International Scientific Journal of Engineering and Management
Volume: 02 Issue: 04 | April – 2023 DOI: 10.55041/ISJEM00396
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
www.isjem.com

POTATO DISEASE CLASSIFICATION USING GRADIENT BOOSTING

A.SAI PRASANNA U19CS071
M. GANESH U19CN113
K. PRADEEP U19CN010
G. NIKITH U19CS338

Dr.B.SELVAPRIYA
Assistant Professor, CSE

Dr.S.MARUTHUPERUMAL
Head Of Department, CSE

1.ABSTRACT

Potatoes are one of the widely consumed foods throughout the world. Usage of potatoes increases day by day. India is the second largest country in producing potatoes. It is good if we predict the disease earlier. By this wastage of potatoes decreases. Most of the potato disease can be predicted based on condition of leaf. Potato disease are of 2 types – Early blight and Late blight. Dataset is taken from Kaggle website which contains 2000 pictures of healthy and unhealthy potato's leaf. The dataset contains three classes, two disease classes and one healthy potato class. Models are trained by different train-test splits to understand better and get accurate results. To test performance of the data Applied Accuracy Precision, Recall, F1 score and ROC/AUC curve are used. By using Gradient Boosting approach results are better even for mostly effected leaf.

2.INTRODUCTION

Potato is the most important crop in India. India is one of the largest countries which is producing potatoes because it is rich in minerals, nutrients and it contains Vitamin C. Three-fourth of the population in India consumes potato daily. Potato can get many types of diseases. Many of the plant diseases cannot be detected by our naked eye. Farmers has to take care of the crop seriously. In that process, it is important to identify the disease earlier.

Plant has to cross many phases like
weather condition, rain fall. It should also be protected from animals. By predicting the disease earlier farmers can come across huge amount of loss and production of potatoes increases. By using Convolutional neural networks, the process becomes easier. By early detection of disease can avoid wastage of crops and farmers can get profit.

When the disease is identified on time financial loss of a plant decreases. Potatoes of different size and shape are scanned to get accurate results. Detecting potato manually, with huge amount of research is important. Images can be detected by using heavy sensors. Dataset is divided into two parts, training and testing. Approximately 75% of the data is used for training and remaining data is used for testing. Effected leaves should be identified by farmers and give appropriate treatment by using fertilizers to get rid of loss.

© 2023, ISJEM (All Rights Reserved)

Farmers who grow potatoes experiences huge economic losses every year because of various diseases that can happen to a potato plant. There are two common diseases known as early blight and late blight. Early blight is caused by a fungus and late blight is caused by a specific microorganism. If farmer can detect these diseases early and apply appropriate treatment then it can save lot of waste and prevent the economic loss. The treatments for early blight and late blight are little different so it's important to identify the disease accurately what kind of disease is there in that potato plant. To do these images of healthy potato plant leaf and the potato plant leaf which has early blight or a late blight disease. We have taken the data from Kaggle.

3.LITERATURE REVIEW

[3.1] Potato Crop Disease Classification using Convolutional Neural Network.

In this, they proposed deep learning-based approach to detect early blight and late blight diseases in potato by analysing the visual interpretation of the leaf of several potato crops in *2019* .There are four convolutional layers with 32,16 and filters in each respective layer. No feature extraction is done

[3.2] Potato Disease Classification using Neural Networks.

Pre-trained neural network architectures have been used. They have created a simple Image Classification model that categorizes disease using a simple and classic Convolutional Neural Network Architecture. Streamlit is used to build a web-based application and deploy it on Heroku. In *2020* They achieved very low accuracy and no pre-processing is done. If pre-processing is done, they could get accurate results.

[3.3] Artificial Intelligence in Potato Leaf Disease Classification.

The proposed architecture depends on 14 layers, including two main convolutional layers for feature extraction with different convolution window sizes followed by two fully connected layers for classification in *2021*. Pre-trained deep neural network architectures such as Alex Net, VGG16. By using these models computation time decreases.

[3.4] Potato Leaf Disease Classification Based on Distinct Color and Texture Feature Extraction.

This method is divided into three steps: image segmentation, feature extraction and classification. First, RGB leaf image is transformed into L*a*b* colour space and then k-means clustering algorithm is applied to separate background. They combined statistical texture features by using Euclidean distance in *2021*. Only software work is done in approach, it is not tested with real time potato leaves.

[3.5] Potato Plant leaves disease detection and Classification using Machine Learning Methodologies.

Popularly known Plant Village Dataset is taken from Kaggle. For the process of image segmentation, the K-means methodology was considered in *2021*, for feature extraction the grey level co-occurrence matrix concept was used, for classification multi-class support vector machine methodology was used.

Accuracy can be improved by performing normalization and by using ReLU activation functions .

4.EXISTING SYSTEM

In the previous approach, the CNN model is used for disease detection and classification, which includes input layer, convolutional layers, pooling layers and output layers. It is implemented in python using TensorFlow library which is an interface for expressing machine learning algorithms. The architecture consists of four convolutional layers. Convolutional layer is the core layer on the CNN.

It is used to extract the feature map from the input image and then creating a new matrix with smaller size using filter. The next convolutional layers use next two layers. The ReLU activation function was used after the matrix convoluted through all convolutional layers. For designing Keras and TensorFlow libraries were used based on python programming language. The model architecture was implemented on the cloud-based kernel to achieve high computing power and graphical computing power.

4.PROPOSED SYSTEM

Data analysis and Data visualization is done for easy understanding of data. Gradient Boosting, Logistic regression and Support Vector Machine models are used for classification. For feature selection Hyperparameter tuning, GridSearchCV is performed. For testing Precision, Recall and ROC/AUC curve is used.

5.MODULE DESCRIPTION

- **Data Collection**
- **Data Cleaning**
- **Pre-processing of images**
- **Data Augmentation**
- **Analyse and Prediction**
- **Accuracy on test data**
- **Saving the trained model**

[5.1] Data Collection

Collecting data allows you to capture a record of past events so that we can use data analysis to find recurring patterns. From those patterns, you build predictive models using machine learning algorithms that look for trends and predict future changes. Predictive models are only as good as the data from which they are built, so good data collection practices are crucial to developing high-performing models. The data need to be error-free (garbage in, garbage out) and contain relevant information for the task at hand. For example, a loan default model would not benefit from tiger population sizes but could benefit from gas prices over time. In this module, we collect the Potato disease data from Kaggle dataset archives. This dataset contains the information of information about different types of potato diseases

[5.2] Data Cleaning

Data cleaning is a critically important step in any machine learning project. In this module data cleaning is done to prepare the data for analysis by removing or modifying the data that may be incorrect, incomplete, duplicated or improperly formatted. In tabular data, there are many different statistical

ISSN: 2583-6129
www.isjem.com

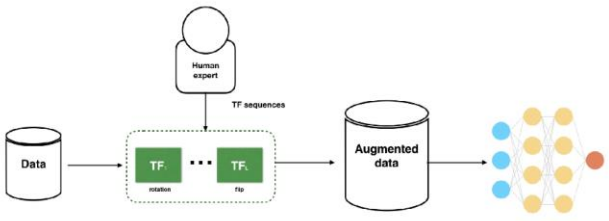
analysis and data visualization techniques you can use to explore your data in order to identify data cleaning operations you may want to perform the rainfall prediction.

[5.3] Pre-Processing of Images

Pre-processing is a common name for operations with images at the lowest level of abstraction — both input and output are intensity images. These iconic images are of the same kind as the original data captured by the sensor, with an intensity image usually represented by a matrix of image function values (brightness). The aim of pre-processing is an improvement of the image data that suppresses unwilling distortions or enhances some image features important for further processing, although geometric transformations of images (e.g., rotation, scaling, translation) are classified among pre-processing methods here since similar techniques are used.

[5.4] Data Augmentation

Data augmentation is useful to improve the performance and outcomes of machine learning models by forming new and different examples to train datasets. If the dataset in a machine learning model is rich and sufficient, the model performs better and more accurately.



5.CONCLUSION

Data Cleaning and Data Preprocessing of the Dataset were performed for better results. The difference between the 3 classes of potato disease has been observed. Data Augmentation is performed for better processing of potato plant leaves.

6.REFERENCES

[1] M. Islam, A. Dinh, K. Wahid, and P. Bhowmik, Detection of potato diseases using image segmentation and multiclass support vector machine 2017, *Can. Conf. Electr. Comput. Eng.*, 8–11.

[2] H. Waghmare, R. Kokare, and Y. Dandawate, Detection and classification of diseases of Grape plant using opposite color Local Binary Pattern feature and machine learning for automated Decision Support System 2016, 3rd *Int. Conf. Signal Process. Integ. Networks*, 513–8.

[3] S. R. Maniyath et al., Plant disease detection using machine learning 2018, *Proc. 2018 Int. Conf. Des. Innov. 3Cs Comput. Commun. Control.*, 41–5.

[4] U. Shruthi, V. Nagaveni, and B. K. Raghavendra, A Review on Machine Learning Classification Techniques for Plant Disease Detection 2019, 2019 5th *Int. Conf. Adv. Comput. Commun. Syst.*, 281–4.

[5] R. Kaur and S. S. Kang, An enhancement in classifier support vector machine to improve plant disease detection, Proc. 2015 IEEE 3rd Int. Conf. MOOCs, Innov. Technol. Educ.,135–40.

[6] V. Pooja, R. Das, and V. Kanchana, Identification of plant leaf diseases using image processing

© 2023, ISJEM (All Rights Reserved)

ISSN: 2583-6129
DOI: 10.55041/ISJEM00396
www.isjem.com

techniques 2018, Proc.- 2017 IEEE *Technol. Innov. ICT Agric. Rural Dev., 1,* 130–3.

[7] J. P. Shah, H. B. Prajapati, and V. K. Dabhi, A survey on detection and classification of rice plant diseases 2016, 2016 *IEEE Int. Conf. Curr. Trends Adv. Comput.*

[8] A. Akhtar, A. Khanum, S. A. Khan, and A. Shaukat, Automated plant disease analysis (APDA): Performance comparison of machine learning techniques 2013, Proc - 11th *Int. Conf. Front. Inf. Technol. FIT* 2013, 60–5.

[9] Z. N. Reza, F. Nuzhat, N. A. Mahsa, and M. H. Ali, Detecting jute plant disease using image processing and machine learning 2017, 2016 3rd *Int. Conf. Electr. Eng. Inf. Commun. Technol.*

[10] S. Kaur, S. Pandey, and S. Goel, Plants Disease Identification and Classification Through Leaf Images: A Survey 2019, *Arch. Comput.* Methods Eng., 26, 507–30.

© 2023, ISJEM (All Rights Reserved)