Pesticide Recommendation based on Pests Dectection

Himanshi[1], Ritwik Duggal[1], Abhishek Kumar[1], and Rohit Thakur[2]

[1]Computer Science, NIT Hamirpur
[2]Chemical Engineering, NIT Hamirpur

June 6, 2021

Abstract

Detection and Recognition of eight different pests in Tomato Leaves. Pests are one of the major problems faced in farming. Recognition of pests would help save a lot of resources as the exact method to deal with them would be known. The aim of the proposed work is to provide recognize the pest in the crop in real time, which is done by using Machine Learning techniques and deployment using an app.

Keywords

Machine Learning, KNN, Pests, Tomato Leaves, Agriculture

1 Introduction

Pests are the insects which attack the crops and cause harm, these can often be the reason for ruined crops. Currently, farmers use a variety of pesticides to be sure that pests do not attack their crops, but the use of excessive pesticides cause the soil to lose its natural fertility and often most of these pesticides are not even essential as most farmers do not know which pests infest their crops. Use of specific pesticides instead of this can be very helpful in order to deal with this problem, the aimed approach will help solve the problems mentioned.

There have been many approaches in this field, such as The white flies were detected using the Relative Difference in Pixel Intensities (RDI) algorithm in [1], An algorithm for distinguishing between white fly and aphid is proposed in [2], detecting white flies is to measure the size of the white flies and count them using background subtraction of images containing white flies [3], the main issue with these traditional methods is that they are inaccurate when dealing with multi-class pests.

A setup aimed at detecting whiteflies with GLCM feature extraction yields very low accuracy with multi-class classification; additionally, the use of GLCM is decreasing nowadays due to its low efficiency in [4]; parallel processing is more cost effective, and deep learning yields more accurate and better results than GLCM. To address the issue of multi-class classification, a machine learning model is trained in the proposed work, with pesticide recommendations based on pest detection rather than traditional single-class classification methods, and fertiliser recommendations based on soil and climate conditions.

Our aimed method ensures high accuracy even when detecting multiple classes of pests infesting tomato leaves along with real-time deployment, using a dataset of [5] images to train the model using k-Nearest Neighbors (KNN), this method can be developed further can potentially help to detect multiple pests in multiple crops.

2 BACKGROUND & RELATED WORKS

The agriculture not only provides food but also a big source for the economy for any country and the problem of pests in agricultural field and the use of pesticides is very common. There has been some earlier work in the field of autonomous pest detection and fertilizer recommendation in crops. The white flies were detected using the Relative Difference in Pixel Intensities (RDI) algorithm in [1]. It also counts white flies to estimate the density of white flies in the field. This algorithm is effective for greenhouse and agricultural crops. It evaluates 100 images and has a 97 % accuracy rate. When dealing with overlapping white, it works well.

An algorithm for distinguishing between white fly and aphid is proposed in [2]. It also includes a method for distinguishing between affected and unaffected leaves. This algorithm employs a support vector machine to extract various image features for use as input in classifying them.

Another method for detecting white flies is to measure the size of the white flies and count them using background subtraction of images containing white flies [3]. The Sobel edge detection operator is then used to detect the edges of whiteflies in the image so that they can be easily distinguished. This algorithm detects three times faster and covers three times as many leaves as the previous algorithm.

Another method for detecting white flies is to first find the image background, then image regional properties with the use of GLCM (grey level co-occurrence matrix) for whitefly classification in [6]. Image background was computed using morphological operators, then the background was deleted from the original image to get the ROI and SVM was used for classification after feature extraction using GLCM, yielding 97 % accuracy.This algorithm detects overlapping whiteflies, but using GLCM reduces its efficiency in multi-class classification.

Another approach was to recommend fertilisers in [7] using ANN (Artificial Neural Network) and IoT (Internet of things). Arduino and WSN (Wireless sensor network) are used to collect soil and PH data . ANN was used to normalise the cost and improve the efficiency of the proposed model's accuracy. For recommendation, SVM was used. This method assists farmers in producing a high yield by planting crops that are appropriate for the soil profile.

In [8], another approach was to develop an ontology-based recommendation system for crop suitability and fertiliser recommendation. The model was trained using previously collected ontology data from the state of Maharashtra. Fertilizer recommendations are made based on the soil's nitrogen, phosphorus, and potassium (NPK) content. The model also makes crop suitability recommendations. The recommendation system employs the random forest algorithm as well as the k-means clustering algorithm.

In [9], a model for disease detection and pesticide recommendation in agriculture was proposed based on the concept of the internet of things (IOT). Temperature, humidity, and moisture content in the soil and atmosphere were measured using sensors. Image processing techniques were used to preprocess the images, and SVM was used for classification and pesticide recommendation. On detection of climate conditions and plant disease, this system sprinkles water and fertiliser on diseased leaves.

The main issue with these traditional methods is that they are inaccurate when dealing with multi-class pests. A setup aimed at detecting whiteflies using SVM in [6] yields very low accuracy with multi-class classification.To address the issue of multi-class classification, a machine learning model is trained in the proposed work, with pesticide recommendations based on pest detection rather than traditional single-class classification methods, and fertiliser recommendations methods based on soil and climate conditions.

3 Performance analysis

After gathering the data, it is fed into the machine learning model. Decision Trees, Ensembles, and k-Nearest Neighbors (KNN) are used for classification. Both models have been trained and tested using the same data set. The data set is divided into two halves at random, one for training and one for testing. To compare SVM and KNN algorithms, the Receiver Operating Characteristic Curve (ROC) and area under its curve (AUC) are used.

To begin, the following terms are defined in order to clarify the exhibition estimates used for correlation of the calculations:

- True Positive (TP)-A true positive result is one in which the model correctly predicts the positive class.
- True Negative (TN)-A true negative result is one in which the model accurately predicts the negative class.
- False Positive(FP)-A false positive is a result where the model mistakenly predicts the positive class.
- False Negative(FN)-A false negative is a result where the model erroneously predicts the negative class.
- Accuracy: It is the ratio of the number of correct forecasts to the total number of information tests, where Accuracy = (TP + TN) / Total.
- Precision: Precision denotes the number of positive recognizable pieces of proof that were truly correct, where Precision = TP /(TP + FP).

A ROC curve is a diagram that depicts the presentation of a characterization model at all grouping edges. This bend represents two boundaries: True Positive Rate (TPR) and False Positive Rate (FPR). TPR is defined as TPR=TP/(TP+FN). FPR is defined as FPR=FP/(FP+TN).

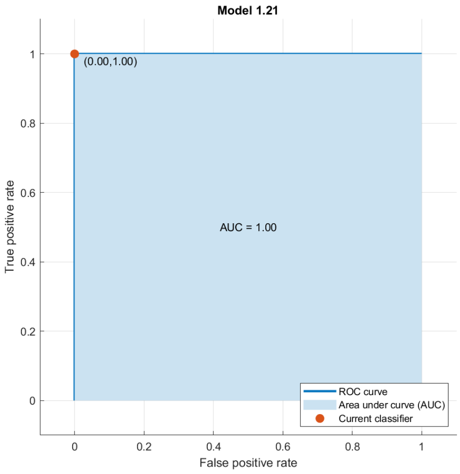

Figure 1: ROC curve of Ensemble model

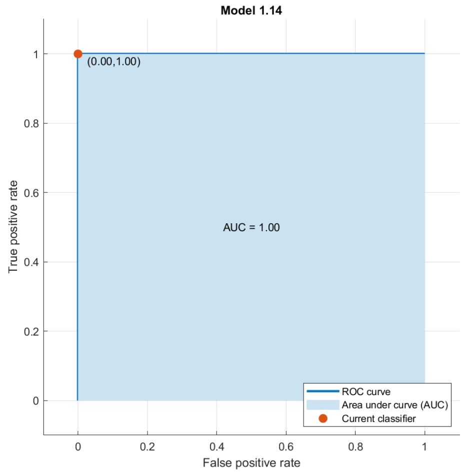

Figure 2: ROC curve of K-Nearest Neighbours model

The AUC estimates the entire two-dimensional region under the entire ROC bend from (0,0) to (1,1). AUC represents the total proportion of execution across all possible characterization limits. The likelihood that the model positions an irregular positive model more profoundly than an arbitrary negative model is one method of deciphering AUC. The value of AUC ranges from 0 to 1.A model with 100% incorrect forecasts has an AUC of 0.0; one with 100% correct expectations has an AUC of 1.0 [10].

Ensemble learning is a powerful machine learning algorithm that has demonstrated clear advantages in a variety of applications. In the context of machine learning, an ensemble is a machine learning system that is built with a collection of individual models that work in parallel and whose outputs are combined with a decision fusion strategy to produce a single answer to a given problem.

Ensemble Bagged Trees: Bagging is an abbreviation for bootstrap aggregation. Averaging together multiple estimates is one way to reduce the variance of an estimate.
f(x) = 1/M $\sum_{m=1}^{M} f_m(x)$, where M is the number of different trees
To obtain data subsets for training the base learners, bagging employs bootstrap sampling. Bagging uses voting for classification and averaging for regression to aggregate the outputs of base learners [11].

Figure 1 shows the ROC curve of Ensemble bagged tree. It also includes the AUC which comes out to be 1.0. It can be observed that the true positive rate remains nearly constant between 0.0 and 0.1.

KNN is a classification method in which the function is only approximated locally and all computation is postponed until the function is evaluated. Because this algorithm relies on distance for classification, normalising the training data can significantly improve its accuracy if the features represent different physical units or come in vastly different scales [12].

K-Nearest Neighbours: KNN makes predictions directly from the training dataset. Predictions for a new instance (x) are made by searching the entire training set for the K most similar instances (neighbours) and summarising the output variable for those K instances. This could be the mean output variable in regression, or the mode (or most common) class value in classification. A distance measure is used to determine which of the K instances in the training dataset are most similar to a new input. The most commonly used distance measure for real-valued input variables is Euclidean distance. The square root of the sum of the squared differences between a new point (x) and an existing point (xi) across all input attributes j is used to calculate Euclidean distance.
EuclideanDistance(x, xi) = sqrt(sum((xj – xij)^2)) [13]

Figure 2 represents the ROC curve of K-Nearest Neighbours model along with its area which comes out to be 0.1. The AUC is closely equal to 1 which represents high accuracy. The line remains nearly constant from 0.0 to 1.0 FPR at 1.0 TPR.

Next, the outcomes and correlation of the different machine learning algorithms are discussed.

Table I summarises the outcomes of all three algorithms. Based on the table, it is clear that KNN is the best method for predicting pesticides. The distinction between Ensemble and KNN is negligible. Because decision trees pro-

	Decision Tree	Ensemble	KNN
Accuracy	83%	94.1%	94.7%

Table 1: Comparative results of machine learning models.

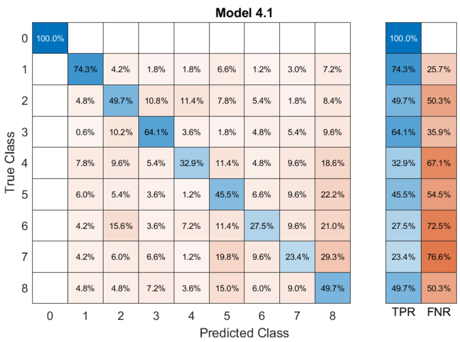

Figure 3: Confusion Matrix for Decision Tree Model

vide the least accuracy, they should be avoided.

A confusion matrix is a specific table design that allows the perception of the exhibition of a calculation in the field of ML and explicitly the issue of measurable arrangement. Each line of the confusion matrix addresses cases in a hypothetical class, while each segment addresses examples in a real class (or the other way around). The name comes from the fact that it makes it easy to see if the framework is confusing two classes (for example, mislabeling one as another). It's a unique type of possibility table, with two measurements ("real" and "anticipated") and indistinguishable arrangements of "classes" in the two measurements. It's a table with two rows and two columns that show the number of false positives (1,0), false negatives (0,1), true positives (0,0), and true negatives (0,0). (1,1).

Next, Figure 3 shows the detailed confusion matrix with values for Decision Tree model. Total false results are more than total true results and hence we should avoid the use of Decision Tree model for pesticide recommendation.

Next, Confusion matrix for KNN is represented in Figure 4 which shows that total true results are much higher than false results. This shows that KNN model can be preferred for pesticide recommendation.

When the information was tested, the Decision Tree model had an accuracy of 83%, the Ensemble model had an accuracy of 94.1%, and the KNN model had an accuracy of 94.7%. Both Ensemble and KNN perform admirably, producing nearly accurate results, with KNN slightly outperforming Ensemble. It is also possible to conclude that Decision Tree is unsuitable for pesticide recommendation. In comparison to other works, such as Gondal d. [6] SVM had a binary classification accuracy of 97%. The result took approximately 22.19s per image to generate, which was far too long for an early-warning system. Our KNN model achieves much higher accuracy on multi classification on a larger number of test cases in less than 5 seconds.

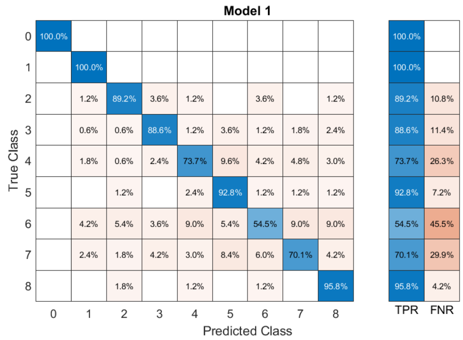

Figure 4: Confusion Matrix for KNN Model

4 Discussion

This is where you interpret your results, pointing out interesting trends within your data and how they relate to your initial hypothesis. This is also the place to justify your methodology, if you're so inclined (i.e. Why did you specifically use a certain statistical test over another? Why this tool over that tool?). Lastly, you're going to want to discuss potential sources of error. Make sure to make explicit reference to figures/tables when discussing your data; it can be helpful to walk the reader through your own personal interpretation of each figure in order.

Explain observations, patterns and trends as represented in graphs using scientific theory and reasoning. The discussion of your results should comprise an interpretation of the results in terms of their significance in relation to your original objective.

5 Conclusions

Use this section to briefly summarize findings, comment on the significance of these findings, and discuss potential future directions for study. This section should always be included in review articles, but is optional in original research articles, since conclusions — without comments — on the significance of findings or suggestions for future work can be included at the end of the discussion section. However, it is strongly recommended.

What are the long-term implications of your findings? Wrap up your discussion succinctly while pointing out the significance of your work as well as it what it means for the fields you examined as much as possible. Lastly, suggest ideas for future studies that could build on your work, and justify why they might be useful. Otherwise, you're all done!

6 Acknowledgements

In the Acknowledgements section, the author(s) acknowledge or thank any persons or institutions who helped support the work in any way. In particular, this section must disclose any funding for the research completed, including the grant number (if a grant was awarded).

Anyone to thank/credit for helping your team along the way? This is the place to do it.

References

[1] Huddar SR, Gowri S, Keerthana K, Vasanthi S, Rupanagudi SR. Novel algorithm for segmentation and automatic identification of pests on plants using image processing. 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT12). 2012:1–5.

[2] Mundada RGMRG. Detection and Classification of Pests in Greenhouse Using Image Processing. IOSR Journal of Electronics and Communication Engineering. 2013;5(6):57–63.

[3] Srinivas K, Usha B, Sandya S, Rupanagudi SR. Modelling of edge detection and segmentation algorithm for pest control in plants. 2011:293–295.

[4] Humeau-Heurtier A. Texture Feature Extraction Methods: A Survey. IEEE Access. 2019;7:8975–9000.

[5] Huang ML. A database of eight common tomato pest images. mendeley. 2020 May. Available from: https://data.mendeley.com/datasets/s62zm6djd2/1.

[6] Gondal D, Khan Y. Early Pest Detection from Crop using Image Processing and Computational Intelligence. FAST-NU Research Journal ISSN: 2313-7045. 2015 02;1.

[7] Preethi G SSMLSDVBB Rathi Priya. Agro based crop and fertilizer recommendation system using machine learning. European Journal of Molecular amp; Clinical Medicine. 2020;7(4):2043–2051.

[8] Chougule A, Jha V, Mukhopadhyay D. Crop Suitability and Fertilizers Recommendation Using Data Mining Techniques. 2019 07:205–213.

[9] IOT based Leaf Disease Detection and Fertilizer Recommendation. International Journal of Innovative Technology and Exploring Engineering. 2019;9(2):132–136.

[10] Huang J, Profile V, Ling CX, of Ottawa CMAJHU, of Ottawa JHU, Profile AJHV. Using AUC and Accuracy in Evaluating Learning Algorithms. IEEE Transactions on Knowledge and Data Engineering. 2005 Mar. Available from: https://dl.acm.org/doi/10.1109/TKDE.2005.50.

[11] Smolyakov V. Ensemble Learning to Improve Machine Learning Results. Medium. 2019 Mar. Available from: https://blog.statsbot.co/ensemble-learning-d1dcd548e936.

[12] K-nearest neighbors algorithm. Wikipedia. 2021 Jun. Available from: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm.

[13] Brownlee J. K-Nearest Neighbors for Machine Learning. Machine Learning Mastery. 2020 Aug. Available from: https://machinelearningmastery.com/k-nearest-neighbors-for-machine-learning/.