Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P

EARLY DETECTION AND CLASSIFICATION OF APPLE LEAF DISEASE-USING MODELS

NARESH KUMAR TRIVEDI
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India,
nareshk.trivedi@chitkara.edu.in
ABHINEET ANAND
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India,
abhineet.anand@chitkara.edu.in
ANKIT KHARE
School of Computer Science, University of Petroleum & Energy Studies, Dehradun, UK, India
ankit.khare86@gmail.com,
UMESH KUMAR LILHORE
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
umesh.lilhore@chitkara.edu.in
SARITA SIMAIYA
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
sarita.simaiya@chitkara.edu.in

Abstract

Plants are considered important due to their role in the human energy supply. Disease has the potential to have a huge impact on the amount of yield a plant can produce, resulting in enormous losses in revenue for the market. Prevention is important for agriculture since the timely discovery of disease is essential for protecting crops. To control plant diseases, extensive study is required, and hence time and talent are needed to handle this problem. To study trends in illness in plants, deep learning is being applied. An example of this is identifying plants based on the overall shape, size, height, and width. A deep learning model was proposed and compared with other classification model in this research for the purpose of identifying and diagnosing apple plant leaf disease. Biotic diseases such as fungal and bacterial infections were studied here. This model shows excellent performance, achieving up to 98.1% classification accuracy.

Introduction

As a fast-developing country, India's agricultural sector is fundamental to its early-stage development. Agriculture has become a key factor in the development of the economy. Farmers use several criteria to determine the most appropriate crop for a certain environment, such as the type of soil, climatic conditions, and the profit potential. Due to a rise in population, weather fluctuations, and political uncertainty, new ways for enhancing food production were sought. Agricultural sector, however, faces many challenges, the most prominent of which is significant loss in crop yield. Using this increases the productivity of researchers, who have to hunt for innovative, efficient and precise technologies.

One of the significant reasons for plant loss due to disease is plant disease occurrence. Identification of plant diseases is a challenging task even in the agricultural area. Nonetheless, in larger farms, this procedure is more costly and less accurate. Because of this, it may be harder for farmers in some countries, like India, to demonstrate a specimen to experts. As a result, it is both time-consuming and more expensive

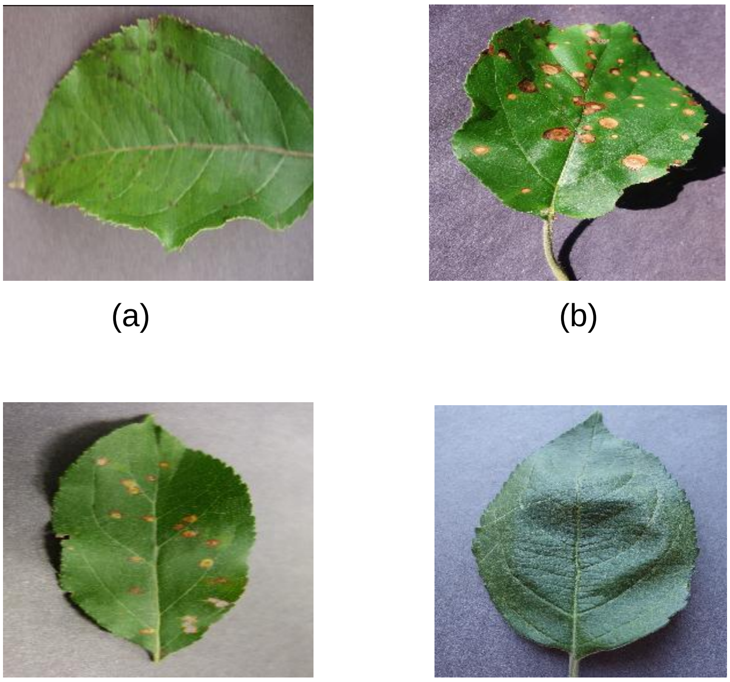
High output on the farm can be obtained by precision agriculture and information technology, which collects information and data on farms. New technologies have been developed that use advanced methodologies to yield improvements on farms. It is feasible to accomplish economic growth in agriculture by using new technologies. It is useful for numerous purposes, like determining the presence of pests in plants, detecting weeds, increasing crop yields, and identifying plant diseases. To protect the crops from illness, to reduce the population of pests, and to boost production, farmers employ pesticides. Crops suffer from illnesses, and this is having an impact on productivity and the economic well-being of farmers and agricultural companies. Because identification of disease and severity is now needed, disease must be identified.

Additionally, ongoing monitoring is required. So, machine learning. Predicting diseases of plant leaves caused by fungus, bacterium, and virus is done by applying accurate detection methods to machine learning results. Although the accuracy of disease prediction using classification algorithms appears to be variable for varied input data, such prediction is difficult. In this study, apple plant disease detection and classification research contributions are summarised and contrasted.

CLASSIFICATION OF PLANT DISEASES

Leaf diseases include rust, mildew, bacterial blight, Downey mildew, brown spot, powdery mildew, and many others. In figure 1, we see how bacterial, fungal, and viral diseases are classified. In order to examine disease on different types of plants, researchers such as J. D. Pujari, R. Yakkundimath, and A. S. Byadgi used Artificial Neural Network, Probabilistic Neural Network, and Support Vector Machine for disease identification [1-4]. The computerised Fuzzy-Relevance Vector Machine classifiers constructed by BalasubramanianVijayalakshmi and Vasudev Mohan [5-7] employed training features and labels as input values. 8-10 prophesied the development of Phytophthorainfestans disease diagnosis on tomatoes by utilising Artificial Neural Networks (X. Wang, M. Zhang, J. Zhu, and S. Geng). Minimum Distance Classifier (or distance matrix classifier) was proposed by Dong Pixia and Wang Xiangdong. An algorithm for classifying illnesses of plants, including jackfruit, tomato, etc., was proposed by S. Arivazhagan, R. NewlinShebiah, S. Ananthi, and S. Vishnu Varthini and this includes SVM classifier as part of the classification method [16-18].

Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P



(a) (b)

(c) (d)

Figure 1: a) Apple scab b) Apple black rot c) Apple rust d) Apple healthy

Proposed system

The technology primarily helps to detect diseases in plant to prevent the loss in product yield. The proposed work focuses on disease detection in plants using image processing and machine learning techniques to attain few objectives such as: i) To detect diseased leaf, ii) To build a prediction model for disease type classification. The images of appleplantare given as input affected with diseases category like scab, cedar rot, and healthy etc. The flow diagram of the proposed work is depicted in Fig.2.

Image Import	Pre-processing (Image embedding)	Segmentation (Apply classification model)	Evaluate Model	Confusion Matrix

Process Flow of the Proposed Disease Detection System

An image of a plant leaf is given as input to detect whether the plant is infected or not and classification is performed if the plant is an infected one to detect the

Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P

disease type. The first thing to do is to import the image via the Import images widget. This widget works as file widget for image. Import Images widget accepts a directory instead of a file.

Image Embedding widget is the most important for the entire **Image Analytics** package as this is where the magic happens. For your information, classification and regressions tasks requires data in the form of numbers and there isn't a good way to perform such tasks with images unless we represent it in the form of numbers. This is where **Image Embedding** widget works by converting it to a vectors of numbers. **Image Embedding** widget reads images and uploads them to a remote server or evaluate them locally. The most important parameters for the **Image Embedding** interface is the Embedder. Squeeze Net embedded was used as it gives result as well as Alex net and uses less no of parameter.

For detection and classification of apple leaves following well known classification algorithms models were used and their outcomes are compared.

K-Nearest Neighbours (KNN): Simple linear regression and classification problems often employ the K-Nearest Neighbours (KNN) technique. Classify new data points based on the similarities measure (e.g. distance function).

The majority vote classifies its neighbours. This class gets the data based on its immediate neighbours. Accuracy is improved the more nearby neighbours you include.

Support Vector Machines (SVM):This machine learning method is called supervised. machine learning algorithms learn from prior input data and anticipate the output Support vector machines are classifiers and regressors that help in supervised learning.For example: The system is trained to recognise apples leaves. Once this data is analysed, it always gives apples leaves as the output.

AdaBoost:AdaBoost is a Boosting approach that is used as an ensemble method in machine learning. This adaptive boosting, or re-assigning weights to each occurrence, is used because inaccurate classifications are weighted higher. Boosting is used to combat bias and variability for supervised learning. It uses a simple sequential growth model. There are a total of four levels of learner in the entire process. Plainly stated, weak learners become strong learners.

Naïve Bayes: Naive Bayes is a binary and multi-class classification technique. For a better understanding, consider using categorical or binary input values. In order to make the computation of probabilities simple, Bayesian methods (dubbed 'naive' or 'idiot' Bayes') simplify the calculations. Rather than trying to determine each attribute value P(d1, d2, d3|h), they are presumed to be conditionally independent and calculated as P(d1|h) * P(d2|H) and so on. This is a pretty strong assumption, namely that the properties do not interact. Nonetheless, the technique performs well even when this assumption is invalid.

Neural Networks: Three layers of neurons can be used to make neural networks: the input layer, the hidden layer(s), and the output layer. In-between layers consist of several neurons, and interconnections are made between them. The more the

Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P

neural network learns, the more accurate its predictions become as the strength of the connections between neurons is fine-tuned.

The test and score widget tests learning algorithms. Different sampling schemes are available, including using separate test data. The widget does two things. First, it shows a table with different classifier performance measures, such as classification accuracy, area under the curve, F1 score, and Recall value. Second, it outputs evaluation results, which can be used by other widgets for analysing the performance of classifiers, such as ROC Analysis or Confusion Matrix.

The Confusion Matrix gives the number/proportion of instances between the predicted and actual class. The selection of the elements in the matrix feeds the corresponding instances into the output signal. This way, one can observe which specific instances were misclassified and how. The widget usually gets the evaluation results from Test & Score.

The overall workflows setup model to detect and classify apple leaves based on the diseases given figure 3.

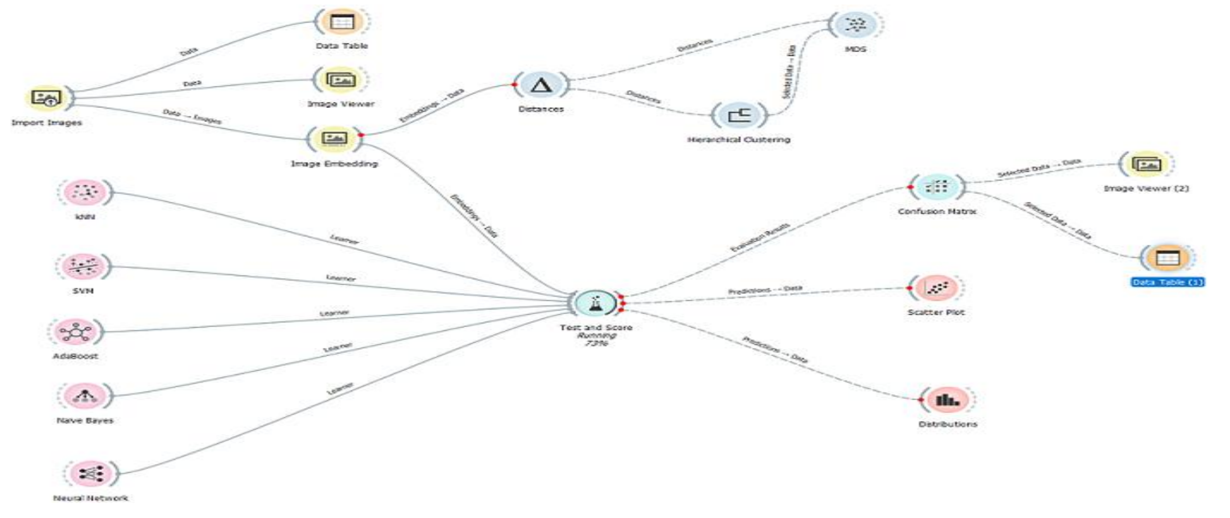


Figure 3

Image analytics by using visual programming. This toolset includes visual programming. Assembly of analytic workflows is enabled for end users by providing a toolkit that makes use of several components, such as image loaders, embedders, and picture profile analyzers. Orange is a data mining framework already including components for clustering, classification, and interactive data and model representations. This picture-specific extension is available in Orange's add-on for image analytics. Open-source and freely available via Orange's home page are supplied for both Orange and the proposed extension Workflows in Orange are used for data analysis. A workflow is a set of widgets—components that may process, model, or show data. Widgets take data as input and show or send results as output. Workflows in Orange are configured by choosing widgets and connecting them together. For instance, the workflow in Fig. 3 loads a group of photos from the directory of your choice, then embeds the images with feature vectors, calculates the

Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P

distances between these vectors, then clusters and visualizes image similarities in the multidimensional scaling plot. Every stage of the Orange procedure may be watched and inspected by users For instance, they can view the photos that have been loaded (the Load Images step in Fig. 3), visually examine the results of hierarchical clustering in a dendrogram, and even display the branch selection or selection of images in a multi-dimensional scaling plot. The selected images can be examined in the dendrogram as well (achievable by connecting the Data Table widget to any of the widgets in the workflow; for brevity not shown in Fig. 3). Checking and inspecting the outcomes at every phase of the analysis pipeline helps users acquire trust in results and understanding of analysis techniques. Educators can also use it to show analysis methods to trainees.

Interactive data visualizations.

Orange widgets are interactive, and when any change is made to widget parameters or any option is made in the graphical presentation, results appear almost instantly. An example of a user-friendly widget found in Fig. 3 is the hierarchical clustering widget. The Hierarchical Clustering widget provides the selected branch's data for image display (Image Viewer (2) in Fig. 3). Any change in the selection of branches in the dendrogram propagates through the workflow and instructs, for instance, Image View (2) to display the photos relevant to the user's selection in the dendrogram.

Results

Using various multi-color image sets of 2668 images of apple leaves data sets from keggle data collection repository (Fig. 4). We employed data mining approaches that are able to deal with noise and discrepancies. Activities such as data clustering, data projection (unsupervised data mining), and the construction of prediction models are the norm in biomedical data mining (supervised data mining). The color orange makes it easier to do these jobs. Using only processes with only a few data processing and visualization elements, users may easily perform these tasks. Figure 1 depicts a workflow for classification models such k-Nearest Neighbors (kNN), Support Vector Machines (SVM), AdaBoost, Naïve Bayes, and Neural Networks. In the workflow, the images are loaded into vector space, their vectors are computed, and the classification is done.

Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P

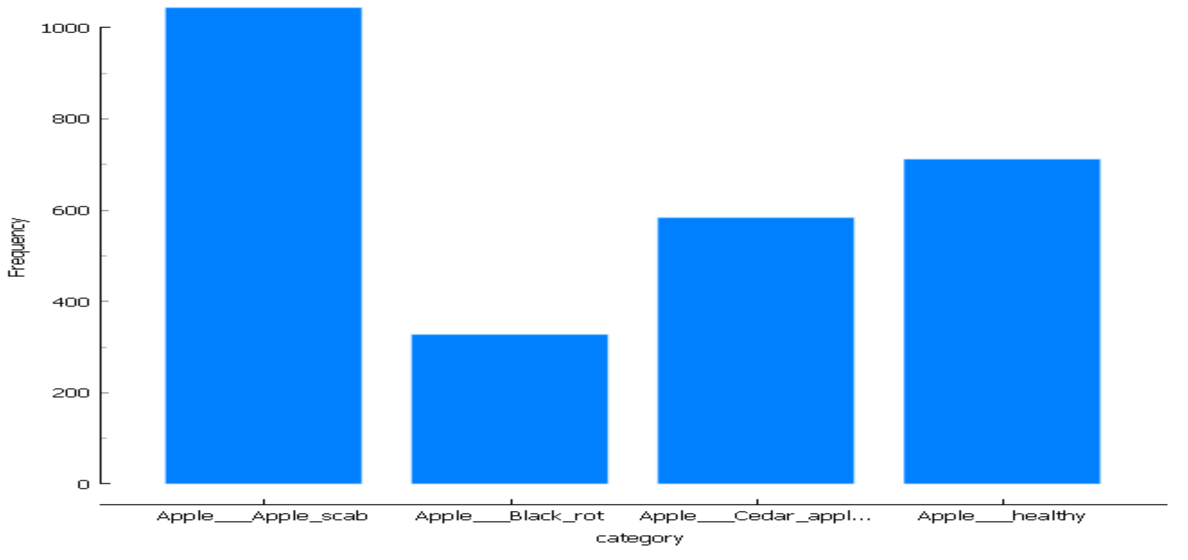
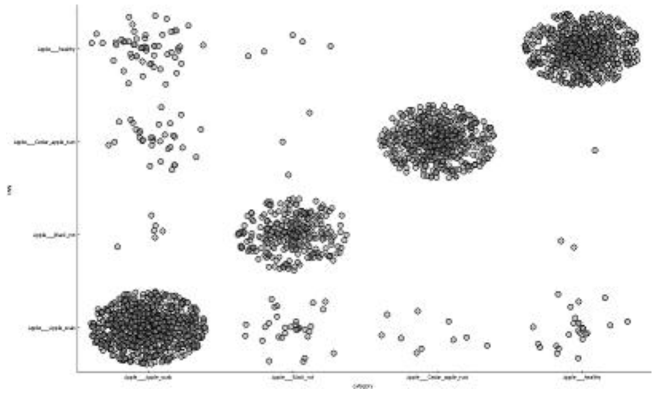


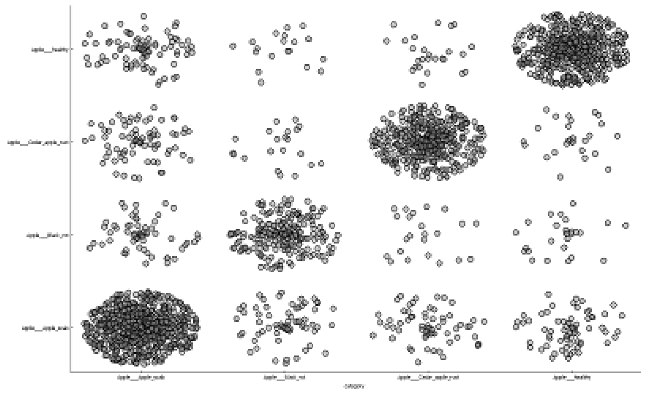
Figure 4: Category wise apple leaves data set

By running the workflow in Fig. 3, the analysis of the leaf names, size, height, and width, see Figure 4. The model above can also be used to develop classifiers that make accurate predictions about the classifications of these apple leaves. First, we can learn from a training set to measure the models' correctness. Then, we can test the models on a different test set to see how well they've learned. To achieve accuracy, the process shown in Fig. 3 conducts these operations and employs cross-validation. The data came from an image embedding matrix, therefore we used k-Nearest Neighbors for modeling, then cross-validation to estimate accuracy. While 92 percent of the photos and three categories from the apple leaves were correctly categorized, only 96 images and 4 categories from the leaves were incorrect, giving in a remarkably high accuracy of 93.5 percent. In order to learn from the picture embedding matrix, SVM is used, and cross-validation is employed in order to estimate the accuracy. The relatively high accuracy of 97.9 percent was due to only 57 out of the 2668 pictures and 4 categories from the apple leaves being categorized incorrectly. The accuracy percentages for Adaboost 566, Naïve bayes classification model 484, and neural network classification model 50 are 78.8%, 78.8%, and 98.1%, respectively refer figure 5.

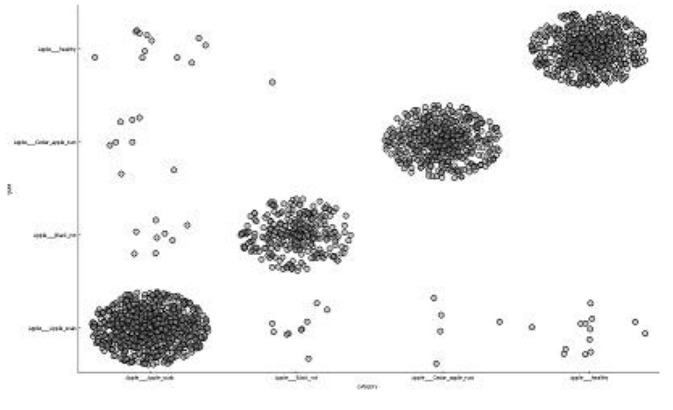
Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P

Different scatter plot diagram with misclassified prediction of apple plant leaves



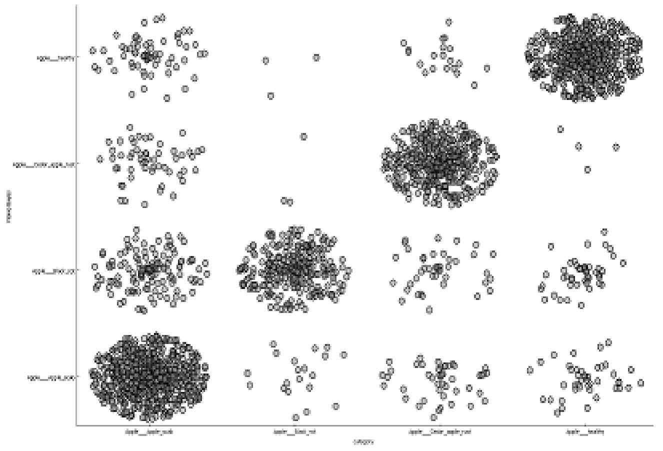


kNNAdaBoost

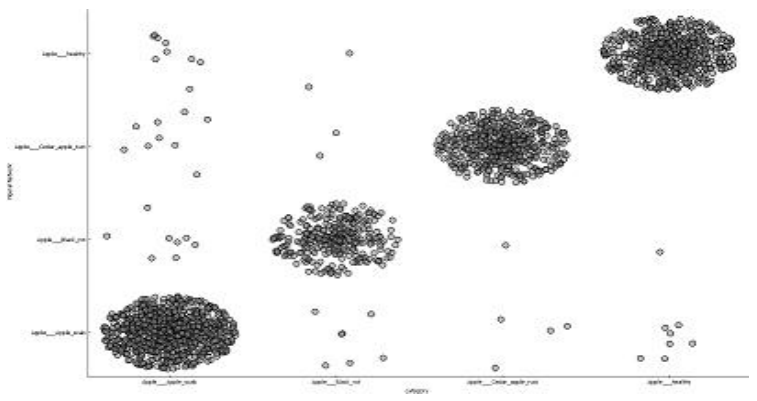


Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P

SVM



Naïve bayes



Neural Network

Figure 5:

The use of the workflow in Fig. 3 is also advantageous since it provides increased accuracy while offering exploration and comprehension of the results in an interactive confusion matrix. Another way to state this is that if you follow the method, you will see correctly labeled apple leaves in the image viewer. Classification models in other classifications had the same methodology applied to them. KNN, SVM, Adaboost, Naïve Bayes, Neural Network for the apple leaves classification all yielded good cross-validated accuracy, with F1 values of 0.935, 0.979, 0.788, and 0.981 correspondingly. Additionally, F1 values of 0.95 were found for apple disease classification refer figure 6.

Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P

Confusion matrix for kNN (showing number of instances)

	Predicted				
Actual	Apple___Apple_scab	Apple___Black_rot	Apple___Cedar_apple_rust	Apple___healthy	∑
Apple___Apple_scab	948	6	33	57	1044
Apple___Black_rot	31	289	3	5	328
Apple___Cedar_apple_rust	10	0	574	0	584
Apple___healthy	25	2	1	684	712
∑	1014	297	611	746	2668

Confusion matrix for AdaBoost (showing number of instances)

	Predicted				
Actual	Apple___Apple_scab	Apple___Black_rot	Apple___Cedar_apple_rust	Apple___healthy	∑
Apple___Apple_scab	824	60	77	83	1044
Apple___Black_rot	68	220	21	19	328
Apple___Cedar_apple_rust	65	23	470	26	584
Apple___healthy	70	27	27	588	712
∑	1027	330	595	716	2668

Confusion matrix for SVM (showing number of instances)

	Predicted				
Actual	Apple___Apple_scab	Apple___Black_rot	Apple___Cedar_apple_rust	Apple___healthy	∑
Apple___Apple_scab	1016	8	8	12	1044
Apple___Black_rot	10	317	0	1	328
Apple___Cedar_apple_rust	5	0	579	0	584
Apple___healthy	13	0	0	699	712
∑	1044	325	587	712	2668

Confusion matrix for Naive Bayes (showing number of instances)

	Predicted				
Actual	Apple___Apple_scab	Apple___Black_rot	Apple___Cedar_apple_rust	Apple___healthy	∑
Apple___Apple_scab	789	134	55	66	1044
Apple___Black_rot	20	302	3	3	328
Apple___Cedar_apple_rust	46	43	475	20	584
Apple___healthy	43	47	4	618	712
∑	898	526	537	707	2668

Confusion matrix for Neural Network (showing number of instances)

	Predicted				
Actual	Apple___Apple_scab	Apple___Black_rot	Apple___Cedar_apple_rust	Apple___healthy	∑
Apple___Apple_scab	1018	8	9	9	1044
Apple___Black_rot	7	317	2	2	328
Apple___Cedar_apple_rust	4	1	579	0	584
Apple___healthy	7	1	0	704	712
∑	1036	327	590	715	2688

Figure 6:

The results of our pilot investigation support the notion that general embedders might perform adequately well, even when given the benefit of specialist embedders. An excellent illustration of this is our work in which we discovered that F1 cross-validated accuracies were higher than predictions from neural networks.

Conclusion

The overall accuracy was approximately 0.981, resulting in an F1 score of approximately 0.981. Though impressively precise models were constructed from

Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P

features using a deep neural network that was trained on 70,433 photos, this wasn't expected, since deep neural networks are generally unable to reconstruct images. According to these findings, even easy transfer learning techniques can be used to tackle a large range of picture analytics applications. They point out that domain-specific networks may have a low marginal utility and their additional precision will not be sufficiently compensated by the effort required to build them.

References

[1] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, “Identification and classification of fungal disease affected on agriculture/horticulture crops using image processing techniques,” *IEEE International Conference on the Computational Intelligence and Computing Research*, 2014.

[2] BalasubramanianVijayalakshmi and Vasudev Mohan, “Kernel based PSO and FRVM: An automatic plant leaf type detection using texture, shape and color features,” *Computer and Electronics in Agriculture*, vol. 125, pp. 99-112, 2016.

[3] X. Wang, M. Zhang, J. Zhu and S. Geng, “Spectral prediction of Phytophthorainfestans infection on tomatoes using artificial neural network (ANN),” *International Journal of Remote Sensing*, pp. 1693–1706, 2008.

[4] Dong Pixia and Wang Xiangdong, “Recognition of Greenhouse Cucumber Disease Based on Image Processing Technology,” *Open Journal of Applied Sciences*, vol. 3, pp. 27-3, Mar. 2013.

[5] S. Arivazhagan, R. NewlinShebiah, S. Ananthi and S. Vishnu Varthini,“Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features,” *Commission Internationale du Genie Rural(CIGR) journal*, vol. 15, no. 1, pp. 211-217, 2013.

[6] HarshalWaghmare, RadhaKokare and YogeshDandawate, “Detection and Classification of Diseases of Grape Plant Using Opposite Colour Local Binary Pattern Feature and Machine Learning for Automated Decision Support System,” *3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, 2016.

[7] T. H. Jaware, R. D. Badgujar and P. G. Patil, “Crop disease detection using image segmentation,” *Proceedings of Conference on Advances in Communication and Computing (NCACC'12)*, April, 2012.

[8] Dheeb Al Bashish, Malik Braik, and SuliemanBani-Ahmad, “A Framework for Detection and Classification of Plant Leaf and Stem Diseases,” *International Conference on Signal and Image Processing*, 2010.

[9] A. A. Joshi and B. D. Jadhav, “Monitoring and Controlling Rice Diseases Using Image Processing Techniques,” *2016 International Conference on Computing, Analytics and Security Trends (CAST) College of Engineering Pune*, Dec 2016.

Tianjin DaxueXuebao (ZiranKexueyuGongchengJishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online): 0493-2137
E-Publication: Online Open Access
Vol:54 Issue:08:2021
DOI 10.17605/OSF.IO/X8J6P

[10] Bauer, S. D. Korc and F. Forstner, "Investigation into the classification of diseases of sugar beet leaves using multispectral images," *Wageningen Academic Publishers,* pp. 229-238, 2009.

[11] K. R. Gavhale, U. Gawande and K. O. Hajari, "Unhealthy region of citrus leaf detection using image processing techniques," *International Conference on Convergence of Technology I2CT*, pp. 1–6, 2014.

[12] S. B. Jadhav and S. B. Patil, "Grading of Soybean Leaf Disease Based on Segmented Image Using K-means Clustering," *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, vol. 4, no. 6, Jan 2015.

[13] K. Renugambal and B. Senthilraja, "Application of image processing technique in plant disease recognition," *International journal of engineering research and technology*, vol. 4, no. 03, Mar. 2015.D. Luna, G. Robert G. Renann, Baldovino, E. A. Cotoco, Anton Louise P. de Ocampo, C. Ira,Valenzuela, A. B. Culaba, and E. P. DadiosGokongwei, "Identification of philippine herbal medicine plant leaf using artificial neural network," *9th IEEE International* Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), pp. 1-8, 2017.

[14] Sandika Biswas, BhushanJagyasi, Bir Pal Singhy and Mehi Lal, "Severity Identification of Potato Late Blight Disease from Crop Images Captured under Uncontrolled Environment," *IEEE Canada International Humanitarian Technology Conference - (IHTC)*, 2014.

[15] Bin Liu, Yun Zhang, Dong Jian He and Yuxiang Li, "Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks," *Symmetry,* vol. 10, no. 11, 2017.

[16] HabibollahAghAtabay, "Deep residual learning for tomato plant leaf disease identification," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 24, 2017.

[17] S. Gaikwad and K. J. Karande, "Image Processing Approach For Grading And Identification Of Diseases On Pomegranate Fruit," *International Journal of Computer Science and Information Technologies (IJCSIT)*, vol. 7, no. 2, pp. 519-522, 2016.

[18] Wan MohdFadzil, Shah Rizam, R. Jailani and M. T. Nooritawati, "Orchid leaf disease detection using Border Segmentation technique, *IEEE Conference on Systems, Process and Control (ICSPC)*, vol.1, pp.168-179, Dec. 2014.