Smart Agriculture: Detecting and Classifying Leaf Diseases with ML Techniques

[1] Ritesh Kumar Mishra, [2] Satyam Kumar, [3] Abhishek Kumar Saxena

1 Information Technology, Bansal Institute of Engineering and Technology

2. Information Technology, Bansal Institute of Engineering and Technology

3 Information Technology, Bansal Institute of Engineering and Technology

---***---

Abstract:

The concept of savvy cultivating is picking up pace in the agrarian segment these days, and it utilizes sensors and an extent of machine learning-based advances. As per later overviews, 56 percent of the agrarian segment is enduring overwhelming misfortunes due to illnesses shaping plant clear-outs. It's imperative to screen the spread of the infection and progress agrarian yields. We are required to recognize the malady to begin with on time in order to maintain a strategic distance from it spreading. Hence, we can address this issue by executing a few calculations to identify affliction and clear it out. In this paper, we have compared support vector machines (SVM), K-Nearest Neighbor (KNN), and convolutional neural networks (CNN). The three different models are presented and examined in this ponder, and they are competent at distinguishing eight different leaf illnesses. The CNN demonstration has come to a 96 percent precision when preparing the pictures of the soybean leaf illness dataset, outperforming the KNN and SVM models, which have exactness of 64 percent and 76 percent, respectively.

Introduction:

India is the second-largest country in the world, and feeding a huge people is not a basic errand. Other than that, we are, moreover, confronting a nourishment emergency along with a steep increment in nourishment prices. The sooner we distinguish the malady, the longer we have to remedy it and maintain a strategic distance from the misfortune of the edit. It was a moderate preparation already to distinguish which malady existed in the plant, and by that point the illness had spread to the entirety. In order to dodge the misfortune of the trim, we have to grasp current advances like AI and machine learning. Soybean leaf infections of diverse sorts are caused by temperature varieties or other bacterial maladies, as per the dataset utilized in our consideration [1].

This is a multi-class classification issue due to the huge number of names to be classified. The maladies being classified are as follows:

Bacterial Curse: It is a predominant soybean infection that is more likely to happen in cool and damp climates. The infection is fundamentally shown in moo levels, and the disease can be spread through seeds.

• Brown Spot: This illness is brought on almost by a bacterial or parasitic disease on the takeoffs, as well as unpredictable watering of the plant. The ailment can be analyzed by the expansive number of spots that are unmistakable on it.

• Copper Phytotoxicity: The condition is brought around by over-the-top levels of copper in plant tissues, which are habitually showered over a wide field, as well as inadequate precipitation at the same area.

• Fleece Mold: It is a fungal-like living being foliar malady. It is transmitted from plant to plant through airborne spores. It is a wet-weather infection since the contamination is encouraged by amplified leaf moisture.

• Solid: This category has a collection of solid clears that can be utilized to classify the leaf when it is solid.

• Fine Buildup: The white shape is caused by tall stickiness and a requirement of wind stream. Planting your vegetation too close together, anticipating palatable conversation or almost circulation, or overwatering your crate or arranging soil might lock in the advancement of white mold.

• Soybean Mosaic Contamination: This sickness shows up in the center of the winter and at that point vanishes in the center of summer. If the infection shows up and can be made, common conditions that are favorable for aphid change can energize this disease.

Methodology:

Two vital examination gadgets were utilized to conduct the bibliometric examination: bibliophagy and VOS Watcher. The fundamental inspirations for utilizing the two as of now said examination-defiant are as follows. Bibliophagy's predominance stems from the truth that, not at all like other bibliometric gadgets, it offers a comprehensive set of quantifiable strategies and visualizations that permit execution examination and conceptual mapping of the subject of investigation [55]. Other than that, Bibliophagy is discharged as an open-source R bundle with a web-based graphical interface.

2. RELATED WORKS:

Paper [2] briefs around the leaf disclosure in tomato plants utilizing CNN; for that, they utilized the exchange learning concept and imported ResNet-50. They wrapped up with an exactness of 97%, and due to this tall aggregate of exactness, the outline can recognize the infection inside the most compelled period. In showing disdain toward the reality that in [3], they executed utilizing CNN-based Alex and compared it with VGG-16 and Lenet-5 models utilizing a dataset comprising, for the most part, 7000 pictures. They achieved an exactness of 96.7% and utilized a few principal ML calculations like SVM and KNN, but they got a lower precision than VGG-16 and Lenet-5. In the paper, a few of the afflictions are portioned utilizing Otsu's strategy and utilized Adjoining Twofold Plans (LBP) and amass to disengage different highlights from it. Classified the information utilizing the SVM strategy and wrapped up with a precision of 94.6% utilizing a polynomial bit. So early-stage range of plant sickness will offer assistance to change from crushing its yield.

3. Machine learning strategies:

3.1 Support Vector Machines

When utilizing SVM, it is one of the critical models of machine learning that for straight information, one of the most vital hyperplanes is made, and for non-linear information, it depends upon bits. After you donate the outline and the input photographs, a computation is done to select the number of highlights that can be recovered from the information. The information is at that point put on a chart, and a few hyperplanes are made [6].

Parts, which are utilized for SVMs, are by and large non-linear, which move data from a lower-dimensional space to a higher-dimensional one by a hyperplane keeping to data. There are, on an outstandingly basic level, two sorts of bits that can be utilized for this sort of data; choosing the most sensible bit can be done in the following ways: Understanding the concept of hyperparameter tuning. In this concept, all bits are utilized to start with to check how adjusted the data is; the one with the most principal exactness is named the best bit for that data [7, 8]. There are 2 sorts of kernels.

- Polynomial Kernel

- Spiral Premise Work Part (RBF kernel)

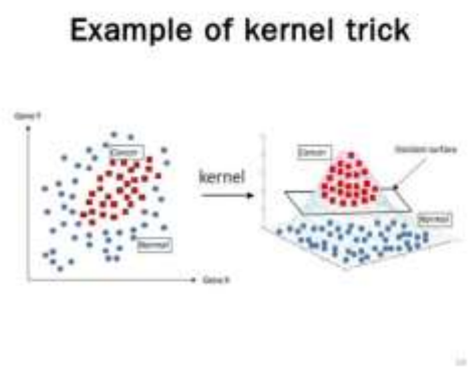


Fig. 1. Kernel procedure

3.2. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is an administered learning strategy that classifies the information focuses into classes concurring with their closest neighbors. When an unused information point is included, the demonstrator finds the K closest focuses from the preparing information set. The most visited course among the K neighbors is utilized to classify the unused information point. It is executed for all the modern information focuses to classify them in like manner [10].

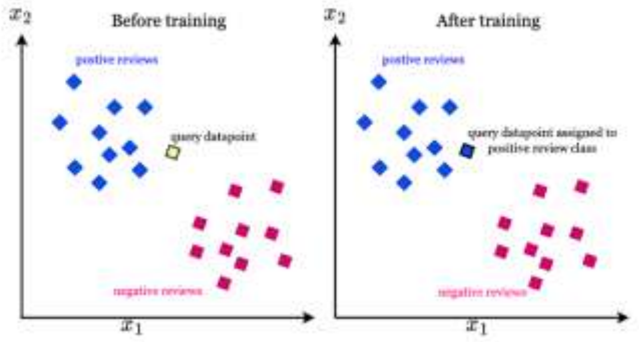


Fig. 2. K – Nearest Neighbors

To choose the ideal estimate of K, the demonstration can be endeavored for different values of K. It is regular to select an odd estimate of K so that there are no ties when partitioning information into two classes. Execution of different values of K can be compared based on classification precision and can be delineated utilizing plots for understanding more about the demonstrated behavior.

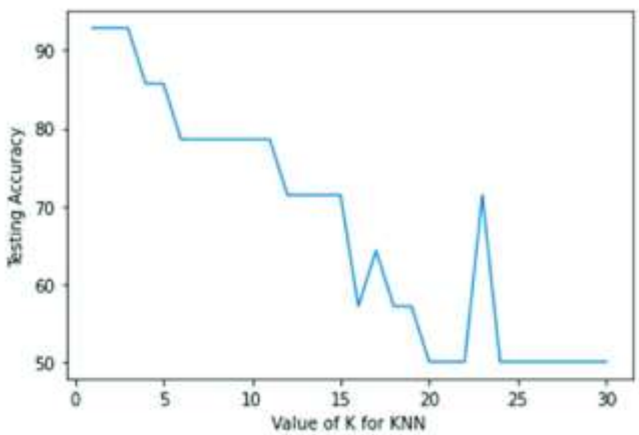


Fig. 3. Accuracy plot for best K-NN

3.3. Convolutional Neural Network

It is a subarea of profound learning calculations where such calculations, when actualized in real-time, see the world as people do and handle it as people handle it, and all the conceivable ways in which a human would think about almost a specific circumstance beneath all conceivable circumstances are prepared into such calculations. These machines have advanced into all angles of life, shifting from our typical versatile phones to exact supercomputers. They are supplanting people with advanced machines to improve the efficiency of the labor [13].

There are 3 layers in this CNN:

• Input Layer

• Hidden Layer

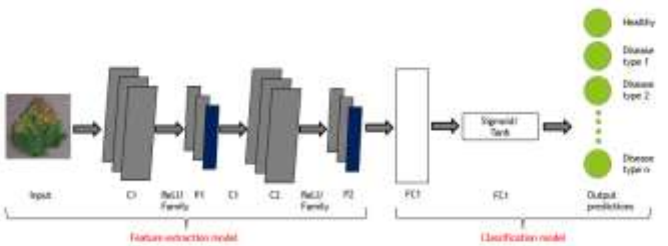
• Output Layer

• Input Layer: The input layer comprises pictures given as inputs to the show; here, picture data is put away in pixels and kept in hubs, and all handling operations in CNN happen in nodes.

•Hidden Layer: This covered-up layer in this layer is mindful for performing calculations such as information preparation, including extraction and information change. The higher the number of covered-up levels in the CNN engineering, the more complex the structure. The information handling in the covered-up layer permits the show to learn highlights and react rapidly to real-time data, which makes the show closer to the current-day CNN models.

Throughout this extent, we utilized the highlights determined from the covered-up layer to identify and foresee the nearness of disease on the leaves.

•Output Layer: The yield layer is a completely associated layer where the covered-up layer yields are smoothed and are utilized as an input. It gets input and changes it over into the wanted classes [14].



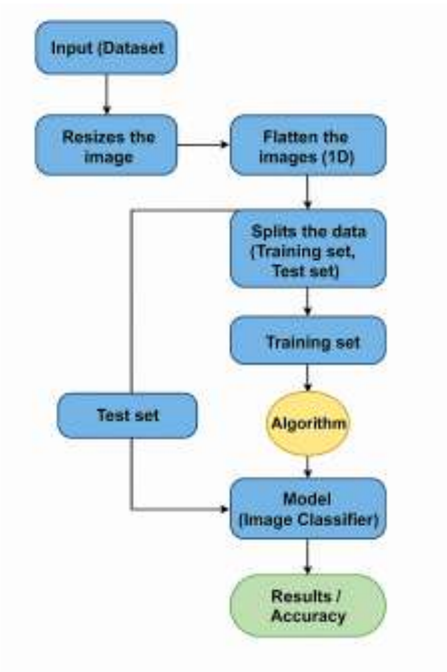
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

4. Implementation

The soybean dataset that was attained from reference [13], distinctive machine learning calculations are imported, and at that point, SVM, KNN, and CNN calculations executed in Python are employed to decide the kind of infection passing in the flake.

4.1. Back Vector Machines

Outlines how SVM is employed. After bringing in the dataset, all cinema is resized and smoothed, converting 2D cinema to 1D cinema. The dataset is at that point part of prepared information and test information in a proportion of 7030, with the prepared information being employed as an input to the calculation and the test information being a test of the fineness of the calculation before running the SVM calculation. Once a cast of perfection to fête infection in the flake, test cinema is given as illustration input, and (b) shows the show parameters for fitting the cinema in the SVM model.



5.(a) SVM operation

4.2. K-Nearest Neighbors (KNN)

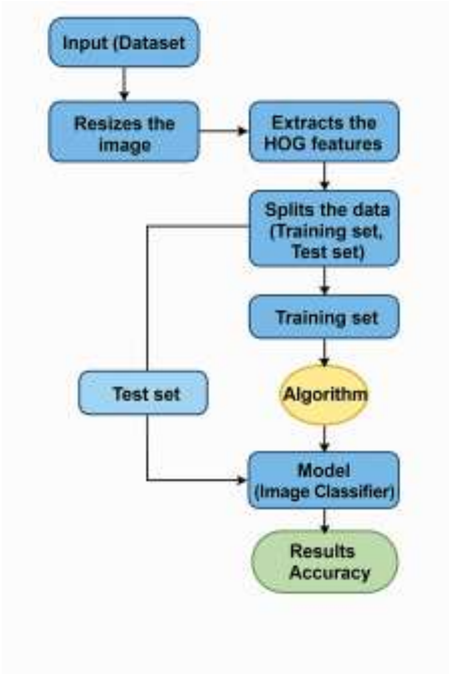


Fig. 6 illustrates how KNN is utilized. Or maybe of the pictures, KNN portions overeater highlights utilizing Euclidean remove and Manhattan remove from them and two parts of the dataset in the same way that does; still, KNN calculation is utilized, or maybe of SVManda7525 part between prepare and test information, Manhattan remove gives lesser delicacy in the KNN calculation when compared to Euclidean remove. To forecast the complaint, test pictures from fromtest information are utilized as test input.

4.3. Convolutional Neural Networks (CNN)

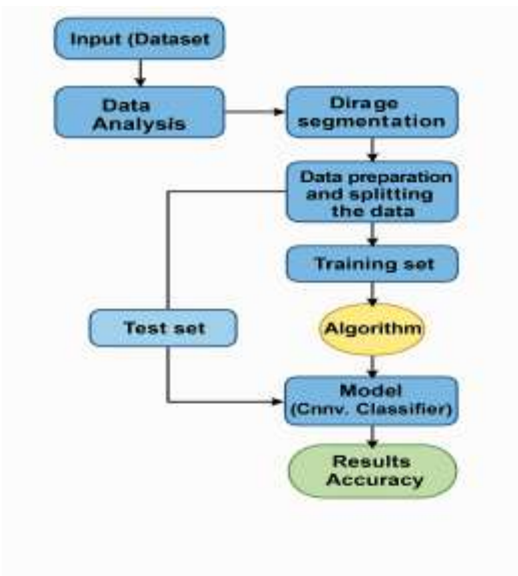


Fig. 7. (a) CNN Implementation;

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 72, 72, 32)	896
max_pooling2d (MaxPooling2D)	(None, 36, 32)	0
conv2d_1 (Conv2D)	(None, 36, 36, 128)	36,992
max_pooling2d_1 (MaxPooling2D) 2D)	(None, 18, 18, 128)	
conv2d_2 (Conv2D)	(None, 18, 18, 256)	295,168
max_pooling2d_2 (MaxPooling2D)	(None, 9, 9, 256)	0
flatten (Flatten)	(None, 20736)	0
dense (Dense)	(None, 128)	2.654.336
dense_1 (Dense)	(None, 64)	8,256
dense_2 (Dense)	(None, 10)	650

(b) CNN model simulation parameters

The flowchart depicts a CNN-based image classification pipeline. From the input dataset, it performs data analysis and image segmentation, then data preparation and train-test split. The training set feeds an algorithm to build the CNN model, followed by the built model's testing on the test set to produce accuracy results.

5. Results

We worked with and compared three machine learning algorithms in our study on the detection of soybean leaf disease: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Convolutional Neural Network (CNN). We have trained the models using a data set containing images of eight leaf types, like different diseases and normal leaves. We conducted experiments that made CNN far more accurate in precision than the remaining two algorithms, which achieved a precision of 96%. SVM and KNN, however, clocked 76% and 64%. This is because CNN has the ability to learn complex features based on its deep structure; hence, it is better suited for image classification tasks. This discovery shows the potential of CNN in real-time disease detection systems so that early diagnoses can be made and farmers can take timely action to prevent crop loss. Therefore, CNN is the most efficient model for multi-class classification when it is a matter of soybean leaf disease identification.

6. Conclusion

Here in this study, we have analyzed the application of machine learning models, that is, SVM, KNN, and CNN, to classify and detect eight various soybean leaf diseases. Our findings depicted that CNN was better than other models with 96% accuracy rate. As compared to SVM and KNN models, the accuracy rates were 76% and 64%, respectively. The justification for the excellence of CNN's performance is based on its deep learning ability, which allows it to extract complex features from images, hence being highly suitable for classification of plant disease. The research shows that CNN is a viable method for the early diagnosis of soybean leaf disease, allowing farmers to prevent crop loss through early action. This study emphasizes the application of sophisticated machine learning approaches in smart farming, which can significantly enhance agricultural yield and mitigate the impact of plant diseases on yield.

7. Future Scope

The regions of future investigation for this ponder contain optimization of CNN models with more design complexity, such as ResNet, or beginning to accomplish more exactness. Real-time malady discovery utilizing portable apps or IoT sensors can encourage prompt mediation by agriculturists. Scaling up the dataset sizes to incorporate a more noteworthy assortment of plants and natural conditions will progressively demonstrate generalizability. Combining picture investigation with natural information can encourage upgrading the exactness of forecasts. In addition, mechanized illness treatment recommendations and edge computing for nearby, speedier handling are ranges prepared for future breakthroughs, with tall potential for benefits to worldwide agrarian efficiency and nourishment security.

References

[1] https://www.digipathos-rep.cnptia.embrapa.br/jspui/simple-search?filterquery=Soja+%28Soybean%29&filtername=crop&filtertype=equals

[2] N. K. E., K. M., P. P., A. R., and V. S., "Tomato Leaf Disease Detection using Convolutional Neural Network with Data Augmentation," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 1125-1132.

[3] P. B. R., A. Ashok, and S. H. A. V., "Plant Disease Detection and Classification Using Deep Learning Model," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 1285-1291.

[4] M. E. Pothen and M. L. Pai, "Detection of Rice Leaf Diseases Using Image Processing," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 424-430.

[5] J. N. Reddy, K. Vinod, and A. S. Remya Ajai, "Analysis of Classification Algorithms for Plant Leaf Disease Detection," 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2019, pp. 1-6. [6] N. R. Bhimte and V. R. Thool, "Diseases Detection of Cotton Leaf Spot Using Image Processing and SVM Classifier," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 340-344. [7] P. Krithika and S. Veni, "Leaf disease detection on cucumber leaves using multiclass Support Vector Machine," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017, pp. 1276-1281. [8] Y. K. Dubey, M. M. Mushrif, and S. Tiple, "Superpixel based roughness measure for cotton leaf diseases detection and classification," 2018 4th International Conference on Recent Advances in Information Technology (RAIT), 2018, pp. 1-5. [9]https://www.google.com/search?q=svm+non+linear%2C+linear%2C+and+kernal+decision+boundary&tbm=isch&ved=2ahUKEwjKycPnpYX5AhXhoukKHYNdA9YQ2cCegQIABAA&oq=svm+non+linear%2C+linear%2C+and+kernal+decision+boundary&gs_lcp=CgNpbWcQAzoECCMQJ1C6C1jcNGDNmgBcAB4AIAB1QGIAY4VkgEGMi4xOS4xmAEAoAEBqgELZ3dzLXdpei1pbWfAAQE&sclient=img&ei=atHWYsrsJOHFpgeDu42wDQ&bih=714&biw=1536&rlz=1C1UEAD_enIN991IN991#imgrc=xj-P7V0nvaWoUM [10] M. P. Vaishnnave, K. S. Devi, P. Srinivasan, and G. A. P. Jothi, "Detection and Classification of Groundnut Leaf Diseases using KNN Classifier," 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019, pp. 1-5. [11] S. Veni, R. Anand, D. Mohan, and P. Sreevidya, "Leaf Recognition and Disease Detection using Content-Based Image Retrieval," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 243-247. [12] https://images.app.goo.gl/7y9H518kMhF7RcZT6 [13] A. Jenifa, R. Ramalakshmi, and V. Ramachandran, "Cotton Leaf Disease Classification using Deep Convolution Neural Network for Sustainable Cotton Production," 2019 IEEE International Conference on Clean Energy and Energy Efficient Electronics Circuit for Sustainable Development (INCCES), 2019, pp. 1-3.

[14] S. Kumar, K. Prasad, A. Srilekha, T. Suman, B. P. Rao, and J. N. Vamshi Krishna, "Leaf Disease Detection and Classification based on Machine Learning," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2020, pp. 361-365.