IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

INTELLIGENT RECOMMENDATION SYSTEM FOR AGRICULTURAL CULTIVATION

G. Sai Krishna[1], D. Calvin Antony[2]and, Dr. G. Sivashankar[3]

[1]*Department of Computational Intelligence,*
SRM Institute of Science and Technology, Chennai, India
[2]*Department of Computational Intelligence,*
SRM Institute of Science and Technology, Chennai, India
[3]*Department of Computational Intelligence,*
SRM Institute of Science and Technology, Chennai, India

Abstract

When growing crops, farmers face a number of difficulties, including erratic irrigation, subpar soil, etc. A significant portion of farmers, particularly in India, lack the knowledge necessary to choose the right crops and fertilizers. Additionally, both farmers and consumers suffer significant losses when crops fail due to disease. The use of Deep Learning has not been adequately investigated, despite recent advancements in the automated identification of certain diseases utilizing Machine Learning approaches. Such models are also difficult to employ due to the good quality data used in their learning, a absence of processing capacity, and the algorithms' weak generalization capabilities. In order to do this, we develop an open-source, user-friendly online application to address some of these problems.

1. INTRODUCTION

We are all aware that farming is the backbone of the Indian society. Indian agriculture is very significant. We are all aware that agriculture is the backbone of the Indian society. Indian farming is very significant. More over 60% of the land is chosen for farming, which provides food for 1.3 billion people.. Farming is the growing of plants and animals. Agriculture aided in the development of India's civilization. We require soil in order to cultivate crops. A crucial component of agriculture is soil. To produce food efficiently, the soil must be in good condition. It provides the roots with support, water, oxygen, and essential nutrients. The soil on which all plants utilized in food production are cultivated is also the system's base. There are numerous varieties of soil in India.. They are red soil (corn), black soil (sugarcane, sunflower), and alluvial soil (cotton, rice).To automate complex processes, artificial intelligence systems are used in a similar manner to people. The foundation of machine learning is data, such as records of transactions, people, or images. The data is collected and evaluated to act as training material for the machine learning system. With more data, the software produces better results. The system is then trained to recognize patterns or forecast outcomes on its own by the developer, who then selects a machine learning (ML) model to use, inputs the data, and runs the model. Farmers can utilize the environmental data collected by remote sensors to guide their decisions and keep an eye on their crops once it has been processed by algorithms and statistical statistics. The number of inputs and statistical data collected boosts the algorithmic rule's prediction power. Farmers are expected to use these technologies to improve the quality of their produce by choosing better fields. The primary system will determine the most productive crop type that the farmer should produce to get the most yield from the crop type that is farmed in a home garden or the respectable land area by utilizing the system of temperature, soil hydrogen ion concentration, and soil moisture detection. The data that was gathered is processed. with a centralized database that is connected to several research modules is given an explicit algorithmic rule.

Artificial intelligence systems are employed similarly to people to automate complicated activities. Data, such records of transactions, individuals, or photos, is the basis of machine learning. The data is gathered and evaluated to act as learning material for the ML system. With more data, the software produces better results. The system is then trained to recognize patterns or forecast outcomes on its own by the developer, who then selects a machine learning (ML) model to use, inputs the data, and runs the model. Farmers can utilize the environmental data collected by remote sensors to guide their decisions and keep an eye on their crops once it has been processed by algorithms and statistical statistics. the bigger the statistical inputs.

3. BLOCK DIAGRAM

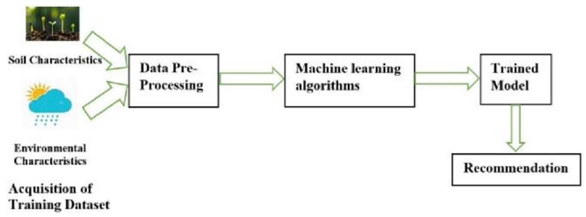

Fig.1. Crop and fertilizer recommendation

4. REQUIREMENT

4.1. CROP RECOMMENDATION

The authors test with Deep Neural network, SVM, Regression trees, and other methods before determining that, for their dataset, Random forests produce the best crop choices. Additionally, they create a system for mobile applications that collects crop yield estimates for particular crops based on GPS location data and suggests crops based on inputs like area and oil quality. trees chosen randomly to suggest crops.

4.2. FERTILIZER RECOMMENDATION

Numerous studies have been conducted on fertilizer recommendations, and the majority of them make use of the soil's N, P, K, and pH values as well as its depth, weather, location, and precipitation. Although rule-based classification is the standard method, some approaches also cluster fertilizer data using K-Means and Random Forests to make recommendations. Plant disease detection has been a very active area of study, and many various methods have been put forth through time, the most recent ones utilizing deep learning techniques.

4.3. DATASET

This study's dataset was obtained via Kaggle. It is a crop suggestion dataset that provides details on different crop kinds and the characteristics that determine which crop is ideal for growing

5. MODULES

5.1. DATASET COLLECTION

The gathering of data enables us to preserve a history of previous occurrences so that you can utilize data analysis to discover recurring tendencies. You can build predictive models from these patterns that follow trends and predict future changes using machine learning techniques. High-performing predictive models must be built using efficient data collecting methods since The effectiveness of predictive models relies on the quality of the underlying data. The data must contain reliable information that is relevant to the task at hand (waste in, garbage out). For instance, a debt default model might benefit over time from increased gas prices but not from an increase in the number of tigers. We gather the crop recommendation data for this module from the Kaggle dataset archives.

5.2. PREPARATION OF DATA

5.2.1. DATA CLEANING

In each machine learning project, data cleansing is a crucial step. In this module, data cleaning is done to get the data ready for analysis by getting rid of or changing the data that might be inaccurate, incomplete, duplicated, or formatted incorrectly. You can study your tabular data using a variety of statistical analysis and data visualization approaches to spot potential areas that require data cleaning activities.

5.3. FEATURE EXTRACTION

This is done to reduce the number of attributes in the dataset, which has the advantage of speeding up training and increasing precision. Machine learning, pattern recognition, and image processing all start with a no of measured data and utilize the feature selection approach to generate values that are destined to be functional and essential. This technique hastens adaptation and learning, and in some situations, it has resulted in expanding interpretations. Dimension reduction and feature extraction are connected. When the input data is redundant and deemed to be too vast to scan the input data can be condensed into a reduced feature set.

5.4. MODEL TRAINING

The machine learning model was implemented with a dataset alluded to as a learning model. It offers number of appropriate type in data that impacts the output and examples of the result. In order to contrast the refined output to the selected output, the learning model is pre-owned to run the selected data through the model. The correlation's insights are simulated to enhance it. Model installation is the name of this evolutionary algorithm. For the model to be correct, the learning dataset or authenticate dataset must really be precise. An ML algorithm can noticed and find the fine values for all correct variables when given data during the model training stage of machine learning. There are lots of types of machinery.

5.5. TESTING MODEL

Using the test dataset, we apply the newly learned machine learning model to the test in this module. Model testing is a process of assessing how a fully trained model will perform on a testing set. In machine learning, the desired behavior and data are habitually input by a programmer, and the machine typical developed the understanding. The primary purpose of machine learning analysis is to validate whether the learnt logic will maintain fairly representative of the number of programmers we summon for the testing.

5.6. PERFORMANCE EVALUATION

In this session, using performance evaluation metrics including F1 score, accuracy, and classification error, we assess the effectiveness of trained machine learning models. If the model doesn't perform well, we increase the machine learning algorithms' performance. An essential component of machine learning is performance evaluation. But it's a difficult task. As a result, thorough research must be done in order for machine learning to be used to radiation oncology or other fields in a reliable manner. Accuracy, precision, and recall are used to assess a classification model for crop recommendation.

5.7. PREDICTION

While assessing the probability of a particular outcome, like customer or not might stir in 30 days, the output of an model that has been learned on past data and then applied to new data is referred to as a "prediction." Every row in the new data will have possible values for an unrevealed feature effectiveness of the algorithm, allowing the model's developer to select the value that is most likely to be given to that feature. Expectations might be misleading. For instance, when you utilize machine learning to find out the optimum course of action for a marketing campaign, you are successfully forecasting a future development. The "prediction" may, however, relate to situations that have already occurred, as to whether a transfer was dishonest or not. In this case, the transaction is already over, but you're still trying to figure out which one was legitimate so you may choose what to do next. We use a trained and enhanced machine to determine whether the crop recommendation is correct.

5.8. RANDOM FOREST ALGORITHM

Favored ML model random forest is a part of the supervised learning approach. It can be utilized to ML matters surrounding both classification and regression. Its foundation is the idea of ensemble learning, a method for merging several classifiers to handle complicated issues and improve model performance. Random Forest, as the name suggests, is a classifier that maximizes the estimated accuracy of the dataset by comparing number of decision trees applied to different subsets of the data collected. Instead of trusting merely on one decision tree, the random forest uses prediction depends on the votes of the more no of outcomes from every regression tree and forecasts the output. the larger the population.

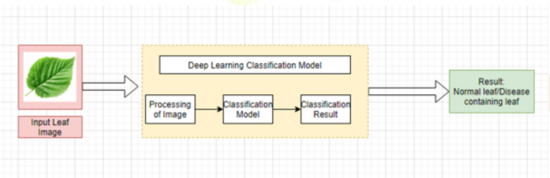

Fig.2. Plant disease detection

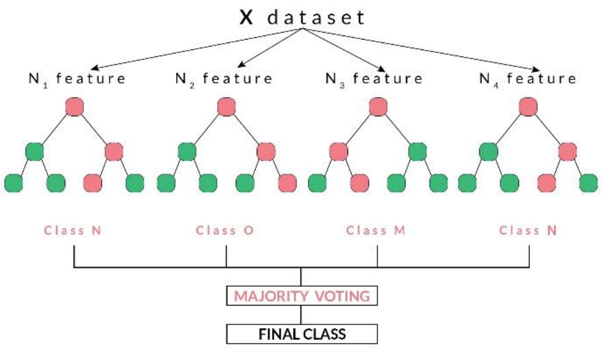

Fig.3. Random Forest Algorithm

Precondition: A training set S := (X1,Y1),..
..(Xn,Yn) features F and number of trees in forest B.
function RANDOMFOREST(S, F)
H → ø
for i € 1,….,B do
$S^{(i)}$ ← A bootstrap sample from S
h_i ← RANDOMIZED TREELEARN($S^{(i)}$,F)
H ← H U {h_i }
end for
return H
end function

function RANDOMIZED TREELEARN(S, F)
At each node:
f ← very small subset of F
split on best feature in f
return The learned tree
end function

6. METHODOLOGY

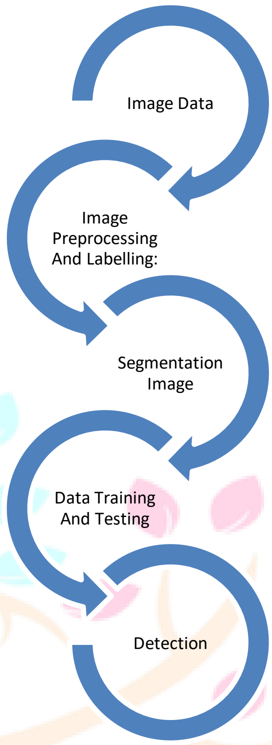

Fig.4. Methodology

6.1. IMAGE DATA

The code was linked to the Kaggle online kernel for fast processing and the study of training loss and validation, and the dataset was acquired from the online Kaggle of the Plant Village dataset.

6.2. IMAGE PREPROCESSING AND LABELLING

Pre-processing typically encompasses the reduction of low-recurrence foundation disturbance, correction of the power of the individual particle images, deletion of reflections, and obfuscated about certain areas of the images. Pre-processing is a method for enhancing information in image. The pre-processing technique for photos also comprised physically manipulating the presumably large number of photographs, drawing a square around the leaves to illustrate the area of inquiry (plant leaves). Pictures possessing a less ambitious goal and measurements that weren't accurately 500 pixels were discarded from the dataset, were not regarded to be important images for the dataset. Furthermore, the dataset was limited to images in which the location of intrigue was nearer the target. Images were assured to include that necessary information for highlight learning in this way. So many resources are found online, but their worth is frequently debatable. In order to confirm the accuracy of the classes in the dataset that was first obtained by a catchphrases search, horticultural authorities previewed leaf photos and classified all of the images with the suitable infection abbreviations Using well characterized images is vital for the production and approval of the dataset, as is common information. An accurate and reliable identifying model can only be built in this manner. At this point, duplicate photos that were still present after the dataset's primary objective of grouping and classifying images was reached.

6.3. SEGMENTATION IMAGE

"Partitioning a digital image into several parts" is the method. We will just separate the photographs into the foreground and background since the only thing we are interested in here is background removal. The image is converted to grayscale over the course of five simple phases. using the largest contour to create a mask, locating the image's edges, and then applying the mask to the original image will eliminate the backdrop.

6.4. DATA TRAINING AND TESTING

During the learning phase, the model's internal weights are adjusted automatically over a number of epochs. The architecture, normalization methods, training strategy, and importance of the hyperparameters are instances of external factors affecting the training process. It is difficult to compare research and their findings to draw conclusions about how to define the training phase because they do not utilize the same data and do not disclose all the information needed to duplicate their experiments. As a result of their experiments not being repeated numerous times to assess the effects of random initializations and training sample ordering, it is also challenging to understand the relevance of the conclusions drawn in these investigations.

6.5. DETECTION

Finding every single instance of a real-world thing, such as a human face, a flower, an automobile, etc., in photos or videos in real-time and with the highest degree of precision is called object detection. The object identification technique recognizes every instance of an item category using derived features and learning techniques. We start by using an image as our input. The image is then divided into several parts. Then, we'll treat each area as a separate image. Send CNN all of these areas (pictures) so they can group them into different categories.

7. ALGORITHM USED

CNN models function better for object classification and recognition in visual databases. CNNs have benefits, but there are also some downsides, such as the lengthy training process and necessity for large datasets. Deep CNN models are essential to extract the delicate and low-level aspects from the images; nonetheless, this exacerbates model training. Transfer methodologies can be used to solve the aforementioned issues. Transfer learning makes use of pre-trained networks to promote the application of model parameters derived on one dataset to multiple problems.

7.1. TRANSFER LEARNING APPROACH

Even on a strong GPU workstation, learning and good a model might take days or even weeks. Learning and developing a algorithm from start take time. A CNN model built from zero using a publicly available plant disease data claimed to reach 25% accuracy in 200 epochs, whereas a pre-learned CNN model implementing a transfer learning strategy obtained 63 percent accuracy in practically half the iterations. The details of the dataset and the pre-learned network algorithm for segmentation determine which transfer learning algorithm should be used.

7.2. CONVOLUTIONAL NEURAL NETWORK

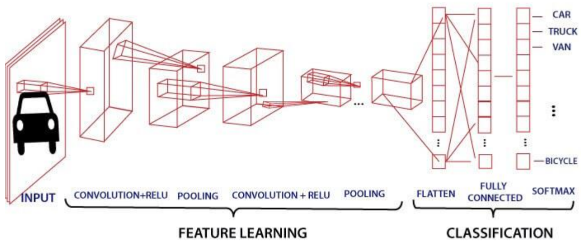

Fig.5. Representation of CNN

Convolutional neural networks are one of the main types of neural networks used in image identification and classification. Artificial neural networks are mostly used in a huge number of applications, such as classification, object detection. A photo is submitted to CNN, which classifies and analyses it using phrases like "dog," "cat," "lion," "tiger," etc. The computer's interpretation of the image as a collection of pixels depends on the image's resolution. CNN will apply a number of convolution layers, pooling, and fully linked layers, filters to each input image (Also known as kernels). Then, an object will be classified using probabilistic values between 0 and 1 and the Soft-max function.

7.3. RESNET

Deep Neural Networks are frequently given additional layers in order to address challenging problems, which raises performance and accuracy. The concept of layering is that as numerous layers are constructed, eventually they will learn elements. As an illustration, different layers learn different parts in the images and so on. The standard CNN model has been found to have a more depth entry. The error rate for learning and validation data for a 20-layers neural network and a 56-layers neural network, respectively, is presented in this graph. There is evidence of the 56-layer error rate.

7.4. RESIDUAL BLOCK

The production of ResNet, or residual networks, which are made up of residual blocks, has solved the issue of learning more deep networks. The direct connection, which omits some intermediate levels, is the first difference (this may vary between models). It's called a "skip connection," or a link in the middle of leftover blocks. In contrast to earlier The skip connection modifies the layer's output. Without this skip link, the input "x" is multiplied by the layer weights before a bias factor is applied.

8. RESULT

In India, farming plays a vital role in the economy. When farmers prosper, the country advances. Our approach will therefore help farmers plant the proper seed depending on soil conditions, increasing national production. In the future, we plan to work with a bigger data set, more attributes, and yield prediction. The suggested crop suggestion mechanism is quite helpful for Indian farmers. The goal is to promote precision farming and inform farmers about modern tools and infrastructure. A sensible decision might directly affect their income. This kind of method is very beneficial for new farmers who are unaware of the nuances of the soil and crops. If more research is done on it, this system has room to expand.

9. CONCLUSION

We suggest the "Farmer's Assistant," a painless web application built on ML and scraping. Our system successfully enables the provision of a number of features, including the recommendation of crops using the Random Forest method, the recommendation of fertilizers using a policy-based classification system, and the identification of crop diseases using the Efficient Net model on leaf images. The user can input data using forms on our user interface and receive responses immediately. Additionally, we use the LIME accountable method to explain our predictions on the disease prediction image, which may help explain why our algorithm makes the predictions it does and allow us to use this understanding to improve datasets and models.

REFERENCES

[1] S. D. Khirade, A. B. Patil, "Plant Disease Detection Using ImageProcessing", 2015 International Conference on Computing Communi_x0002_cation Control and Automation, 2015, pp. 768-771,
doi: s10.1109/IC_x0002_CUBEA.2015.153.

[2] A. V. Panchal, S. C. Patel, K. Bagyalakshmi, P. Kumar, I. RazaKhan, M. Soni, "Image-based Plant Diseases Detection using Deep Learning", Materials Today: Proceedings, 2021, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.07.281.

[3] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database", 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848

.[4] Russakovsky et al., "ImageNet Large Scale Visual Recognition Chal_x0002_lenge", International Journal of Computer Vision, 2015, 115(3), 211- 252.
https://doi.org/10.1007/s11263-015-0816-y.

[5] George A. Miller, "WordNet: A Lexical Database for English", Com_x0002_munications of the ACM, 1995, Vol. 38, No. 11: 39-41.

[6] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", Part of Advances in Neural Information Processing Systems 25 (NIPS 2012)

[7] S pudumalar , E. Ramunujam , C. Kavya Crop recommendation for precision agriculture.

[8] Zeel Doshi , Rashi Agarwal Neepa shah Intelligent crop recommendation system using machine learning algorithms.