiers like RFGBM, J48, Naïve Bayes, Decision Tree, KNN, SVM and RF. Classification

* *E-mail:* **anita.sharma071992@gmail.com (Corresponding Author)**
† *E-mail:* **chiragsharma1510@gmail.com**
§ *E-mail:* **vsharma@gehu.ac.in**
‡ *E-mail:* **svats@gehu.ac.in**

parameters, such as Precision, Recall, F-measure, Error, and Accuracy is used as a measure for effectiveness of the feed forward ANN achieving an accuracy of 86.78%.

Subject Classification: *68T07.*

Keywords: *Machine learning, Deep learning, Disease detection, SVM, ANN, GLCM.*

1. Introduction

Diseases in plants are one of the most significant challenges affecting the quality and quantity of agricultural produce [1]. Nowadays, one of the most essential challenges in agriculture is to detect and classify plant lesions to improve the quality of plant production for economic growth. It is reported that in India the most common disease of Indian mustard is Alternaria blight which affects the crop adversely causing nearly 47% losses in the yield [2]. Monitoring plant health and Pathogen detection at an early stage are essential to reduce spreading of disease and facilitate effective and efficient management practices [3]. Technology has a vital role to play in detection of these diseases in horticulture, floriculture etc.

By 2050, the world's [4] population is expected to reach to nearly 10 billion, which calls for increase in agricultural development somewhere in the range of 50% in contrast to 2013[5]. Between 2005/07 and 2050 every individual would require raising the overall food production by around 70 percent [6]. However, projected amount of crop production needs to be ensured by the farmers across the world by increasing the area of agricultural land or by improving the productivity on available agricultural area by adopting smart farming techniques [7]. This paper is organized into 4 sections as follows: in section 1 the disease detection process in plants is discussed briefly. The phases of the process with various techniques used in that, their pros and cons are reviewed. Section 2 contains the related work in computer vision in plant disease detection. Section 3 gives the results and discussions with respect to all the classifiers at last Section 4 considers the recommendations, challenges, limitations.

1.1 *Disease Detection and Classification Process*

Disease detection is nowadays scan also be carried out with the help of image processing techniques [8][9]. A network of Sensor Nodes (SNs) [23][25] deployed within the agricultural fields are used to collect the data in regular intervals that can be used to analyze the crop health and these systems are also managed using some softwares[28][10][11][12]. The

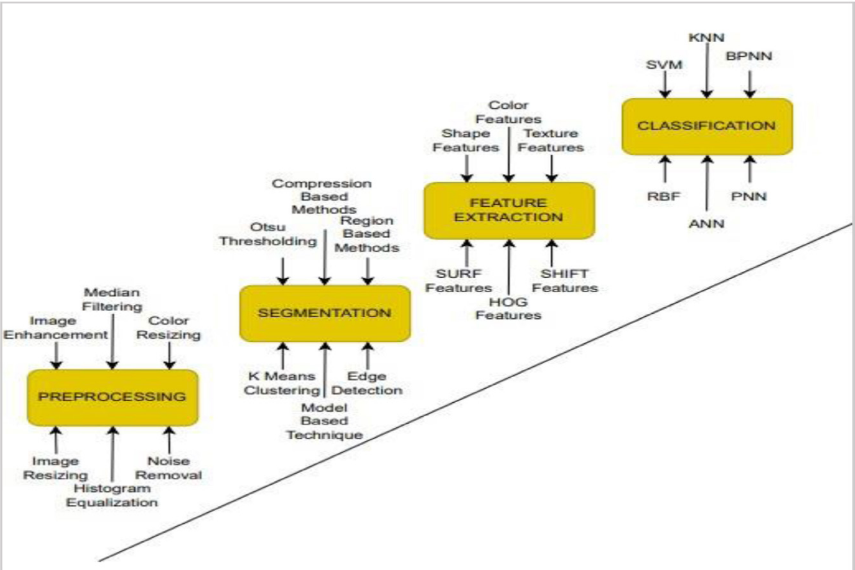


Figure 1

Disease detection techniques in plants

techniques consist of four major steps which includes the preprocessing of data, disease detection followed by the feature extraction and classification [13][14][15][16].

Figure 1 shows the step-by-step process of disease detection in plants. The process is carried out and further the training and testing data is divided from the dataset [17][18][29]. At the end the validations are done with the comparison.

2. Literature Review

Many researchers now have explored the area of disease detection in plants using various techniques like machine learning and deep learning[26]. Jitesh P Shah et al. carried out a survey on 19 research papers including the present research carried out on rice diseases, other plants, and fruits. The survey done by the authors includes criteria like dataset, no. of classes, techniques used for preprocessing, segmentation techniques, feature extraction, edge detection, color space and background of images. A comparative study on various segmentation and classification techniques applied in rice disease detection is carried out. It was observed that the SVM classifier is suitable for high dimensional data whereas neural networks can be applied where a large data set is to be trained. The least used classifier is the rule-based classifier as it is based on production rules. A technique for detection of diseases in rice crop is proposed in this work. Mainly the researchers worked on three diseases i.e., bacterial leaf blight, brown spot, and leaf smut of rice crop.X.E. Pantazi et.al [20] worked on an automated technique for identifying the crop disease on various leaf samples by using Local Binary Patterns (LBPs) for extracting the features and for classification single Class Classification classifier is used. The proposed methodology uses a committed One Class Classifier for individual plant conditions like health, powdery mildew and black rot. Vine Leaves were used for the successful application of the algorithms. The results are said to be true for most of the cases. As mentioned, 44 of the 46 tests plants were classified correctly, 95% accuracy was achieved. The research combines advanced Image processing techniques with Artificial Intelligence (AI) like GrabCut Algorithm to classify the plan conditions.Aakanksha Rastogi et.al [21] proposed a method for identifying the leaf diseases and further grading them by using image processing and machine vision. The process is carried out in two phases. Firstly, pre-processing is done for extracting the leaf from the noises and then Artificial Neural Network based training is done to train the network based on features. The entire process proposed uses Euclidean distance technique, K-Means based segmentation of defected areas, Fuzzy logic system for grading and finally ANN based classification of disease.Raciel Year Toledo et.al [19] proposed Automatic Detection of Nutrition by analyzing the shape and texture of coffee leaves. The research was focused on identification of nutritional deficiency of Boron, Iron, Calcium and Potassium in coffee tea tree leaves. The algorithm used for segmentation was Otsu's Method. To extract the characteristics of the coffee tea tree leaves, Blurred Shape Model (BSM) and Gray-Level Co-occurrence descriptors were applied onto them. To infer the results of deficiency, obtained images are used for training KNN, NN and Naïve Bayes Classifiers. Better accuracy was produced by the proposed algorithm associated to Boron (B) and Iron (Fe) deficiencies.

3. Proposed Work

This paper presents a comparison of AI models on the plant dataset. Various classifiers are used for the comparisons such as RFGBM, J48, Naïve Bayes, Decision Tree, KNN, SVM, ANN and RF.

4. Result and analysis

Among all the classifiers, ANN has the highest performance than all other classifiers. Table 1 shows the comparison of proposed algorithm with other algorithms. The results of the comparisons are also presented in the form of bar graphs from Figure 2.

4.1 *Dataset used*

All images are taken are of 512*512 resolution using a mobile device[22][27] as this makes the dataset a bit less complex and offers low computation scaling the dataset uniformly for accurate extraction of features. The images can also be taken using the aerial view and then the data is sent to the device using wireless services[24]. The images are smoothened using the median filter. In the next step the OTSU technique is applied to segregate the foreground and the background based on the image characteristics. At last, the classifier is used to detect the unhealthy and healthy portion of leaves.

4.2 *Parameters Taken*

In this work a total of 5 parameters are taken. These are: Sensitivity, Accuracy, Precision, F- score and Recall.

Confusion matrix: While dealing with classification problems confusion matrix are of great use as the percentage of correct classified samples and misclassified samples are the measures of the efficiency of the classification model. This n*n matrix evaluates the efficiency of the classifier used for any problem.

Accuracy: It is the measure of predictions that the model got right. It is defined as the ratio of the total observations predicted correctly to the total number of observations predicted. The higher the accuracy of the model, the higher the efficiency of the classifier.

$$\text{accuracy} = \frac{(\text{true positive} + \text{true negative})}{\text{true positive} + \text{true negative} + \text{false positive} + \text{false negative}} \quad (1)$$

Precision: It is the measure of the correct positive predictions made by the classifier to the total positive results predicted by the classifier. The higher the precision of the model, higher is the reliability of the mode in predicting the samples as positive.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \quad (2)$$

Recall: It gives the measure of proportion of actual positives that were predicted right. Higher value of recall indicates that the samples were detected more positively.

$$\text{recall} = \frac{\text{True posistive}}{\text{True positive+False negative}} \tag{3}$$

F-Score: The Harmonic Mean of precision and recall is the F1 Score. F1 Score has a range of [0, 1]. It indicates the exactness and robustness of the classifier.

$$\text{fscore} = \frac{\text{True posistive}}{\text{True positive+(False positive+False negative)}} \tag{4}$$

Specificity: This gives the measure of the true negatives that are correctly identified by the model.

$$\text{specificity} = \frac{\text{True negative}}{\text{True negative+False positive}} \tag{5}$$

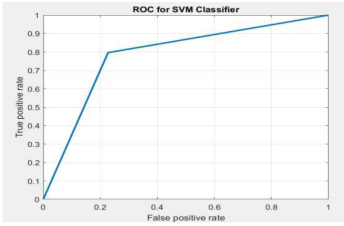
4.3 *Comparison Table*

The comparison of the classifiers is shown in Table 1. Among all the classifiers, ANN gives the best results in terms of accuracy, F-score, specificity, and precision.

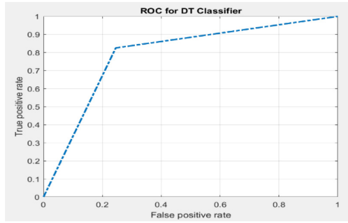
Table 1

Comparison of various existing classifiers based on some parameters

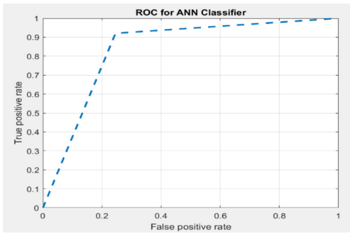
Classifier Name	Accuracy	Precision	Recall	F-Score	Specificity
DT	79.209	73.894	73.612	78.567	78.095
J48	78.314	77.564	77.111	78.267	78.953
RF	83.561	82.341	82.212	83.452	82.223
ANN	86.782	84.121	84.674	86.906	85.123
SVM	82.345	83.897	83.906	83.126	84.323
KNN	79.947	77.764	77.514	75.908	76.987
Naïve Bayes	76.569	76.452	75.589	75.541	76.906
Logistic Regression	77.906	76.908	75.08	77.011	75.987
RFGBM	85.456	83.435	83.231	85.633	84.675



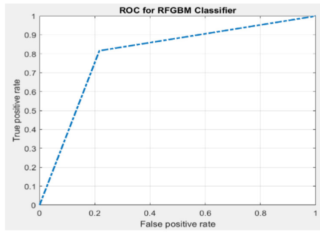
(a)



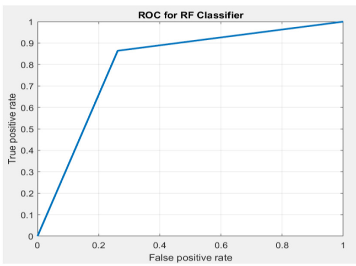
(b)



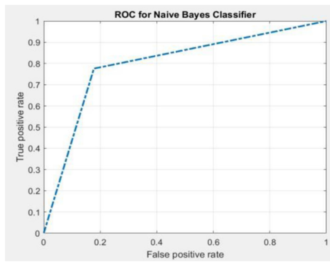
(c)



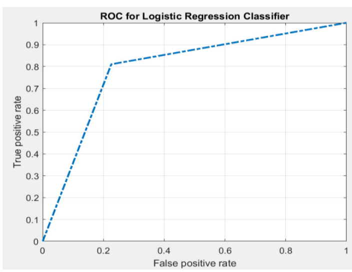
(d)



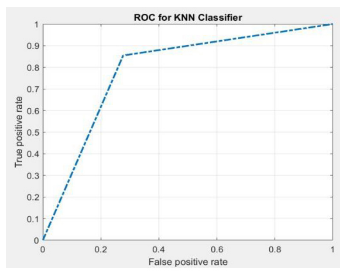
(e)



(f)

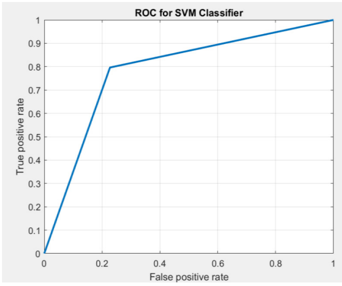


(g)

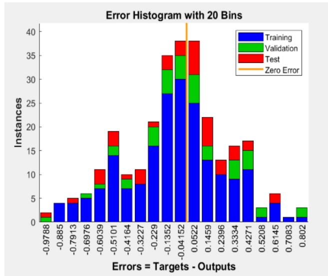


(h)

Contd...



(i)



(j)

Figure 2:

(a) ROC curve for SVM; (b) ROC curve for DT; (c) ROC curve for ANN; (d) ROC curve for RFGBM; (e)ROC curve for RF; (f) ROC curve for Naïve Bayes; (g) ROC curve for LR; (h) ROC curve for KNN; (i) ROC curve for SVM; (j) Histogram with 20 bins

Figure 2 (a) to (i) shows the ROC curve for all the. The ROC curve for ANN among all the classifiers shows the least variation form the threshold i.e from true positive rate and false positive rate.

All the comparisons in Figure 3 show that the ANN performs the best when the mustard leaves are given as input for the classification of healthy and unhealthy leaves. ANN outpowers all other classifiers with an accuracy of 86.78%, precision of 84.121%, recall of 84.674%,F-score of 86.906% and specificity of 85.123%. The same has been presented in the form of bar graphs. All the classifier's validation parameters are presented in Figure 3 which shows ANN performs best among all others.

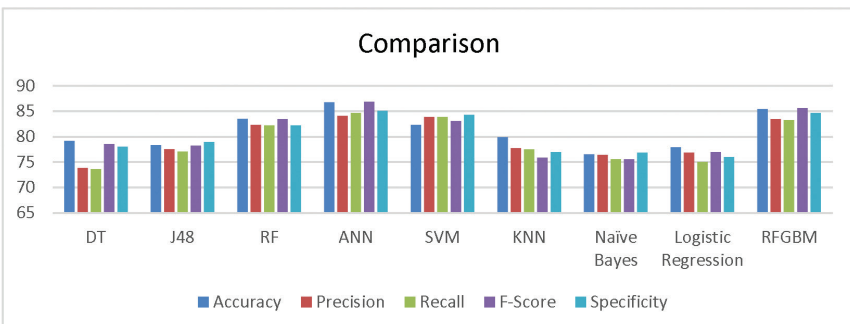


Figure 3

Comparison of RFGBM, J48, Naïve Bayes, Decision Tree, KNN, SVM, ANN and RF

5. Conclusion and future work

In this paper various classifiers were used and after comparing the results it was found that the performance of ANN is better with the accuracy of 86.726 among all other on the mustard leaf images. As ANN works with feed forward principle. More improvements can be made in each phase of disease identification as refinement in segmentation is needed to get more accurate results. Also, some metaheuristic techniques can be used to improve the efficiency of the classifiers by hyperparameter tunning.

Conflicts of Interest

The authors declare that they have no competing interests.

References

[1] "The state of food and agriculture," "Crop description : Global scenario :," pp. 1–19 (2009).

[2] G. Polder, G. W. A. M. van der Heijden, J. van Doorn, and T. A. H. M. C. Baltissen, "Automatic detection of tulip breaking virus (TBV) intulip fields using machine vision," *Biosyst. Eng.*, vol. 117, no. C, pp. 35–42 (2014), doi: 10.1016/j.biosystemseng.2013.05.010.

[3] T. U. Rehman, M. S. Mahmud, Y. K. Chang, J. Jin, and J. Shin, "Current and future applications of statistical machine learning algorithms for agricultural machine vision systems," *Comput. Electron. Agric.*, vol. 156, no. December 2018, pp. 585–605 (2019), doi: 10.1016/j.compag.2018.12.006.

[4] B. S. Bhandari, J. P. Rai, N. Singh, D. K. Gaur, and U. Rani, "Agricultural Statistics at a Glance," pp. 26–36 (2013), [Online]. Available: http://agricoop.nic.in/agristatistics.htm.

[5] M. Bhange and H. A. Hingoliwala, "Smart Farming: Pomegranate Disease Detection Using Image Processing," *Procedia Comput. Sci.*, vol. 58, pp. 280–288 (2015), doi: 10.1016/j.procs.2015.08.022.

[6] S. S. Chouhan, U. P. Singh, and S. Jain, "Applications of Computer Vision in Plant Pathology: A Survey," *Arch. Comput. Methods Eng.*, vol. 27, no. 2, pp. 611–632 (2020), doi: 10.1007/s11831-019-09324-0.

[7] A. Paul, S. Ghosh, A. K. Das, S. Goswami, S. Das Choudhury, and S. Sen, A Review on Agricultural Advancement Based on Computer Vision and Machine Learning, vol. 937. Springer Singapore (2020).

[8] G. Sambasivam and G. D. Opiyo, "A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks," *Egypt. Informatics J.*, vol. 22, no. 1, pp. 27–34 (2020), doi: 10.1016/j.eij.2020.02.007.

[9] M. A. F. Azlah, L. S. Chua, F. R. Rahmad, F. I. Abdullah, and S. R. W. Alwi, "Review on techniques for plant leaf classification and recognition," *Computers,* vol. 8, no. 4 (2019), doi: 10.3390/computers8040077.

[10] V. Sharma et al., "OGAS: Omni-directional Glider Assisted Scheme for autonomous deployment of sensor nodes in open area wireless sensor network," ISA Trans. (Aug. 2022), doi: 10.1016/j.isatra.2022.08.001.

[11] V. Sharma, R. B. Patel, H. S. Bhadauria, and D. Prasad, "Policy for planned placement of sensor nodes in large scale wireless sensor network," KSII Trans. *Internet Inf. Syst*., vol. 10, no. 7, pp. 3213–3230 (2016).

[12] V. Sharma, R. B. Patel, H. S. Bhadauria, and D. Prasad, "Deployment schemes in wireless sensor network to achieve blanket coverage in large-scale open area: A review," *Egypt. Informatics J.,* vol. 17, no. 1, pp. 45–56 (2016).

[13] S. Vats, S. Singh, G. Kala, R. Tarar, and S. Dhawan, "iDoc-X: An artificial intelligence model for tuberculosis diagnosis and localization," *J. Discret. Math. Sci. Cryptogr.,* vol. 24, no. 5, pp. 1257–1272 (2021).

[14] S. Vats, B. B. Sagar, K. Singh, A. Ahmadian, and B. A. Pansera, "Performance evaluation of an independent time optimized infrastructure for big data analytics that maintains symmetry," *Symmetry* (Basel)., vol. 12, no. 8 (2020), doi: 10.3390/SYM12081274.

[15] A.Singh, SVN Sreenivasu ,U.S.B. K. Mahalaxmi , H.Sharma , D. Patil, and E. Asenso, "Hybrid Feature-Based Disease Detection in Plant Leaf Using Convolutional Neural Network, Bayesian Optimized SVM, and Random Forest Classifier," *Journal of Food Quality*, vol. 2022 (2022).

[16] P. Sharma, P. Hans, and S. C. Gupta, "Classification of plant leaf diseases using machine learning and image preprocessing techniques,"

Proc. Conflu. 2020 - 10th Int. Conf. Cloud Comput. Data Sci. Eng., pp. 480–484 (2020), doi: 10.1109/Confluence47617.2020.9057889.

[17] A. Cruz et al., "Detection of grapevine yellows symptoms in Vitis vinifera L. with artificial intelligence," Comput. Electron. Agric., vol. 157, no. December 2018, pp. 63–76 (2019), doi: 10.1016/j.compag.2018.12.028.

[18] W. jie Liang, H. Zhang, G. feng Zhang, and H. xin Cao, "Rice Blast Disease Recognition Using a Deep Convolutional Neural Network," Sci. Rep., vol. 9, no. 1, pp. 1–10 (2019), doi: 10.1038/s41598-019-38966-0.

[19] M. Vassallo-Barco, L. Vives-Garnique, V. Tuesta-Monteza, H. I. Mejía-Cabrera, and R. Y. Toledo, "Automatic detection of nutritional deficiencies in coffee tree leaves through shape and texture descriptors," *J. Digit. Inf. Manag*., vol. 15, no. 1, pp. 7–18 (2017).

[20] S. Kumar, K. Prasad, A. Srilekha, T. Suman, B. P. Rao, and J. N. Vamshi Krishna, "Leaf disease detection and classification based on machine learning," Proc. Int. Conf. Smart Technol. Comput. Electr. Electron. ICSTCEE 2020, pp. 361–365 (2020), doi: 10.1109/ICSTCEE49637.2020.9277379.

[21] X. E. Pantazi, D. Moshou, and A. A. Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers," Comput. Electron. Agric., vol. 156, no. November 2018, pp. 96–104 (2019), doi: 10.1016/j.compag.2018.11.005.

[22] Fragkakis, Michail, Nikolaos Alexandris, and Fotis Georgiakodis. "A survey on the trust and security models of mobile agent platforms." *Journal of Discrete Mathematical Sciences and Cryptography* vol.15, no. 1, pp. 31-47 (2012).

[23] Joshi, Ashish, and Amar Kumar Mohapatra. "Authentication protocols for wireless body area network with key management approach." *Journal of Discrete Mathematical Sciences and Cryptography,* vol. 22, no. 2, pp. 219-240 (2019).

[24] Guo, Fuchun, Yi Mu, and Zhide Chen. "Multi-identity management for identity-based cryptography." *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 11, no. 6, pp. 639-672 (2008).

[25] Khan, Tayyab, and Karan Singh. "Resource management based secure trust model for WSN." *Journal of Discrete Mathematical Sciences and Cryptography* 22.8 : 1453-1462 (2019).

[26] Bathla, Gourav, Himanshu Aggarwal, and Rinkle Rani. "Using deep learning to improve recommendation with direct and indirect social trust." *Journal of Statistics and Management Systems* 22.4 : 665-677 (2019).

[27] Ali, Syed Salman, and B. V. V. S. Prasad. "Secure and energy aware routing protocol (SEARP) based on trust-factor in Mobile Ad-Hoc networks." *Journal of Statistics and Management Systems* 20.4 : 543-551 (2017).

[28] Harisalo, Risto, and Jari Stenvall. "Trust management in the Finnish ministries: Evaluation of management systems." *International Journal of Public Administration* 26.8-9 : 915-940 (2003).

[29] Li, S.Cheng, J.Cui, C.Li, Z.Li, C.Zhou and C.Lv, "High-Performance Plant Pest and Disease Detection Based on Model Ensemble with Inception Module and Cluster Algorithm," Plants, vol. 12 (2023), https://doi.org/10.3390/plants12010200.