SOIL CLASSIFICATION AND CROP PREDICTION USING DEEP LEARNING AND MACHINE LEARNING

DHATRIKA BHAGYALAXMI[1], B. SEKHAR BABU[2]

Dhatrika Bhagyalaxmi Research Scholar,, Department of CSE, Scholar Koneru Lakshmaiah Education Foundation, Guntur Andhra Pradesh ,India E-mailAddress: bhagyakmp519@gmail.com

B. Sekhar Babu , Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India E-mailAddress: sekharbabu@kluniversity.in

Abstract – Image classification is a significant application in the domain of Machine Learning and Deep Learning. This has been carried out in various domains. The work in this paper is the application of a deep learning technique for soil classification and a machine learning technique for crop prediction. The primary focus of agricultural soil monitoring for increasing crop production efficiency is on the cultivation and enhancement of versatile soil parameters. Population pressures, land constraints, and the abandonment of traditional soil management approaches have all contributed to a decline in soil nutrients in developing nations such as India. Crop health is an important part of today's agriculture's productive system. Implementing the appropriate crop health business strategy can result in a significant rise in crop production. Efficiency could be accomplished through efficient soil resource supervisors and corrective energy and nutrient application measures. The crop forecasting and classification of soil concerns are effectively managed in this "soil classification and crop prediction," a Machine and Deep learning-based web application. It is implemented with two algorithms, the VGG-19 network for Soil classification and the KNN model for Crop prediction.

Keywords: Machine Learning,Deep Learning,VGG-19,KNN

1. INTRODUCTION

Images are easy for humans to classify with a glance, but the same is not true in the case of machines. Computers cannot classify images as easily as humans do. They do not understand the image as it is. It takes a lot of computation to make the computers understand the class to which a particular image belongs. It is precisely what Image classification is about. It is a technique where computers can analyze an image and tell which class they belong to. A class is anything, a car, an animal, or a type of soil in this regard. The simplest example is if you give an image of a dog, the computer can tell you it is a dog. But the way the computers can do this is interesting. Because the concept of Moravec's paradox is true for machines, which states that things which are easy for humans are difficult for Artificial Intelligence to do. The same goes for image classification. It is done by the computers based on the raw pixel data. Every image is broken down into many pixels. Images are mathematically interpreted by computers to find the underlying patterns. The biggest challenge is that no two pictures are the same even though they look like one. They vary in sizes, perspectives, colors, hues and shades, illumination, positioning of objects, angles, poses, and a wide variety of other factors. This is where Machine Learning and Deep Learning techniques play a prominent role. Deep Learning, in particular, provides better accuracy in classifying images with the help of Neural networks [1].

Neural networks consist of different layers. They have an input layer, an output layer, and many hidden layers as shown in Fig 1 [2].

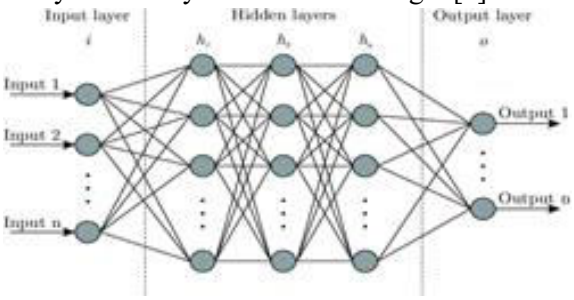

Fig 1: Deep Neural Network

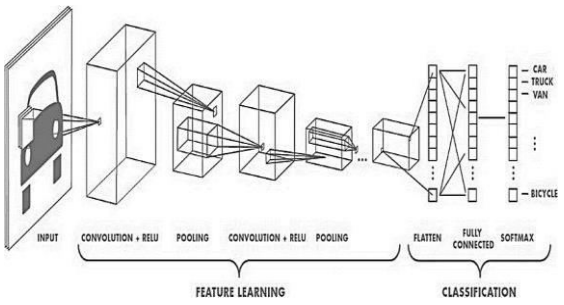

Fig 2: Architecture of CNN

The Convolutional Neural Network (CNN) shown in Fig 2 [3] is an important Deep Learning algorithm that is prominently used for image classification. In this, an image is taken as input and able to differentiate them from one another. One of the major advantages of CNN is that the pre-processing required is also very less. There are different layers in the Convolutional Neural networks. Convolution layers, Pooling layers, RELU correction layers, and Fully Connected layers are particularly important. The Convolution layer is the first layer. It finds the set of features from the input. Convolution filtering is used to do this. It moves like a sliding window on the image and calculates the convolution product. The Pooling layer is present between two convolution layers. This layer reduces the number of parameters and calculations. The efficiency of the network is improved and over-learning is avoided. The RELU layer is an activation function. The negative values received as inputs are replaced by zeroes. The Fully Connected layer is the last in the network. This layer produces an output vector after receiving an input vector. All these layers perfectly work together to provide the best results. There are various architectures of CNN that provide much better results [4]. They are LeNet, AlexNet, VGGNet, GoogleNet, ResNet, ZFNet, etc. In this paper, the VGG-19 Network, a variant of the VGG Network is used to classify the soil images.

The rest of the paper is organized into different sections. Section 2 explains Related work. Section 3 of this paper explains the methodology implemented. In the next section 4, the results are given. Section 5 concludes the paper.

2. RELATED WORK

In this section, we will look at some of the important crop estimation research that has been done in crop production. The publishers of [5] focused on the use of data mining algorithms in crop monitoring. Because data mining is a new emerging technology, the authors investigated the problem of forecasting agriculture production. The authors discussed how the main goal of this work was to find the desired database systems that provide pinpoint efficiency and generality in terms of yield overall. The authors evaluated these various types of data mining algorithms on different datasets.

The authors proposed a process called the Crop Selection Procedure (CSM) [6]. The researchers describe a method for determining crop challenges, increasing yield potential net yield rates over weather conditions, and maximizing economic growth. The novelists looked into the different influential dimensions that can be used by various predictive model types for crops. The authors also describe algorithms and various Machine Learning methods. Crops were defined as seasonal crops, entire crops, short-term horticultural products, and long-term plantation crops in the proposed Agriculture Selection. The planned method employs a support vector algorithm to generate the fully reflects probability locations. The pixel-wise probability maps obtained are refined using the proposed KNN median filter, which is based on matching but instead approximating non-local areas of the city. The model does not require any special segmentation or efficiency strategies because it employs the stochastic basic concept of real images via KNN and provides competitive categorization with calculation. Several tests were carried out on multispectral satellite data sets, and the results reveal that the improved method's class labels are exactly equivalent to the many proposed new spatial-temporal image classification models [5].

Soil classification has been the subject of fascinating science and engineering. The

engineers must have some basic information about the type and nature of the soil. Previously used soil classification techniques included the normal penetration test (SPT), uniaxial compressive test (CPT), tension meter test (PMT), but instead vane shear experiment (VST). For precise information, the initial processes were time - consuming process and required the occurrence of an expert.

With the advancement of machine learning techniques, many methods have been developed over time as part of Machine Learning (AI) to simplify this same soil classification. Decision Trees (DT), k-Nearest Networks (k-NN), Artificial Neural Networks (ANN), and Support Vector Machines (SVM) are among the technologies developed. Soils are classified using machine learning based on discernable features including soil water content, soil nutrients, soil properties, crop yields, soil pH, and soil structure.

Land classification using image processing based on remote sensing uses pixels from the image to describe land features. Land can only be classified using this method according to its use, such as agricultural, forest, or urban. For such classification, different algorithms and methods are used. The Fuzzy based approach from [7] is one such example. It has been observed that the fuzzier is fixed between the inputs and the outputs. The Binary classification methods also exist, which can be considered an easier approach. But, Statistical binary classification is defined as a problem domain with only two categories available for classification. SVM is also popular.

2.1 Disadvantages of Existing System

The time and space complexities for Binary Classification are less, but they do not seem fit for real-world scenarios. While fuzzy approaches are more accurate, designing them with a rule base structure is complex and time-consuming [8]. The remote sensing methods may be suitable for the broad classification of the land but are not suitable for soil classification. Another popular approach of Support Vector Machine (SVM) requires large space and a long time for computation.

3. PROPOSED METHODOLOGY

In the proposed system machine learning and deep learning methods are used. Based on pre-trained data, the machine can predict the type of soil and predict the crop.

CNN-based VGG 19 model is used to train soil images and used to predict the type of soil.

KNN algorithm is used to train soil parameters to predict the type of crop.

3.1 Proposed System Design

In this project, there are five modules and each module has its own functions, such as:

1. Data Collection
2. Preprocessing
3. Train Test Model Fit
4. Classifiers
5. Flask Framework

3.1.1 Data Collection

For training and testing the model, data from different sources is collected. We collected a total of 4 types of soil class images with each image of 30 images. For training purposes, 80% of images of each class are used and the rest of the images are utilized for testing purposes. For crop prediction temperature, humidity, potassium, nitrogen, and phosphorus values are taken as features and the type of crop as the label.

3.1.2 Preprocessing

The images in the dataset are preprocessed before going to the next step. In the preprocessing step, the image is transformed into a grayscale image because the RGB color image contains so much redundant information that is not necessary for soil detection. The RGB color image contains 24 bits per pixel. In contrast, grayscale images contain 8 bits per pixel and are sufficient for classification. Then, we reshaped the images into (64×64) shape to maintain the uniformity of the input images to the architecture. Then, the images are normalized. It improved the learning algorithm's performance by allowing it to learn more quickly and capture the required features in the images.

3.1.3 Train Test Model Fit

We have now divided our dataset into training and testing sets. Our goal in doing this split is to

evaluate the effectiveness of the algorithm on unknown data and to examine how well it has specialized in learning algorithm. Following that, a model fitting is performed, which is an important phase in model development. Where the model is saved in the system for soil prediction

3.1.4 Classifiers:

The two models implemented are CNN and KNN. CNN VGG-19 model is implemented to classify which class of soil the image belongs to, whereas the KNN model is used to predict which class of crops to plant for the soil.

VGG-19: VGG is an advanced and powerful CNN Architecture that is based on deep learning. It is a successor to AlexNet Architecture. It is one of the best models to classify images. The Architecture of VGG-19 is shown in Fig 3 [9].

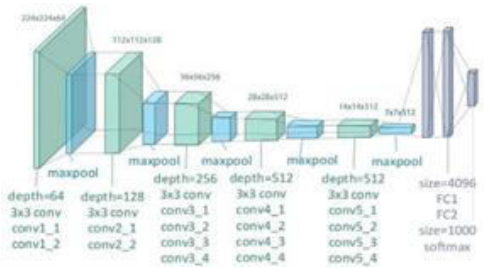

Fig 3: Architecture of VGG-19

The VGG-19, with 19 layers is a variant of VGG and an extension of VGG-16 which is 16 layers deep. It has 16 Convolutional layers, 3 Fully Connected layers, 5 MaxPool layers, and 1 SoftMax layer. The primary objective of this method is to classify the images of soil.

K Nearest Neighbors (KNN): KNN is one of the simplest algorithms used for classification. It is shown in Fig 4

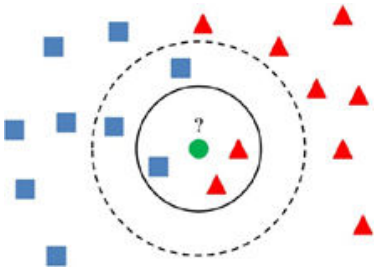

Fig 4: K Nearest Neighbor

Available data is stored in KNN and a new data point is classified based on similarity.

The training data is provided and it is classified into groups identified by the attribute. The objective of this algorithm is to train the soil parameters to predict the type of suitable crops.

3.1.5 Flask Framework:

In this step trained model is deployed into the flask framework and a web application is developed. In the first step, the user registers and login to the application uploads the image, and predicts the result by loading the trained model. The output of soil type is taken as input and soil values are displayed to the user which is given as input to the KNN algorithm which was trained with crop values and the result of the type of crop is displayed to the user.

4. ARCHITECTURE

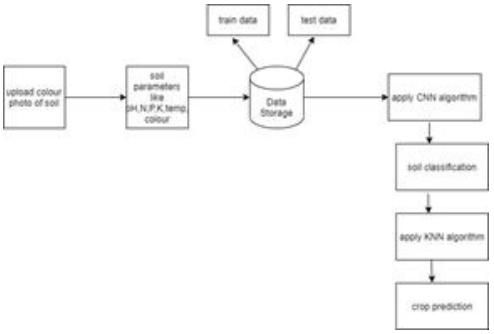

Fig 5: Framework of system

This System falls under the client-server architecture which provides a suitable crop for the farmer. This system is composed of several functional components of login, registration for the farmer, soil classification, and crop prediction modules. The farmer is the user of our website. First, he/she logs into the system then uploads a colour photo of the soil, and waits for the suitable crop to be displayed as output, then the farmer needs to select one crop among them. After the successful completion of actions, the farmer needs to log out of the system. For the soil classification module, after the farmer uploads the color photo of the soil, apply the CNN algorithm (classification technique) for soil classification. For the crop prediction module, after extracting soil features. apply the KNN algorithm (prediction technique) for crop prediction.

5. RESULTS

The Accuracy of the implemented Algorithms is calculated to measure the performance and the ability of their functioning.

5.1 CNN (VGG-19) Algorithm Accuracy:

The Accuracy of the CNN (VGG-19) Algorithm is as follows:

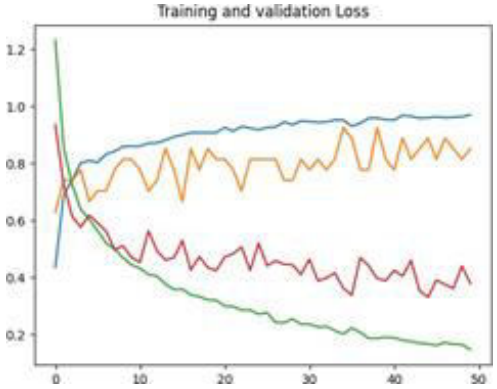

Fig 6: Accuracy of CNN (VGG-19)

Validation Accuracy: 0.8519
Validation Loss: 0.3791
Training Accuracy: 0.9705
Training Loss: 0.1486

Thus, the model is having 85% accuracy in classifying the soil images.

5.1.2 KNN Algorithm Accuracy:

The Accuracy of the KNN Algorithm is as follows:

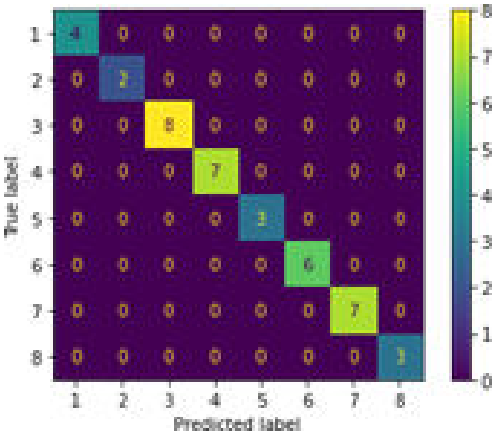

Fig 7: KNN Confusion Matrix

	precision	recall	f1-score	support
1	1.00	1.00	1.00	4
2	1.00	1.00	1.00	2
3	1.00	1.00	1.00	8
4	1.00	1.00	1.00	7
5	1.00	1.00	1.00	3
6	1.00	1.00	1.00	6
7	1.00	1.00	1.00	7
8	1.00	1.00	1.00	3
accuracy			1.00	40
macro avg	1.00	1.00	1.00	40
weighted avg	1.00	1.00	1.00	40

Fig 8: KNN Classification Report

The Accuracy Score is 1.0.

Thus, the model is having 100% accuracy in crop prediction.

6. CONCLUSION

Soil testing is a technique of gaining access to soil constituents that are useful in determining soil fertility and acidity. This system aids in determining the type and pH of the rhizosphere that must be implemented. Farmers' importance is on soil pH and nutrient value because plant and vegetable growth depends on the pH factor present in the soil. Soil pH and soil nutrients are typically measured manually in government laboratories. There are numerous labs, but the equipment required to accomplish this isn't available. As a result, we must use digital image processing to calculate soil pH and nutrients.

Soil classification is one of the important aspects of agriculture work, construction work, or any other line of work. There are many existing classification systems using Decision trees, Naïve Bayes Classifiers, Support Vector Machines, and different Neural Networks. This paper implemented the classification and prediction techniques using CNN and KNN. There is a scope for much more advancement and the arrival of new techniques in this domain.

7. REFERENCES

- Krishna, M & Neelima, M & Mane, Harshali & Matcha, Venu. (2018). Image classification using Deep learning. International Journal of Engineering & Technology. 7. 614. 10. 14419 / ijet. v7i2. 7.
  10892.
- Bre, Facundo & Gimenez, Juan & Fachinotti, Víctor. (2017). Prediction of wind pressure coefficients on building surfaces using Artificial Neural Networks. Energy and Buildings.158.10.1016/j.enbuild.2017.11. 05.
- Alamsyah, Andry & Apriandito, Muhammad & Masrury, Riefvan. (2019). Object Detection Using Convolutional Neural Network To Identify Popular Fashion Product. Journal of Physics: Conference Series. 1192. 012040. 10.1088/1742- 6596/1192/1/012040.
- K. Chatterjee, M. S. Obaidat, D.

Samanta, B. Sadoun, S. H. Islam, and R. Chatterjee, "Classification of Soil Images using Convolution Neural Networks," 2021 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), 2021, pp. 1-5, DOI: 10.1109/CCCI52664.2021.9583192.

- S.Pudumalar, "Crop Recommendation System for Precision Agriculture",2016 IEEE Eighth International Conference on Advanced Computing (ICoAC).
- Rakesh Kumar, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique"2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy.
- B. Bhattacharya, and D.P. Solomatine "An algorithm for clustering and classification of series data with constraint of contiguity", Proc. 3T"d nt. Conf: on Hybrid and Intelligent Systems, Melbourne, Australia, 2003, pp. 489-498.
- Gordon, A.D. "A survey of constrained classification", Computational Statistics & Data Analysis, vol. 21, pp. 17-29, 1996.
- Zheng, Yufeng & Yang, Clifford & Merkulov, Aleksey. (2018). Breast cancer screening using convolutional neural network and follow-up digital mammography. 4. 10.1117/12.2304564.
- Alaliyat, Saleh. (2022). Video based Fall Detection in Elderly's Houses.
- Zeel Doshi, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms" in 2018 IEEE.
- Rohit Kumar Rajak, "Crop Recommendation System to Maximize Crop Yield using Machine Learning technique", International Research Journal of Engineering and Technology (IRJET) e- ISSN: 2395-0056 Volume: 04 Issue: 12 Dec-2017.
- Nidhi H Kulkarni, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique" 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018 ISBN: 978-1-5386-6078-2
- Lakshmi. N, "Crop Recommendation System for Precision Agriculture," International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887 Volume 6 Issue V, May 2018
- J.K.Periasamy and B.Latha, 2020, An Enhanced Secure Content De-duplication Identification and Prevention (ESCDIP) Algorithm in Cloud Environment", in Neural Computing and Applications (in Springer), ISSN Electronic 1433-3058 | ISSN print: 0941-0643, Impact Factor - 4.664, Volume 32, Issue 2, pp.485-494,January 2020.
- Periasamy J.K, 2018, "Probabilistic Broadcast Protocol for Data Gathering using neighbor nodes in Mobile IoT", in Indian Journal of Public Health Research & Development Volume. 9, Issue 3, 2018 pp.779-782 Print ISSN:0976-0245. Online ISSN: 0976-5506.
- Alberto Gonzalez-Sanchez, "Predictive ability of machine learning methods for massive crop yield prediction",Spanish Journal of Agricultural Research 2014 12(2): 313-328 ISSN: 2171-9292

Thenmozhi, K., & Srinivasulu Reddy, U. (2019). Crop pest classification based on deep convolutional neural network and transfer learning. Computers and Electronics in Agriculture,164,104906.https://doi.org/10.10 16/j.compag.2019.104906