Pakistan Journal of Engineering and Technology, PakJET
Multidisciplinary | Peer Reviewed | Open Access
Volume: SI, Number: 01, Pages: 126- 130, Year: 2020

Cotton Crop Disease Detection using Machine Learning via Tensorflow

Nimra Pechuho*, Qaisar Khan, and Shoaib Kalwar

Electronic Engineering Department, Mehran University of Engineering and Technology SZAB Campus Khairpur, Pakistan
*Corresponding author: Nimra Pechuho (nimrapechuho109@gmail.com)

***Abstract*- World population is expected to be 10 billion in 2050. With more mouths to feed, agriculture needs to boost up to meet the food requirements. However, developing countries like Pakistan has seen a decline in their production of the crops. One of the main reasons behind declined in the production of the cotton crop is the damage caused by cotton diseases. Our model is giving farmers an easy and efficient method to diagnose cotton diseases and will recommend the usage of pesticides. It is based on machine learning, which learns with every use. Agriculture needs innovative ideas to increase its yield. CottonCare (Cotton Crop Disease Detection using Deep Learning via TensorFlow) is also one of the steps to integrate artificial intelligence into agriculture. The goal of this project is to help the farmers in decreasing the production cost and achieving the higher yield, which is also going to contribute to the country's economy.**

***Index Terms*— Deep learning, Disease detection, Image classification, Machine learning, Transfer learning.**

I. INTRODUCTION

Agriculture majorly contributes Pakistan's economy as this sector contributes around 20% of GDP (Gross Domestic Products) and 65% of the country's exports in which cotton and cotton-based products have a significant share of 73%. But the cotton production has been declined in the past few years due to increasing production cost and damage done by the cotton diseases. The farmers face a major challenge is proper diagnoses to save the cotton crop. So, our project will present a technique to detect the insect-based diseases in the cotton crop by deep learning [1-3].

For the excellent production of cotton, farmers need to have a comfortable and less expensive method to detect cotton crop diseases. But the methods used in previous research like SVM (Support Vector Machine), SVM-GA (Support Vector Machine – Genetic Algorithm), means clustering is a low speed, need bulky computational devices and are limited for the number of diseases costly, which increases the overall cost of the system [4-8].

Detection of cotton diseases is done with fewer calculations and execution according to the devices used. The given method/technique has multiple frontends i.e. Could be written in python, C++, JavaScript (JS) etc. the user may note using one's expertise. Growth of economy of the country by increasing cotton production and quality by detecting its diseases and controlling at early stages be checked at the different times of the day, to see how the project is working when the temperature of the day is changed, and the angle of the sun differs [9-11].

A. Applications

This project will provide services to the farmers and will help the cotton industry commercially. It will have many applications in cotton crop. Few are mentioned below:
Industrial: This project will take the cotton industry to automation. The industries must send their field officers to every crop of the country for the Inception. But with our model, the process of diagnosing and recommendation will be done through the cotton care model making process the automated [12].

Commercial: The pesticide companies sell their products through the shopkeepers, which increases the cost of pesticide products. This will give options to the companies to place their products on the cotton care's online store.

Research: This model will be helpful for the researchers who are working on the cotton crop. It will give them valuable information for their research purpose

B. Benefits

This project is going to help the farmers mainly. From increasing the productivity to cost-saving, this model is going to help the farmers of the country mainly.

Cost Saving: The primary reason due to the production of the cotton crop has decreased over time is the high production cost of the crop for the farmers. So, this project is going to help the farmers in solving this problem by giving them an online purchase option from the pesticide's companies.

Time-Saving: The traditional method of diagnosing app, which is done by field officers is time consuming as they come from

other cities, and one city has many fields in it. So sometimes it takes one or more days field officer to go and check the health of the plant. There are few diseases in cotton which can destroy the cotton plant in a day. So, this project will give the diagnosis and recommendation at the same time and the biological and chemical control.

Higher Accuracy: The cotton care model will be more efficient in diagnosing the cotton diseases the traditional methods. The cotton care model is based on machine learning, so it will have higher accuracy than the human experts.

Recommendation System: The model will recommendation and advises on the usage of the pesticides in crop. It will also give the monitoring suggestions from the start for the care of cotton crop until the yield.

Higher Production: This system will ultimately be going to help to increase the production of the cotton crop, which will be beneficial for the farmers as well as for the country.

C. Previous Methods

Crop diseases are an immediate risk to be eliminated. The cotton crop in Pakistan is especially suffering from diseases like bacteria, virus, fungi and insects. Those diseases are affecting efficiency as well as the production of cotton. The method uses CNN to detect diseases. CNN works on the specific regions for pattern recognition. P.R. Ruth and R.V. Kshirsagar introduced a "Cotton leaf disease identification using pattern recognition techniques" which use snake segmentation in which HU moments used for distinctive attribute and average classification is to be 85.52%.

Yogita K Dubey, Milind M. Mushrif & Sonam Tiple proposed a "Super pixel-based roughness measure for cotton leaf diseases detection and classification", which uses superpixel segmentation and SLIC algorithm and SVM for feature extraction. The accuracy obtained was 94% for Alternaria, bacterial and whiteflies' diseases.

Sivasangari, K. Priya & K. Indra introduced "Cotton leaf disease detection and recovery using genetic algorithm" used colour transformation, thresholding and edge detection through the android app and GA-SVM classifier used to detect disease with 97% accuracy on fungal, viral and insect diseases.

Sujatha R, Y Sarvan Kumar and Garine Uma Akhil introduced "Leaf disease detection using Image Processing", that state segmentation/superpixels, conversion to RGB to HIS and for classification K-means algorithm in the last SVM for comparison with given dataset. These steps are carried on MATLAB.

D. Diseases

Bacterial Blight: This disease also known as angular leaf spot. Bacterial blight happens due to bacteria and attacks the cotton crop during mid-season when cotton starts bearing fruit (ball).

Symptoms:

- The shady lesions between the veins in cotton plant.
- It establishes black veins underside of the cotton leaf. The ball becomes water-soaked.
- It creates circles like cigarette burns.

FIGURE 1: Bacterial Blight Infected Leaf

Cotton Leaf Curl Virus (CLCV): Earth's Cotton leaf curl is a diseased caused by the virus classes' begomoviruses. These viruses spread due to whitefly scientific name is Bemisia tebaci. These viruses are icosahedral in shape Fig. 1.

Symptoms:

- Leaf is curled.
- The veins of cotton are swelled.
- Leaf enation or outgrowth of leaf happens means the leaf as appearance occurs on the downward side of the leaf or leaf on leaf structure appears.
- The stunted growth happens means poor yield we get.

FIGURE 2: CLCV Infected Leaf

Whiteflies: Whitefly is an insect of cotton which suck plant juice and produce honeydew, a sticky substance. Due to honeydew fungal diseases appear on leaves and the plant quickly becomes extremely weak and may be unable to carry out photosynthesis. Leave wilting, turning pale or yellow are the

main symptoms of its attack and thus, growth of the plant will be stunted Fig. 2.

Symptoms

- The leaf becomes curly.
- This also spread CLCV.
- The fruit means cotton becomes black if it remains up to this stage.

FIGURE 3: Whiteflies Infected Leaf

Mealy Bug: Diffuse Mealybug bug is a big menace in many parts of the world, particularly in Pakistan where it makes genuine harm an assortment of harvests. This bug assaults more than 300 plant species, including cotton. It starts with patches, and movement can be assisted by wind, rainfall, irrigation machinery, birds, people moving through the crop. Its growth rate is fast, and even one mealybug can destroy a crop of first two months. Mealybug sucks pest from leaf, stem and the fruit and spreads honeydew that resists the photosynthesis process also degrades the lint quality Fig. 3.

Symptoms

- It looks white powder is spread on leaf, stem or fruit.
- Patches of stunted or dead plants noticed far from the field.
- The blackish leaves and stems

Figure 4: Mealy Bug Infected Leaf

THRIPS: Reflected Thrips are insects having wings. They can attack the cotton crop at early stages. They lay eggs are in kidney shape. The female thrips make an opening in plant tissue before laying eggs. After hatching larva begins to feed on plant tissue. There are four stages of thrips cycle. These damage by sucking the pest of the plant and also resist photosynthesis by damaging the cells and tissues of plants Fig. 4.

Symptoms

- Leaf becomes silver-grey
- Leaf weakens and starts falling.
- Leaf cones merge upside.

FIGURE 5: Thrips Infected Leaf

JASSID: Heat is the cotton jassid is a serious sucking pest of cotton in Pakistan. Adults are small, like the tip of a lentil, and flat. Adults are usually yellowish-green or white with black spots on the front wings. They jump and fly away at the slightest disturbance. The cotton jassid sucks sap from the underside of leaves and leaf buds. When jassids are abundant, cotton growth is stunted, the leaves turn downwards, and massive fruit loss may occur on pre-flowering plants. Leaves turn pale, and a rust-red colour develops at the edges. During a severe attack, leaf shedding occurs shown in Fig. 5.

Symptoms

- Yellowing and upward curling of leaves.
- Later brownish discolouration occurs from margins.
- It is shedding of dried leaves shown in Fig. 6.
- Stunted growth.

Figure 6: JASSID Infected leaf

II. METHODOLOGY

Machine learning is used to create models for applications having high accuracy that deals with images, text, video or audio with the help of TensorFlow. TensorFlow is a framework given by Google to work with deep learning. A deep-learning subset of ML whose algorithms are used for extensive data processing that uses CNN. The developers quickly deploy the applications via TensorFlow. Our model uses a transfer learning technique. Transfer learning means that the model is trained already on another problem and retrain it on our data accordingly. Because of this, our time and efforts are saved to start it from scratch. The Deep Neural Network inception-V3 pre-trained model is used trained on the ImageNet dataset. ImageNet is taught in 1000 different classes as Dalmatian and dishwater. Before retraining an Inception V3 it is necessary to run inference on the pre-trained model to understand its input and SoftMax layer.

This image of a panda is included in the Inception of data-file. The Inception model is quite confident that this image shows a panda, with a classification score of about 89% and the next highest score being only about 0.8% for an indri, which is another exotic animal. The output of the Inception model is a so-called SoftMax-function. The SoftMax-outputs are sometimes called probabilities because they are between zero and one, and they also sum to one - just like possibilities. But they are not probabilities in the traditional sense of the word, because they do not come from repeated experiments. It is perhaps better to call the output values of a neural network for Classification scores or ranks, because they indicate how strongly the network believes that the input image is of each possible class.

INFERENCING INCEPTION-V3: Unlike training, inference doesn't re-evaluate or adjust the layers of the neural network based on the results. Inference applies knowledge from a trained neural network model and uses it to infer a result. So, when a new unknown data set is input through a trained neural network, it outputs a prediction based on the predictive accuracy of the neural network. Inference comes after training as it requires a trained neural network model. But the trained model is still on our localhost (laptop) to make it available for client inference it needs to be deployed on a cloud server shown in Fig. 7. This is where services like Heroku kicks in.

RUNNING INFERENCE ON RETRAINED INCEPTION-V3: It's easy to run inference on the retrained model, the following step will be followed to run a simple inference on locally available (laptop) retrained model.

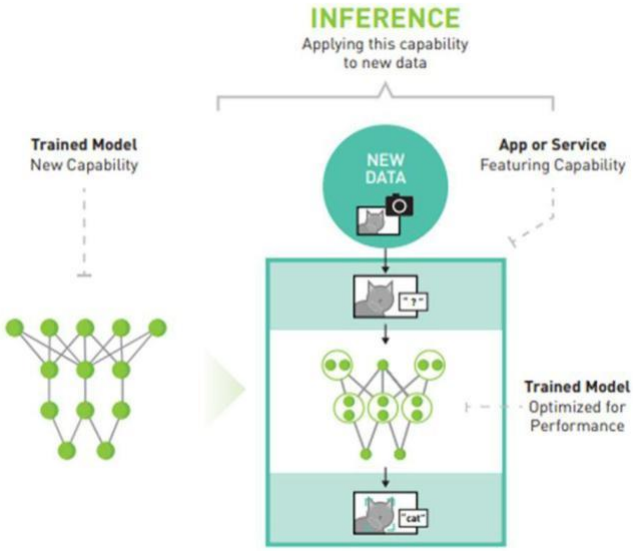

FIGURE 7: High Level Explanation of Inference

III. RESULTS AND DISCUSSION

During training, a single pass of the data through the model is called an epoch. After every single epoch, the same data would be passed repeatedly through multiple epochs. This repeated process is when the model would be learning. The model is first initialized its set with arbitrary weights also recall that we said at the end of the network. There are some important hyperparameters of the convolutional neural network to tune to achieve good performance on disease classification. Results can be improved by altering the details of the learning process.

TRAINING STEPS: The rate of improvement in the accuracy slows the longer model is trained, and at some period it will stop altogether (or even go down due to overfitting), we had 3 experiments 500, 1500 & 1000 to see what works best for our model.

HYPER-PARAMETERS: Certain other parameters can be used to adjust the model to see if they help with results. The learning rate hyperparameter controls the value of weights which get passed to the final layer to be updated while training. Intuitively smaller the learning rate greater time it will take while training, but it can end up helping the overall precision. That's not always the case though, so it must be carefully selected to see what works for our case we have come to find it 0.01. The train batch size, as the name suggests it select training batch, which controls the number of images is weighed during each training step to estimate the updates to the final layer. We had experimented with 5 & 10 images and it comes out 10 images per batch works perfectly with our model.

TRAINING, VALIDATION, AND TESTING SETS: Training data should be split into three distinct data sets these data sets will consist of a training, validation and test set. The training set is usually the largest, which are all the images fed into the

convolutional neural network during training, with the results used to update the model's weights

TRAINING SET: The training set is the set of data that is used to train the model during each epoch (iteration). The model will be trained repeatedly using the same data from the training set and will proceed to learn more about the features.

VALIDATION SET: This dataset is distinct from the training dataset. It is used to validate the model while training to avoid model overfitting. This validation process helps give information that may help in adjust the model hyperparameters. While training, the model will be trained on the training set, simultaneously validating the data in the validation set.

TESTING SET: After the model has been successfully retrained, this set of data is used to test the model. This set is distinct from training and validation sets given in Tab. 1.

Table I: Results of Retraining Model (Accuracy Graph)

Parameters	1st Run	2nd Run	3rd Run
Training Steps	500	1500	1000
Learning Rate	0.01	0.01	0.01
Training Batch Size	10	05	10
Misclassified Pictures	11	9	5
Testing %age	5	5	5
Validation %age	10	10	10
Final Test Accuracy	86.3%	88.6%	91%

II. CONCLUSION

This PAPER aimed to study and understand the diseases in cotton crop and to develop a model based on machine learning which can diagnose the cotton diseases. The work starts with the background and introduction of the cotton crop, the problems the cotton crop is facing and the objectives of our project, which are going to help to solve these problems. A literature review was done after to have a good idea on the recent researches done on diagnoses of the diseases in cotton crop. After the literature, the thesis focuses on the methodology used in making of the model; it includes the project description followed by the components used in and their physical properties. Following that is the experimental result where it concludes that in this thesis work and project designing, the model was successfully developed, which can easily be used on cross-platform. After the designing of the model, the accuracy levels were checked on different cotton diseases. The readings were taken more than ninety percent accuracy level, which shows the efficiency of the project as this model is based on machine learning so it will learn with each new image taken by the model, which will make it more accurate with each use. On the other section, the mobile which based on this will give the option to purchase the pesticides online. After online pesticide shop, the mobile will provide recommendations to the farmers about the pesticide usage, and monitoring advises preventing the diseases in cotton crop.

RECOMMENDATIONS

The world population is increasing with the rate of 1.08% in every year. The demand for food will improve with time. So, we must get maximum production from agriculture fields to satisfy the requirements. Besides, global warming has become a significant challenge world is facing today and agriculture is also one of the causes of global warming. So, to prevent agriculture causing global warming and to increase the production. Monitoring, diagnosing, and fair use of pesticides has become necessary in agriculture. Cotton care project will help in diagnosing the diseases and monitoring of the cotton crop with the help of machine learning. But other crops and plants in the world are also needed to be monitored for better production and less use of chemicals. The artificial intelligence is bringing a revolution in every sector of life, so artificial intelligence should also be integrated with agriculture to get more production from different fields.

REFERENCES

[1] Dubey, Yogita K., Milind M. Mushrif, and Sonam Tiple. "Superpixel based roughness measure for cotton leaf diseases detection and classification." In *2018 4th International Conference on Recent Advances in Information Technology (RAIT)*, pp. 1-5. IEEE, 2018.

[2] Dhinesh, E., and A. Jagan. "Detection of Leaf Disease Using Principal Component Analysis and Linear Support Vector Machine." In *2019 11th International Conference on Advanced Computing (ICoAC)*, pp. 350-355. IEEE, 2019.

[3] Khirade, Sachin D., and A. B. Patil. "Plant disease detection using image processing." In *2015 International conference on computing communication control and automation*, pp. 768-771. IEEE, 2015.

[4] Rothe, P. R., and R. V. Kshirsagar. "Cotton leaf disease identification using pattern recognition techniques." In *2015 International Conference on Pervasive Computing (ICPC)*, pp. 1-6. IEEE, 2015.

[5] Devaraj, Abirami, Karunya Rathan, Sarvepalli Jaahnavi, and K. Indira. "Identification of Plant Disease using Image Processing Technique." In *2019 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0749-0753. IEEE, 2019.

[6] Gupta, Vishal Mani Tiwari&Tarun. "An Exploration on the Identification of Plant Leaf Diseases using Image Processing Approach." (2016).

[7] Zhang SW, Shang YJ, Wang L. Plant disease recognition based on plant leaf image. J. Anim. Plant Sci. 2015 Jan 1;25(3):42-5.

[8] Hang J, Zhang D, Chen P, Zhang J, Wang B. Classification of Plant Leaf Diseases Based on Improved Convolutional Neural Network. Sensors. 2019 Jan;19(19):4161.

[9] Jamil, Mohsin, Asim Waris, Syed Omer Gilani, Bilal A. Khawaja, Muhammad Nasir Khan, and Ali Raza. "Design of Robust Higher-Order Repetitive Controller Using Phase Lead Compensator." IEEE Access 8 (2020): 30603-30614.

[10] Bashir N, Jamil M, Waris A, Khan MN, Malik MH, Butt SI. Design and Development of Experimental Hardware in Loop Model for the Study of Vibration Induced in Tall Structure with Active Control. Indian Journal of Science and Technology. 2016 Jun;9:21.

[11] Jamil M, Arshad R, Rashid U, Ayaz Y, Khan MN. Design and analysis of repetitive controllers for grid connected inverter considering plant bandwidth for interfacing renewable energy sources. In2014 International Conference on Renewable Energy Research and Application (ICRERA) 2014 Oct 19 (pp. 468-473). IEEE.

[12] U. Saleem, M. Khan, and G. Abbas, "Design Layout and Installation Methodology of Cable Trays in a Distribution Substation of Pakistan", Pakistan J Engg & Tech, vol. 2, no. 2, pp. 6-11, Dec. 2019.