PADDY AND RAGI CROP YEILD PREDICTION AND RICE BLAST DISEASE DETECTION USING ML

Yashavanth H R[1], Varoon N[1], Yashwanth Kumar M U[1], T Venkata Amruth[1], Akshatha M[2]

Final year students[1], Assistant Professor[2]
Department of Computer Science and Engineering,
Vidyavardhaka College of Engineering, Mysore, Karnataka.

Abstract

There are numerous issues facing the agricultural industry today, including how to enhance crop production, maintain correct agricultural parameters, maintain proper fertilizer, decrease the incidence of farmer suicides, and raise farmer profitability, among others. Since humans cannot survive without food, it is possible that food is the most significant part of life. Crop risk management and transportation decisions are guided by crop yield predictions, which also aid in managing crop storage. As a result, food storage becomes absolutely necessary for all nations, especially emerging ones. Storing enough food for the long term is a crucial goal for the governments of these nations, especially in the event of a catastrophic calamity. Consequently, estimating food is exceedingly difficult.
The world's population demonstrates the need for increased production... Productivity can be increased by protecting the crops and preventing them from diseases and pests. Identification of diseases plays a major role in prevention or protection of crops. Here, Oryza Sativa (Rice or Paddy Crop) is considered. A very common disease which attacks this crop is the Rice Blast. The identification of this disease can be done using Deep learning CNN. It is very important to detect this disease in paddy crops as rice is one of the main food items which is required by everyone in the world.

Keywords-Supervised Learning, Naïve Bayes Algorithm, KNN, Deep learning CNN, etc.

I. INTRODUCTION

India is commonly an agricultural country, with agriculture imparting employment to the bulk of the population. Agriculture is having a brilliant effect at the country's economy. In the preceding decade, India has visible intense herbal screw ups consisting of drought and flood. Crop manufacturing suffers a huge loss because of such calamities, and farmers go through as a result. Many farmers are committing suicide because of their economic losses. If herbal calamities aren't gift, then there can be surprising pest assault destroying the crop. In any scenario, the farmer and the crop are continuously at the verge of disaster. There are regulations in area with the aid of using the authorities, however they're insufficient. Crop manufacturing prediction earlier can help farmers and authorities' organizations in making plans for storage, selling, placing minimal help prices, importing/exporting, and different activities.

Protecting the yield from diseases is also important. As maximum yield of a crop is important, the maximum utilization of that yield is also important. Even if a small portion of the crop yield is infected by a disease, then it will be a huge waste. Here we talk about the paddy crop (rice) as it is one of the most important parts of the food industries globally. A very common disease which attacks this crop is the Rice Blast. This disease is caused by a plant pathogenic fungus named Magnaporthe Oryza. This disease is responsible for 30-40% loss of rice yield globally, which is the equivalent of feeding 60 million people. As the loss is high because of the Rice blast it is very important protect the yield by any means as much as possible.

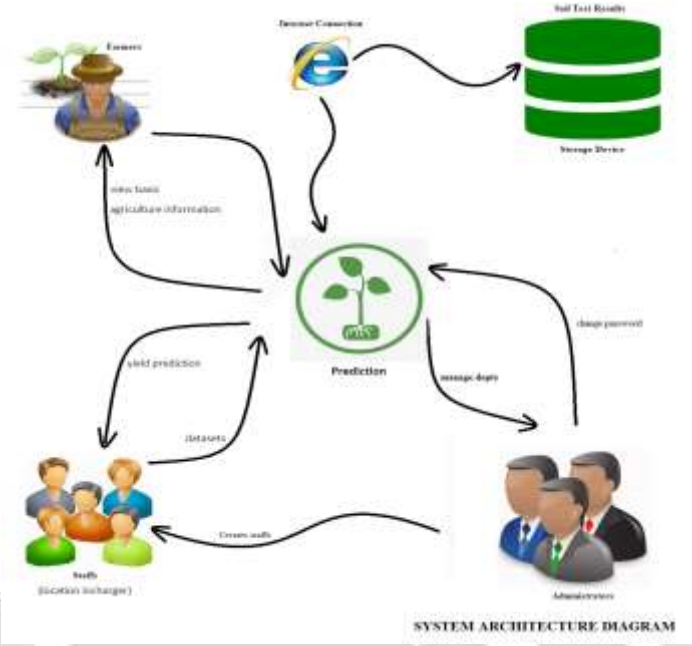

Figure 1: System Architecture Diagram

Machine Learning

Machine mastering is a statistics-pushed technique of inspecting a system. Machine mastering is a department of statistics technological know-how wherein statistics is processed the use of system mastering algorithms.

For Example, To learn how to differentiate between spam and inbox messages, for instance, ML can be applied to email communications. Only if the programme performance improves with experience is it said that a computer programme has learned from experience E with respect to some task T and some performance P. E. Machine learning (ML) is a subfield of artificial intelligence that uses statistical, probabilistic, and optimization techniques to find patterns in massive, complicated data sets and learn from prior experiences.

Machine Learning Types:

Supervised learning is a gadget studying hobby that entails studying a feature that converts an enter to an output primarily based totally on examples of enter-output pairs (SL). From a fixed of education examples and tagged education data, it infers a feature. In supervised studying, every instance includes an enter object (regularly a vector) and a favoured output value (additionally referred to as the supervisory signal).

Unsupervised learning the use of synthetic intelligence (AI) structures to discover styles in information units which includes information factors which can be neither classified nor labelled is referred to as unsupervised studying.

Reinforcement learning is a sort of machine learning in which favourable behaviours are rewarded while undesirable behaviours are punished. In general, a reinforcement learning agent can monitor and understand its environment, respond, and learn through trial and error.

II. RELATED WORK

1. IEEE PAPER TITLE: Prediction of Crop Cultivation

YEAR OF PUBLICATION: 2019

AUTHORS: Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, Wolfgang Bein.

METHODOLOGY: linear regression with polynomial features, and support-vector regression using a Radial Basis Function (RBF) kernel.

LIMITATIONS:

- Linear regression and support vector regression generates outputs graphically which is difficult to analyze.

- Not suitable in real time.
- Small Data-set used for prediction.

2. IEEE PAPER TITLE: Crop Yield Prediction and Efficient use of Fertilizers.

YEAR OF PUBLICATION: 2019

AUTHORS:S.Bhanumathi, M.Vineeth and N.Rohit.

METHODOLOGY: Random Forest and Back propagation algorithm used for implementation.

LIMITATIONS:

- Less parameters used for yield prediction
- Based on fertilizers, system predicts crop yield, but not considering all agriculture parameters.

3. IEEE PAPER TITLE: Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications.

YEAR OF PUBLICATION: 2020

AUTHORS: DHIVYA ELAVARASAN AND P. M. DURAIRAJ VINCENT.

METHODOLOGY USED: The proposed work constructs a Deep Recurrent Q-Network model which is a Recurrent Neural Network deep learning algorithm over the Q-Learning reinforcement learning algorithm to forecast the crop yield.

LIMITATIONS:

- Uses neural network techniques.
- Huge data required.
- More time required for prediction.

4. IEEE PAPER TITLE: Plant Disease Detection using CNN

YEAR OF PUBLICATION: 2019

AUTHORS: Adnan Mushtaq Ali Karol, Drushti Gulhane, Tejal Chandiwade.

METHODOLOGY: Convolution Neural Network

LIMITATIONS:

- Less accuracy

5. IEEE PAPER TITLE: Application for Detection of Plant Disease

YEAR OF PUBLICATION: 2018

AUTHORS: Tanvi Ingle, Ravina Shingade

METHODOLOGY: Image Processing and Developed a web app

LIMITATIONS:

- Less efficiency

Kumar and Srinivasan, [17] introduce a new force and level set functions to improve on Chan-Vese (C-V) active contour without edge method. These functions were constructed using square difference formula. The improved Chan-Vese (C-V) active contour without edge method reduced

III. PROPOSED SYSTEM

Here we are using Supervised Learning for yield prediction

It's a predictive version it's utilised for sports that require predicting one cost from a fixed of different values. The labels for supervised learning are probably predetermined. It assigns a label to an object depending on the parameters of 1 in each of a set of labels. In supervised learning, we have got numerous algorithms to assemble models, consisting of KNN, Naive Bayes, Decision Tree, ID3, Random Forest, SVM, Regression methods, and so on.... We choose out the finest approach for predictions based absolutely on the requirement, labels, parameters, and statistics set. In the presence of uncertainty, an set of guidelines is used to expand a model that generates predictions based mostly on evidence. In this project, we rent the "naive bayes set of rules and KNN method" for yield prediction, which can be each green and paintings nicely for a number of parameter combinations. It additionally produces unique results.

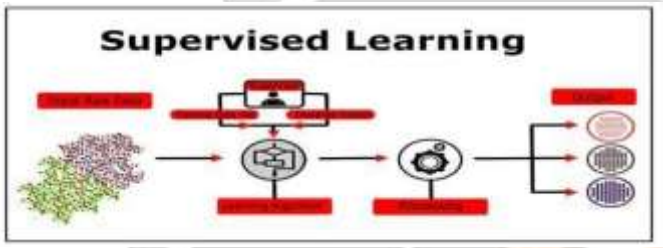

Figure 2: Supervised Learning

Simple Bayes algorithm

It is a classification method built on the Bayes Theorem and predicated on the idea of predictor independence. A Naive Bayes classifier, to put it simply, believes that the presence of one feature in a class has nothing to do with the presence of any other feature.

Step 1: Scan the dataset (storage servers)

Step 2: Calculate the probability of each attribute value. [n, n_c, m, p]

Step 3: Apply the formulae

P (ai/ vj)=(n_c + mp)/(n+m)

Where:

ai = attribute value

vj = subject value

n = the number of training examples for which v = vj

n_c = number of examples for which v = vj and a = ai

p = a priori estimate for P(ai/ vj)

m = the equivalent sample size

Step 4: Multiply the probabilities by p

Step 5: Compare the values and classify the attribute values to one of the predefined sets of class.

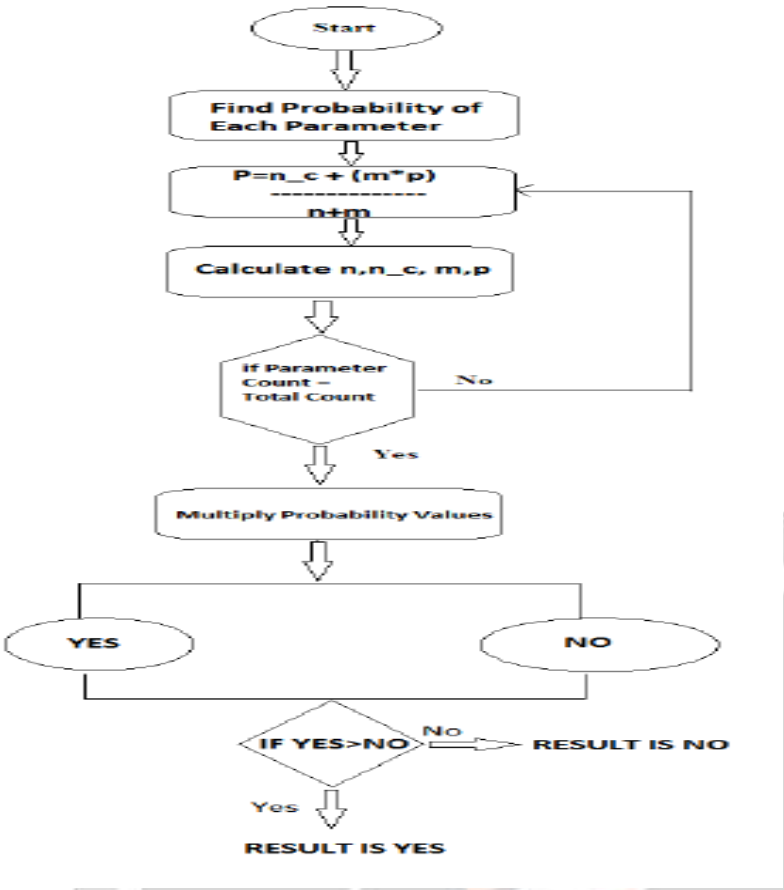

Figure3: Flow of Naïve Bayse Algorithm

Here we are using CNN for Rice Blast disease detection

CNN24 is a supervised mastering multi-layer neural community that normally includes parts: a characteristic extractor and a trainable classifier. The characteristic extractor has characteristic map layers and makes use of convolutional filtering and down sampling25 to extract discriminating traits from uncooked pictures. The most important feature of CNN, convolutional filtering, has vital properties: neighbourhood receptive area and shared weights. Convolutional filtering may be concept of as a neighbourhood characteristic extractor that identifies the institutions among pixels in an uncooked photo with the intention to extract the best and applicable high-stage capabilities to enhance a CNN model's generalisation capabilities.

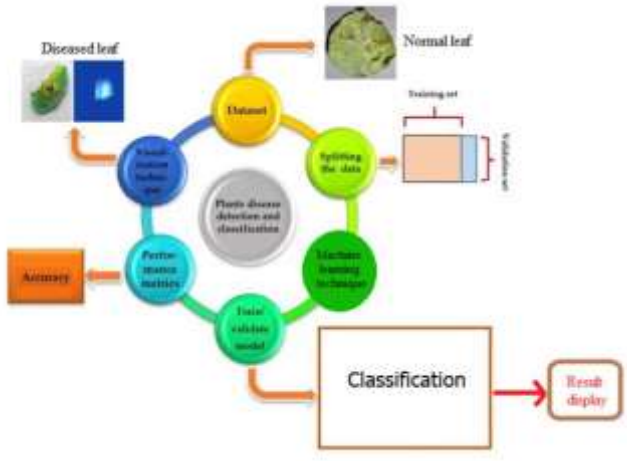

Figure 4: Work Flow of Disease Detection

Steps of Image Processing

1. Data Pre-processing:

- It is the process of deleting the corrupted images from the dataset.

- Data Augmentation: A crucial stage in developing an image recognition model is image transformation. Additionally known as data augmentation. Transformation of the model's picture is necessary to prevent overfitting. Images can be changed in a variety of ways, including by scaling, cropping, squashing, and padding. Squishing and padding, however, remove the photos' original information while, respectively, adding more pixels. As a result, resizing the photos randomly produces effective outcomes.
- Implementation of the Vgg16 deep learning algorithm: pre-trained weights from the vgg16 model should be loaded in order to train our model.
  A pre-trained CNN model is applied to the training of this image dataset. Transfer learning is the name given to this strategy. Jeremy advises utilising pre-trained models in order to speed up training and improve accuracy. This is particularly relevant to issues with computer vision.

2. Model Training Phase:
   - Training Process: Training the model through one complete cycle traversing through complete Convolutional Neural Network layers
   - Plot Confusion Matrix: Different techniques can be used to verify model performance. Utilizing the confusion matrix is one of the common techniques. The matrix's diagonal values represent accurate forecasts for each class, whereas other cell values represent a number of inaccurate predictions.
   - Exporting the Trained Model as export.pkl: Once the model is trained and satisfied with the outcome, it's time to deploy the model. For deploying the model into production, we need to save our model architecture and the parameters it's trained on. For this, the export method is used. The exported model is saved as a PKL file, which is a file created by pickle (a Python module).
   - Load the Trained Model for Testing: Load the model which have been exported named export.pkl. It is like a mathematical equation. Since the model is trained already no need to train again and again just feed the input it will give you an output.
   - Prediction Phase: In this phase, we will pass the image as an input and we will obtain the corresponding disease name as an output.

VGG16 Architecture:

Figure 5: VGG16 Architecture

The network receives a dimensioned image as input (224, 224, 3). The same padding and 64 channels with a 3*3 filter size are present in the first two layers. Then, two layers have convolution layers of 128 filter size and filter size following a max pool layer of stride (2, 2) (3, 3). The next layer is a max-pooling stride (2, 2) layer that is identical to the layer before it. There are then 256 filters spread across 2 convolution layers with filter sizes of 3 and 3. There are then two sets of three convolution layers, followed by a max pool layer. Each filter has the same padding and has 512 filters of size (3, 3).

The stack of two convolution layers then receives this image. As opposed to AlexNet's and ZF-11*11 Net's and 7*7 filters, we use 3*3 filters in these convolution and max-pooling layers. It also employs 1*1 pixels in some of the layers to adjust the amount of input channels. To prevent the spatial characteristic of the image, 1-pixel padding is applied after each convolution layer.

We obtained a (7, 7, 512) feature map after adding a convolution and max-pooling layer to the stack. This output is flattened to create a (1, 25088) feature vector. There are then 3 fully connected layers; the first layer uses the most recent feature vector as input and produces a vector of size (1, 4096); the second layer also produces a vector of size (1, 4096); however, the third layer produces a vector of size (1, 1000), which is used to implement the SoftMax function to classify 1000 classes. ReLU is used by every hidden layer as its activation function. Because ReLU promotes quicker learning and lessens the likelihood of vanishing gradient issues, it is more computationally efficient.

IV. RESULT

Initially, Paddy and Ragi datasets are trained the we pass the testing data sets for prediction which provides 96% accuracy for paddy and 98% accuracy for ragi yield prediction.

Figure 6: Paddy Yield Analysis

Figure 7: Ragi Yield Analysis

After training the model we will pass the image as an input for disease detection and obtain the disease name as an output.

Figure 8: Rice Blast Disease Detection

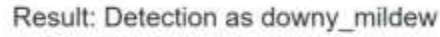

Figure 9: Downy_Mildew Disease

The inputted image contains Downy_Mildew disease which is a sub disease of rice blast disease.

Confusion Matrix:
A confusion matrix is a table that displays the effectiveness of a classification system. The output of a classification algorithm is shown and summarized in a confusion matrix.

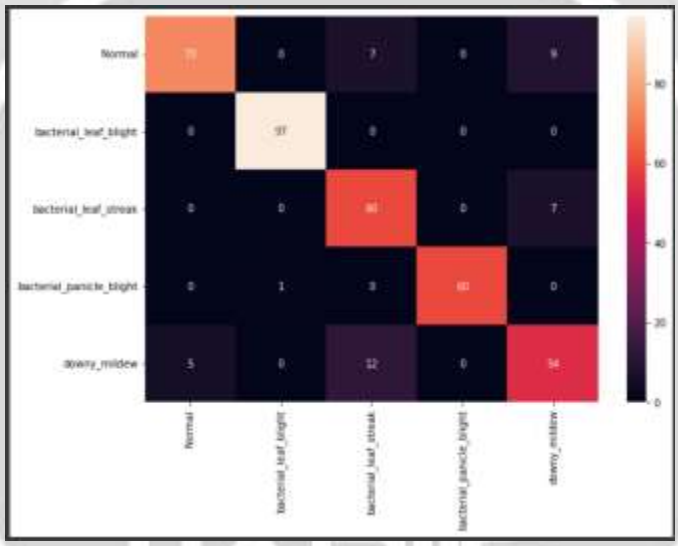

Figure 10: Confusion Matrix

V. CONCLUSIONS

Prediction of crop yield and detection of diseases is an essential factor in the present-day world. By using the above- mentioned methodologies, we have built a software which will predict the plant yield and identify the major diseases.

India is a nation where agriculture and industries associated with it are the main sources of income for the people. The country's economy primarily depends on agriculture. India is one of the nations that frequently experiences severe natural disasters like floods or droughts that ruin crops. Because of this, farmers suffer enormous financial losses and commit suicide.

We built this project using machine learning methods to forecast the outcome.

For yield prediction, we utilised the Naive Bayes algorithm, which provides accuracy of 96% for paddy and 98% for ragi crop, and for disease detection, we used the Convolutional Neural Network.

REFERENCES

1. Georg Ruß, Rudolf Kruse, Martin Schneider, andPeter Wagner. Estimation of neural network parameters for wheat yield prediction. In Max Bramer, editor, Artificial Intelligence in Theory and Practice II, volume 276 of IFIP International Federation for Information Processing. Springer, July 2008.

2. Adamchuk, V. I., Hummel, J. W., Morgan, M. T., &Upadhyaya, S. K. (2004). On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture, 44*, 71-91

3. Adnan Mushtaq Ali Karol , Drushti Gulhane , Tejal Chandiwade “Plant Disease Detection using CNN & Remedy” , IJAREEIE, Vol. 8, Issue 3, March 2019, pp 622-626.

4. Tanvi Ingle, Ravina Shingade “Application for Detection of Plant Disease” IRJET, Vol. 05, Issue: 04, Apr 2018, pp 2299-2231.

5. P. Srinivas and P. Santhuja, "Utilization of Data Mining Methods to Investigate Crop Yield Forecast," *2019 International Conference on Emerging Trends in Science and Engineering (ICESE)*, 2019, pp. 1-4, doi: 10.1109/ICESE46178.2019.9194693.

6. S. H. OU. In: Rice Diseases. Kew Surrey, England, Commonwealth Mycological Institute, 1972

7. C. Eric. In, “Report on Study Tour to Examine Symptoms of Rice Diseases in Southern USA and California”, Principal Research Scientist NWS Agriculture, Agricultural Institute, Orange NSW, 2002.

8. R. K. Webster, “Rice Blast Disease Identification Guide”, Davis, Dept. of Plant Pathology, University of California, 2000.

9. Amit Kumar Singh, Rubiya. A, Senthil Raja, “CLASSIFICATION OF RICE DISEASE USING DIGITAL IMAGE PROCESSING AND

10. SVM CLASSIFIER”, ‘International Journal of Electrical and
11. Electronics Engineers ISSN- 2321-2055 (E) IJEEE, Volume 07, Issue 01’, Jan- June 2015

12. Y.Sanjana, AshwathSivasamy, SriJayanth, “Plant Disease
13. Detection Using Image Processing Techniques”, ‘International
14. Journal of Innovative Research in Science, Engineering and Technology Vol. 4, Special Issue 6, ISSN(Online): 2319 – 8753 ISSN (Print): 2347 – 6710’, May 2015.

15. Y. J. N. Kumar, V. Spandana, V. S. Vaishnavi, K. Neha and V. G. R. R. Devi, "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector," *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 2020, pp. 736-741, doi: 10.1109/ICCES48766.2020.9137868.

16. Shivani. S. Kale and P. S. Patil, "A Machine Learning Approach to Predict Crop Yield and Success Rate," 2019 IEEE Pune Section International Conference (PuneCon), 2019, pp. 1-5, doi: 10.1109/PuneCon46936.2019.9105741.

17. S. Ramesh and D. Vydeki, "Rice Blast Disease Detection and
18. Classification Using Machine Learning Algorithm," 2018 2nd International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), 2018, pp. 255-259, doi: 10.1109/ICMETE.2018.00063.

19. jie. Sun, Z. Lai, L. Di, Z. Sun, J. Tao and Y. Shen, "Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S. Corn Belt," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5048-5060, 2020, doi: 10.1109/JSTARS.2020.3019046.2014.