Adaptive Based Clustering with R-CNN (ACR-CNN) for Plant Disease Detection

[1]Dr.S.Nirmala Devi,
Associate Professor,
Guru Nanak College,
Velachery, Chennai-600042

[2]Dr.P.V.Kumaraguru,
Associate Professor & CEO
Guru Nanak College,
Velachery, Chennai-600042

Abstract

Detecting plant disease in an earlier stage is one of the tasks where these can lead to heavy loss in cultivation. Artificial intelligence is the intelligence of machines where human involvement and activities become very less such that manpower can be saved for important terms such as the development of agriculture and farming. In such a way the research challenges are more likely to relate to the detection and diagnosis in plant diseases based on spots affected by many viruses that can be handled using deep learning techniques. This work focuses on providing solutions by classifying, segmenting input from the images. By applying Convolutional neural network the comparison analysis with the enhanced model has been introduced. The implementation of an Adaptive based clustering with R-CNN (ACR-CNN) is used to enhance the clarity of plant images in various shapes. When there is the identification of affected portions such as leaves, stems by automation and monitoring plants can have high development. By segmentation of affected areas from the kaggle dataset, the proposed algorithm can maintain the threshold of complete appearance in detecting using the proposed framework. Also the implemented results show the detection gives better results and evaluation of performance is also higher in comparison with existing algorithms.

Keywords: Artificial intelligence, Clustering, Neural network, Plant Disease, machine learning.

1. Introduction

The economy of India depends mostly on agricultural activities as agriculture contributes majorly to the export of goods. This makes the need for the production of crops that are free from diseases thereby increasing the economy of our country. The change in climate is one of the factors that govern the disease protection in plants thereby altering the development of pathogen rates. Resistance of the host is also modified leading to physiological changes. Complicated situations arise as these diseases start transferring globally. Using pesticides in an inexperienced way produces pathogen resistance where fighting ability is reduced [1]. Agriculture improves by diagnosing the disease at the correct time and in an accurate way. Due to these changes in climate and pathogen resistance issues, financially the nation is affected [2].

When these diseases are not given attention, they might create a huge loss for farmers and for the global economy [3]. Diseases such as late blight may cause huge damage to the potatoes

thereby affecting the production of it by 15%. Annual loss due to crop disease contributes to 20% globally, thereby standing as a major threat to farmers as they are financially less powerful and also their resources seem to be limited. The diseases need to be detected at the earlier stages as farmers can find ways for preventing the failure of crops to evade the upcoming damages. Inspecting the plant diseases in person needs experts in that field for monitoring the farms. But this approach requires more expenditure in case the size of the farms is huge. Stress level is also an important factor in the process of disease detection in plants with the help of techniques used for analyzing the genetic levels as well as methods of thermal imaging. But these approaches suffer from pitfalls such as huge expenditure along with more expertise which is not affordable for farmers who are smallholders [4].

Preventing these diseases at early phases in an environment where there are several changes is significant. Plant pathology detection is of several ways [5]. Certain diseases lack significant symptoms, or they are identified late and in such situations analysis has to be done in an effective way. When plants get infected by diseases, the yield of crops gets affected and the farmer's profit also decreases [6]. To resolve the problem, detecting plant diseases at an early stage had to be done. Farmers or scientists can detect these diseases manually. But, the time taken for detection manually is more and is also a challenging task. Methods in molecular biology, as well as immunology, were used for detecting the diseases at the advanced phases [7].

In order to handle such situations, detecting plant diseases in an automatic way by using machine learning algorithms as well as deep learning. The parameters used for training are more in number in the traditional systems [8]. The time taken for training and prediction is more, which creates a need for the usage of machines with more power for computation. The human brain consists of neurons which is the main inspiration for the techniques in deep learning. CNN stands for Convolutional Neural Networks and RNN stands for Recurrent Neural Networks are employed for identifying the structures that are hidden in the case of data provided [9].

The techniques in deep learning have more advantages compared to the ones of machine learning. Absence of modules for extracting the features while using the raw data is a notable advantage [10]. When the datasets are huge in size because of their dimensions, the time taken for processing them is reduced in the domain of deep learning. Several applications in the domain of computer vision use the techniques of deep learning such as Convolutional Neural Networks as they work effectively for data related to images [11]. Operations such as convolution are employed for the extraction of features while using image datasets. Classification of images that are taken as input for processing can be done using CNN [12].

To remove drawbacks in existing system many system have been proposed to overcome those drawbacks by using different techniques. In the next section this paper tries to present those proposed systems in meaningful way. The management of crops required close inspection especially for management of disease infected crop that can affect the quality and quantity of crop. Image processing is a best technique for agricultural application. Image processing can

detect a pest's attack from the image of plant. The detection and classification of plant diseases are important task to increase plant productivity. The implementation of an Adaptive based clustering with R-CNN (ACR-CNN) is used to enhance the clarity of plant images in various shapes. When there is the identification of affected portions such as leaves, stems by automation and monitoring plants can have high development.

2. Background Study

For the purpose of detection of diseases in plants, the entire leaf is not considered and only the regions with spots are being used [5]. This simplifies the process of identification as the existence of various diseases can be explored where data augmentation is performed by splitting the images of leaves into several other images. The researchers made a deep study in detecting the diseases in plants as well as approaches for identification with the help of techniques in image processing [6]. Several steps used in the model are acquisition of image, pre-processing of images, segmenting images, extraction of features and classification methods. Approaches used for detection are compared for solving the detection problem along with discussion on benefits and pitfalls associated with this problem [7]. A framework for identifying disease in tea plants used SLIC called the Simple Linear Iterative Cluster at the initial phase. The samples of input are converted as blocks using SLIC and for extraction of key points, the Harris technique is utilized [8]. Approaches such as convex hulls are employed for contour purposes and both for key point extraction and training GLCM called as Gray Level Co-occurrence Matrix is employed. Classification in this model is carried out using an SVM classifier. A hybrid model proposed combines CAE [9] called Convolutional Auto encoder and CNN for the detection of plant diseases in an automated manner. Peach plants have bacterial spots in their leaf images and the hybrid model is employed for identifying this. The major goal of CAE is the reduction of parameters used for training. Convolution operation is utilized for extraction of features using the input data and the classification process is carried out by CNN. Image dimensionality is reduced by employing CAE. Among various crops in India, tomato is one of the important crops as it is useful for kitchen activities [10]. The ranking of India in the production of tomatoes is second, but this can get affected if no solution is provided when the tomato plants are infected by several diseases. A deep learning approach, CNN is used for detecting and classifying plant diseases. Convolution layers are 3, max-pooling is 3 and fully connected layers are 2 in number for the proposed model.

3. Methodology

Convolution Neural Network (CNN) is an efficient technique for image processing where the Adaptive based Clustering with R-CNN is introduced here in the proposed work. It uses the crop diseases from the kaggle dataset has efficient performance using the classes recognized from the boundaries that coordinate the point as pixel from the affected region.

3.1.ADAPTIVE MEDIAN FILTERING (AMF)

Nonlinear filtering algorithm has a type of algorithm called Median Filter (MF), often acts as a second-hand so as to remove the unwanted noise. It is a stage of pre-processing to maximize the efficiency of processing of images. It is widely utilized in the image processing areas in all circumstances. It preserves the edges at the same duration as a minimizing noise. It acts as a powerful filter and it is widely grouped and utilized as a smoother modelled for the processing the image extended to process the signals. An important gain of MF of the filter belong to linear type removes the noise that ranges with a large magnitude. AMF algorithm is formulated for the removal of noise in the direction that maximizes the image's quality with the noise reduction carried out in images [22].

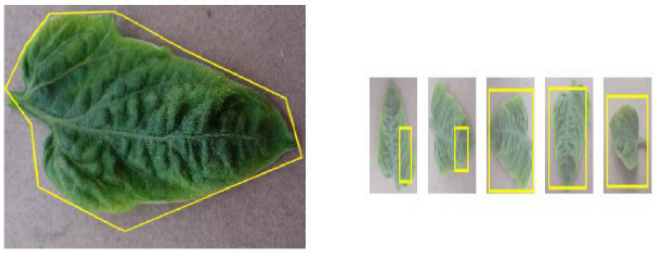

Fig1. Infected Leaf

Figure.1 describes the collection of data for training and testing according to the detection model as an enhanced version of 80% training and 20% testing in comparison with existing systems with Recurrent Neural Network (RNN).Figure 2. Describe about the framework of ACR-CNN.

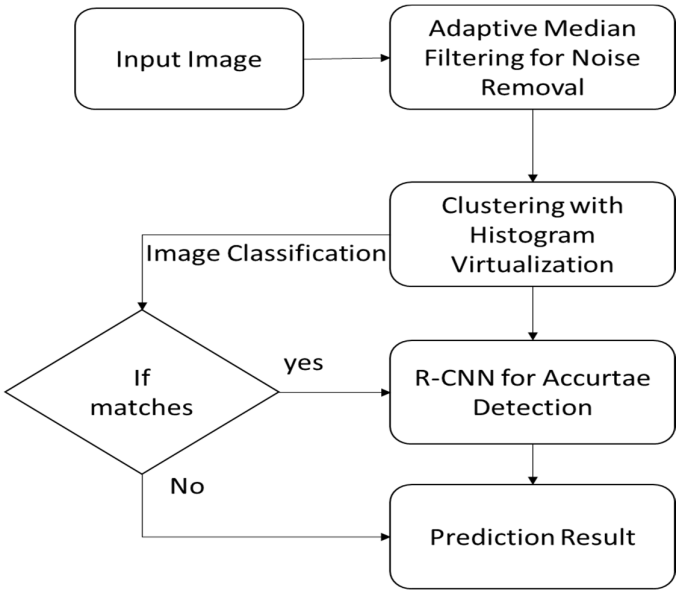

Fig2. ACR-CNN Framework

3.2. Clustering in Machine Learning

In real world, not every data we work upon has a target variable. This kind of data cannot be analyzed using supervised learning algorithms. The help of unsupervised algorithms. One of the most popular type of analysis under unsupervised learning is Cluster analysis. When the goal is to group similar data points in a dataset, then we use cluster analysis. In practical situations, it can use cluster analysis for customer segmentation for targeted advertisements, or in medical imaging to find unknown or new infected areas.

3.3. Proposed ACR-CNN

The Proposed model is imported with the library packages and necessary utilities which load the plant village dataset and train the model according to the diseases and their variation. Once the sequential learning is done, the leaf detection parameters based on trainable parameters transfer the prototype to the transfer learning model. The dataset consists of all rotation images and flipping images to find the classes from the normal data to annotated classes. And then convert class images to XML annotated images for reading the same. The proposed model **ACR-CNN** is performed four classification tests and produced 94% of accuracy based on labeled classes compared with the performance metrics by evaluation of Accuracy and loss.

Which performs the better results in comparison of mean, standard deviation and error mean which can show the earlier detection of plant diseases when compare to other algorithms. The implementation of an Adaptive based clustering with R-CNN (ACR-CNN) is used to enhance the clarity of plant images in various shapes. When there is the identification of affected portions such as leaves, stems by automation and monitoring plants can have high development. By segmentation of affected areas from the kaggle dataset, the proposed algorithm can maintain the threshold of complete appearance in detecting using the proposed framework. Also the implemented results show the detection gives better results and evaluation of performance is also higher in comparison with existing algorithms.

4. Result and Discussion

Across many implementation results, proposed version frameworks from the kaggle dataset are clustered based on color, threshold value. We monitored the variation of grey-scaled values in our comparison. To understand the leaves difference after the grey scaled version we have segmented based on region and instance from the pixel. Table.1 represents the algorithm comparison for accuracy and loss value for independent sample test. The sample size was 105 that was applied using the proposed algorithm to achieve better accuracy in enhancing the prediction system. By transferring the training and testing model based on the mean variation and sample size are bounded. Image classification and segmentation was successfully identified the affected region from the dataset.

Results comparison metrics for performance evaluation are accuracy, and loss. Results to estimate other performance characterizations such as precision recall curves using the usual definition:The accuracy is how close a measured value is to a standard or known value. Accuracy is also associated with the weighted arithmetic mean of the precision and inverse

precision (weighted by the bias) and the weighted arithmetic mean of the recall and inverse recall (weighted by the prevalence)

Accuracy = (Number of relevant images/total of retrieval images)*100(21)

Table 1. Performance Comparison of ACR-CNN with RNN and R-CNN Algorithm

			Accuracy	**Loss**
Algorithm		Sample Size	105	105
	RNN	Mean	77.14	22.87
		Standard Deviation	8.504	8.518
		Error Mean	2.96	2.657
	R-CNN	Mean	78.73	39.27
		Standard Deviation	10.292	10.292
		Error Mean	2.987	2.757
	ACR-CNN	Mean	80.73	41.27
		Standard Deviation	11.292	12.292
		Error Mean	3.657	3.657

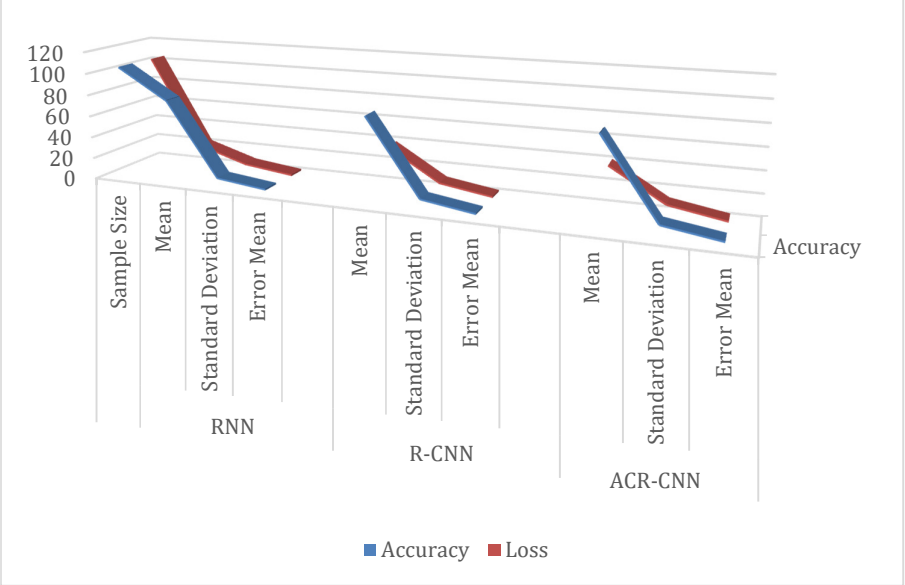

Fig3. Result of ACR-CNN

Figure 3 portray the analysis of the performance measures in terms of Accuracy and loss with respect to three methods such as RNN, R-CNN and ACR-CNN.The proposed ACR-CNN framework shows the higher accuracy when compared to other algorithms and gives better prediction of diseases.

5. Conclusion

Major impacts and threats might get affected the farming and its surrounding without proper maintenance for agriculture. To maintain and monitor the system efficiently which increases the performance rate using ACR-CNN algorithm. This algorithm based feature analysis helps in major classification and segmentation for finding the unhealthy plant from the kaggle dataset and identifies the result with better accuracy. Based on class and label such as virus, bacteria, plant spot, affected region diseases are categorized. The implementation of an Adaptive based clustering with R-CNN (ACR-CNN) is used to enhance the clarity of plant images in various shapes. Also the implemented results show the detection gives better results and evaluation of performance is also higher in comparison with existing algorithms. The future enhancement based on better prediction using high dimensional view of data with accurate prediction.

References

[1] Thomee, B., Shamma, D.A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., Li, L.J.: The new data and new challenges in multimedia research. arXiv preprint arXiv:1503.01817 (2015)

[2] L. Dong, J. Su and E. Izquierdo, “Scene-oriented Hierarchical Classification of Blurry and Noisy Images,” IEEE Trans. Circuits and Systems for Video Technology, vol.21, no.5, pp.2534-2545, 2012.

[3] L. Dong and E. Izquierdo, “A Biologically Inspired System for Classification of Natural Images,” IEEE Transactions on Image Processing, vol. 17, no. 5, pp.590-603, 2007.

[4] Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. Communications of the ACM. 2008;51(1):107–113.

[5] M. H. Almeer, “Cloud Hadoop mapreduce for remote sensing image analysis,” Journal of Emerging Trends in Computing and Information Sciences, vol. 3, no. 4, pp. 637–644, 2012. View at Google Scholar

[6] W. S. Zhu and P. Wang, “Large-scale image retrieval solution based on Hadoop cloud computing platform,” Journal of Computer Applications, vol. 34, no. 3, pp. 695–699, 2014. View at Google Scholar

[7] K. Wiley, A. Connolly, and S. Krughoff, “Astronomical image processing with Hadoop,” in Proceedings of the 20th Conference on Astronomical Data Analysis Software and Systems, pp. 93–96, Astronomical Society of the Pacific, San Francisco, Calif, USA, July 2011.

[8] Y. M. Zhu, “Image classification based on Hadoop platform,” Journal of Southwest University of Science and Technology, vol. 26, no. 2, pp. 70–73, 2011.

[9] Ranger, C., Raghuraman, R., Penmetsa, A., Bradski, G., Kozyrakis, C.: Evaluating MapReduce for Multi-core and Multiprocessor Systems. In: Proc. of 13th Int. Symposium on High-Performance Computer Architecture (HPCA), Phoenix, AZ (2007)
[10] Lammel, R.: Google's MapReduce Programming Model - Revisited. Science of Computer Programming 70, 1–30 (2008)
[11] Sweeney, C., Liu, L., Arietta, S. and Lawrence, J., 2011. HIPI: a Hadoop image processing interface for image-based mapreduce tasks. Chris. university of Virginia, 2(1), pp.1-5.
[12] Epanchintsev, T. and Sozykin, A., 2015. Processing large amounts of images on hadoop with OpenCV. In CEUR Workshop Proceedings. Vol. 1513: Proceedings of the 1st Ural Workshop on Parallel, Distributed, and Cloud Computing for Young Scientists (Ural-PDC 2015), pp.137-143.
[13] L. Dong, J. Su and E. Izquierdo, "Scene-oriented Hierarchical Classification of Blurry and Noisy Images," IEEE Trans. Circuits and Systems for Video Technology, vol.21, no.5, pp.2534-2545, 2012
[14]Sozykin, A. and Epanchintsev, T., 2015, MIPr-a framework for distributed image processing using Hadoop. In 2015 9th International Conference on Application of Information and Communication Technologies (AICT) ,pp. 35-39.
[15] Dong, L., Lin, Z., Liang, Y., He, L., Zhang, N., Chen, Q., Cao, X. and Izquierdo, E., 2016. A hierarchical distributed processing framework for big image data. *IEEE Transactions on Big Data*, *2*(4), pp.297-309.
[16] Bao, S., Huo, Y., Parvathaneni, P., Plassard, A.J., Bermudez, C., Yao, Y., Lyu, I., Gokhale, A. and Landman, B.A., 2018, March. A data colocation grid framework for big data medical image processing: backend design. In *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications* (Vol. 10579, p. 105790A). International Society for Optics and Photonics.
[17] H. Kocalkulak, and T. T. Temizel, A Hadoop Solution for Ballistic Image Analysis and Recognition, in: Proc. International Conference on High Performance Computing and Simulation (HPCS 2011), Istanbul, July 2011, pp. 836-842.
[18] Yamamoto, M. and Kaneko, K., 2013. Parallel image database processing with MapReduce and performance evaluation in pseudo distributed mode. " International Journal of Electronic Commerce Studies", 3(2), pp.211-228.
[19] Liu, T., Liu, Y., Li, Q., Wang, X.R., Gao, F., Zhu, Y.C. and Qian, D.P., 2015. SEIP: system for efficient image processing on distributed platform. Journal of Computer Science and Technology, 30(6), pp.1215-1232.
[20] Kumar, P., Bodade, A., Kumbhare, H., Ashtankar, R., Arsh, S. and Gosar, V., 2019. Parallel and Distributed Computing for Processing Big Image and Video Data. In Multimodal Analytics for Next-Generation Big Data Technologies and Applications (pp. 337-360). Springer, Cham.