Drawbacks Of Dataset For Tomato Leaf Diseases Recognition

Debabrat Bharali[1] Dr.Kanak Chandra Bora[2]
Dr.Rashel Sarkar[3]

[1]Phd scholar,Department of Computer Science ,University of Science and Technology,Meghalaya
Shillong,India
Email id: dbharali15@gmail.com
[2]Associate Professor ,Department of Computer Science,University of Science and Technology,Meghalaya
Shillong,India
Email id: boopborabora@yahoo.in
[3]Associate Professor ,Department of Computer Science and Engineering,Royal Global University
Guwahati,India
Email id: sarkarrashel@gmail.com

Abstract

This paper brings out the drawbacks of dataset for Tomato leaf diseases. Diseases in tomato leaf causes major loss in economy and production as well as it reduces the quality and quantity of products. As a result farmers face a lot of problem to control and monitoring of Tomato leaf health is treated by a number of factors and out of those factors is leaf disease and because of those factors it is difficult to detect Tomato leaf diseases in the early stage. Deep learning techniques have been used to detect Tomato leaf diseases, but the existing system failed to detect leaf diseases because existing datasets are trained and tested and on few images of leaf of a particular region. Very few diseases such as Early blight, late blight are covered in the existing dataset and segmentation is missing such as in which stage leaf is healthy or not, percentage of affected area of infected leaves in the existing dataset

Keywords—dataset, segmentation, ages, deep learning, early blight, late blight.

1.INTRODUCTION

Diseases in leaf causes major loss in economy and production as well as it reduces the quality and quantity of products [1]. The estimated annual crop yield loss due to leaf disease is about 16% globally, which is the major cause of food shortage and increased food production costs [2]. According to the Food and Agriculture Organisation Report (FAO), the world's population will reach approximately 9.1 billion by 2050. For a steady food supply, about 70% of food production growth is required [3]. The factors affecting the leaf and their products are categorised as diseases and disorders. The important factors are the diseases caused by algae, fungi, or bacteria, whereas the biotic factors include are rainfall, moisture, temperature and nutrient deficiency [4]. In existing method to diagnose leaf diseases; one of the primary approach is a visual estimation. The traditional leaf disease diagnosing techniques depend on the farmer's experience, which is most unreliable. Compared to the conventional leaf disease diagnosing techniques, the researchers have introduced the spectrometer to diagnose the plant leaves as healthy and infected [5]. An other method is to extract the leaves DNA by using the polymerase chain reaction [6] or real-time polymerase chain reaction [7]. Such techniques are challenging, expensive and time-consuming, require a highly expertise, experiment condition and massive use of crop protection products. The recent trends in Artificial Intelligence (AI), Machine Learning (ML) and Computer Vision (CV) technologies allow developing the automated leaf disease detection techniques. These techniques can eaccurately detect leaf diseases in a limited time without human intervention. It has been observed that Deep learning has been the most prominent usage in agriculture [8].Chen et al. [11] proposed a deep learning model that counts the apples and oranges from the real-time images. Dias et al. [12] proposed an apple flower semantic segmentation using the convolutional neural network (CNN) counting the number of flowers from the plants. Ubbens et al. [13] conducted a study to estimate the plant leaves using the CNN model. Recently, numerous types of deep learning architecture has been proposed for leaf disease classification. The most prominent technique is the convolutional neural network (CNN). A convolutional neural network is a class of deep nueral network inspired by the biological nervous system and vision system with significant performance compared to other models. As compared to Artificial Neural Network (ANN), CNN requires few neurons and multilayer convolution layers to learn the features, but it required an extensive dataset for training [10]. In the last few decades, several techniques have been developed to detect leaf diseases in various crops [14,15]. In most of the techniques, features were extracted using the image processing techniques, then extracted features were fed to a classification technique. Deepa and Nagarajan [16] proposed a plant leaf disease detection technique. The authors first applied the Kuan filter for noise removal and applied a Hough transformation to extract the colour, shape and texture features. A reweighted linear program boost classification was applied to classify the plant leaf disease. The PlantVillage dataset was used to evaluate the performance of the proposed technique. From the above introduction covered we found that the existing dataset was imbalanced, therefore, training the model was very much challenging. Very diseases were covered and segmentation is not present to check in which stage the leaf are infected or not infected as well as percentage of affected area of infected leaf of Tomato

2.RELATED WORKS

In the last few decades, many researchers worked on multiple crops, including tomatoes; their focus was not on the single tomato crop diseases [20–22]. The models were trained on specific region dataset (PlantVillage [23]), which was developed in the USA and Switzerland. The diseases of tomato vary from other regions due to the difference in leaf shapes, varieties and environmental factors [24]. Geetharamani and Pandian [20] proposed a deep CNN model to differentiate between healthy and unhealthy leaves of multiple crops. The model was trained using the PlantVillage dataset with 38 different types of crops with disease leaf images, healthy leaf images and background images. The focus of the model was not on single tomato crop diseases. The model is also trained in specific regiondataset USA and Switzerland, which failed to detect the region tomato leaf diseases. Kamal et al. [21] developed leaf disease identification models named Modified MobileNet and Reduced MobileNet using depthwise separable convolution instead of convolution layer by modifying the MobileNet [25]. The proposed model was trained on multiple crops of the PlantVillage dataset, where the leaf images were collected from a specific region of the world. Khamparia et al. in [22], proposed a hybrid approach to detect leaf disease using the combination of CNN and autoencoders. The model was trained on the PlantVillage dataset for multiple crop diseases and specific region diseases. In [26], Liang et al. proposed a leaf disease diagnosis and severity estimation network based on a residual structure and shuffle units of ResNet50 architecture [27]. The PlantVillage dataset was also used to detect the multiple crop diseases of a specific region. Ferentinos [28] investigated AlexNet [28], Overfeat [29], AlexNetOWTBn [30], VGG [31] and GoogLeNet [32] deep learning-based architectures in order to identify the normal or abnormal leaf from leaf images. The researchers performed the transfer learning approach using the PlantVillage dataset to detect the specific region's multiple crops diseases. Many researchers worked on tomato crops diseases but also trained the models on a specific dataset PlantVillage. Khalifa et al. [33] proposed a CNN model to detect early blight and late blight diseases along with a healthy class. Sanjeev et al. [35] proposed a Feed-Forward Neural Network (FFNN) to detect early blight, late blight diseases along with healthy leaves. The proposed method was trained and tested on the PlantVillage dataset. Barman et al. [36] proposed a self-build CNN (SBCNN) model to detect the early blight, late blight potato leaf diseases, and healthy class. The PlantVillage dataset was also used to train the model, which is for a specific region. They did not validate their model on unseen test data. Tiwari et al. [37] used a pre-trained model VGG19 to extract the features and used multiple classifiers KNN, SVM and neural network for classification. The model also trained on the PlantVillage dataset to detect the early blight and late blight disease of tomato leaves. They did not test their model on unseen data. Lee et al. [38] developed a CNN model to detect the early blight, late blight diseases, and healthy leaves of tomato. The researchers also used the PlantVillage dataset belonging to a specific region. The model was not tested on unseen data. Islam et al. [39] proposed a segment-based and multi-SVM-based model to detect tomato diseases, such as early blight, late blight and healthy leaves. The summary of the related work is given below

Reference	Methodology	Drawbacks	Findings
[20]	CNN	The focus of the model was not on single tomato crop diseases. The model is also trained in specific region USA and Switzerland, which failed to detect the tomato leaf diseases since the existing dataset have high inequality class	To differentiate the healthy and unhealthy leaves of multiple crops
[21-25]	Modified MobileNet and Reduced MobileNet using depthwise separable convolution	The proposed model was trained on multiple crops of the existing dataset namely PlantVillage dataset, where the leaf images were collected from a specific region of the world.They failed to achieve a high accuracy and segmentation is missing in the existing dataset	Identify the different types of leaf diseases using Modified MobileNet and Reduced MobileNet using depthwise separable convolution.
[22]	Combination of CNN and autoencoders	The model was trained on the PlantVillage dataset i.e existing dataset for multiple crop diseases and specific region diseases.Only few diseases are covered and segmentation is missing in the dataset.	Identify the various types of leaf diseases using the combination of CNN and autoencoders
[26-27]	Residual structure and shuffle units of ResNet50 architecture	The model was trained on the PlantVillage dataset i.e existing dataset for multiple crop diseases and specific region diseases.Only few diseases are covered and segmentation is missing in the dataset.	Leaf diagnosis using Residual structure and shuffle units of Resnet50 architecture
[28-32]	Deep learning architecture based on AlexNet, Overfeat, AlexNetOWTBn, VGG, GoogLeNet	The researchers performed the transfer learning approach using the PlantVillage i.e existing dataset to detect the specific region's multiple crops diseases. The existing datasets contain lack of training samples with imbalanced class samples	Identify the normal or abnormal leaf from leaf images.
[33]	CNN	The model also trained on the PlantVillage dataset to detect the early blight and late blight disease of tomato leaves. Only few diseases are covered and segmentation is missing in the dataset. They did not validate their model on unseen test data	Detect the early blight and late blight disease of tomato leaves
[35]	Feed-Forward Neural Network (FFNN)	The model also trained on the PlantVillage dataset to detect the early blight and late blight disease of tomato leaves. Only few diseases are covered and segmentation is missing in the dataset. They did not validate their model on unseen test data	Detect the early blight and late blight disease of tomato leaves
[36]	self-build CNN (SBCNN) model	The model also trained on the PlantVillage dataset to detect the early blight and late blight disease of tomato leaves. Only few diseases are covered and segmentation is missing in the dataset. They did not validate their model on unseen test data	Detect the early blight and late blight disease of tomato leaves
[37]	pre-trained model VGG19 to extract the features and used multiple classifiers KNN, SVM and neural network for classification.	The model also trained on the PlantVillage dataset to detect the early blight and late blight disease of tomato leaves. They did not test their model on unseen data. Only few diseases are covered and segmentation is missing in the dataset.	Detect the early blight and late blight disease of tomato leaves
[38]	CNN	The researchers also used the PlantVillage dataset belonging to a specific region. The model was not tested on unseen data. Only few diseases are covered and segmentation is missing in the dataset.	Detect the early blight and late blight disease of tomato leaves
[39]	segment-based and multi-SVM-based model	The researchers also used the PlantVillage dataset belonging to a specific region. The model was not tested on unseen data. Only few diseases are covered and segmentation is missing in the dataset.	Detect the early blight and late blight disease of tomato leaves

Table 1: Summary of Related Work

4. CONCLUSION

From the above related work we conclude that there are lot of problems exist in the related work due to improperrecognition of leaf diseases, differences in tomato diseases, dataset and ecological factors. The existing system have a bouncy faulty estimate to admit tomato leaf diseases in the particular region. The existing tomato leaves disease datasets contain lack of training samples with imbalanced class samples. Another problem is that the recent methods have a low concurrence speed due to the large number of trainable parameters needs to be improved. The next problem is that the existing dataset consists of not many images, whereas to train the CNN model dataset must be enormous. The existing dataset consists only 160 images i.e healthy leaf images. Let say if dataset split into training, validation and testing by 70%, 15% and 15% respectively, then the normal leaf class will be reduced for training. In that case, the existing methods have incomplete training in particular class. Existing dataset have an inequality class since the late blight and early blight class have 800 images each. Therfore we can say that only few diseases such as early blight and late blight is covered.in existing dataset. Such a method failed to achieve high accuracy The last problem in the related work is the non-availability of the tomato leaf segmentation technique such as in which stage the leaf are infected or not as well as percentage of affected area.

REFERENCES

[1] Iqbal, Z.; Khan, M.A.; Sharif, M.; Shah, J.H.; ur Rehman, M.H.; Javed, K(2018). An automated detection and classification of citrus plant diseases using image processing techniques: A review. Comput. Electron.Agri 153

[2] Oerke, E.C(2006) Crop losses to pests. J. Agric. Sci.

[3]Golhani, K.; Balasundram, S.K.; Vadamalai, G.; Pradhan, B.(2018) A review of neural networks in plant disease detection using hyperspectral data. Inf. Process. Agric. 5, 354–371.

[4] Chouhan, S.S.; Kaul, A.; Singh, U.P.; Jain, S(2018) Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology. IEEE Access , 8852–8863.

[5] Sasaki, Y.; Okamoto, T.; Imou, K.; Torii, T(1998).Automatic diagnosis of plant disease-Spectral reflectance of healthy and diseased leaves. IFAC Proc. Vol. 1998, 31, 145–150.

[6] Koo, C.; Malapi-Wight, M.; Kim, H.S.; Cifci, O.S.; Vaughn-Diaz, V.L.; Ma, B.; Kim, S.; Abdel-Raziq, H.; Ong, K.; Jo, Y.K.; et al(2018) Development of a real-time microchip PCR system for portable plant

[7] Henson, J.M.; French, R(1993) The polymerase chain reaction and plant disease diagnosis. Annu. Rev. Phytopathol, 31, 81–109.

[8] Singh, U.P.; Chouhan, S.S.; Jain, S.; Jain, S(2019). Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease. IEEE Access , 7, 43721–43729.

[9]Ferentinos, K.P(2018) Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agri 145, 311–318.

[10] Aamir, M.; Irfan, M.; Ali, T.; Ali, G.; Shaf, A.; Al-Beshri, A.; Alasbali, T.; Mahnashi,(2018) M.H.J.D. An adoptive threshold-based multi-level deep convolutional neural network for glaucoma eye disease detection and classification. Diagnostics, 10, 602.

[11] Chen, S.W.; Shivakumar, S.S.; Dcunha, S.; Das, J.; Okon, E.; Qu, C.; Taylor, C.J.; Kumar, V(2017). Counting apples and oranges with deep learning: A data-driven approach. IEEE Robot. Autom. Lett. ,781–788.

[12] Dias, P.A.; Tabb, A.; Medeiros, H(2018). Multispecies fruit flower detection using a refined semantic segmentation network. IEEE Robot. Autom. Lett., 3, 3003–3010.

[13] Ubbens, J.; Cieslak, M.; Prusinkiewicz, P.; Stavness, I(2018).The use of plant models in deep learning: an application to leaf counting in rosette plants. Plant Methods

[14] Zhang, K.; Xu, Z.; Dong, S.; Cen, C.; Wu, Q(2019) Identification of peach leaf disease infected by Xanthomonas campestris with deep learning. Eng. Agric. Environ. Food 12, 388–396

[15] Lu, J.; Hu, J.; Zhao, G.; Mei, F.; Zhang, C(2017). An in-field automatic wheat disease diagnosis system. Comput. Electron. Agric., 142, 369–379.

[16] Deepa, N.; Nagarajan, N(2020). Kuan noise filter with Hough transformation based reweighted linear program boost classification for plant leaf disease detection. J. Ambient. Intell. Humaniz. Comput.12, 5979–5992.

[17] Karthik, R.; Hariharan, M.; Anand, S.; Mathikshara, P.; Johnson, A.; Menaka, R(2020)Attention embedded residual CNN for disease detection in tomato leaves. Appl. Soft Comput. , 86, 105933.

[18] Zhang, Y.; Song, C.; Zhang, D. Deep(2020). learning-based object detection improvement for tomato disease. IEEE Access 8, 56607–56614

[19] Sambasivam, G.; Opiyo, G.D(2020). A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. Egypt. Inform

[20] Geetharamani, G.; Pandian, A(2019) Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Comput. Electr. Eng. 76, 323–338.

[21] Kamal, K.; Yin, Z.; Wu, M.; Wu, Z(2019). Depthwise separable convolution architectures for plant disease classification. Comput. Electron. Agric 165, 104948.

[22] Khamparia, A.; Saini, G.; Gupta, D.; Khanna, A.; Tiwari, S.; de Albuquerque, V.H.C(2019).Seasonal Crops Disease Prediction and Classification Using Deep Convolutional Encoder Network. Circuits Syst. Signal Process

[23]. Hughes, D.; Salathé, M(2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics.

[24] Baker, N.T.; Capel, P.D(2011). Environmental Factors that Influence the Location of Crop Agriculture in the Conterminous United States; US Department of the Interior, US Geological Survey: Reston, VA, USA

[25] Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H(2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications

[26] Liang, Q.; Xiang, S.; Hu, Y.; Coppola, G.; Zhang, D.; Sun, W(2019). PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network. Comput. Electron. Agric. 157, 518–529

[27] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition(2016). In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 , pp. 770–778.

[28] Krizhevsky, A.; Sutskever, I.; Hinton, G(2012). Imagenet classification with deep convolutional networks. In Proceedings of the Conference Neural Information Processing Systems (NIPS), Lake Tahoe, NV, USA, 3–6 ; Volume 25, pp. 1097–1105.

[29] Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y(2018)Overfeat: Integrated recognition, localization and detection using convolutional networks

[30] Krizhevsky, A(2014) One weird trick for parallelizing convolutional neural networks.

[31] Simonyan, K.; Zisserman, A(2014). Very deep convolutional networks for large-scale image recognition.

[32] Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A(2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 1–9.

[33]. Khalifa, N.E.M.; Taha, M.H.N.; Abou El-Maged, L.M.; Hassanien, A.E(2021) Artificial Intelligence in Potato Leaf Disease Classification: A Deep Learning Approach. In Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges; Springer: Berlin/Heidelberg, Germany.

[34]. Rozaqi, A.J.; Sunyoto, A(2020).Identification of Disease in Potato Leaves Using Convolutional Neural Network (CNN) Algorithm. In Proceedings of the 2020 3rd International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 24–25 pp. 72–76.

[35] Sanjeev, K.; Gupta, N.K.; Jeberson, W.; Paswan, S(2020). Early Prediction of Potato Leaf Diseases Using ANN Classifier. Orient. J. Comput. Sci. Technol. , 13, 2–4.

[36]. Barman, U.; Sahu, D.; Barman, G.G.; Das, J(2020). Comparative Assessment of Deep Learning to Detect the Leaf Diseases of Potato

based on Data Augmentation. In Proceedings of the 2020 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2–4 ; pp. 682–687

[37] Tiwari, D.; Ashish, M.; Gangwar, N.; Sharma, A.; Patel, S.; Bhardwaj, S(2020). Potato leaf diseases detection using deep learning. In Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 461–466.

[38] Lee, T.Y.; Yu, J.Y.; Chang, Y.C.; Yang, J.M(2020).Health Detection for Potato Leaf with Convolutional Neural Network. In Proceedings of the 2020 Indo–Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN), Rajpura, India, pp. 289–293.

[39] Islam, M.; Dinh, A.; Wahid, K.; Bhowmik, P(2017). Detection of potato diseases using image segmentation and multiclass support vector machine. In Proceedings of the 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE), Windsor, ON, Canada.