 ****

**GRAPHES AND STRAWBEDRRY PLANTS DISEASES DETECTION**

**Group Members:**

Abdul Sami Wadho (leader).

Taimoor khan.

**Sir Sajid Majeed:**

Signature.

**ABSTRACT**

Between 30 and 40 percent of Pakistan's yearly strawberry and grape plant diseases are fatal. Plant disease early detection is still challenging because of a lack of lab equipment and experience. In this work, we investigate the potential of computer vision methods for early and scalable plant disease diagnosis. One key obstacle to enabling vision-based plant disease detection is the lack of sufficiently large-scale non-lab data sets. In light of this, we introduce PlantDoc: a dataset for diagnosing plant diseases visually. Our dataset, which includes 5632 data points total from 13 different plant species and up to 6 different disease classes, required 300 person hours to annotate internet-scraped photos. We train three models for the goal of classifying plant diseases in order to demonstrate the effectiveness of our dataset. According to our findings, modeling with our dataset can result in a 31% improvement in classification accuracy. We think that our dataset can contribute to lowering the entry-level difficulty of plant disease identification using computer vision techniques.

**KEYWORDS**

Deep Learning, Object Detection, Image Classification.

**INTRODUCTION**

The demand for all plant products rises by roughly 1.6% annually along with the Earth's population [16]. Meeting the increasing demand for food quality and quantity requires protecting strawberries and grapes from plant diseases [22]. Plant diseases alone cost the world economy over US$220 billion a year in terms of economic value [1]. As stated by Pakistan

According to the Council for Agricultural Research, pests and diseases cause the annual loss of more than 35% of strawberry and grape production [15]. An concerning rise in insect and plant disease outbreaks poses a threat to food security. These illnesses have extensive effects on the economy, society, and environment, endangering food security [5]. For farmers, the timely diagnosis of plant diseases is still a difficult challenge. Other than speaking with other farmers or calling the Kisan helpline, they don't have many choices [17].

To recognize the sick leaves, a person must have a thorough understanding of plant diseases. Moreover, to diagnose a sick leaf, a laboratory infrastructure is typically required. In this work, we investigate the potential of computer vision for efficient and scalable plant disease diagnosis. Recent developments in deep convolutional neural networks have led to remarkable advancements in computer vision. Although it can take a long time to train big neural networks, the trained models can categorize images fast, which makes them useful for consumer applications on smartphones.

The use of image processing to identify plant illnesses creates new opportunities to apply the insights gained from deep learning techniques to practical agricultural issues. This leads to improvements in crop productivity, disease prevention, and agricultural knowledge. The majority of vision-based systems now in use need for high-resolution photos with a simple backdrop. On the other hand, we concentrate on photos in natural environmental settings with non-trivial background noise and provide the best query resolution for grapes and Strawberry plants, since most Pakistani farmers utilize low-end mobile devices with natural background and lighting conditions. In light of this, we emphasize our two primary contributions: i) PlantDoc development:

a collection of 5632 photos from 13 different plant species and 27 classes (17–10, disease-healthy) ii) demonstrating the usefulness of the curated data set for illness identification in uncontrolled settings. As far as we are aware, this dataset is the first of its kind to include information from uncontrolled environments. In order to determine the necessity of a dataset in uncontrolled environments, we assessed our dataset utilizing a variety of classification and object detection architectures that were covered in Section 4. The findings implied that classifying or detecting photos in real-world scenarios was not possible with a lab-controlled dataset. We discovered that the categorization error can be decreased by up to 31% by fine-tuning the models using PlantDoc. Therefore, our dataset may be utilized to develop an application that effectively identifies and classifies 27 Grapes and Strawsberry plant disease/healthy classifications.

**Increased Accuracy:**

Large datasets of annotated photos of diseases affecting strawberries and grapes can be used to train machine learning models, which can let them recognize subtle disease indications that the human eye might miss.

**Efficiency:**

When compared to manual examination, automated systems can evaluate enormous volumes of photos rapidly and consistently, saving time and resources.

.

**Objectivity:**

In order to detect diseases in strawberries and grapes more consistently and objectively, machine learning algorithms rely on facts rather than opinion.

**2 RELATED WORK**

Our related work can be broadly categorized into: i) techniques for plant disease detection; and ii) datasets advancing research in plant disease detection.

**2. Techniques for Disease Detection in Grapes and Strawberries**

**Image** **Processing:**

Techniques like color segmentation, texture analysis, and edge detection can be used to isolate diseased areas from healthy plant tissue in grape and strawberry images.

**Machine** **Learning**:

Supervised learning algorithms like Support Vector Machines (SVMs) and Random Forests can be trained on datasets of labeled images containing various grape and strawberry diseases. These models learn to recognize disease patterns and classify new images accordingly.

**Deep Learning:**

Deep Convolutional Neural Networks (CNNs) are a powerful machine learning subfield that excel at image recognition. CNNs can automatically extract features from grape and strawberry images without manual feature engineering, potentially leading to higher disease detection accuracy.

Prior work by Sankaran et al. [19] proposed using reliable sensors for monitoring health and diseases in plants under field conditions.

However, due to the high cost of hardware and the skill required to operate such sensors, plant disease diagnosis through the use of sensors has the potential to benefit only a small number of farmers. On the other hand, Patil et al.'s earlier research [18] retrieved shape features for sugarcane leaf disease identification, ultimately achieving a final average accuracy of 97-94%. For disease identification on maize leaves, Patil et al. [3] employed texture features, specifically inertia, homogeneity, and correlation, which were generated by computing the gray level co-occurrence matrix on the image and color extraction. According to the morphological characteristics of the veins on leaves, three distinct species of legumes have been identified using neural networks in recent study [8].

Similarly, tea leaf illnesses have been identified using feature extraction and Neural Network Ensemble (NNE), yielding a 91% final testing accuracy [25]. Convolutional neural network variations for disease diagnosis utilizing plant leaf pictures have been the subject of numerous other recent publications [7, 21]. One major disadvantage of these efforts is that they are specific to one Grapes and Strawberry plants. Reproducibility has also been impacted by the fact that the datasets utilized in the works have not been made accessible.

**Datasets for Grapes and plant disease detection**

The Plant Village dataset(PVD) [14] is the only public dataset for plant disease detection to the best of our knowledge. The data set curators created an automated system using GoogleNet [23] and AlexNet [12] for disease detection, achieving an accuracy of 99.35%. However, the images in PlantVillage dataset are taken in laboratory setups and not in the real conditions of cultivation fields, due to which their efficacy in real world is likely to be poor. In contrast, we curate real-life images of healthy and diseased plants to create a publicly available dataset.

**THE PLANTDOC DATASET**

The Plant Village dataset contains images taken under controlled settings. This dataset limits the effectiveness of detecting diseases because, in reality, plant images may contain multiple leaves with different types of background conditions with varying lighting conditions (shown in Figure 1). Against this background, we now describe our curated dataset and discuss the techniques used for curation.

**Data Collection**

To account for the intricacies of the real world, we require models trained on real-life images. This fact motivated us to create a dataset by downloading images from Google Images and Ecosia [6] for accurate plant disease detection in the farm setting. We downloaded images from the internet since collecting large-scale plant disease data through fieldwork requires enormous effort. We collected about 20,900 images by using scientific and common names of 38 classes mentioned in the dataset by Mohanty et al. [14]. Four users filtered the images by selecting images based on their metadata on the website and guidelines mentioned on APSNet [2]. APS compiled a list of peer-reviewed literature corresponding to each plant disease. We referred APS’ prior literature and accordingly classified images. Some of the most important factors for classification were the color, area and density of the diseased part and shape of the species. We removed inappropriate (such as non leaf plant, lab controlled and out-of-scope images) and duplicate images across classes downloaded due to web search. Every image was checked by two individuals according to the guidelines to reduce labeling errors. Finally, to have sufficient training samples, we removed the classes with less than 50 images.

**Plant image classification**

Our main goal was to construct a model which can detect a leaf in an image and then classify it into the particular classes shown in Figure 2. We performed two main experiments, which we discuss after describing our experimental settings. 4.2.1 Experimental settings. For training the networks, we used stochastic gradient descent with momentum 0.9, categorical cross entropy loss, and a learning rate of 0.001. All weights were initialized with the orthogonal initializer. We applied common data augmentation techniques such as rotation, scaling, flipping etc. on the input images. All images were resized to 100 × 100, before feeding into the networks. For pre-trained models, we used the weights provided in Keras trained on Image Net. 4.2.2 Plant image classification using raw images (uncropped). Our first experiments aims to understand classification accuracy on the uncropped Plant Doc dataset. We evaluated the performance of VGG16 [20] using different training sets on Plant Doc as shown in Table 1. 4.2.3 Plant image classification using cropped images. Further, we evaluate the performance of several popular CNN architectures on the Grapes and Strawberry –Plant Doc dataset that have recently achieved state of-the-art accuracies on image classification tasks on the popular datasets, such as Image Net [4], CIFAR-10 [11], etc. Table 2 gives the complete list of the architectures that we used for benchmarking our Cropped-Plant Doc dataset. This experiment was conducted to verify the performance of Plant Village in real-setting.

**Leaf Detection**

The aim of our next experiments is to evaluate the performance of Faster R-CNN with InceptionResnetV2 model and Mobile Net model on our Plant Doc Dataset as shown in Table 3. We use mean average precision (mAP: higher is better) to evaluate the models and compare it with scores on COCO dataset since no evaluation exists in the domain of plant disease.

**RESULTS AND DISCUSSION**

Figure 4 shows saliency map and gradient activation map of the Grapes leaf and Strawberry bacteria Spots respectively. As expected, the neural network is learning to focus on the set of visual features which are correlated with disease such as the blemishes in the leaf (lines in Grapes leaf and spots in strawberry Bacterial Spots). The network even learns the shape of the leaf (shown in second row of Figure 4) to help it distinguish between species. As predicted, the results in Table 1 clearly shows that real case scenarios have low accuracy when processed initially with ImageNet or PlantVillage. Also, Table 1 and Table 2 clearly shows low accuracy achieved by training on PlantVillage and testing on PlantDoc. Model fails to produce accurate results due to background noise, images with leaf from multiple classes in a dataset and lowresolution leaf images. Table 3 shows that Faster R-CNN with InceptionResnetV2 performs the best with an mAP of 38.9. It is interesting to see that MobileNet performance is decreased when pre-trained on COCO+PlantVillage compared to the model where pre-training was done only on COCO. This attributes to the fact that PlantVillage is not contributing towards better results. MobileNet gives an mAP of 22 when evaluated on COCO dataset which has significantly more classes [9].

**APPLICATION BUILDING**

We were able to adapt the above solution to a mobile environment (Figure 5) by using models that very significant reduce complexity, without sacrificing the effective accuracy. This allowed us to achieve the best possible performance, given that the application should predict the bounding boxes and classes in real time in a mobile CPU. We have build application that utilizes Mobile Nets Object Detection Network [10] due to its efficiency and competitive accuracy. The network builds on top of the SSD framework [13].

**LIMITATIONS**

The dataset has been curated with care, but due to lack of extensive domain expertise, there are some images in the dataset which can potentially be wrongly classified (shown in Figure 6). Further, to train highly accurate models for disease detection, we may require a dataset with more number of images in each class. But, due to non-availability of public dataset and lack of real-life scenario for field work, our approach gives a feasible direction to tackle the on-going problem of disease detection.

**CONCLUSIONS AND FUTURE WORK**

In this work, we used cutting edge object identification models to tackle the problem of identifying healthy and diseased leaves in photos. Our work's primary contribution is the suggestion of a brand-new dataset, named PlantDoc, for the detection of plant diseases in strawberries and grapes. Our benchmark tests highlight the importance of real-world datasets like ours by demonstrating the ineffectiveness of models learned on controlled datasets. The dataset may be more useful if leaves are extracted from the photos using image segmentation techniques. This dataset, in our opinion, represents a significant advancement toward scalable computer vision-enabled strawberry and grape plant detection.

**REFRENCES**

[1] GN Agrios. 2005. Plant pathology 5th Edition: Elsevier Academic Press. Burlington, Ma. USA (2005), 79–103.

[2] APSNet. 2019. Resources for Plant Diseases. https://www.apsnet.org/edcenter/ resources/commonnames/Pages/default.aspx

[3] Sanjay B Patil, K Shrikant, and Bodhe . 2011. Betel Leaf Area Measurement Using Image Processing. International Journal on Computer Science and Engineering (IJCSE) 3 (01 2011).

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition. Ieee, 248–255.

[5] Messe Düsseldorf. [n. d.]. SAVE FOOD. ([n. d.]). https://www.messe-duesseldorf. com/cgi-bin/md\_home/lib/pub/tt.cgi/SAVE\_FOOD.html?oid=121&lang=2& ticket=g\_u\_e\_s\_t

[6] Ecosia. 2019. Search Engine. https://www.ecosia.org/?c=en [7] Alvaro Fuentes, Sook Yoon, Sang Kim, and Dong Park. 2017. A robust deeplearning-based detector for real-time tomato plant diseases and pests recognition.

[7] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. 2016. Ssd: Single shot multibox detector. In European conference on computer vision. Springer, 21–37.

[8] Sharada P Mohanty, David P Hughes, and Marcel Salathé. 2016. Using deep learning for image-based plant disease detection. Frontiers in plant science 7 (2016), 1419.

[9] T Mohapatra. 2018. ICAR News July-September 2018. Published in monthly newsletter, https://www.icar.org.in/sites/default/files/ ICARNewsJulySeptember2018.pdf.

[10] E-C Oerke, H-W Dehne, Fritz Schönbeck, and Adolf Weber. 2012. Crop production and crop protection: estimated losses in major food and cash crops. Elsevier.

[11] Government of India. 2019. Kisan Knowledge Management System. https: //dackkms.gov.in/account/login.aspx

[12] Sanjay B Patil and Shrikant K Bodhe. 2011. Leaf disease severity measurement using image processing. International Journal of Engineering and Technology 3, 5 (2011), 297–301.

[13] Sindhuja Sankaran, Ashish Mishra, Reza Ehsani, and Cristina Davis. 2010. A review of advanced techniques for detecting plant diseases. Computers and Electronics in Agriculture 72, 1 (2010), 1–13.

[14] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).