ASSIGNMENT OF ADVANCE MACHINE LEARNING

SUBMITTED BY:-
AMANDEEP(11802285)

ROLL NO:-A13
SECTION:-KOM73

SUBMITTED TO:-
Ms Ankita Wadhawan

Git repo -
https://github.com/adeeep/Cotton-disease-prediction.git

INDEX

ABSTRACT

Cotton is one of the economically significant agricultural products in India, but it is exposed to different constraints in the leaf area. Mostly, these constraints are identified as diseases and pests that are hard to detect with bare eyes. This study focused to develop a model to boost the detection of cotton leaf disease and pests using the Transfer Learning technique. To do so, the we have used common cotton leaf disease and pests such as bacterial blight, spider mite, and leaf miner. K-fold cross-validation strategy was worn to dataset splitting and boosted generalization of the CNN model. For this research, nearly 2400 specimens (600 images in each class) were accessed for training purposes. This developed model is implemented using python version 3.7.3 and the model is equipped on the deep learning package called Keras, TensorFlow backed, and Jupyter which are used as the developmental environment. This model achieved an accuracy of 96.4% for identifying classes of leaf disease and pests in cotton plants. This revealed the feasibility of its usage in real-time applications and the potential need for IT-based solutions to support traditional or manual disease and pest's identification.

INTRODUCTION

In India, agriculture is the basis for national economy from which 50% of livelihood and 20% of total foreign trade comes from this agricultural sector . It is believed that Indiaia suitable for many farmable crops, and one among them is cotton. Cotton is also called “White Gold” and “The King of Fibers.” For growers, processors, exporters, and producing countries, cotton is the earnest point of supply. According to the data of Indian report, total 159 million hectares are harvested with the total production value of about $297 billion. Approximately, 18% of crop yield are lost due to different diseases and pests, which result in the loss of millions of dollars worldwide every year.

Even though agriculture is the backbone of India, so far no advanced technologies have been explored in the development of automation in agricultural science and also there are high problems in production and quality due to different diseases and pests. In recent times, the sophisticated emerging technology has attracted many we in the field of detection and classification of cotton leaf diseases and pests. In India, there are several constraints which reduce the yield and quality of the product. Particularly, identification of potential diseases or pests on Ethiopian cotton is based on traditional ways. There is a wide area of farm suitable for cotton plantation, but only limited research attention is given to cotton crop production. Traditionally, experts detect and identify such plant diseases and pests on bared eyes. Bared eye determination is considered as a loss of low-level accuracy in order to detect any diseases. On high demand, different advanced technologies were aided for structuring the systems to assist nonautomatic recognition of the plant diseases and pests to increase the accuracy for any corrective measures. With the help of advanced technologies, the plant diseases were reduced, thus increasing the productivity which helped to raise the economy via boosting the production. For that reason, the implementation of information technology-based solutions in the sector of agriculture had high level of significance for Ethiopia’s development in monetary, community, and eco-friendly developments by increased cotton crops’ productions.

Among different diseases and pests occurred, about 80–90% were on the leaves of cotton. In India, it is observed that there might be a fiscal destruction around 16% because of plant syndromes. However, without control measures, it can cause 30%–50% of loss. Cotton diseases and pests are difficult to identify through bared eyes.

STATEMENT OF PROBLEM

The cotton plant is susceptible to several disorder (biotic and abiotic constraints) attacks due to temperature fluctuation, diseases, and pests. Indeed, the whole world produced nearly 576 kg per hectare of cotton crops, where only 10% of production loss occurred due to different cotton leaf diseases. The United States of America (USA) is a major exporter of cotton in the world and it obtained 5.1 billion US dollars in 2016, but there are well-known native pests which were the reason for the distraction of cotton farms. And, India has 24 percent of cotton land in the world and got 4.6 billions of dollars in 2016, from which generally 18% of cotton crops' production was lost every year due to different diseases that attacked the cotton plants which had its impacts on losing almost nine hundred thousand of Indian rupees . Presently, in India, nearly 12–15% of cotton crop plants are infected due to different diseases . In Ethiopia, performance evaluation of GTP-I showed that these diseases and pests are the main constraints of the world standards in cotton quality and quantity of production. This results in the downfall of the economy of both the farmer and the country.

Detecting these diseases with bare eyes increased the complexity of cotton crops productivity which decreased the accuracy in identification precision. Even an expert would fail to assess and diagnose the diseases with their bare eyes, and this inadequate technique leads to more wastage of cotton crops. Due to these mistaken conclusions, most of the time, certain unnecessary pesticides which badly affect healthy cotton are applied. Leaving the farm for even a short time interval without production will affect the overall nation's GDP.

The we forwarded the following research questions with consideration of the issues cited in the statement of problems:

(1)What is the suitable technique used for diagnosing cotton disease and pests?

(2)How to develop an automatic cotton disease and pests diagnosis system?

(3)How to determine the acquisition of the model?

Deep learning incorporates image processing and data analysis as a path for more possible findings. As it has been a successful application, it has now entered the domain of agriculture. Today, several deep learning-based computer vision applications such as CNN (convolutional neural network), RNN (recurrent neural network), DBN (deep belief network), and DBM (deep Boltzmann Machine) are performing tasks with high accuracy. However, the most prominent application for this research work is CNN

Nowadays, CNN techniques are used to detect different objects and to perform automatic drawings of instructions for analysis purposes. Deep learning draws attention in order to maximize the performances to classify different tasks which help to promise the human intervention data. In this real world, the usage of deep learning shows the major interest for decoding human brain activities. The problem is faced between intertrial and intersubject variability in electroencephalography signals, an indigenous access for attention-based bidirectional long-short-term memory. Convolutional neural networks were analyzed among different factors that are classified into four classes of electroencephalography motor imaginary functions. Here, the usages of bidirectional long-short-term memory with the attention model accomplished the extraction of different features from the raw electroencephalography signals. Advancement of the clinical translation of the electroencephalography motor imaginary-based brain computer interface technology is applicable for varied requests, where this system supports the paralyzed patients. The unusual achievements include the maximum accuracy and time-resolved predictions .

. To make an efficient and effective interface system, the human plays an important role.Graph convolutional neural networks, a novel deep learning framework, addressed the issues in order to differentiate the four-class motor imaginary intentions by mutually agreeing through the similarity of electroencephalography electrodes. To find the motor imaginary, four tasks are preferred with the prediction of highest accuracy .

DATA COLLECTION

For this study, the we has used purposive or judgmental sampling techniques, selecting three infected and a healthy sample from the population, which is nonprobabilistic. During data collection, 2400 images of data are captured and distributed into four equal classes such as bacterial blight, healthy, leaf miner, and spider mite used to train with balanced dataset

Cotton Images' Sample Digitization

The data acquisition system in this research was used with regard to generate clear, unbiased, and simplified digital images of leaf in the cotton plant sample database for further analysis and processing. The aim was to provide the digitizing system with uniform lightning or balanced illumination. The images captured using a smartphone camera and digital camera are then transferred to a computer, displayed on a screen, and stored on the hard disk in the PNG format as digital color images.

Image Data Preprocessing

putting preprocessed images into a network is the first and basic task in all image processing projects. Common image preprocessing tasks in any image processing project are vectorization, normalization, image resizing, and image augmentation. In this research, these image preprocessing tasks are carried out before going to further deep learning processing using OpenCV library in python. Data augmentation is also used to generate more training datasets from the real sets for data samplings

Feature Extraction

Deep learning solves different shortcomings of machine learning feature extraction such as extracting features manually by using the best and robust technique called a

CNN . The layers are used to learn the knowledge. With the use of filtering mechanism the data are used to match and extract their values.

Tool Selection

To collect cotton leaf images for this research, two image capturing devices were used such as a smartphone and digital camera. The proposed model was implemented using python version 3.7.3 for its usages. Also, the model is trained on the deep learning package called Keras, Version: 2.7.0 tf TensorFlow backed.. To evaluate the performance, many experimental setups were conducted with the help of a graphical user interface using Tkinter. From hardware, training and test was carried out on GPU.

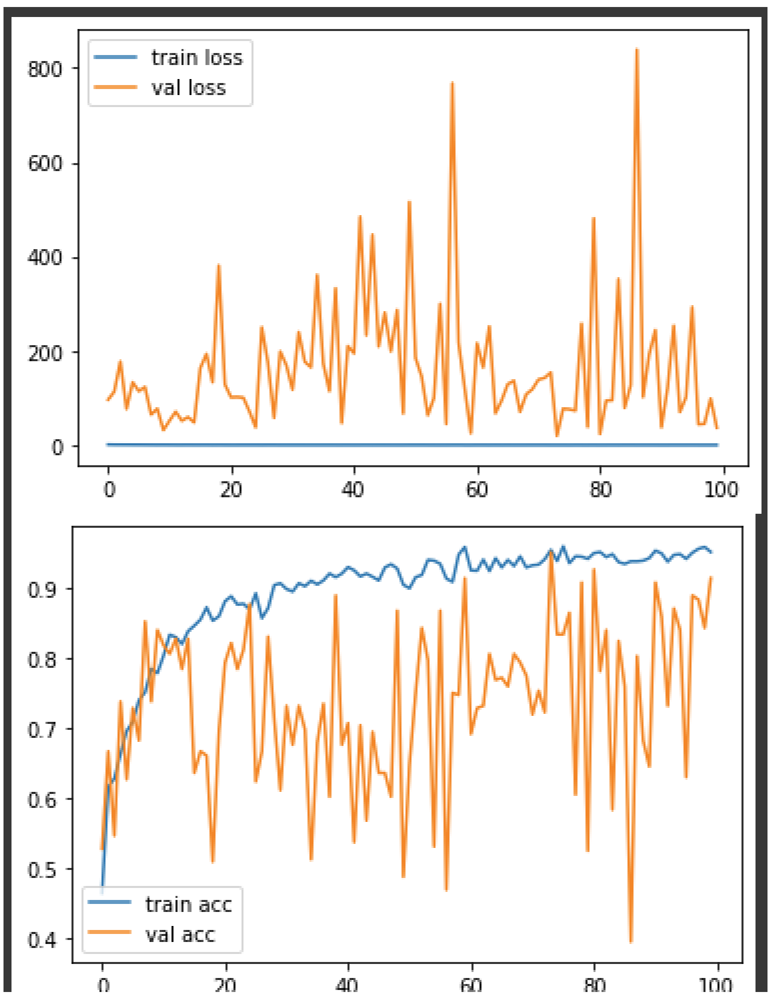
EXPERIMENTAL RESULT

During experimentation, different experiments were undergone to get an efficient model by customizing various parameters that provided different results. Those parameters are dataset color, number of epochs, augmentation, optimizer, and dropout. According to Serawork Wallelign, augmented RGB colored images provided about 15% improvement on accurate than that of not augmented.

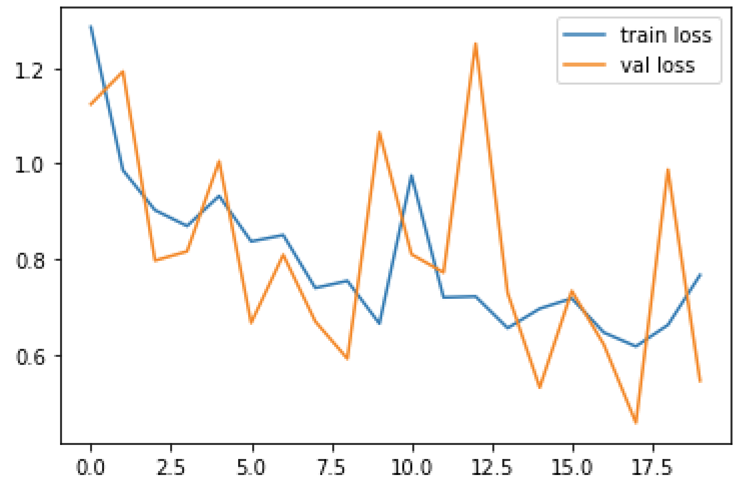
For this experiment we used different transfer learning technique such as ResNet50,Inception V2 and CNN And the model was trained on different epochs such 10,20,50 and 100.

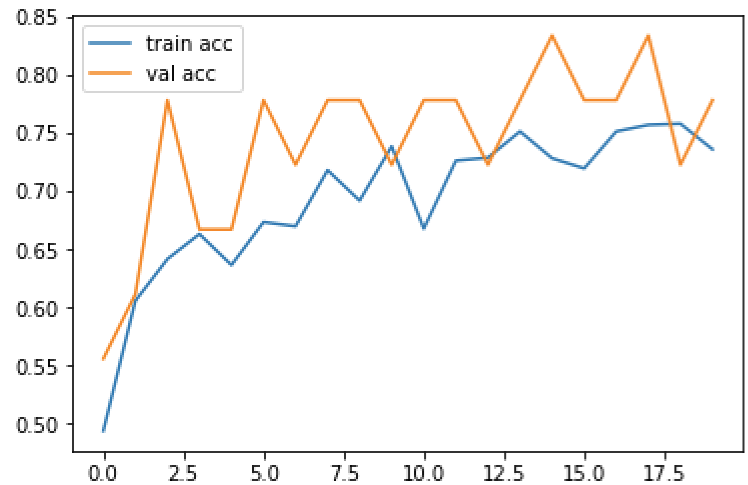
However the model reached the best performance on Inception V3 at 50 epochs and model was 98 percent accurate at identifying the bad and diseased leaves.

CNN:-

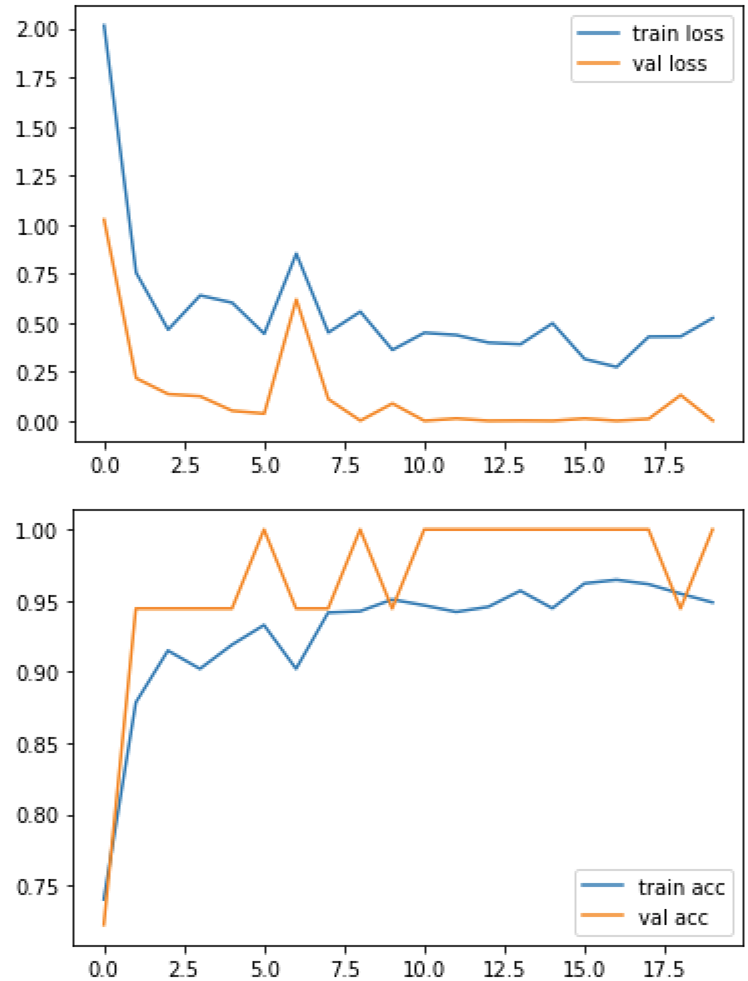

train loss
val loss
800
600
400
200
0
0
20
40
60
80
100
train acc
val acc
0.9
0.8
0.7
0.6
0.5
0.4
0
20
40
60
80
100

Resnet 50:-





Inception V3:-




```
<Figure size 432x288 with 0 Axes>
```

RESULTS AND DISCUSSION

To analyze the performance of the model. RGB-colored image dataset with augmentation provides 15% best performance for the model. We used the transferred learning CNN model and the grayscale dataset achieved 98.6% accuracy. However, color is the main and most decisive feature in cotton detection and classification; therefore, using a colored dataset takes a long time to train the model to add performance even if it is a complex layer. The number of epoch with 50 iterations and the Adam optimization method is very significant to boost the model performance by 10% and 5.2%, respectively. In the end, this developed CNN model achieves 98% of bacterial blight, 94% healthy, 97.6% of leaf minor, and 100% of spider mite, which are correctly classified. Additionally, the We has used different preprocessing techniques for noise removal. The main factors for the misclassification of the result exist between bacterial blight, healthy, and leaf miner. The overall performance of the model, as shown in the confusion matrix, is 96.4% accurate for diagnosis of leaf disease and pests of cotton plants.

CONCLUSION

This deep learning-based model was implemented using Python and Keras packages, and Jupyter was used as a development environment. Different experiments have been undergone in this research study to get an efficient model by customizing various parameters such as dataset color, number of epochs, augmentation, and regularization methods. RGB-colored image dataset with augmentation provided 15% best performance for the model. The numbers of epoch and regularization methods are very significant to boost the model performance by 10% and 5.2%, respectively. The proposed prototype has achieved the highest efficiency of 96.4% for identifying each class of leaf disease and pests in cotton plants. Developments of such automated systems are used to assist the farmers and experts to identify cotton disease and pests by leaf visual symptoms. Obtained results evidence that the designed system for the farmers are much helpful in order to reduce the complexity, time, and cost of diagnosing the leaves from any diseases.

BIBLIOGRAPHY

https://www.researchgate.net/publication/352484184_Deep_Learning-Based_Image_Processing_for_Cotton_Leaf_Disease_and_Pest_Diagnosis

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8124293/

https://www.ijeat.org/wp-content/uploads/papers/v9i3/C5965029320.pdf