Soil classification and plant diseases identification using CNN

Aakash K [a], Ashwath Narayan K S [b]

College of Engineering Guindy
Sardar Patel Rd , Chennai 600025

aakashkumar25112000@gmail.com [a]
ashwath.sharavanan@gmail.com [b]

ABSTRACT

Soil is an important ingredient of agriculture. There are all kinds of soil. Each type of soil can have different kinds of features and different kinds of crops grow on different types of soils. We need to know the features and characteristics of various soil types to understand which crops grow better in certain soil types and can yield better .When we can assess a type of soil and plant relevant or suitable seeds accordingly . One can maximize the yield. This model can help smallholder farmers to classify the type of soil and plant seeds according to it.

" Identify the soil type and plant only relevant seeds in it ".

Plant disease is an ongoing challenge for smallholder farmers , which threatens income and food security.The identification of plant disease is an imperative part of crop monitoring systems. In classifying crop diseases ,the traditional method of human analysis by visual inspection is no longer feasible because it's not easy to classify them .Farmers don't get the exact remedies that can be used to stop those diseases. This machine learning model can identify diseases and also suggest the potential remedies to the user

"Identify any diseases early on and take corrective and preventive measures " .

Index Terms : Soil , Features , Plant Disease , Machine Learning , Convolutional Neural Networks.

---

1. INTRODUCTION

By 2050, global crop production must increase by at least 50% to support the predicted demand . The majority of production currently occurs in Africa and Asia, where 83% of farmers are family run with little to no horticultural expertise . Due to this, yield losses of greater than 50%. Introduction of data mining in the agricultural field has made benefits in the research field.

In this paper, we have proposed a working model that can predict soil series with land type and according to prediction it can suggest suitable crops, further identify plant diseases and provide us with preventive measures . Thus , enabling us to grow crops with better yield. Machine learning is still an emerging and challenging research field in agricultural data analysis

Computer vision and deep learning (DL) techniques have been proven to be state-of-the-art to address various agricultural problems. The recent revolution in smartphone penetration and computer vision models has created an opportunity for image classification in agriculture.Convolutional Neural Networks (CNNs) are considered state-of-the-art in image recognition and offer the ability to provide a prompt and definite diagnosis.

Machine learning techniques can be helpful in this case. The field of machine learning provides well-suited techniques to learn the links between the hyperspectral data and soil texture The crops can be predicted based on a very suitable set of features included in the dataset used , we considerThe Soil dataset , It has images for alluvial, clay, red, and black soils .

We give an overview of the current research in soil classification and plant disease identification. The system design with a split of modules and block diagram is described below. The applied machine learning approaches are introduced with the experimental results . Also contains the evaluation of the different approaches and the result analysis is briefly discussed. Finally, we conclude this study and give an outlook of possible future research ideas.

2. RELATED WORKS

In this section, we briefly review the published research which is related to the presented classification of soil texture and plant disease identification based on hyperspectral data. A first review of geological remote sensing is given by Cloutis (1996). Traditional approaches like nearest mean, nearest neighbor, maximum likelihood, hidden Markov models and spectral angle matching for the classification of soil texture show acceptable results (Zhang et al., 2003, 2005; Shrestha et al., 2005).The increasing popularity of deep learning approaches in many research disciplines has also reached the field of remote sensing. Among different deep learning techniques, the deep convolutional neural network (CNN) has been used mostly for image classification (Krizhevsky et al., 2012; Lu et al., 2017). The CNN model provides a relationship between layers and spatial information of the image and hence it is convenient for the classification of images (Arel et al., 2010). Along this line, there are limited works on plant disease classification using CNN. Lu et al., 2017, investigated the ability of deep CNN technique for classification of different rice diseases. A total of 500 images belonging to 10 categories have been considered and used CNN model with three convolution layers, three stochastic pooling layers and softmax layer at the end. The classification accuracy of 90.48% has been reported.

In our project we have combined the ideas of both the soil classification and plant disease identification and have built a CNN which at one time can predict a soil type and predict suitable recommendations and also identify plant diseases and suggest suitable remedies. The project creates a CNN model , The deep convolutional neural network (CNN) is the most popular and extensively used for image recognition (Lu et al., 2017). It mainly comprises convolutional, pooling and fully-connected layers. The following explains our methodology in this project .

Data pre-processing is crucially important to a model's performance . In this first module the dataset is split into training and testing data and then preprocessing is done i.e.., any background noise or disturbances are removed by augmentation and normalization and the dataset is loaded. After that the important features are extracted which is to be used to train the model. We define the parameters to be used in the CNN model with a series of Convolutional and Max pooling layers with dropouts in between . Now the features extracted from the Extraction module are used to form a model which will be trained using convolutional neural networks to gain knowledge about the dataset, we check for small tweaks in hyper parameters and validate the best suitable accuracy. The dataset will be tested to classify the soil type of the soil based on the information gained from the testing set / will be able to identify diseases if any . Here we build a user interface (application) , such that a common farmer or end user can use the model with ease, In classification , the testing dataset will be able to classify the soil type and suggest suitable crops and provide us about information about the diseases if any present in the plant of the speech. We use a sequential API for model building and create a CNN with layers like convolutional, RelU, Max Pooling, Dropout.We also use softmax function to output a vector.For training the CNN Model, the images in the training set and the testing set are fitted to the sequential model we built using Keras library.

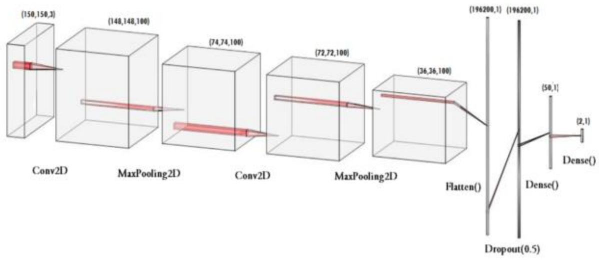


Fig.1 : CNN model

3. SYSTEM DESIGN

This Soil classification and plant disease identification software has two independent models.One is Soil classification model and other is Plant disease identification model. Digital image processing and computer vision approaches can be employed on soil images for classification. Soil texture is a property which has been used in various approaches to determine the soil type mainly clay, alluvial,black and red soil. After the feature extraction, CNN algorithms are used to train models. The trained model is evaluated, tested and saved.The convolution Neural networks is used to train and save the plant model as well .The models are saved and finally it's time to go for general uses.The final block is the interface which uses both models.When soil classification is to be done,a soil image is uploaded, the saved soil model are loaded and the classification is made. The classification gives the suggested crops by referring the dictionary given.When Identify disease is to be done,a leaf image is uploaded, the saved leaf model is loaded and the identification is done

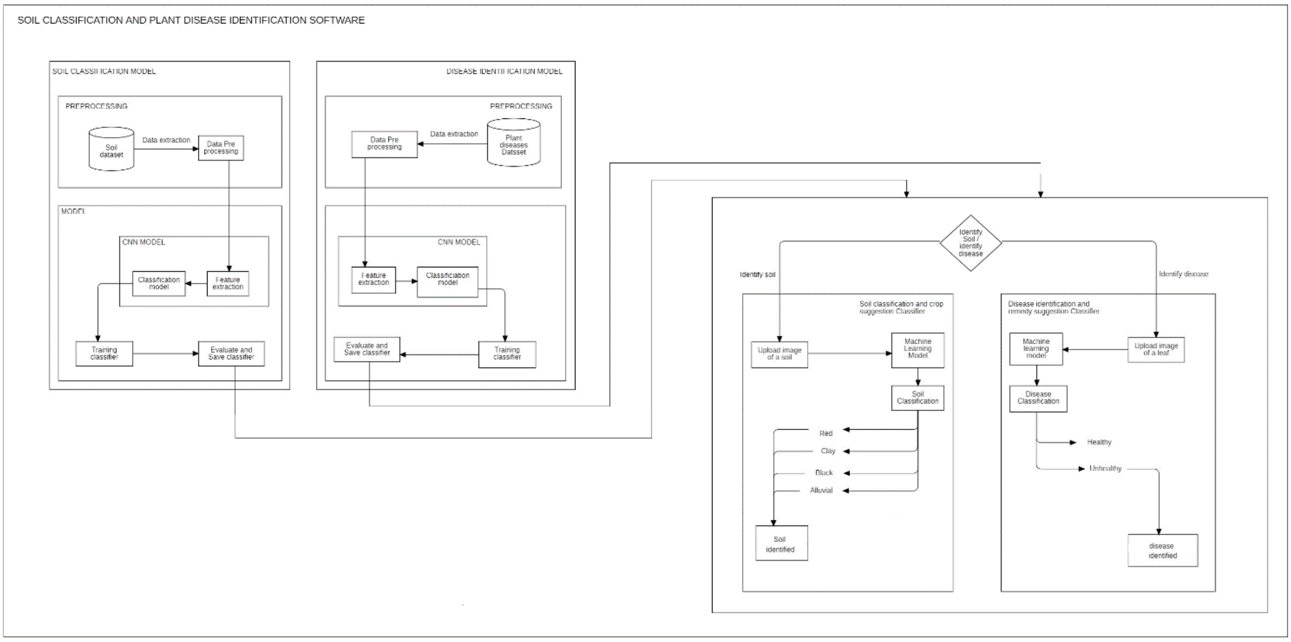


Fig.2 : System Architecture for the project

4. Module split-up

This project has 8 modules, first 4 is related with classification of soil and last 4 is related with plant disease identification.

I SOIL CLASSIFICATION

1 Soil Data set Pre-processing

2 Building CNN MODEL for Soil classification

3 Train ,Test ,Evaluate and save Model

4 SOIL Classification

II PLANT DISEASE IDENTIFICATION

1 Plant disease Data set Pre-processing

2 Building CNN MODEL for Plant disease classification

3 Train ,Test ,Evaluate and save Model

4 Plant disease identification

4.1 MODULE 1 - PREPROCESSING :

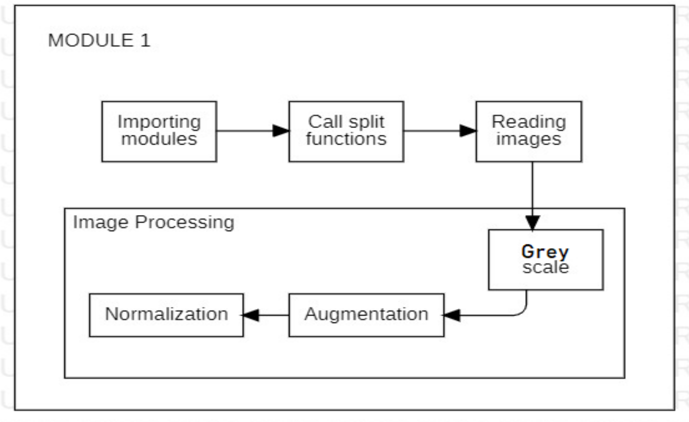


Fig. 3 : Soil and Plant model's - Module 1

Data pre-processing is crucially important to a model's performance .As mentioned in *Fig 3*, the dataset is split into training and testing data and then preprocessing is done i.e.., any background noise or disturbances are removed by augmentation and normalization and the dataset is loaded.

4.2 MODULE 2 – DATA EXTRACTION :

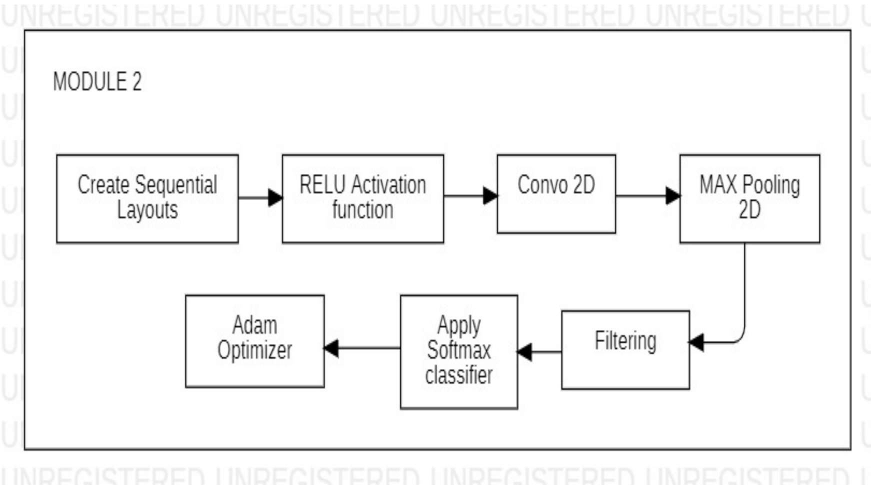


Fig. 4 : Soil and Plant model's Module 2 - CNN model

In this module the important features are extracted which is to be used to train the model. We define the parameters to be used in the CNN model with a series of Convolutional and Max pooling layers with dropouts in between .

4.3 MODULE 3 - TRAINING AND VALIDATION :

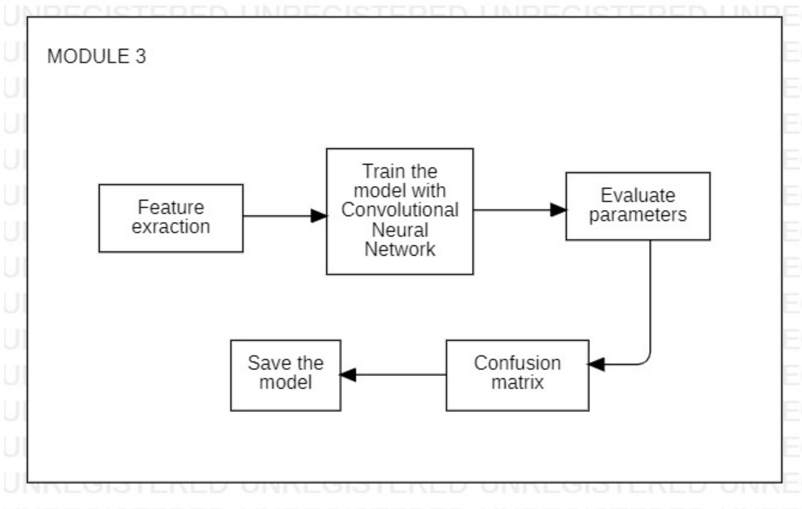


Fig. 5 : Soil and Plant model's- Module 3

In this module the features extracted from the Extraction module are used to form a model which will be trained using convolutional neural networks to gain knowledge about the dataset, we check for small tweaks in hyper parameters and validate the best suitable accuracy.

4.4 MODULE 4 – TESTING AND EVALUATION :

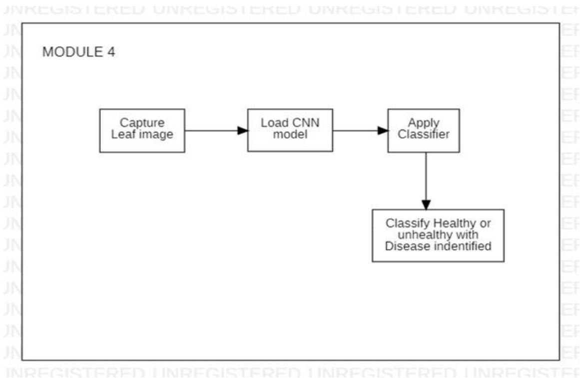


Fig. 6 : Plant disease model's- Module 4

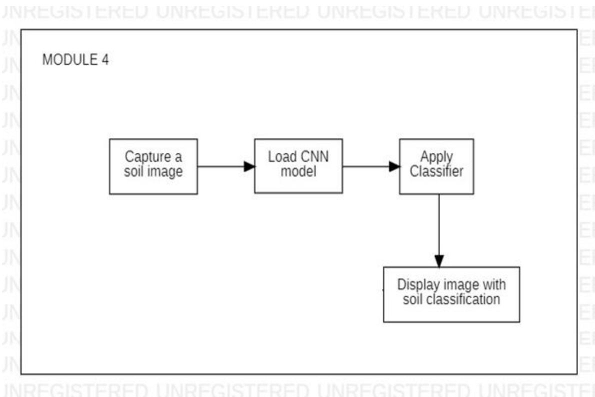


Fig. 7 : Soil model's Module 4

In this module the dataset will be tested to classify the soil type of the soil based on the information gained from the testing set / will be able to identify diseases if any .

5. EXPERIMENTAL RESULTS

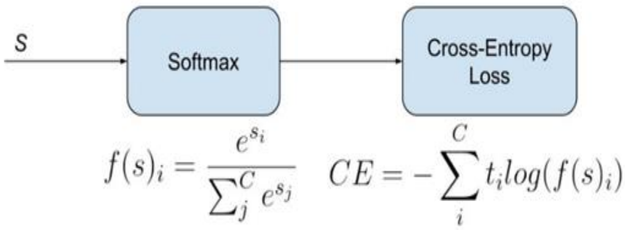
5.1 EVALUATION PARAMETERS :

5.1.1 Accuracy **:**

- Accuracy $= \dfrac{\text{Number of Correct Predictions}}{\text{Total Predictions}}$
- Accuracy $= \dfrac{\text{True positive + True Negative}}{\text{True positive + true negative + false positive + false negative}}$

5.1.2 Loss Function :

It is a Softmax activation plus a Cross-Entropy loss.



$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \qquad CE = -\sum_i^C t_i log(f(s)_i)$$

Fig. 8 : Softmax activation with Cross-Entropy loss

In this loss, we will train a CNN to output a probability over the C classes for each image. It is used for multi-class classification.

5.1.3 Constructing confusion matrix and generating classification report :

- Precision = $\frac{\text{True positive}}{\text{True positive + false positive}}$
- Recall = $\frac{\text{True positive}}{\text{True positive + false negative}}$
- F1 Score = $\frac{\text{2 * True positive}}{\text{2*True positive +false positive + false negative}}$

5.2 CNN LAYERS and their PARAMETERS :

Layer	**Layer shape details**
Optimizer	Adam
Loss function	Categorical cross_entrophy
Metrics	[accuracy]
Conv2D	64 filter,3 x 3filter size,Relu activation function
Max pooling	2 x 2 kernel size

Layer	Layer shape details
Dropout	40%
Conv2D	128 filter,3 x 3filter size,Relu activation function
Max pooling	2 x 2 kernel size
Dropout	30%
Dense	64 neurons,Relu
output	Softmax,4 classes

5.3 SOIL DATASET :

This data-set is created for Soil Type Classification from Image. There are 903 RGB images . The main classifications of the data-set are : "Alluvial Soil", "Red Soil", "Clay Soil" and "Black Soil". The dataset has 2 folders : Test and Train . Train and test , both have 4 main folders representing each classification . Under train , each soil has 180 images and under train each class has 48 images.

5.4 PLANT DATASET :

This dataset is recreated using offline augmentation from the original dataset. This dataset consists of about 87K rgb images of healthy and diseased crop leaves which is categorized into 38 different classes. The total dataset is divided into 80/20 ratio of training and validation set preserving the directory structure. A new directory containing 33 test images is created later for prediction purposes.The dataset has 2 folders : Test and Train. Train and test , both have 4 main folders representing each classification

5.5 EXPERIMENTAL RESULT AFTER TESTING THE MODEL :

The model is set for 100 epoch. Since we have used early call back , it halts where there is a saturation state for the accuracy.Classification reports for the model after running through 25 epochs , the Precision , Recall and f1 score along with their support values for the 4 main soil (alluvial , red , black , clay) listed as 0 , 1 , 2 , 3 are listed.

	precision	recall	f1-score	support
0	0.89	0.81	0.85	48
1	0.94	0.94	0.94	47
2	0.85	0.96	0.90	47
3	1.00	0.96	0.98	46
Accuracy			0.91	188
Macro avg	0.92	0.92	0.92	188
Weighted avg	0.92	0.91	0.91	188

6. Final Results :

6.1 SOIL IDENTIFICATION input and output screenshots :

Fig 9. Input image of a black soil

This is a clear image of a black soil which is taken from our data set and is given as input to the CNN model. There are no filters or disturbance in image , hence a precise output is expected

Fig. 10 : a sample soil image is given as input to the model and output is obtained as BLACK SOIL

6.2 PLANT DISEASE DETECTION tested image :

Fig 11. Input image of a Apple scab disease

In this test case we pass a clear image of a leaf disease , apple scab , taken from our data set. When this is passed to our model we expect a precise output because there are no added filters and image is clear.

TEST CASES

INPUT : leaf image

OUTPUT : disease classification

TEST CASE 1

```
image_path = "leaf/dataset/test/Apple___Apple_scab/test1.JPG"

image = load_img(image_path,target_size=(224,224))
image = img_to_array(image)
image = image/255
image = np.expand_dims(image,axis=0)

result = np.argmax(leaf_model.predict(image))
print("Classification is :", healthType[result])
```

```
Classification is : Apple___Apple_scab
```

Fig. 12 : a sample leafimage is given as input to the model and output is obtained

7. RESULT ANALYSIS

7.1 PERFORMANCE METRICS :

Accuracy Vs Epoch :

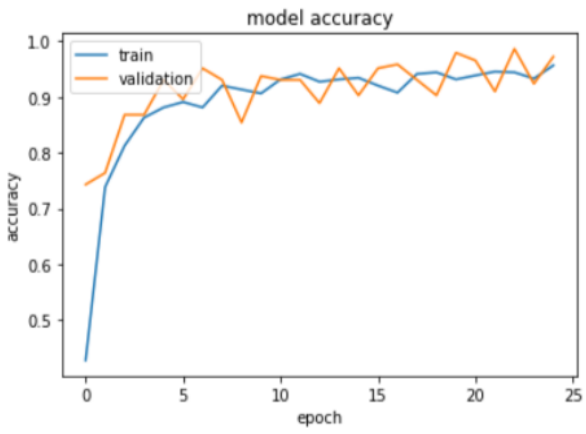


Fig. 13 : Accuracy vs Epoch graph for the model

Fig 13 shows , A graph between accuracy and epoch. The accuracy of the model raises and remains almost unchanged after a particular epoch. The accuracy of the model during training and validating comes out to be 95% and 97%.

Loss Vs Epoch :

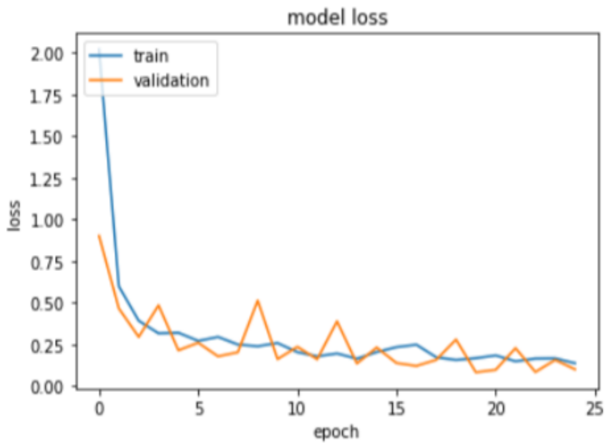


Fig. 14 : Loss vs Epoch graph for the model

A graph between Loss and epoch is plotted in fig 14. The Loss in the model lowers and remains almost unchanged after a particular epoch. The Loss in the model is lowered and an exceptional loss less than 10% is achieved.

7.2 Table of inference :

	Epoch 25	**Epoch 50**	**Epoch 75**	**Early Callback**	**Dropout Layer – 0.4**
Train Accuracy	**93%**	95.8%	97%	89%	85%
Train Loss	**11.9%**	12.8%	4%	26%	36%
Precision	**91%**	91%	94%	87%	85%
Recall	**90%**	90%	94%	86%	82%
F1 score	**91%**	90%	94%	86%	83%
Result	**Better than callback**	Improvement with computation time	Overloaded with more computation	Decent result	More Loss

	Added layer	**Regularized**	**Low train data**	**Existing Model**	**Proposed model**
Train Accuracy	91%	82%	93%		92%
Train Loss	21%	26%	27%		11%
Precision	87%	88%	86%	87%	91%
Recall	86%	87%	84%	84%	90%
F1 score	85%	87%	84%		91%
Result	Train accuracy abruptly increases.	Underfitting	Overfitting		

8. CONCLUSION :

The project aims at creating a model that efficiently classifies the soil instances and maps the soil type to the crop data to get better prediction of suitable crops to be grown with higher accuracies. We have also incorporated the model to predict plant diseases according to the leaf image as input and suggest remedies for the disease identified.

Soil prediction involves types of crop classifications and geographical attributes. It also aims at creating a system that processes the real-time soil data to predict the crops with higher accuracy. The proposed system was developed taking in mind the benefits of the farmers and agricultural sector .The developed system can detect disease in plants and also provide the remedy that can be taken against the disease. By proper knowledge of the disease and the remedy can be taken for improving the health of the plant .

This project has shown the application of an Artificial Neural Network for soil classification and disease identification using a program developed for this purpose. Using actual soil data, it provided an accurate way to classify soils according to series with an accuracy level of 93% and using plant disease dataset , we have achieved an accuracy of 95%. Such high accuracy together with the program developed may provide an important tool for farmers and government agriculture personnel for real world applications .

9. REFERENCES

1. Image-based plant disease identification by deep learning meta-architectures. Muhammad Hammad Saleem , Sapna Khanchi , Johan Potgieter , Khalid Mahmood Arif. Published - 27 October 2020

2. Soil texture classification with 1d convolutional neural networks based on hyperspectral data F. M. Riese , S. Keller Published - 14 June 2019

3. A.-K. Mahlein, T. Rumpf, P. Welke et al “Development of spectral indices for detecting and identifying plant diseases” Remote Sensing of Environment, vol. 128, pp. 21–30, 2020. View at Publisher · View at Google Scholar · View at Scopus

4. Meiqin Zhang, Shanqin Wang*, Shuo Li, Jing Yi,Peng Fu, 2019. Prediction and Map-making of SoilOrganic Matter of Soil Profile Based on Imaging Spectroscopy

5. Cloutis, E. A., 1996. Review article hyperspectral geological remote sensing: evaluation of analytical techniques. International Journal of Remote Sensing 17(12), pp. 2215–2242.

6. Zhang, X., Vijayaraj, V. and Younan, N. H., 2003. Hyperspectral soil texture classification. In: IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 2003, pp. 182–186. Zhang, X., Younan, N. H. and O’Hara, C. G., 2005. Wavelet domain statistical hyperspectral soil texture classification. IEEE Transactions on Geoscience and Remote Sensing 43(3), pp. 615– 618.

7. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, pp. 1097-1105.

8. Arel, I., Rose, C., and Karnowski, T. (2010). Deep Machine Learning - A New Frontier in Artificial Intelligence. IEEE Computational Intelligence Magazine. 5(4), pp. 13-18.

9. Muhammad Hammad Saleem , Johan Potgieter , andKhalid Mahmood Arif . Plant Disease Detection and Classification by Deep Learning. Published: 31 October 2019.

10. Joe G Lagarteja. Android-based Soil Series Classifier Using Convolutional Neural Network. , FEBRUARY 2020.