A Feature Extraction and Machine Learning Based Approach for Automated Identification of Crop Diseases

Kamalsingh Solanki[1] Prof. Devkinandan Nagar[2]

Abstract: **With machine learning being used in agricultural applications, a new domain of science has emerged which is termed as precision agriculture. It is the amalgamation of data science, analytics, AI and ML technologies for enhancing conventional agricultural practices. This paper addresses the challenge of identifying blight (late and early) based on a machine learning approach. In this approach, the image is first pre-processed to convert from RGB to Grayscale and subsequently denoised. Next the statistical features of the image are computed to train a machine learning models based on a probabilistic approach employing the Bayes Theorem of conditional probability. A penalty factor is included for the training purpose termed as regularization which optimise the weight updated mechanism. The final classification accuracy is computed based on the TP, TN, FP and FN rates which yield a classification accuracy of 97.69%.**

Keywords— **Potato Leaf Disease (blight), Image Denoising, Feature Extraction, Deep Neural Networks, Classification Accuracy.**

1. INTRODUCTION

Machine learning and deep learning based approaches are being extensively used for identification of blight (early and late) in potato crops which happens to be a staple in various regions of the wold. To automate the process of blight detection, machine learning and deep learning based approaches have been explored. An effective collection of tools for the early identification of potato leaf blight is provided by machine learning techniques. ML algorithms may be trained to discriminate between healthy and diseased potato leaf classes based on subtle visual signals including discolouration, lesions, and leaf morphology. This is accomplished by training models on massive datasets of labelled photos of potato leaves. Figures 1(a) and 1(b) depict the typical normal and blight infested images.

Fig.1(a) A typical healthy image

Fig.1(b) A typical blight image

A type of machine learning called deep learning has been a game changer for image analysis jobs, such as plant disease identification. One type of deep learning models called neural networks is particularly good at automatically learning hierarchical representations of picture features; this eliminates the requirement for feature extraction that is done by hand. However, completely bypassing the feature extraction part may have its own disadvantages which are:

1. Need to extensively copious datasets to effectively train deep learning models.
2. Lessened accuracy of classification due to variations in image texture and background.
3. Possibility of vanishing gradient and overfitting.

This is the reason why the proposed approach tries to incorporate image denoising (to filter out noise effects), feature extraction and subsequent classification using a deep neural network model.

2. METHODOLOGY

The proposed methodology consists of 3 major parts:

Image Pre-Processing
Image Feature Extraction
Classification

Pre-Processing: The pre-processing parts consists of the RGB to Grayscale conversion as well as denoising the image using the DWT. The mathematical analysis is presented here:

For the images, convert RGB to Grayscale using the following relation:

$$\boldsymbol{Iy = 0.333fr + 0.5fg + 0.1666fb} \quad \textbf{(1)}$$

Where.
Fr, Fg and Fb are the intensity of R, G and B component respectively and
Iy is the intensity of equivalent gray level image of RGB image.

The benefit of this process is the fact that it converts the function of 3 variables to one variable and renders homogeneity.

The next step is the denoising of the image based on the DWT process which tries to filter out the image in the transform domain using wavelet decomposition. The approximate low frequency components are used to retain the actual information while the detailed high frequency components are discarded to remove noise effects.

Feature Extraction: The feature extraction process is necessary to compute important statistical features of the images for the final classification process. The features computed in this work are energy, mean, median, standard deviation, variance, entropy, skewness, kurtosis, contrast, correlation, homogeneity, smoothness and rms value. These feature are then then demarcated for the target variable. In order to overcome the difficulties associated with picture classification, the computation of image statistical features is essential. These features are vital for creating precise and dependable classification models because they capture important traits, improve discriminative power, guarantee robustness, and allow efficiency and interpretability. To fully realise the potential of picture-based classification systems, advanced feature extraction techniques must be included as we navigate the ever-expanding field of image analysis.

Final Classification: The final classification is based on the design of the deep neural network model which classifies the image as:

A) Healthy
B) Blight (early) or blight (late)

For this purpose, the computed and fed to the deep neural network. The image statistical features are measurable attributes that are taken from images and represent different facets of its texture, spatial relationships, and pixel intensity distribution. These characteristics enable efficient differentiation between several groups or categories by offering insightful information about the underlying patterns and structures inside images. mage statistical traits provide resilience against changes in lighting, noise, and geometric alterations. Higher-level properties that are more resistant to distortions are encoded via statistical features, in contrast to raw pixel values, which are susceptible to such alterations. Classifiers generate succinct yet useful representations of visual content by computing statistical parameters including mean, variance, skewness, and kurtosis. These characteristics strengthen the discriminative ability of classification models by encapsulating important statistical characteristics that separate one class from another. As there is no clear demarcation among the normal and blighted potato leaf images, hence a probabilistic classifier is design and used for the final classification based on the Bayes Rule:

The weights of the network are updated such that the condition for maximization is satisfied of a new sample bearing a conditional probability defined as:

$$\boldsymbol{P}\left(\frac{\boldsymbol{X}}{\boldsymbol{X}_{i,}\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{M}}\right)=\frac{\boldsymbol{P}\left(\frac{\boldsymbol{X}_i}{\boldsymbol{X},\boldsymbol{k}_2,\boldsymbol{M}}\right)\boldsymbol{P}\left(\frac{\boldsymbol{X}_i}{\boldsymbol{k}_1,\boldsymbol{M}}\right)}{\boldsymbol{P}(\frac{\boldsymbol{X}}{\boldsymbol{k}_1,\boldsymbol{k}_2,\boldsymbol{M}})} \tag{2}$$

Here,
P denotes the probability of occurrence of an event.
X_i denotes the vector corresponding to the bias and weight values of the network.
X denotes the training data set

The training rule for the approach is based on the Bayes theorem of conditional probability which is effective for classifying overlapping feature vectors, based on a penalty $\boldsymbol{\rho}=\frac{\boldsymbol{\mu}}{\boldsymbol{v}}$. The weights are updated based on the modified regularized cost function:

$$\boldsymbol{F}(\boldsymbol{w})=\boldsymbol{\mu}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{w}+\boldsymbol{v}[\frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}}(\boldsymbol{p}_{\boldsymbol{i}}-\boldsymbol{a}_{\boldsymbol{i}})^{\mathbf{2}}] \tag{3}$$

If $(\pi \ll v)$: Network error are generally low.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

else if ($\pi \geq v$): Network errors tend to increase, in which case the weight magnitude should be reduced so as to limit errors (Penalty). The system flowchart is presented next:

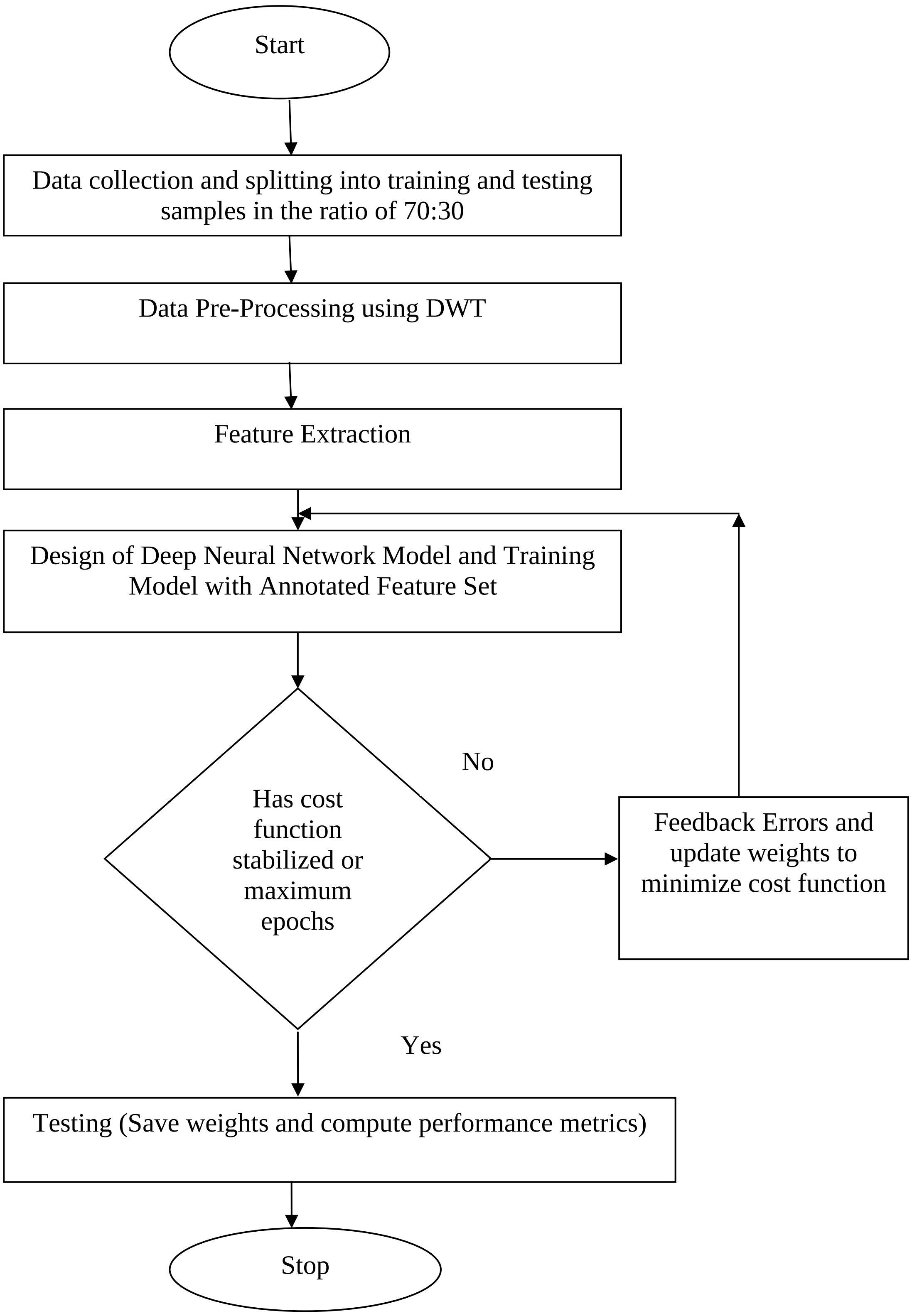

Fig.2 Flowchart of Proposed System

The accuracy of classification is calculated as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

The next section presents the results associated with the proposed approach.

3. RESULTS

The results obtained are resented in this section sequentially:

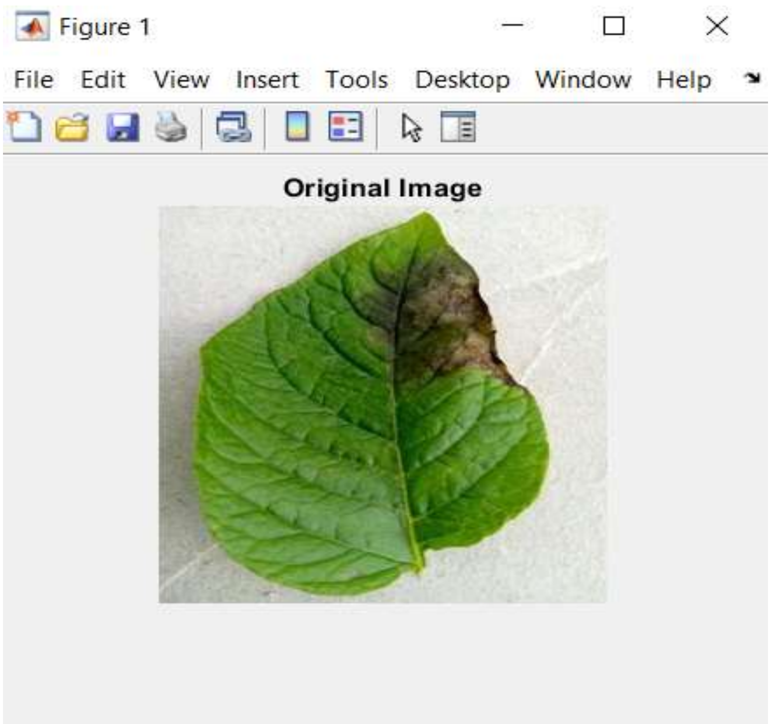

Fig.3 Original Image

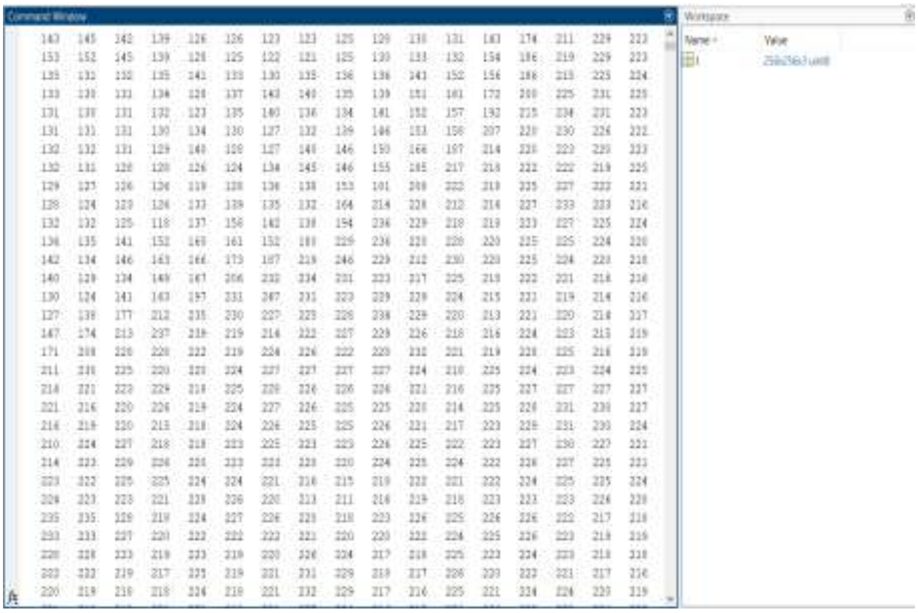

Fig.4 Reading Pixel Values of Image

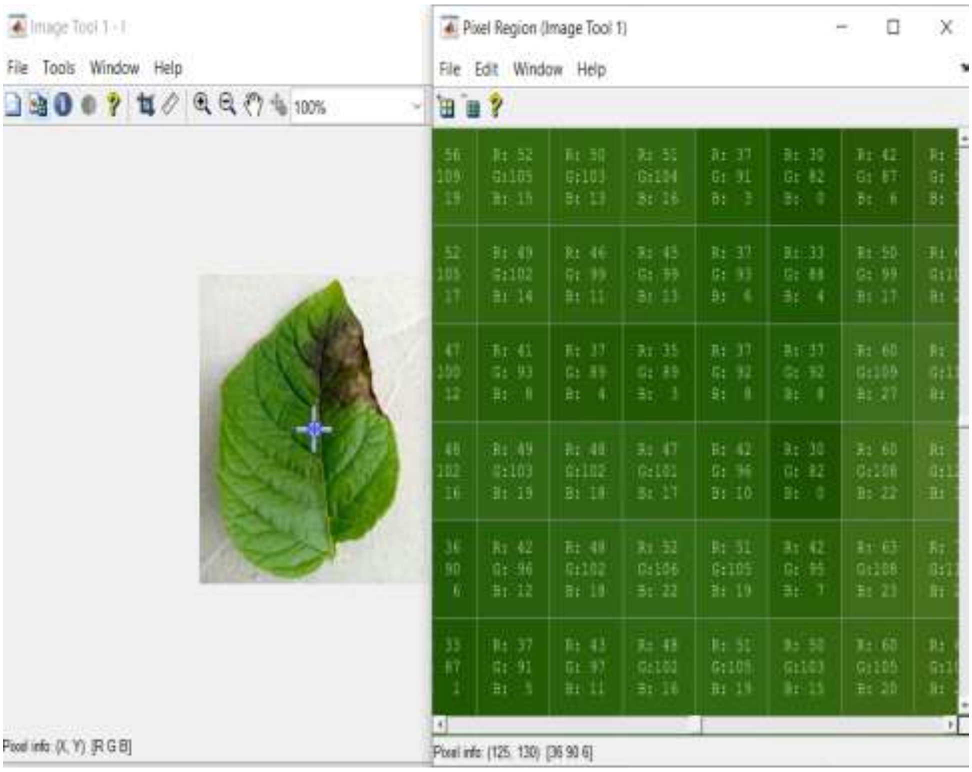

Fig.5 Analysing Pixel Regions

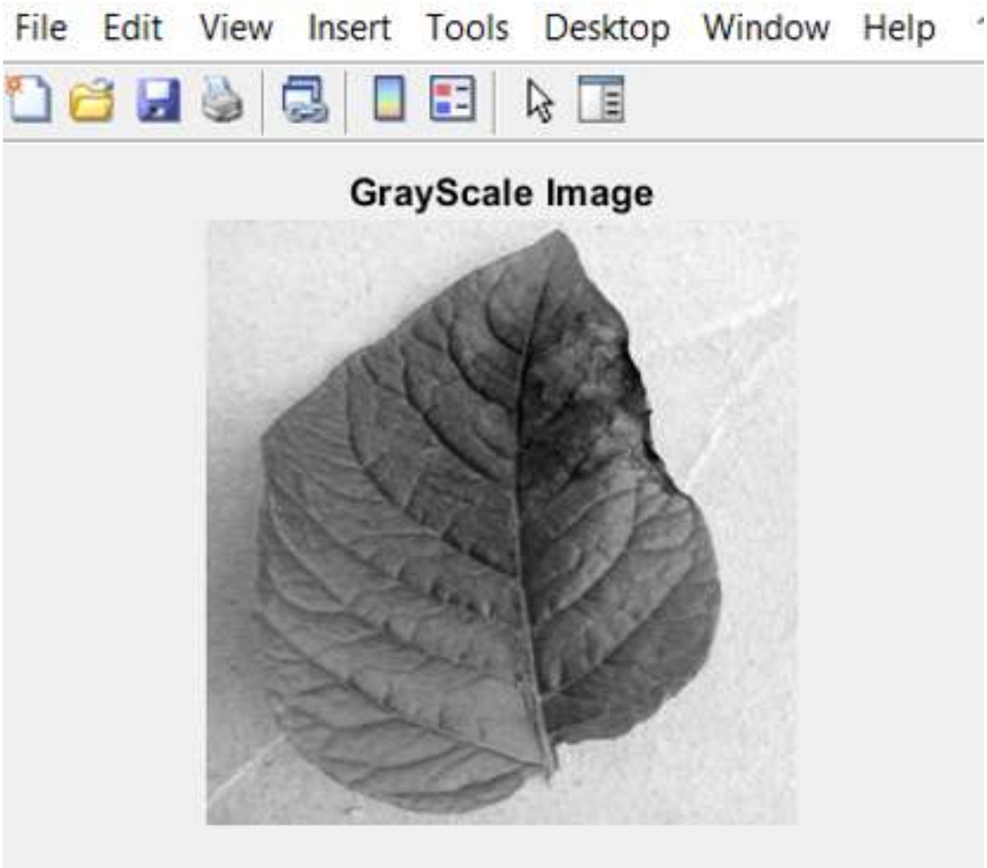

Fig.6 Grayscale Image

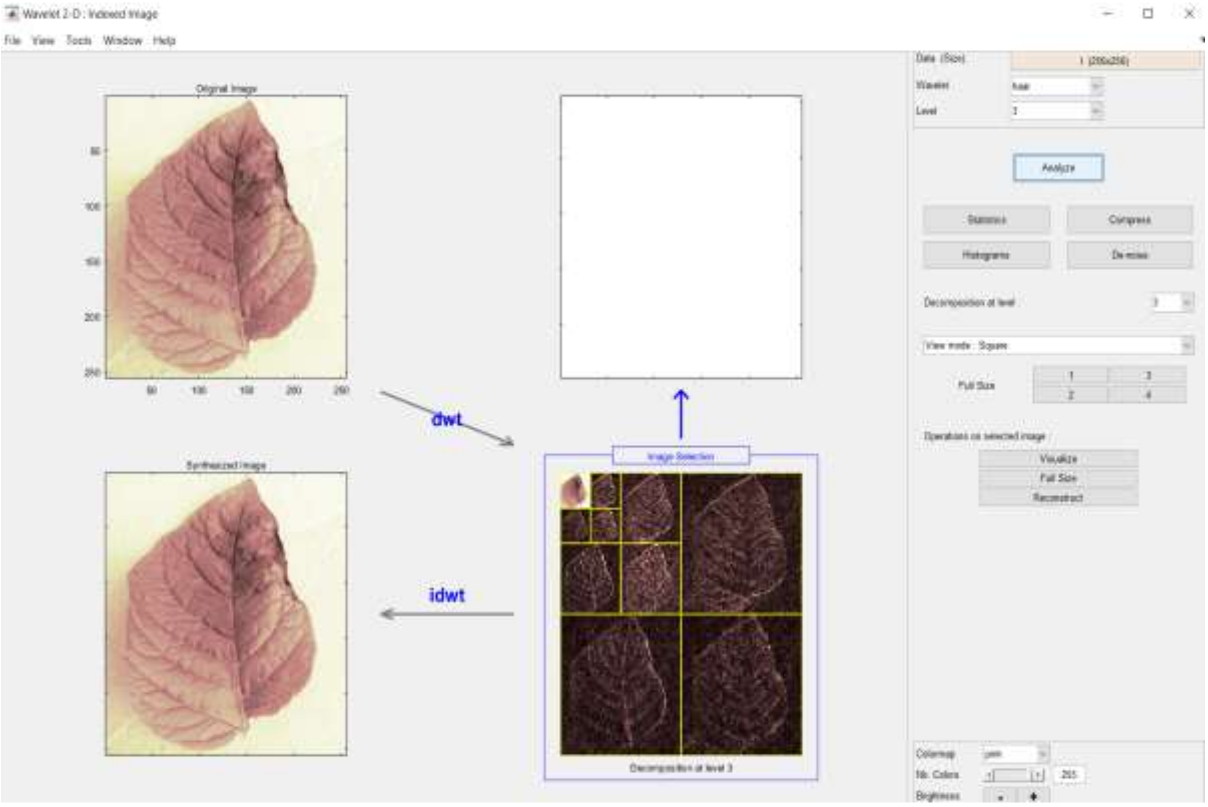

Fig.7 Wavelet Analysis of Image (3rd Level)

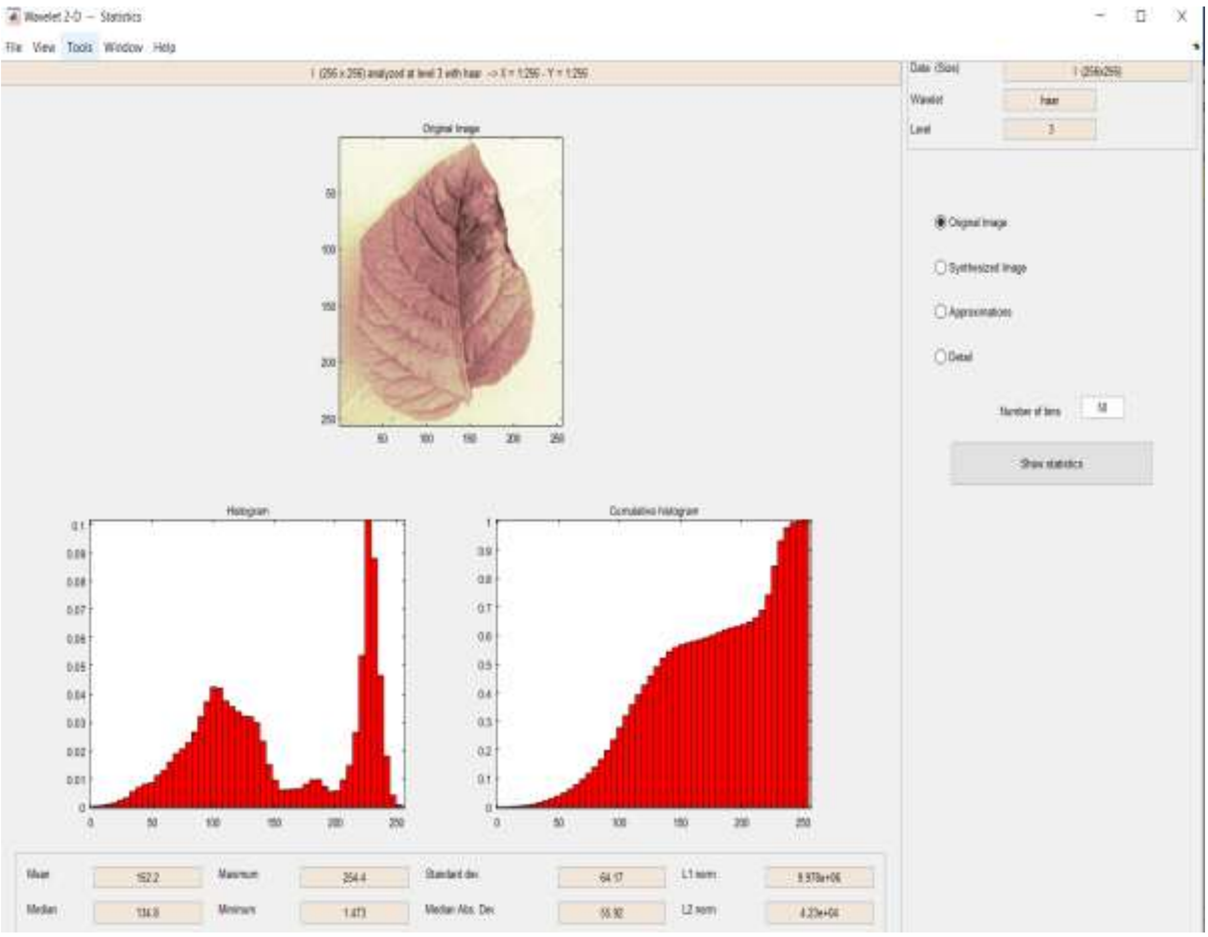

Fig.8 Histogram and Cumulative Histogram of Original Image at 3rd level

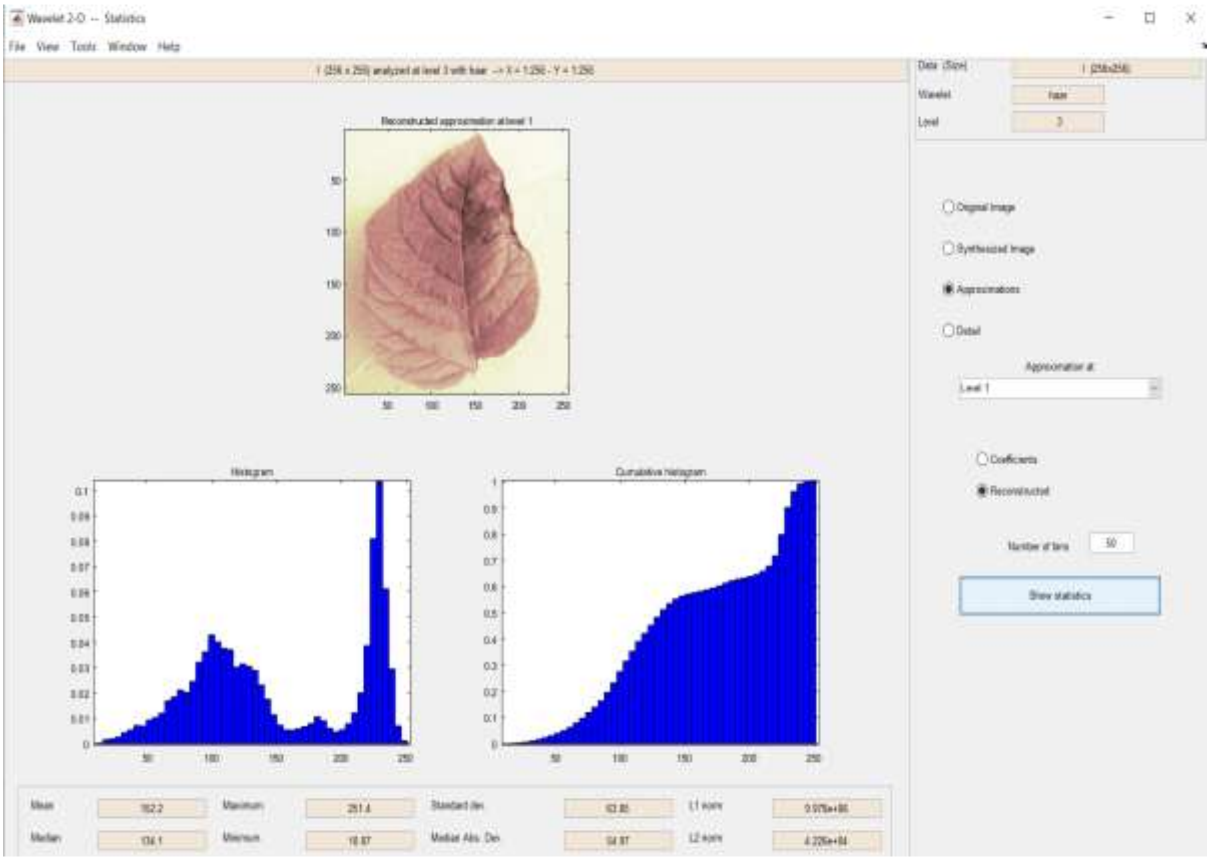

Fig.9 Histogram and Cumulative Histogram of Approximations (3rd Level)

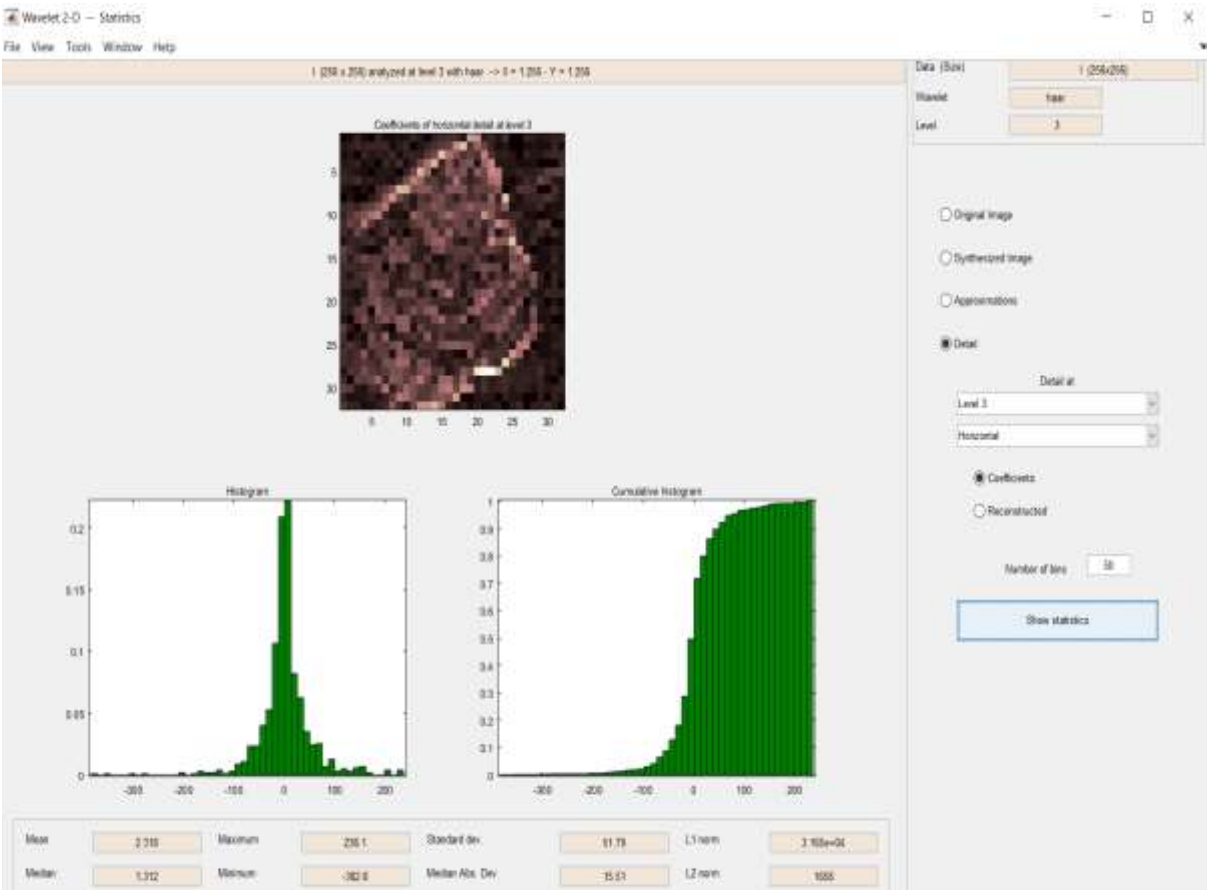

Fig.10 Histogram and Cumulative Histogram of Details at Level 3 of Haarlet

Table 1 Tabulation of data statistical values for original image 'I'

S.No.	**Parameter**	**Value**
1.	Maximum	254.4
2.	Minimum	1.47
3.	Mean	152.2
4.	Median	134.8
5.	Standard Deviation	64.17
6.	Medium Absolute Deviation	55.92
7.	L1 Norm	9.97×10^6
8.	L2 Norm	4.23×10^4

Table 2 Tabulation of data statistical values for Approximations

S.No.	**Parameter**	**Value**
1.	Maximum	1937
2.	Minimum	249.7
3.	Mean	1218
4.	Median	1056
5.	Standard Deviation	499.5
6.	Medium Absolute Deviation	407.9
7.	L1 Norm	91.24×10^6
8.	L2 Norm	4.27×10^4

Table 3 Tabulation of data statistical values for Details

S.No.	**Parameter**	**Value**
1.	Maximum	236.1
2.	Minimum	-382.8
3.	Mean	2.318
4.	Median	1.312
5.	Standard Deviation	51.79
6.	Medium Absolute Deviation	15.51
7.	L1 Norm	3.165×10^4
8.	L2 Norm	1658

Table 1 depicts the statistical DWT features of the original image. Table 2 depicts the statistical DWT features of the approximations and table 3 depicts the statistical DWT features of the detailed co-efficients. The observation which can be made from tables 1, 2 and 3 are the fact that the values for the original image are closer to the approximations while completely different from the details. This clearly indicates the statistical dissimilarity of the details w.r.t. the original image, and hence can be considered as exogenous noise effects which can be filtered through the DWT approach. The total number of images for the classification purpose have been considered as 130 (with a 30% testing split for the overall 430 images.

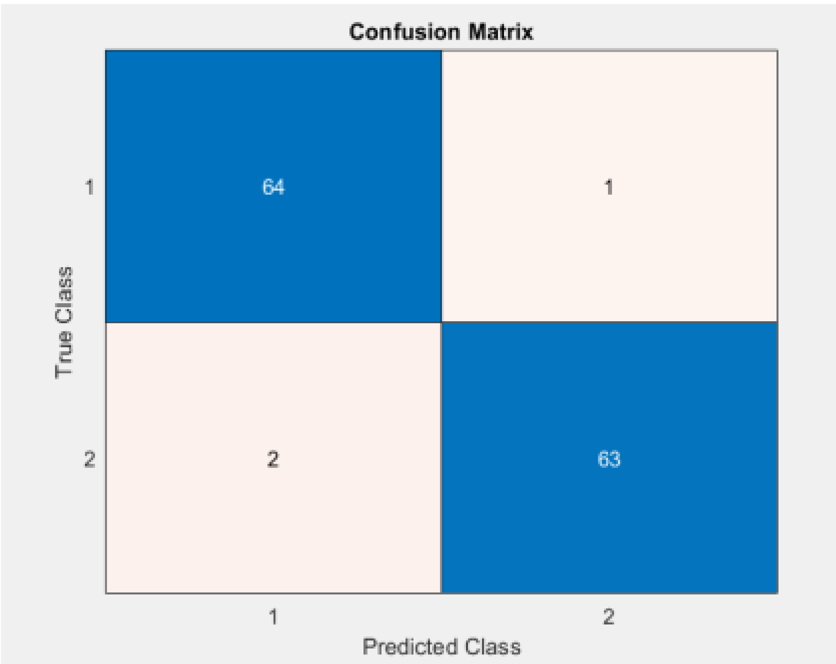

Fig.11 Confusion Matrix

The testing accuracy is computed as:

$$\boldsymbol{Accuracy} = \frac{\mathbf{64 + 63}}{\mathbf{64 + 63 + 1 + 2}} = \mathbf{97.69}\%$$

The accuracy of the proposed approach is thus 97.69% for the proposed approach.
A summary of the results is presented next:

Table 4. Summary of Results

S.No.	**Parameter**	**Value**
1	Data Source	https://data.mendeley.com/datasets/v4w72bsts5/1
2	Image Type	jpg
3	Split Ratio	70:30
4	Feature Extraction	12 statistical features
5	ML Model	Neural Network
6	Algorithm	Back Propagation with Bayesian Regularization
7	Accuracy: Bonik et al., 2023 [6]	94.2%
8	Accuracy: Singh et al., 2022 [7]	94.07%
9	Accuracy: A.K. Singh et al., 2022 [7]	95.9%
10	**Accuracy** **(Proposed Work)**	**97.69%**

CONCLUSION:

In conclusion, it can be said that the potato plant (especially the leaf) is prone to blight disease. If left untreated, potato leaf blight, which is brought on by fungi like Phytophthora infestans, can seriously harm potato crops all over the world and result in large yield losses. Agronomists' subjective and time-consuming visual inspection is the foundation of traditional disease detection techniques. However, there is a chance to completely transform the identification and treatment of potato leaf blight with the introduction of machine learning (ML) and deep learning (DL) approaches. This paper presents not only a machine learning based approach, but rather integrates it with image denoising and statistical feature extraction to train a deep neural network which attains a classification accuracy of 97.69%. The Back Propagation with Bayesian Regularization has been designed to train the probabilistic neural network model with annotated statistical features.

References

[1] Kodandaram, M. & Saha, Sujoy & Rai, Awadhesh & Naik, Prakash. (2013). Compendium on Pesticide Use in Vegetables.

[2] Ahmed F, Al-Mamun HA, Bari A, Hossain E, "Classification of crops and weeds from digital images. A support vector machine approach Journal of Crop Production", Elsevier, 2023, vol.40, pp: 98-104.

[3] Bakhshipour A, Jafari A, Nassiri SM, Zare D, "Weed segmentation using texture features extracted from wavelet sub-images. Biosystems Engineering, "Elsevier 2017, vol.157, pp: 1-12

[4] Bakhshipour A, Jafari A, "Evaluation of support vector machine and artificial neural networks in weed detection using shape features", Computers and Electronics in Agriculture, Elsevier 2018, vol.145, pp: 153-160

[5] Fawakherji M., Youssef A., Bloisi D, "Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation", Proceedings in 2019 Third IEEE International Conference on Robotic Computing (IRC), IEEE 2019, pp: 146-152

[6] C. C. Bonik, F. Akter, M. H. Rashid and A. Sattar, "A Convolutional Neural Network Based Potato Leaf Diseases Detection Using Sequential Model," 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, 2023, pp. 1-6

[7] AK Singh, SVN Sreenivasu, U Mahalaxmi, H Sharma, DD Patil, E Asenso, "Hybrid feature-based disease detection in plant leaf using convolutional neural network, bayesian optimized SVM, and random forest classifier", Artificial Intelligence in Food Quality Improvement Hindawi 2022, Article ID 2845320.

[8] A Singh, H Kaur, "Potato plant leaves disease detection and classification using machine learning methodologies", Proceedings in IOP Conference Series: Materials Science and Engineering, 2022, IOP Conference Series: Materials Science and Engineering, vol. 1022, pp.1-9.

[9] H Afzaal, AA Farooque, AW Schumann, N Hussain, "Detection of a potato disease (early blight) using artificial intelligence", Remote Sensing, MDPI, 2021, vol.13, pp.1-17.

[10] M. A. Iqbal and K. H. Talukder, "Detection of Potato Disease Using Image Segmentation and Machine Learning," 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Chennai, India, 2020, pp. 43-47.

[11] D. Tiwari, M. Ashish, N. Gangwar, A. Sharma, S. Patel and S. Bhardwaj, "Potato Leaf Diseases Detection Using Deep Learning," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 461-466

[12] M. I. Tarik, S. Akter, A. A. Mamun and A. Sattar, "Potato Disease Detection Using Machine Learning," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 800-803.

[13] J. Akther, M. Harun-Or-Roshid, A. -A. Nayan and M. G. Kibria, "Transfer learning on VGG16 for the Classification of Potato Leaves Infected by Blight Diseases," 2021 Emerging Technology in Computing, Communication and Electronics (ETCCE), Dhaka, Bangladesh, 2021, pp. 1-5

[14] B Darwin, P Dharmaraj, S Prince, DE Popescu, "Recognition of bloom/yield in crop images using deep learning models for smart agriculture: A review", Agronomy, MDPI 2022, vol.11., no.4, Art No.646.

[15] E Omrani, B Khoshnevisan, S Shamshirband, "Potential of radial basis function-based support vector regression for apple disease detection", Journal of Measurement, Elsevier 2014, Volume-55, pp. 512-519

[16] S.Haykin, "Neural Networks and Learning Machines", 3rd Edition, *Pearson Publications*.

[17] M.Hagan, "Neural Network Design", 2nd Edition, *Cengage Publication*.

[18] Machine Learning Notes: Stanford University: http://cs229.stanford.edu/materials.html