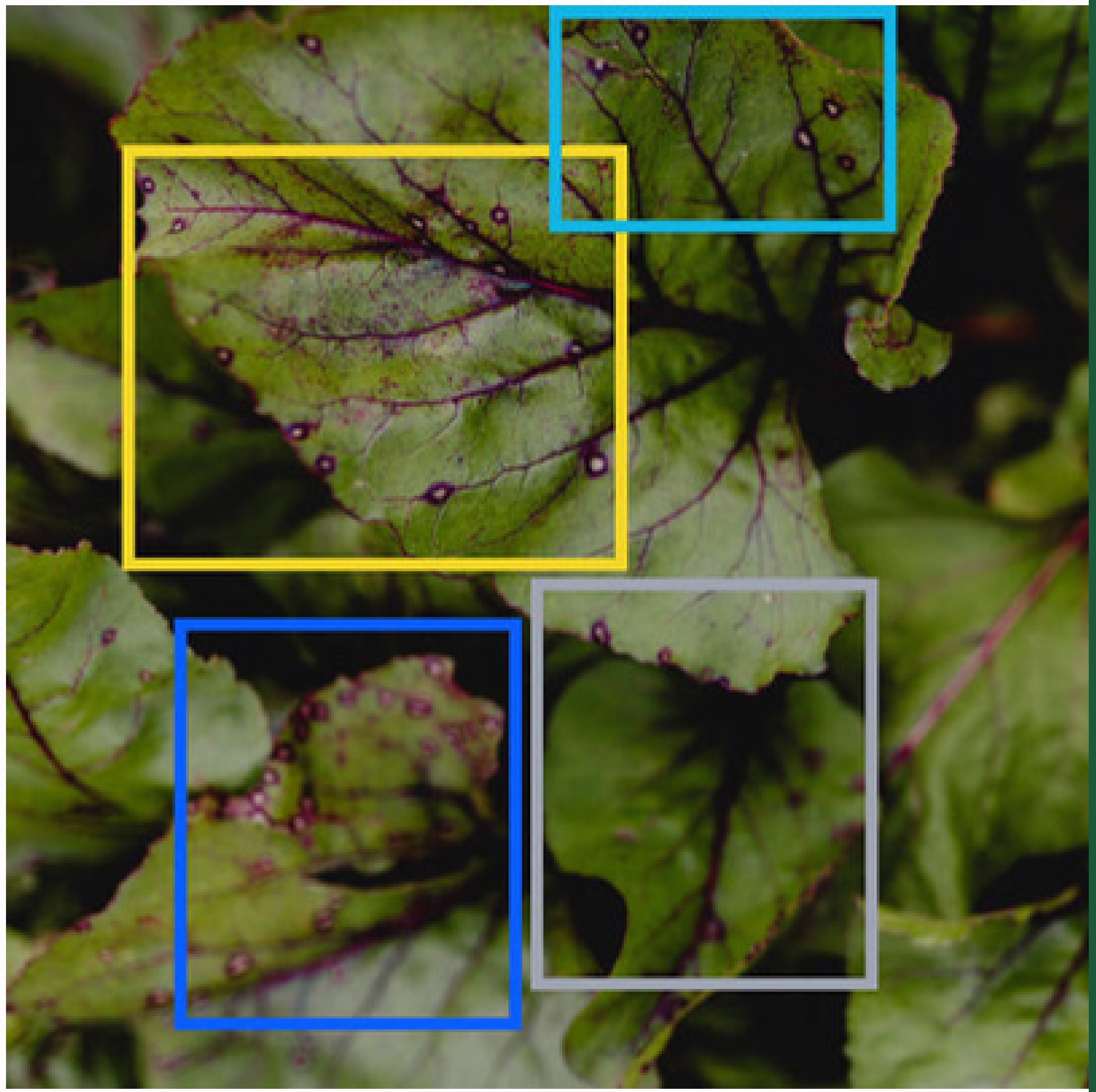
A MACHINE LEARNING ENSEMBLE APPROACH FOR ENHANCED PLANT DISEASE CLASSIFICATION



Our Team

P.Sannith Rao

AP21110010150

Y.Narasimha Rao

AP21110010190

G.Sivamani

AP21110010195

Ch.Sahithya

AP21110010181

Under the guidance
Dr. M Naveen Kumar

Content

01 Introduction

02 Literature Review

03 Methodology

04 Results

05 Conclusion

06 References

AP21110010181

Introduction

- In the contemporary landscape of agriculture, precision plant disease detection stands as a critical point, revolutionizing crop health monitoring through image processing technologies.

- Image processing involves manipulating digital images, starting with diverse source image acquisition and preprocessing steps like filtering for visual enhancement. The core lies in image analysis, utilizing algorithms for feature extraction, recognition, and classification, offering powerful means for diverse practical applications.

- This research focuses on precision plant disease detection using Histogram of Oriented Gradients (HOG) and histogram features.

- Using the Plant-Village-Dataset, our model distinguishes between Diseased and Healthy Apple Leaves.The data preprocessing pipeline ensures model accuracy, involving RGB to BGR conversion, then BGR to HSV for robustness. Image segmentation refines the dataset, separating leaf color from the background. Aimed at contributing to precision agriculture, this research establishes a secure framework for reliable plant disease detection in the digital era.

Literature Review

[1] Introduces a method for plant disease detection through machine learning and image processing, exploring disease classification, image processing, and picture capture, achieving high accuracy in identifying tomato illnesses.

[2] Presents a deep learning-based approach for plant disease diagnosis, exploring classification and feature extraction, demonstrating effective identification of apple illnesses using apple leaf photos.

[3] Explores a plant disease identification method using CNNs, emphasizing feature extraction and categorization, and demonstrating high accuracy in identifying tomato illnesses from leaf photos

[4] Introduces a technique with DCNNs and transfer learning for plant disease identification, exploring classification and feature extraction, achieving effective identification of grape diseases using grape leaf photos

[5] Combines SVMs with color and texture data for accurate plant disease identification, exploring the method's efficacy with a dataset of soybean leaf photos.

[6] Presents a strategy combining transfer learning and machine learning for plant disease identification, exploring feature extraction and classification, achieving high accuracy in identifying grape diseases.

[7] Introduces a strategy combining transfer learning and machine learning for plant disease identification, exploring feature extraction and classification, achieving high accuracy in identifying potato diseases.

[8] Combines transfer learning and Machine learning approaches for plant disease detection, exploring feature extraction and classification, demonstrating high accuracy in identifying maize diseases.

Data Preprocessing

Image Loading: Commencing with our research, we loaded 800 images for each class (Diseased and Healthy) from the dataset. This initial step forms the foundational stage for subsequent analyses.

Color Space Conversion

To meet OpenCV library requirements, we converted images from RGB to BGR and subsequently to HSV, leveraging HSV's ability to separate luma from chroma, crucial in computer vision applications.

Image Segmentation

To distinguish the leaf from its background, we executed image segmentation techniques. This step was pivotal in extracting the color information specifically related to the leaf structure.

Global Feature Extraction

Global features extracted with shape, texture, and color descriptors.

Label Encoding and Dataset Splitting

Encoded image labels numerically for machine interpretability. Dataset split into training/testing sets, maintaining an 80/20 ratio.

AP21110010150

Methodology

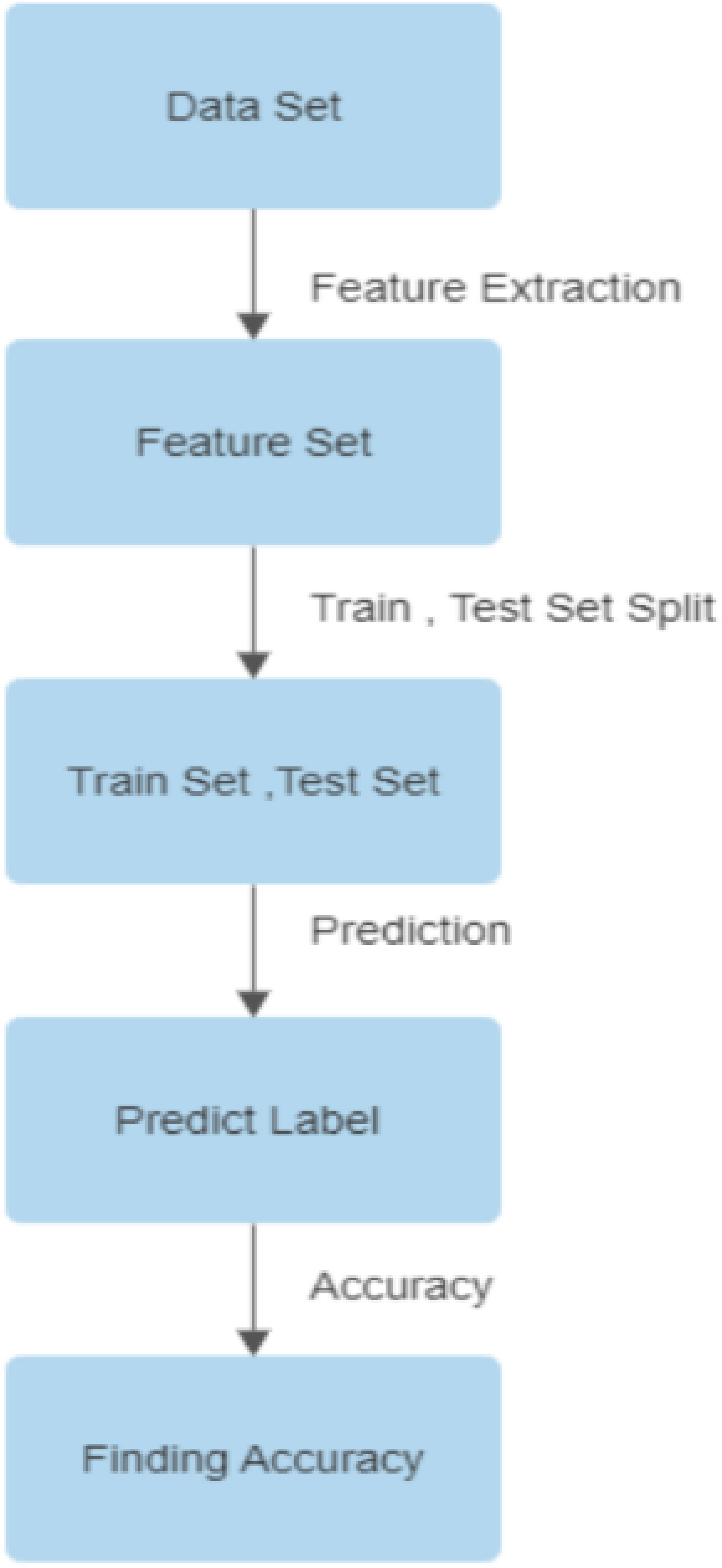
- Two classes healthy and diseased were included in the image data set that was used for the study. Each class had 800 photos taken for study. We now take the features out of the images and add them to the respective feature labels. After normalization, the features are prepared for the train and test split.

- We split the dataset into train and test splits for the model implementation (the test portion is 20% of the entire dataset). The predicted label for the test class will be found by applying the machine learning model. We will determine the accuracy of the data set over the designated ML model by using this predicted label.

- Several machine learning models, including Gaussian Naive Bayes (GNB), K-Nearest Neighbours (KNN), Random Forest, Decision Trees, Support Vector Machines, and Logistic Regression, are used in this research project.

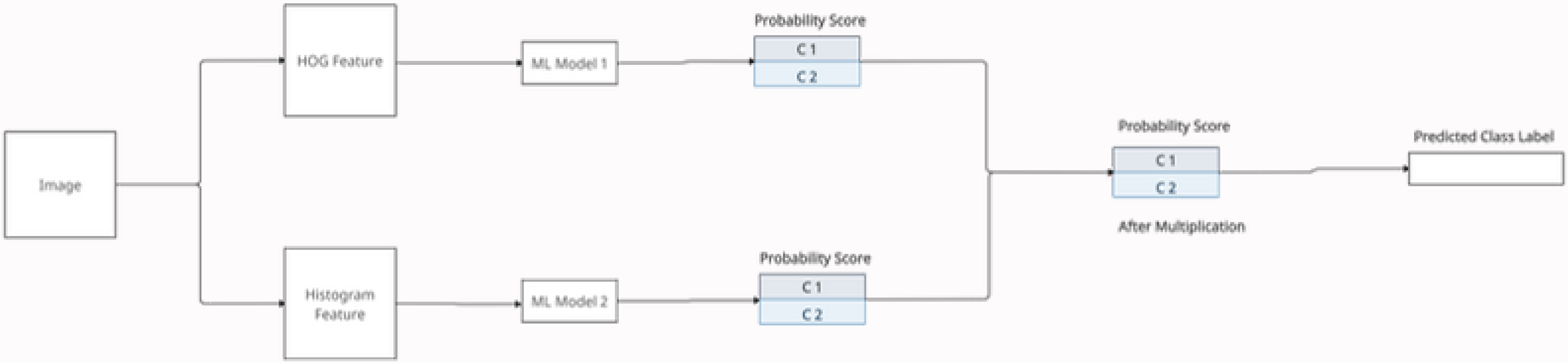
- Four features are extracted from the image in this research project.They are Haralick, Hu Moments, Histogram, and HOG (Histogram of Oriented Gradients)

- Before implementing the proposed approach, we determine the accuracy of each individual feature. HOG sometimes depends on the pixel size per cell; therefore, testing the accuracy of pixel sizes per cell (8*8, 16*16, and 32*32) was also necessary in order to choose the appropriate pixel size.



AP21110010150

Proposed Method



The Proposed method is Fusing of HOG and Histogram features at decision level.

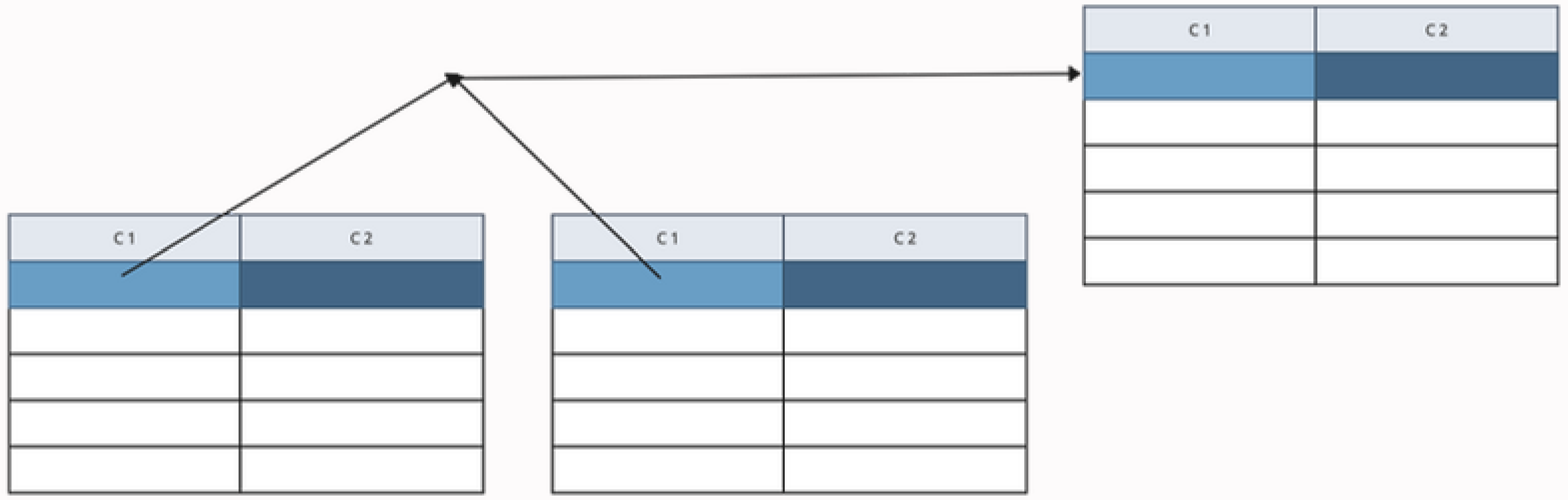
For Every images we Extract the Histogram and HOG features and stored in their respectively feature's labels.Train the ML model for each Feature label. We will get probability score for each image. Then the fusion of both probability scores of 2 models and then predict the class label this is the breif overview of our proposed method

AP21110010190

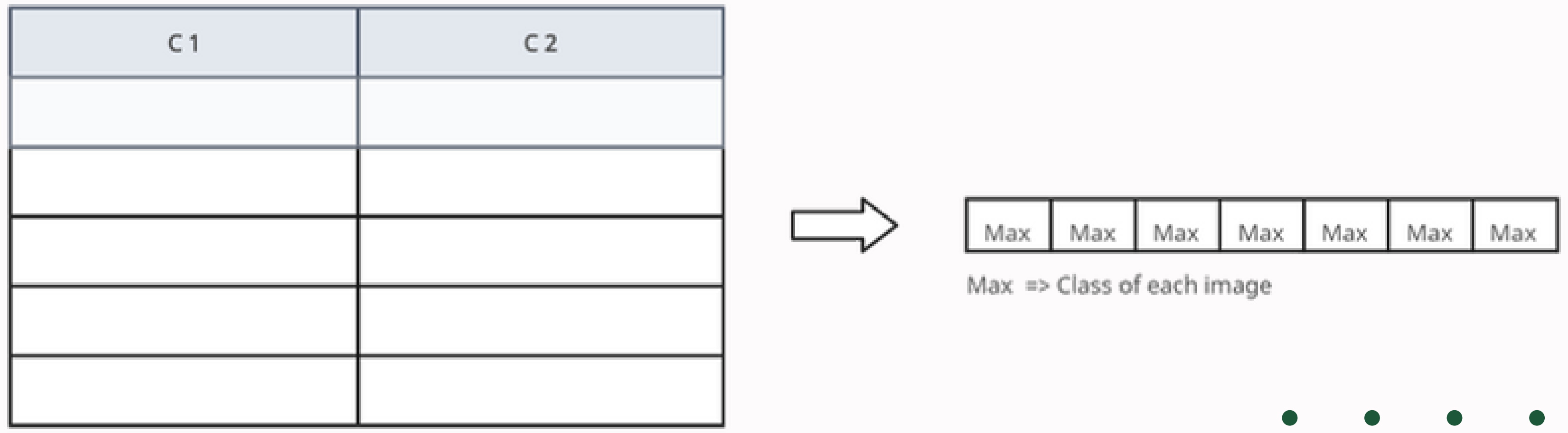
Proposed Method

Probability prediction:
From the two ML models, we will predict the probability that the image belongs to the particular class; its size is (1*2).

Fusing the two models:
After obtaining the probability score of two models, the element-wise multiplication of two matrices will perform, and we will get a 1*2 resultant matrix.



Selection of the Max element from the result matrix:
Now, after multiplication, we get a (1*2) matrix, and now we select the maximum probability from the two classes of each image and add it to the predicted label.



Finding Accuracy:
Finding the accuracy of the predicted label on various ML models.

We got the same accuracy mentoined in the reference research paper with only 2 features, only using random forest as a model for both features.

AP21110010190

Results

HISTOGRAM					
RF	**LR**	**DTC**	**K Nearest**	**SVC**	**Gaussian**
97.50%	94.68%	92.81%	92.81%	96.25%	84.68%

Histogram Accuracy of Different Models

HOG						
Pixel Size Per cell	**RF**	**LR**	**DTC**	**K Nearest**	**SVC**	**Gaussian**
8*8	81.25%	82.81%	66.87%	76.25%	82.18%	65.93%
16*16	83.43%	85.31%	72.18%	76.87%	84.06%	67.81%
32*32	83.75%	84.37%	67.81%	79.68%	83.43%	73.75%

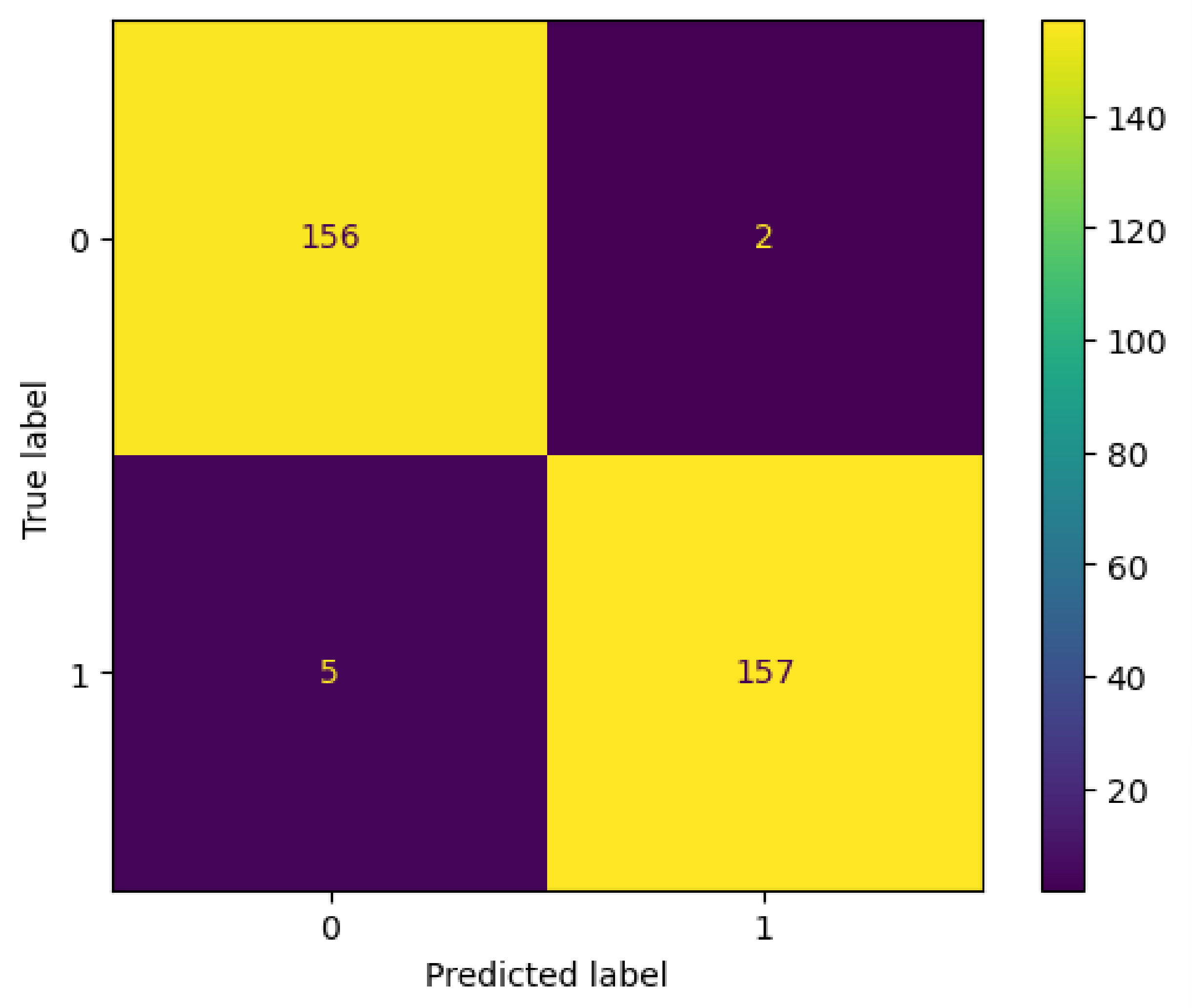
HOG Accuracy of Models of Various Pixel size per cell

AP21110010195

Results

HOG					
RF	**LR**	**DTC**	**K Nearest**	**SVC**	**Gaussian**
83.43%	85.31%	72.18%	76.87%	84.06%	67.81%

HOG Accuracy of different Models



Confusion matrix for Histogram-HOG Model

AP21110010195

Results

Accuracy Table For Proposed Method

HOG	HISTO →	RF	LR	DTC	K nearest	Gaussian	SVC
RF		97.8125%	93.125%	92.8125%	95.00%	84.6875%	95.625%
LR		93.125%	90.00%	92.8125%	94.375%	84.6875%	95.00%
K nearest		92.5%	87.8125%	93.4375%	94.062%	86.25%	91.5625%
Gaussian		70.625%	65.9375%	86.5625%	82.187%	81.875%	65.9375%
SVC		96.5625%	93.4375%	92.8125%	96.25%	84.6875%	96.25%
DTC		69.3%	66.875%	77.1875%	75.625%	75.00%	66.875%

We got the same accuracy with only two features only

Accuracy of Proposed model is 97.8125%

AP21110010195

Conclusion

- We In our proposed method, we achieved equivalent accuracy with only two features, whereas the reference paper used three features to attain the same accuracy. Notably, we introduced the Histogram of Oriented Gradients (HOG) feature, absent in the reference paper. The fusion of HOG and histogram at the decision level contributed to the comparable accuracy.
- In conclusion, this research contributes to the domain of precision agriculture, showcasing the potential of machine learning in revolutionizing plant disease detection and fostering sustainable agricultural practices.

Reference

[1] Singh, A., & Misra, S. (2018). Plant disease detection using image processing and machine learning techniques: A review. In 2018 3rd International Conference on Computing Methodologies and Communication (ICCMC) (pp. 119-123). IEEE.

[2] Sladojevic, S., Arsenovic, M., Anderla, A., & Culibrk, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. Computational Intelligence and Neuroscience, 2016.

[3] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419.

[4] Zhang, Y., Zhang, D., & Zhu, X. (2018). Deep convolutional neural networks for grape leaf disease detection using transfer learning. Plant Disease, 102(12), 2339-2347.

[5] Yang, C., Liu, X., & Zhao, C. (2017). Soybean leaf disease recognition based on color and texture features and support vector machines. Journal of Intelligent & Fuzzy Systems, 33(4), 2181-2190.

[6] Zhang, Y., Zhang, D., & Zhu, X. (2019). Grape leaf disease recognition using deep learning and transfer learning. Plant Disease, 103(2), 228-238.

[7] Zhang, Y., Zhang, D., & Zhu, X. (2019). Potato leaf disease recognition using deep learning and transfer learning. Plant Disease, 103(3), 444-453.

[8] Zhang, Y., Zhang, D., & Zhu, X. (2019). Maize leaf disease recognition using deep learning and transfer learning. Plant Disease, 103(4), 666-675.

THANK YOU