Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

An Ensemble Learning Framework for Predicting Tea Leaf Diseases Using Stacking Classifiers

T. Pramod Kumar[1*], K. Vijaya Lakshmi[2]
[1*]Research Scholar, Department of Computer Science, Sri Venkateswara University, Tirupati, A.P., India
[2]Professor, Department of Computer Science, Sri Venkateswara University, Tirupati A.P., India

ARTICLE INFO

Received: 7 Nov 2024

Revised: 18 Dec 2024

Accepted: 25 Dec 2024

ABSTRACT

Diseases in tea crop form one of the greatest obstacles in farming that can result in a decrease in yield and economic profit. The paper introduces a machine learning based technique of early detection and classification of tea crop diseases using stacking ensemble technique. The system envisioned will combine various information sources such as data on environmental sensors (temperature, humidity and soil moisture), weather and leaf images. Upon preprocessing and feature extraction, a wide range of base classifiers, including: Random Forest, Support Vector, Xgboost, K-Nearest Neighbor, and Convolutional Neural Networks are trained to obtain various patterns of disease symptoms. These base models are then aggregated to form the final classification by a meta-learner and the performance of the model is measured by using K-fold cross validation and various measures of performance such as accuracy, precision, recall, and F1-score. The results of the experiment prove that stacking ensemble works better when compared to single classifiers, and it gives stable performance with unbalanced and noisy data. The methodology provides the solution to smart agriculture in a scalable, real time method and contributes to proactive disease control in tea plantations.

Keywords: Tea crop disease detection, Stacking ensemble learning, Agricultural machine learning, Smart farming systems, Image Classification, Convolutional Neural Network (CNN), Random Forest (RF), Support Vector Machine (SVM), Feature Extraction, Agricultural Automation.

I. Introduction

Tea is among the most economically important and widely used drinks across the globe as millions of lives rely on its production especially in such countries as India, China, Sri Lanka, and Kenya. Nevertheless, tea crops are very vulnerable to various foliar infections including blister blight, red rust, and grey blight which may have a devastating effect on the quality and quantity of produce. The conventional techniques of detecting plant diseases are based on manual observations and are conducted by the professional; it is not only time-consuming and labor-intensive, but also susceptible to inconsistencies due to human error. But also cannot be scaled easily or provide the timely diagnosis. Recent progress in machine learning (ML) and computer vision has created new possibilities to automatize the detection of plant diseases through the use of digital images. These technologies can be used to identify the symptoms of the disease in time and make timely and specific interventions by analyzing visual features obtained in the form of images of leaves. Machine learning models, once trained on large and varied datasets have shown high accuracy in the classification of plant diseases. When trained in environments with specific climatic and soil conditions, i.e. moderate temperatures, high humidity and well-drained acidic soils, tea growing is a long-term process. Nevertheless, tea plantations are very susceptible to diverse biotic stresses especially fungal and bacterial diseases, which may interfere with the yield and the quality of the leaves.
The most frequently occurring and harmful disease of tea crop are:
a) The healthy tea leaf is the basis of the production of the high-quality tea. It appears as a bright green in

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

young leaves and as dark green in matures ones.
b) The tea leaf blight is normally due to fungal pathogens.
c) The pathogen Cephaleurosparacid causes the algal leaf disease in tea, which is also known as Red Rust, and which affects the leaves, young stems and branches of tea plants as well as causing typical reddish-orange, velvety spots and resulting in a decrease in photosynthetic activity, defoliation and low crop yield particularly in humid, poorly-ventilated environments.
d) Tea bud scab is a comparatively less known but significant disease of tea which infests the developing buds and young leaves resulting in low quality and production.
e) Tea bud blight is a major infection that mostly occurs on the tender buds as well as young shoots of tea plants resulting in lower yield and quality of crops.
f) Tea grey blight is a prevalent and harmful plant disease of the foliage, which affects the leaves, decreases the photosynthetic area, and consequently yield and quality.

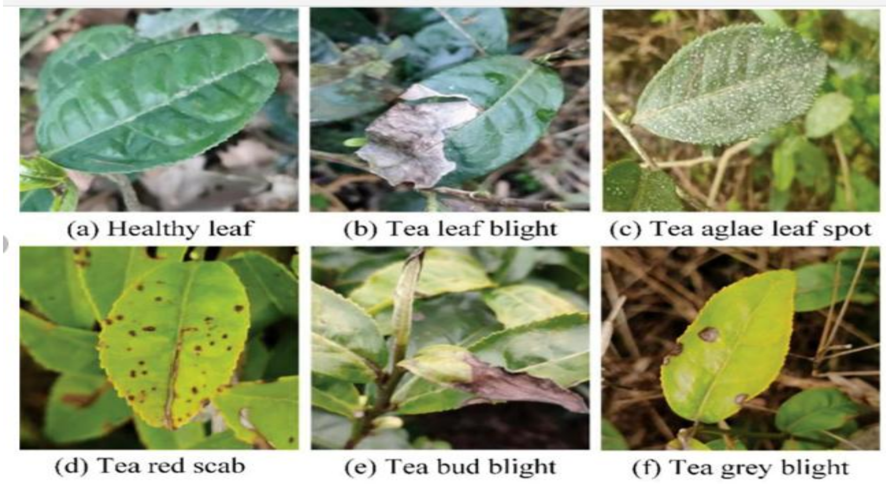

Fig. I (i) Tea leaf disease data set presentation

One of the major problems that tea farmers are facing is the prevalence of plant diseases which negatively impact on the production and quality of leaves. Blister blight, red rust, grey blight as well as algal leaf spot are some of the commonly diagnosed diseases in tea leaves. The causes of these diseases are fungi, algae or bacteria and are usually presented by discoloration and spots or deformities on the leaves. Otherwise, they may cause serious losses in crops and economic returns. The disease detection techniques are largely dependent on manual examination by trained personnel which is both time consuming and labor intensive besides being subjective. Also, the initial symptoms may be mild and may be easily ignored hence making it hard to intervene in good time. As digital agriculture continues to evolve, the application of machine learning (ML) and image processing frameworks in automated crop disease detection and classification is becoming popular as a research topic. This paper aims to design an intelligent system that can be used to detect and classify tea leaf diseases without human interference. The research is aimed at enhancing the accuracy and efficiency of the disease detection with the help of the two traditional ML models. The ensemble approach in which the various classifiers are stacked is also investigated to exploit the advantages of using multiple classifiers and improve the overall performance. Traditional techniques of plant disease detection include human detection that may be time-consuming, unequivocal, and incapable of detecting the disease at early stages. To overcome this issue, our effort utilizes methods in machine learning to allow automatic identification and classification of the diseases in tea leaves.
We consider the stacking-based ensemble method which consists of SVM, RF, and CNN models and enhances the diagnostic accuracy. The current work is meant to assist in real-time, field ready solutions in assisting farmers to handle the crop health efficiently and sustainably.
With the introduction of precision agriculture and the increasing accessibility of sensor data and systems based on image monitoring, Machine Learning has become one of the most common solutions in different areas as a means of automating the process of detecting crop diseases. Nevertheless, single-model methods

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

can be inadequate when using heterogeneous agronomical data that comprises of environmental parameters, time-series sensor observations, and visual leaf symptoms. In order to deal with these shortcomings, we suggest a very detailed approach to predict the disease in tea crops by utilizing a stacking platform of machine learning. The method integrates several forms of input data, including weather conditions, soil parameters, and high-resolution leaf images and uses diverse base learners such as RF, SVM and CNN. The effectiveness of the proposed approach is assessed based on cross-validation and various performance metrics- accuracy, precision, recall, and F1-score to provide and extensive information on the predictive behavior of the model and enhance the interpretability and confidence of the agricultural stakeholders. The work is relevant to the formation of smart, scalable, and data-driven decision support systems to support sustainable tea production.

II. LITERATURE REVIEW

The use of machine learning (ML) and computer vision in the agricultural field, especially in it as a method of plant disease diagnosis, has been growing at a very fast pace in the last decade. The identification of diseases in leaf images has been automated with different algorithms and strategies, and the researchers have shown that this problem can be found in a wide range of studies.

One of the first papers to discuss the potential of deep CNNs to recognize plant diseases in leaf images is the work of Sladojevic et al. (2016).Their model was very precise on a variety of plant species, which preconditioned the further studies in detecting diseases depending on crops[1].

On the same note, Brahimi et al. (2017) used CNNs to classify tomato leaf disease and underscored the efficacy of deep learning to identify complex visual features that were otherwise overlooked by the traditional machine learning models. Their analysis has indicated the importance of preprocessing including the background removal and augmentation to improve the performance of the models [3].The other researchers have concentrated on the ensemble learning methods. An example is the study by Azimi et al. (2011) whereby the authors used the Random Forest and Gradient Boosting classes since they observed that combination of multiple weak learners resulted in better accuracy and robustness in practical agricultural setting [4]. In the current paper (Vijai Singh, Namita Sharma, Shikha Singh,2020), the author presents a review of different methods of identifying plant diseases with a more prominent focus on the imaging techniques that would help with their early detection. It analyses existing trends and issues in the field of plant disease detection with the help of computer vision and advanced imaging methods, such as thermal, hyperspectral, fluorescence, multispectral, and 3D imaging. It also gives an overview of classification algorithms like SVM, K-means clustering, deep learning and KNN. They emphasize that it is necessary to have cost-effective and efficient techniques and reliable sensors to check the health of plants. The next round of research should be on the creation of more sound and precise early detection systems [5]. The current paper (Huajia Wang, Jinan Gu and Mengni Wang, 2023) explores how the concept of computer vision and machine learning has been implemented in the tea industry within the past decade. Deep learning (DL) has popularity in performing functions such as tea harvesting and plantation navigation because it is present in detection and segmentation. The more widely used in the disease control and processing is the trend of using traditional machine learning (TML), which is preferred by its interpretability and real-time image classification. The use of depth cameras is not fully exploited and it is scarcely used in harvesting. The main issues that were observed are accurate picking of tea buds, multi-disease, and a standard evaluation measure, and the necessity of open datasets. The paper (Wenxia Bao, Tao Fan, Gensheng Hu, Dong Liang andHaidong Li,2022) asserts that Tea leaf diseases have a pronounced effect on the yield and quality and therefore the need to detect them in an accurate and automated manner through the use of depth cameras and UAVs. CNNs that are traditional are unable to detect in natural scenes because of complicated backgrounds and thick foliage. This paper presents AX-RetinaNet, a better target detection network which improves multi-scale fusion of features and uses attention to eliminate noise. Additional performance and overfitting reduction measures are the data augmentation. The results of the tests confirm the fact that AX-RetinaNet is more effective than such old models as SSD, YOLO, and EfficientNet and can be heavily relied on to identify tea leaf disease with high accuracy in practice.

The article of (Nishmitha&Nikitha Shetty,2023) points towards the efficacy of the image-based classification

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

with the help of CNN and SVM. CNN recorded a classification accuracy of 92% whereas SVM got 88.9% and this means that CNN is a better tool to be used to detect tea leaf diseases. Reliance on High-Quality Data on Image. CNN and SVM models can only work effectively when presented with large datasets of high-resolution well-labeled images. Lighting or background mismatch, or blocking Inconsistent lighting, backgrounds or occlusions, and such inconsistencies can severely degrade accuracy. Lack of Generalization, Most models are trained on a particular dataset and might not be able to easily generalize to actual real-world conditions, e.g., with different leaf conditions, backgrounds or disease stage. Although they are precise, CNNs are computationally complex and might not be useful in real-time detection of leaf diseases on devices with low resources that are common in rural or field environments. The majority of studies investigate only a few leaf diseases. Unseen or rare diseases might be overlooked in absence of them in the training data. Labeled datasets need to be trained with human annotations by professionals, which may not be efficient and may have errors, particularly distinguishing between similar disease symptoms. Small Amount of Multimodal Data implies that most models use visual data only. It is typically assumed that other forms of data (environmental conditions, soil health, sensor data, etc.) can be more exploited to increase prediction robustness, although they are often overlooked to lead to lower accuracy in image-based classification problems [8].

According to research studies, there is undisputed innovation of the traditional machine techniques of learning to more advanced deep learning techniques and ensemble techniques in the detection of plant diseases. Nevertheless, a significant deficit of comparative assessments on the use of various algorithms on tea-specific datasets is observed. It is through this that there is the necessity to have in-depth research such as this one on which various machine learning methods are compared to a greater degree in which the future generation of automated tea disease detection can be informed. Conventional approaches likeSVM and KNN have been widely adopted, which are based on features extracted manually such as texture, color histograms and shape descriptors. Although these are computationally efficient, the quality of feature selection is very critical to their accuracy. Hybrid and stacking models have been on the rise in the past few years with the models combining several classifiers to use their respective advantages. Such models, particularly with some handcrafted elements, have demonstrated the possibility of dealing with such issues as scarce data and changes in lighting or background.

Despite these contributed achievements in applying machine learning and deep learning to tea disease classification, a number of challenges and issues remain unsolved: Comparative analysis between different algorithms based on tea-specific data is still incomprehensive, and it is hard to establish which models work best in this field since it may require differentiating the light and other features in the background. Conventional models of machine learning such as SVM and KNN heavily rely on features extracted by humans such as color, shape, and texture, and their performance on tea disease classification is highly dependent on the quality and consistency of feature selection. In addition, Even though hybrid and ensemble models are more accurate by integrating the capabilities of several classifiers, they have a complexity of computation and a longer training duration and are not applicable in real-time systems. Moreover, large labeled datasets that are particular to tea leaf diseases are limited making this a major challenge, especially to deep learning models, which need a lot of data to be effective. The problem of over fitting also cannot be ignored particularly in models that are trained on small amounts of data without the correct kind of regularization strategies. Those drawbacks point to the necessity to have more powerful, scalable and data-efficient solutions that would address the specifics of tea crop monitoring.

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

III. PROPOSED METHODOLOGY

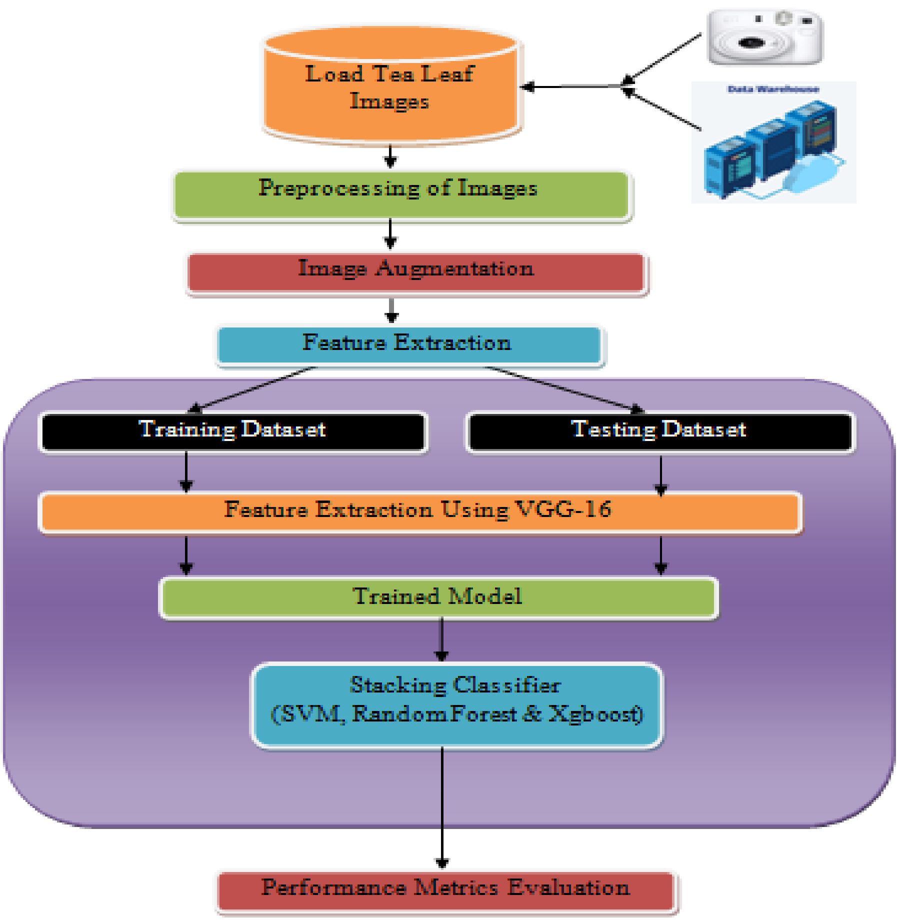

Fig.III(i) Diagram of Proposed Methodology to Predict Tea Crop Diseases

The steps that were undertaken to create and test the stacking method of disease detection and classification in tea crops include the following phases: data collection, preprocessing, feature extraction, model training and performance evaluation. The following model diagram illustrates how an image of a tea leaf can be loaded into a camera or repository to be subject to the disease detection system:

(i)Loading Tea Leaf Images from Camera or Repository:

The tea leaf pictures would have to be obtained either in real time by a camera or in a picture database in order to carry out an automated process of detecting the disease.

From Camera (Real-Time Capture):

Images of tea leaves in the field are captured using a digital camera or a mobile device, and directly inputted into the detection system using USB, Wi-Fi, or embedded software. Image capture settings such as resolution, lighting, and focus can be optimized to enhance the accuracy of detection. Useful in real-time monitoring as well as on-site diagnosing various diseases.

From Repository (Dataset):

These datasets are loaded through programming libraries (OpenCV, TensorFlow, or PyTorch) to ensure that high-quality and consistent images are used to train and evaluate machine learning models in different conditions and different diseases. Both sources guarantee that high-quality images can be used in the detection system to extract features and classify them.

(ii) Preprocessing of Loaded Tea Leaf Images

Image processing is an important step in the development of a successful disease detection system because it plays a significant role in determining the accuracy and reliability of the prediction of the model. The procedure is usually initiated by the image acquisition process, which involves the photography of tea leaves

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

with high-resolution digital cameras or depth cameras frequently in the natural field conditions. These raw images can be different in size and quality, so the initial work of preprocessing them is to resize them to a standard resolution, like 224x 224 or 512x 512 pixels, so that they all have the same model input. The reduction of noise is then done through strategies such as Gaussian or median filtering to eliminate the irrelevant pixel changes and bring out the significant features. Since tea leaves are usually photographed on a complicated and messy natural background, to isolate the leaf area, segmentation or background removal methods (color thresholding, GrabCut, or deep learning-based segmentation) are applied to the image to isolate the leaf area.

Color space conversion is used to improve the pattern of diseases identification, commonly to convert RGB pictures to HSV, Lab, or grayscale, depending on the features that should be identified. A normalization process is also required, during which pixel intensities are adjusted to some standard range (typically 0 to 1 or -1 to 1), to enhance the rate of convergence, and to minimize the impact of light variations across the data set. Moreover, enhanced model robustness can be achieved by means of data augmentation. These are random rotations, flips, translations, zooms and brightness/contrast adjustments, all of which aid in simulating more diverse resemblances of real-world conditions and decreases overfitting through augmenting the variety of the dataset.

A few systems also use morphological operations or edge detection algorithms to add more to the disease spots and patterns on leaves. Finally, proper labeling should be performed, and every image that is preprocessed, should be labeled with its respective disease (e.g., tea blight, white scab) or should be healthy. Such processed and labeled images are subsequently trained on machine learning or deep learning models so that they learn discriminative features to classify them with high precision.

(iii) Image Augmentation:

Image augmentation can play a very important part in enhancing the resilience and precision of the deep learning models applied to detect tea leaf disease, especially in the case where the given data size is small or skewed. In the natural environment, the images of tea leaves are highly variant with respect to the lighting conditions, orientation, and background noise, and geometric transformations, including rotation (e.g., 90 - degree or random angles), flipping (horizontal and vertical), scaling, and random cropping, are used to replicate the variations and increase the model generalization capacity. They are particularly applicable in situations in the field where tea leaves can be located at several angles or partially obstructed.

Photometric augmentations alter the properties of images (e.g. contrast, brightness, saturation, hue) to mimic different levels of lighting that can be normally seen outdoors.Noise can be injected (see, e.g., Gaussian noise), as well as blurring, to enhance resistance to camera quality and interference. Shearing and translation are examples of affine transformations which are used to model spatial distortions that may happen in actual plantation environments. More sophisticated augmentation Cutout techniques to make the model learn to work with incomplete data, random parts of an image are removed to make it more resilient.

Besides these conventional techniques, color jittering and channel shuffling are also applied to feature learning in multiple color spaces that are vital in situations where diseases reflect themselves with minor color variations on leaves. Elastic distortions can be used also to simulate natural leaf distortions due to pests or wind. Secondly, such approaches as Mix-up or Cut Mix that involve mixing two images and their labels have demonstrated encouraging performance in enhancing model regularization and generalization.

Even more complex models such as synthetic image generation with Generative Adversarial Networks (GANs) are used to generate realistic disease samples, especially in cases where the training data is scarce, e.g. rare disease classes. This does not only serve to balance the dataset but also subjects the model to disease patterns which it would not otherwise experience. In general, image augmentation is a potent method of enhancing accuracy, reliability, and generality of tea leaf disease classification models in various and unpredictable real-life scenarios.

(iv) Feature Extraction:

It is a very basic step in detecting and categorizing tea leaf diseases, since it entails determination of the most appropriate patterns or features of the images, which differentiate diseased leaves and healthy ones. The traditional machine learning methodology applies such forms of manual features to the visual characteristics of tea leaves as color frequency maps, texture descriptors (e.g., Local Binary Patterns or GLCM), edge descriptors (e.g., Sobel, Canny) and form descriptors (e.g., shape-based features). The resulting

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

characteristics are subsequently submitted to classification algorithms such as the SVM, KNN or decision trees to classify the disease. Nevertheless, the handcrafted feature extraction lacks in the ability to capture complex patterns and variations that are witnessed in the real world conditions.
Conversely, deep learning models, particularly Convolutional neural networks (CNNs) can extract features automatically without doing any pre-processing step such as preliminary feature extraction and are trained to learn hierarchical patterns directly out of raw image pixels. CNNs learn step by step higher-level features (like spots, blight patterns, or leaf deformation) in deeper layers by using various layers of convolutional filters to detect more basic level features (edges, color changes, etc.) at lower levels. This compound method enables CNNs to learn the global and local features thus proving very useful in distinguishing minor differences induced by various diseases. Also, such techniques as transfer learning with pretrained models (e.g., VGG, ResNet, and EfficientNet) are also employed to use powerful feature representations obtained with large-scale datasets and reconsidered to match the needs of the individual task of classifying tea leaf diseases and detecting them in a timely and accurate way. As a whole, robust feature extraction, whether manual or automated, is also a key to increase the accuracy of classification and allows detecting the tea leaf diseases in a prompt and accurate way.

(v) Training a Tea Leaf Disease Detection Dataset Using the Stacking Method:

Data preparation notebooks The Stacking or stacked generalization is a machine learning technique that integrates the predictions of multiple machine learning models to generate a stronger and more accurate overall model. The basic principle of the Stacking machine learning process is that the multiple base models (also known as level-0 models) are trained on the same data, and their outputs are used as inputs by an additional higher-level model (also called a level-1 model, or ensemble model) to produce the ultimate prediction. When applied to tea leaf disease detection, stacking can be made to a much more advanced stage of accuracy enhancement by the ability to The images are subjected to feature extraction methods like the use of convolutional layers in CNNs or manually engineered features like texture and color histograms. They are then inputted into the base models to train them and a common stacking ensemble can comprise of a combination of models such as CNN, SVM, Decision Trees, and KNN. All these models could be good at recalling various details of the leaf disease images. As an example, CNNs can automatically learn hierarchical features in an image, whereas SVM may have better classification of minor differences in a disease symptom. The main benefit of stacking is that the weaknesses of each model will be overcome, combining the merits of both models. An example is that CNNs can be used on complex features of images, but SVM can be used to further refine the final classification using the margin between the classes as input and via a meta-learner, a higher level model that learns how to blend the results of the base models. Typically, a meta-learner can be a logistic regression, another neural network or even basic classifiers such as the Random Forest. Another significant advantage of stacking is the ability to reduce bias and variance as the stacking learns the models perform better in specific circumstances and how to give each model prediction.Overfitting and underfitting are frequent problems in machine learning and are reduced through stacking. As an example, one model can be very biased and overlook significant features of a disease or too much biased and overfit the training data and cannot be applied to new images. However, the stacked model is able to reduce both of them by taking into account different points of view of each base model. Moreover, the imbalanced data are more easily handled by stacking. The dataset of tea leaf disease may contain underrepresented disease classes and the base models may not be effective with such minority classes. Completely resolving these imbalances could be achieved by the meta-learner focusing on models that are more effective with underrepresented classes, which is why cross-validation is typically used during training. This is done by dividing the data into several folds, and training the base models on one of the folds, and then testing them on the other folds. The cross-validation will make sure that the meta-model is trained on the valid, impartial predictions, thus giving birth to a stronger final model. Moreover, one may apply k-fold stacking, in which the k subsets of the data are averaged or weighted by the prediction of each base model. After training the stacking ensemble, it is more accurate and reliable in giving the tea leaf disease classification prediction. This model can cope with different challenges that could be caused by the natural scene images, including the color of leaves difference, the level of disease process, and varying environmental aspects.
The strength of stacking to combine diverse models and form an ensemble that is capable of handling complex, noisy, and unstructured data like tea leaf images is a potent technique that can be used to automate

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

the process of disease detection in tea leaf images and it may also be utilized to transform the current tea farm management practices.

(vi) Testing the Tea Leaf Disease Detection Model Using the Stacking Method:

After the stacking model has been trained on tea leaf disease dataset, its performance is measured in the testing phase by using a separate test dataset, which is a set of images that the model has never seen in the course of the training session and therefore, its performance serves as a measure of how well the model can generalize to new unseen data. The testing phase starts with the performance of the same preprocessing work which was performed during the training process including resizing, noise reduction and feature extraction to ensure consistency between the training dataset and the testing dataset. The test images are preprocessed then sent to the individual base models of stacking ensemble.

Each of the base models gives its own predictions that are typically presented as probabilities or class labels.. Such predictions then are inputted into the ensemble learner, which combines the outputs of the base models in order to produce the decision. The role of the ensemble-model is important because it gets to learn how to balance the predictions of the base models in the best way in order to enhance the performance of the ensemble. In an example, a given base model would be effective in detecting a particular number of diseases and the other model would be effective in detecting other types of diseases, the meta-model will integrate their predictions in a manner that improves the overall classification accuracy.

In order to determine the effectiveness of the model, there are several important metrics. Accuracy gives the general percentage of correct classifications of the test data. Precision is the degree to which model accurately finds positive cases of disease and recall (or sensitivity) is the degree to which the model finds all the real positive cases. The confusion matrix is especially useful since the F1-score alone does not give a complete picture of how the model performs on a given dataset, the confusion matrix shows the True Positives, False Positives, True Negatives, and False Negatives of each disease category, and, as such, helps locate the patterns of misclassification. Moreover, the Area under the Curve (AUC-ROC) is a useful assessment measure used to determine the probabilistic predictions of the model to ensure that it did not overfits the available training data and is generalizable to the unseen inputs. Cross-validation was also implemented in the process of training the model to ensure that it did not overfit the training data and can be generalized to the unseen inputs. This serves the purpose of making sure that the performance of the model on the test set is not contrived and is a true measure of its generalization capability. Stacking also minimizes the chances of overfitting as the base models are trained on various folds before being tested on the remaining data which results in the model being more robust in testing.

Once the model is tested, a review of the errors is carried out to determine whether they are of a recurrent nature or any trend in the errors. To illustrate, some diseases that have similar symptoms can be often confused, or the model cannot work with the images that have less manifested symptoms of the disease. Errors are examined to provide a fine-tuning of the model or more data augmentation processes can be introduced to enhance the accuracy. Furthermore, their testing stage will be important in determining whether or not the stacked model can effectively generalize to monitor unknown tea leaf images under different environmental sceneries due to the lack of the specific classes of disease. The testing process will assist in confirming that the stacking ensemble model is robust, accurate and available to be deployed in the real world by confirming through extensive performance measures, error analysis and cross-validation that the model demonstrates strong robustness and accuracy, making it suitable for real-world deployment in tea leaf disease detection.

(vii) Feature Extraction Using VGG-16:

The technique of extracting features to detect tea leaf disease with VGG-16 is the utilization of a pre-trained VGG-16 model, which is good at extracting hierarchical visual features of images. First, tea leaf images are predefined by the necessary size (224x224 pixels) needed by VGG-16 and normalize their pixel values to match the environment where the model was trained (i.e. using the mean and standard deviation of ImageNet). The tea leaf images are processed first and then sent through the VGG-16 network in which the convolutional layers automatically isolate features; which in turn are the textures, edges, and patterns representing various characteristics of the disease. The fully connected layers of the model are abandoned and feature maps generated by the final convolutional layer are processed further. These acquired patterns are then converted to a one dimensional vector which captures the important information to classify a

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Journal of Information Systems Engineering and Management
2024, 9(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

disease. These feature vectors extracted may then be fed to different machine learning models, i.e., SVM or KNN, to identify the tea leaf diseases. Alternatively, VGG-16 can also be re-trained on the tea leaf disease dataset, which means that the model is re-trained to reshape the already learned characteristics to the particular task of detecting tea leaf disease. The method offers an effective and strong means of obtaining meaningful features of tea leaf images, which can be used in precise classification of diseases.

(viii) Trained Model:

Once the training process is completed with the feature extraction and potentially stacking of classifiers, the final model is capable of automatically classifying the tea leaf diseases. The training process of the model that would be used in detecting the tea leaf diseases begins with the collection and preprocessing of the data. This includes acquiring a large image dataset of tea leaves, including healthy and diseased ones that are then run through preprocessing, i.e. resizing to a constant size, pixel normalization and data augmentation (i.e. random rotations and flips) to reduce overfitting and enhance generalisation of the model.

An already trained VGG-16 model is employed in extracting features when it comes to capturing hierarchical and high-level representations of the images. This step involves freezing the convolutional layers of the VGG-16 so that they do not change during training but the fully connected layers are trained on the tea leaf dataset. Additional manual feature extraction methods like texture or edges can be used in case necessary to provide additional features to the feature set. After extracting the features, stacking technique is applied to merge the features. This is done by training base models, including SVM, Random Forest (RF) or KNN to the features extracted. The predictions of these base models are then combined with an ensemble-learner (e.g. the Logistic Regression or Gradient Boosting) to achieve the final output of whether or not a leaf is healthy or diseased, which is evaluated using evaluation measures (e.g. accuracy, precision and recall). In case the model is effective, the initial images of tea leaves that have not been seen previously can be predicted regarding their health condition. Deep learning (including VGG-16) and stacking serves as an effective approach to guarantee the high accuracy and stable prediction of the disease, even under a problematic condition. The stacking classifier approach is a potent method of enhancing the predictive performance of the machine learning model with the help of various machine learning models. To identify diseases in tea leaves, this method works in combination with the peculiarities of SVM, RF and XGBoost as the base models. Each of these models has different ways of processing the data, with SVM all the more efficient in high-dimensional space, RF more efficient at representing complex patterns with decision trees and XGBoost being more effective with a robust gradient boosting model which gives the model the ability to learn characteristics that distinguish between healthy and diseased tea leaves. Each of the models separately tests the data and makes its predictions by feeding the extracted features to the base models. The meta-model is fed with these predictions and the probability thereof, in most cases these are the predictions of the base models, which are then in most cases the Logistic Regression, but other models such as the Gradient Boosting can also be used to decide the most accurate result. This is to enable the stacking classifier to make use of the complementary advantages of the various underlying models, which perform better than trying to use any of the underlying models individually. Since it is possible to correct the weaknesses of individual models- e.g. SVM is sensitive to noise, Random Forest overfits in certain instances, and XGBoost is computationally expensive- the stacking method will provide a more balanced and precise solution to the problem of tea leaf disease classification. Moreover, the fact that the stacking method can implement a combination of different models will ensure that the solution is more robust, hence more adaptable to the variations in the tea leaf images, i.e. change in lighting conditions or the direction of the leaf. Once the training is done, a model is tested on a different set of data to test its accuracy and capability to predict unseen data. The model obtained then can be stored in real-time prediction, which can offer tea growers an efficient, automated process of detecting leaf diseases and make informed decisions regarding the management of crops. The combination of different algorithms into one framework will increase the accuracy of detection of the disease, reduce the errors associated with the different conditions, and make sure that the tea producers can guarantee the quality and the high yields because of the quick identification and control of the disease.

(ix) Performance Metrics Evaluation

A tea leaf disease detection system is tested on a set of performance metrics that determine the accuracy, reliability, and robustness of the system based on a stacking classifier. These measures assess the capability of the model to correctly distinguish healthy and diseased tea leaves and the level of effectiveness of the model

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

in practical agricultural environments.

IV. CONCLUSION

Stacking classifier does not only enhance the performance of classification but it is also flexible and allows the choice of model and the adjustment to various types of tea leaf diseases and imaging conditions. It is due to the combination of various base models: SVM with its margin-based decision making, Random Forest due to its resistance to overfitting, and XGBoost due to its gradient boosting that it becomes ensured that the model can be applied to various data complexities and noises. The ensemble model, which is trained on the outputs of the base learners, acts as a second-level learner and absorbs the collective patterns of the base learners and corrects the weaknesses of the base learners. In addition, stacking with deep feature extraction features like VGG-16, mean that high-level and abstract features which capture the critical visual patterns of disease symptoms, say, spots, discoloration or blights are captured. This adds to the power of the system to identify the slightest changes in the appearance of the leaves that would otherwise have been ignored by the traditional methods. The outcome is a complete system which could effectively perform valid and very precise classification of tea leaf diseases. This plays an important part in the agricultural sector by facilitating early detection, minimizing human work in the manual inspection as well as supporting the use of informed strategies in disease management. The stacking technique is therefore an encouraging research and implementation prospect on precision agriculture and smart farming systems in the future.

V. REFERENCES

[1] P. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep neural networks based recognition of plant diseases by leaf image classification,” Computational Intelligence and Neuroscience, vol. 2016, Article ID 3289801, 2016.

[2] N. Brahimi, S. Arsenovic, S. Laraba, K. Boukhalfa, and S. Moussaoui, “Deep learning for plant diseases: detection and saliency map visualisation,” Human and Machine Learning, vol. 8, pp. 93–117, 2017.

[3] L. Liu, W. Zhang, and Q. Wang, “Automatic tea leaf disease classification using deep learning,” Journal of Applied Science and Engineering, vol. 23, no. 1, 2020.

[4] Azimi, A. Fern, and R. Tadepalli, “Robust image classification using ensemble of local and deep features,” In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1242–1249,2011.

[5] Vijai Singh, Namita Sharma, Shikha Singh,“A review of imaging techniques for plant disease detection”2020

[6] http://www.keaipublishing.com/en/journals/artificial intelligence-in-agriculture

[7] Huajia Wang, Jinan Gu and Mengni Wang.,“A review on the application of computer vision and machine learning in the tea industry” Volume 7 – 2023 https://doi.org/10.3389/fsufs.2023.1172543 Food Syst., 20 April 2023

[8] Wenxia Bao, Tao Fan, Gensheng Hu*, Dong Liang &Haidong Li, “Detection and identification of tea leaf diseases based on AX-RetinaNet”,2022

[9] K.R.Aravind, P.Raja1, K.V.Mukesh, R.Aniirudh, “Disease Classification in Maize Crop using Bag of Features and Multiclass Support Vector Machine”,Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018) IEEE Xplore Compliant - Part Number:CFP18J06-ART, ISBN:978-1-5386-0807-4; DVD Part Number:CFP18J06DVD, ISBN:978-1-5386-0806-7

[10] Archana Chaudharya, Savita Kolheb, Raj Kamal, “An improved random forest classifier for multi-class classification”,www.sciencedirect.com Information Processing In Agriculture 3 (2016) 215–222 journal homepage: www.elsevier.com/locate/inpa

[11] Dr. Rizwan Rehman, “Disease Detection in Tea Leaves Using Computer Vision and Deep Learning Techniques- a Review”, International Journal of Creative Research Thoughts (IJCRT) www.ijcrt.org

[12] Sunil S. Harakannanavara ,, Jayashri M. Rudagi , Veena I Puranikmath, Ayesha Siddiqua , R Pramodhini, “Plant leaf disease detection using computer vision and machine learning algorithms”,2666-285X/© 2022 The Authors. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[13] Lili Li, Shujuan Zhang ,Andbinwang, “Plant Disease Detection and Classification by Deep Learning A Review”, March 24, 2021, date of publication April 8, 2021, date of current version April 19, 2021. Digital Object Identifier 10.1109/ACCESS.2021.3069646

[14] Cezary Szczepanski,K.R.Aravind, P.Raja1, K.V.Mukesh, R.Aniirudh, “Disease Classification in Maize Crop using Bag of Features and Multiclass Support Vector Machine”,Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018) IEEE Xplore Compliant - Part Number:CFP18J06-ART, ISBN:978-1-5386-0807-4; DVD Part Number:CFP18J06DVD

[15] Munaf Mudheher Khalid, Oguz Karan, “Deep Learning forPlant Disease Detection”, International Journal of Mathematics, Statistics, and Computer Science Vol. 2, 2024 ISSN: 2704-1077 eISSN 2704-1069, DOI: 10.59543/ijmscs.v2i.8343

[16] Nishmitha & Nikitha Shetty, “Tea Leaf Disease Detection” International Research Journal of Modernization in Engineering Technology and Science Volume: 05/Issue:05/May-2023 Impact Factor-7.868

Copyright © 2024 by Author/s and Licensed by JISEM. This is an open access article distributed under the Creative Commons Attribution License which permitsunrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.