A Smart Agricultural AI Model For RGB Image-Based Disease Detection On Apple Trees

Priyadarshini Pattanaik[1*], Nguyen Manh Cuong[2], Navya Gubbi Sateeshchandra[3], Kaddour Chelabi[4], Archana Balaji[5]

[1*]Faculty of Computer Science and Informatics, Berlin School of Business & Innovation (BSBI), 12043 Berlin, Germany, ppattanaik055@gmail.com. *Corresponding Author
[2,3,4]Faculty of Economics and Business Administration, Berlin School of Business & Innovation (BSBI), 12043 Berlin.
[5]Student of Computer Science and Informatics, Berlin School of Business & Innovation (BSBI), 12043 Berlin, Germany.

Abstract— *The early and accurate detection of plant diseases is crucial because of its contribution to socioeconomic growth in agricultural productivity and worldwide food security. Traditional methods of plant disease detection often depend upon time-consuming, intensive research surveys and onsite field inspections, which are time-consuming and liable to human error. In the last few decades, the incorporation of imaging technology with automated artificial intelligence (AI) algorithms has appeared as a promising answer, allowing speedy and accurate early identification of plant diseases. In this work, an automated framework is developed to identify and classify diseases in apple plants at the right time to reduce financial loss and human labor. However, advancements in sensor technology, information analytics, and artificial intelligence algorithms continue to enhance smart agriculture. In this work, we have used a multispectral dataset analyzing grayscale and RGB sample images with preprocessing, and classification to discover apple leaf illnesses. Color spatial capabilities have been recognized as crucial for assessing the severity of apple plant species infections. Our findings indicate that blue channel color space supplied better clarity and noise-unfastened outputs, making them more effective for detecting diseased leaves than other color space channels and grayscale images. Two AI-based models, Random Forest and Convolutional Neural Networks (CNNs) were fine-tuned and used for disease detection. The CNN model outperformed Random Forest, achieving an accuracy of 89.05%, precision of 90.71%, remember of 89.05%, and an F1 score of 89.87%. These effects underscore the high functionality of CNNs to hit upon and classify plant diseases with precision while minimizing false positives and negatives. The integration of CNNs into RGB channel color space detection workflows facilitate early diagnosis and timely interventions, improving plant control, safeguarding yield, and promoting agricultural sustainability.*

Keywords— *Convolutional Neural Networks, RGB Images, Plant diseases, Deep Neural Networks, Infectious Disease, Plant Pathology, Smart Agriculture, Agricultural Sustainability*

INTRODUCTION

The global population is anticipated to rise to 9.8 billion with the aid of 2050 (United Nations 2015). To feed the populace, agricultural manufacturing will want to grow closely over current levels, efficiently using assets. Furthermore, the entire crop yield discount because of all crop pests and illnesses reaches 40% (Oerke et al. 1994), causing worldwide food security to be undermined (Chandler et al. 2011). The use of traditional plant protection products to shield crops against pests and sicknesses sadly has poor effects on the environment, bio range, and human health and should therefore be largely reduced. Ways to make crop safety more sustainable are required and incorporated pest management (IPM) is promoted as the first-class manner forward making a distinction in this attempt. IPM emphasizes the growth of a healthful crop with the least feasible disruption to agro-ecosystems (Lamichhane et al. 2016). To enhance productivity farmers, adopt numerous techniques to suitable crops and determine suitable insecticides to support plant increase. The incidence of plant diseases substantially affects the quality and quantity of agricultural production. In the study of plant pathology, scientists have a look at plant sicknesses, making a specialty of figuring out and knowledge of the visually observable patterns exhibited by way of affected plants. Accurate assessment of plant health and disease status is crucial and essential for successful agricultural practice. Historically, the exam and prognosis of plant illnesses trusted guide evaluation through professional people, a procedure that turned into exertions-in-depth and time-consuming. In modern-day agriculture, image analysis techniques are extensively applied for plant ailment identity, primarily focused on signs determined at the stems, vegetables, and leaves.

Smartphones can be used by farmers to take images of their crops, which they can then upload to cloud-based platforms where machine learning (ML) algorithms process the photos and produce useful information. For example, (Mrisho et al. 2020) presented Nuru, a deep-learning model for diagnosing cassava diseases on Android devices. With

a 65% accuracy rate, this device shows the reliability of reasonable arrangements that utilize current cell phone innovation for in-field diagnostics.

Enormous scope crop observation is likewise made conceivable by high-goal cameras introduced on ground-based stages and automated elevated vehicles (UAVs). Plant pathology has been changed by late improvements in the man-made reasoning (man-made intelligence) area, particularly in machine vision (Pattanaik et al., 2022). Machine vision permits machines to grasp their environmental elements by joining camera frameworks and PC calculations to remove data from images or recordings (Shin et al., 2023). By giving ranchers admittance to constant information to help with all-around informed direction, this innovation can change farming (Tian et al., 2020).

Machine vision has been utilized to recognize bugs (Sena Jr. et al., 2003), distinguish illnesses (Karthik et al., 2020; Sethy et al., 2020), and screen crops for signs of pressure (Kacira et al., 2002; Khotimah et al., 2023; Nhamo et al., 2020). Huge yield misfortunes are brought about by vermin and sicknesses. Accordingly, imaging innovations like satellites, drones, UAVs, and cell phones are being utilized increasingly to follow crop medical images and recognize ailments. Ranchers can answer rapidly and successfully because of these stages (Hafeez et al., 2022). On a more extensive scale, satellite sensors give sagacious data by distinguishing sickness episodes and assessing crop wellbeing across the whole field (NASA Science Mission Directorate, 2010; Yang, 2020).

The utilization of AI (ML) methods to naturally recognize patterns and irregularities reminiscent of harvest infections is developing (Pattanaik et al., 2021). Convolutional brain organizations (CNNs), repetitive brain organizations (RNNs), and autoencoders are instances of cutting-edge profound learning models that have shown extraordinary commitment in the ID of illness designs from images (Pattanaik et al., 2024). For instance, (Mohanty et al. 2016) prepared a profound convolutional brain organization (DCNN) that could distinguish 14 harvest species and 26 sicknesses utilizing a dataset of 54,306 photographs of solid and unhealthy plant leaves. The model's astounding 99.35% exactness on the test set exhibited the adequacy of profound learning methods for precise and versatile yield illness determination. A broad examination of image-based techniques for crop illness discovery is given in this survey. It evaluates the significance of man-made reasoning models that have been reported in the audit work, featuring critical snags and exploring the capability of different sensor advances in sickness identification. The audit inspects the impacts of environmental change on crop wellbeing and features how relevant elements, like climate and field the executive's procedures, influence sickness ID.

Numerous crops and diseases are systematically analyzed, emphasizing machine learning techniques catered to regional and species needs. Image-based surveillance systems for extensive monitoring to identify new infections are also examined in the review. Technologies for real-time monitoring are cited as crucial resources for early disease detection, allowing for prompt interventions and self-sufficient crop management. The efficiency of collaborative mapping approaches in raising the accuracy of disease detection is also investigated. An examination of automated technologies intended for implementation in low-income and developing nations, to offer affordable and easily accessible answers to the world's agricultural problems, rounds out the review.

Images serve as the basis for the extraction and interpretation of visual information, making them a fundamental data type for image processing applications (Li, H. et al. 2025). Two-dimensional arrays are commonly used to represent digital images, with each element corresponding to a distinct coordinate (r, c) and storing information about the brightness or intensity at that location (Pérez-Rodríguez et al. 2019).

The smallest controllable and addressable unit in digital imaging is the pixel, which is distinguished by its intensity, which varies based on the kind of image. The methodical arrangement of pixels in a two-dimensional matrix makes manipulation and analysis easier. A pixel's brightness is determined by its intensity, and this information is essential for several image processing tasks.

- **Binary Image:** The most basic type of digital image, binary images are made up of just two-pixel values, usually black and white, denoted by the numbers 0 and 1. Because only one binary digit is needed to encode each pixel, these images are known as 1-bit per pixel images. Applications where clarity and simplicity are essential, such as document scanning and object segmentation, frequently use binary images.
- **Grayscale Image:** Grayscale images are colourless and monochromatic, representing information about brightness. In an 8-bit representation, each pixel in a grayscale image has a single intensity value, usually between 0 (black) and 255 (white), resulting in 256 distinct shades of Gray. In computer vision, photography, and medical imaging, grayscale images are frequently utilized because they highlight.

- **Indexed Image:** A colormap matrix and a pixel array make up the two-part representation used by indexed images. The colormap defines the actual RGB colour values, and the pixel array contains indices pointing to rows in the colormap. Each row of the colormap, which is a $m\times3$ m×3 matrix with floating-point values in the interval [0,1], indicates the red, green, and blue components of a given color. Maps, diagrams, and some forms of medical imagery are examples of images with limited color palettes that can be effectively represented by index images (Pérez-Rodríguez et al. 2019).

- **RGB Image:** RGB images utilize three parts — red, green, and blue — to address every pixel to catch full-variety data. Generally, 8 pieces are utilized by every part, for a sum of 24 pieces for each pixel. More than 16 million tonnes are conceivable with this organization, making for various purposes, contemporary image handling strategies exploit these image kinds of characteristics. For preprocessing undertakings like thresholding, edge recognition, and morphological tasks, parallel images are fundamental. In computationally requesting applications like surface examination and facial acknowledgment, where a variety of information is pointless, grayscale images are now and again. Images with restricted variety ranges, like warm guides and topical guides in geospatial examination, can be proficiently put away and delivered thanks to list images. RGB images are fundamental for undertakings like item discovery, semantic division, and creative delivery that call for a rich variety of data. Multispectral and hyperspectral images, which go past the RGB range to integrate additional groups like infrared or bright, are one more illustration of a high-level image design. These images can be utilized in material examination, remote detection, and farming. Rich otherworldly information is caught by these configurations, giving a more exhaustive understanding of natural and surface qualities. Creative arrangements across businesses are made conceivable by image handling calculations that utilize these different image types to boost execution for undertakings like further developing differences, fragmenting locales of interest, or grouping objects.

MATERIALS AND METHODS
Analysis of chlorophyll content by RGB indices calculation
During data collection, crop indices (Cis) were computed and dissected for every yield with different spectral data from red (R), green (G), and blue (B) parts. Color is the most important feature that plays a vital role in regularizing plant species and displaying spectral marks on these different color channels.

Spectral data from red (R), green (G), and blue (B) groups were dissected for each yield to compute crop indices (CIs) on dates. Color is a regularly utilized strategy to recognize plant species, as various yields display one-of-a-kind spectral marks in these groups. Exactly when CIs are gotten from RGB gatherings, they give unequivocal phenotypic information that directs the various assortments of harvests or plants. Since RGB images consolidate both soil and reap leaves, the dirt part was covered before processing CIs (Figure 2 and Figure 3). This covering was performed using a managed image with the assistance of the help vector machine (SVM) strategy, known for its high accuracy (Banerjee et al., 2018). Spectral marks for crops were extricated as shown in Figure 3. RGB-based crop records have been comprehensively utilized and shown power for extricating grouped phenotypic data under moving advancement conditions (Gracia-Romero et al., 2017). In this audit, 16 are not set in stone for each yield discernment date using the "raster" group in R Studio v4.2.2. Clear estimations for all VIs were figured for each collection across different image-getting dates. Tukey's Clear Gigantic Differentiation (HSD) test was coordinated in SPSS v16 to assess immense differences between pack suggests.

Late movements in RGB test assessment have furthermore overhauled how we could decipher crop phenotyping under different regular circumstances. For example, concentrates by (Zhao et al. 2020) and (Ahmed et al. 2021) included the ability of significant standard RGB imagery in exactness agribusiness, highlighting its part in crop noticing and yield assumption. Such techniques continue to progress with work on computational gadgets and artificial intelligence strategies, offering more conspicuous precision and viability in agrarian investigation.

A. Data Processing using RGB Indices
Exploratory data analysis (EDA) is a basic push toward the data science work process. It gives a profound comprehension of the information, which is fundamental for settling on informed conclusions about information preprocessing, including designing, and model determination (Vigni, M.L et al. 2013). By directing exhaustive EDA, it is ensured that the ensuing Machine Learning and Deep Learning algorithms demonstrating processes are based on

strong groundwork, prompting more exact and dependable outcomes. This cycle incorporates gathering, cleaning, investigating, dissecting, and finally getting bits of knowledge. In this work, the distribution of each feature with statistical metrics like mean, median, mode, and range are used, and they help us in choosing appropriate preprocessing techniques and models. For the data preprocessing stage, the original dataset is used as shown in Figure 1. Figure 1 showcases the different distribution of plant species with their metrics of sample distribution.
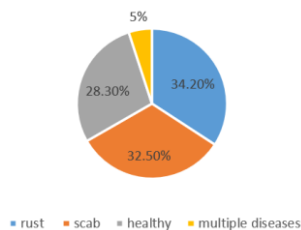
Figure 1: Illustrates the original dataset chart and different plant species sample distribution.

B. Methods

Random Forest (RF) is one of the machine learning algorithms that show good performance in plant pathology (Hatuwal, B.K 2020). Random Forest is characterized by its robustness and reliability. Usually, the Random Forest approach develops many small decision tree structures, offering decision outputs after training and learning for better prediction. The RF model will allow predictive analysis employing lowering overfitting through the power of the wide variety of decision branches. The RF belongs to the powerful ensemble approach using the electricity of many classifiers for introduced accuracy and is well-located for extremely complex classification obligations, along with the character of identifying diverse plant ailment classifications from images. In this phase, we are going to discuss the implementation of the Random Forest set of rules implemented around plant pathology through image class. Now we can present this class implementation that separates the occurrence of various foliar illnesses and other not-unusual crop fitness issues. Here many snapshots of Apple tree leaves of plant pathology have been captured, aiming to develop a model for early detection and a friendly tool for farmers and agriculture specialists, and, in turn, get progressed advanced automated analysis of plants to get higher yields. Here, AI AI-based automated method for facts training, image preprocessing, training, and testing RF is used. Along with this, we have included the performance metrics to be carried out at some point in the overall version assessment, in addition to presenting the version's benefits and drawbacks. In this work, we exhibit how the RF approach may be adapted to boost the outcomes and accuracy of faster plant disease detection, so we can aim towards advanced and automatic approaches inside the subject of agricultural pathology.

Convolutional Neural Networks (CNNs) are a sort of profound deep learning model especially viable in image classification works (Pattanaik et al. 2024). They succeed at recognizing designs and spatially ordered progressions in images, making them appropriate for diagnosing plant sicknesses. By utilizing various layers, CNNs consequently and adaptively gain complex elements straightforwardly from crude pixel information, successfully dealing with the high-layered highlight space ordinary in plant pathology samples. In this work, we fostered a CNN model to group plant pathology images. The objective was to make a device able to precisely recognize different foliar disease plant samples, for example, apple scab, rust, and others, as well as separate these from solid healthy leaves. Utilizing an enormous, publicly available labeled dataset, the model furnishes farmers and horticultural experts with an early advanced disease detection framework for illness acknowledgment and intercession.

The CNN model incorporates a few basic parts intended to successfully concentrate and cycle highlights from input image samples: The CNN deep learning model comprises many layers and can be used to extract and process features from the input image samples. The first layer of CNN helps to convolute the input image samples and detect several features. After the input layer, the next layer is the max pooling layer i.e. MaxPooling2D used to reduce the spatial dimensions and to direct on important relevant features. The feature maps can be flattened and into one-dimensional vector processing into dense layers after convolutional and pooling layers. To assess the model, we utilized measurements like review, accuracy, F1 score, and others. This gave a thorough evaluation of the model's assets and limits when applied to real-time farming scenarios.

II. EXPERIMENTAL SETUP ANALYSIS

A. Datasets

A multispectral dataset of plant pathology images is expected to foster a powerful imaging model. The dataset is a collection of different yields impacted by various infections. It will guarantee a different portrayal of true situations. For this task, we utilized a dataset from Kaggle (The Plant Pathology 2020 FGVC7 dataset) which involves sample images of apple leaves showing various sorts of sicknesses connected with apple trees, for example, apple scab, rust, and numerous illnesses, such as well as solid healthy leaves. Apple leaf infections include normal illnesses like scab, rust, and fine mold or powdery mildew (Thapa et al. 2020). Among these, the apple scab, brought about by a parasitic microorganism, stands apart as quite possibly one of the most financially critical contagious illnesses influencing apple crops internationally (Agarwal et al. 2019). The sickness appears as noticeable parasitic designs on the leaf surface. Likewise, rust illness can incur extreme harm under helpful ecological circumstances. Impacted leaves show little yellow spots, as seen in plants experiencing rust. Figure 2 and Figure 3 outline healthy and unhealthy apple plant leaf image samples respectively. Similarly, an entire infected leaf image is shown in Figure 4.
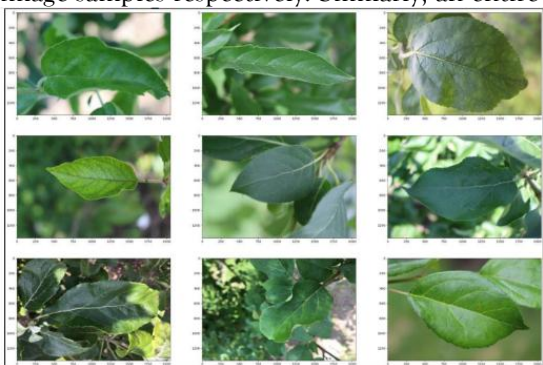
Figure 2. Sample images of healthy apple tree leaves
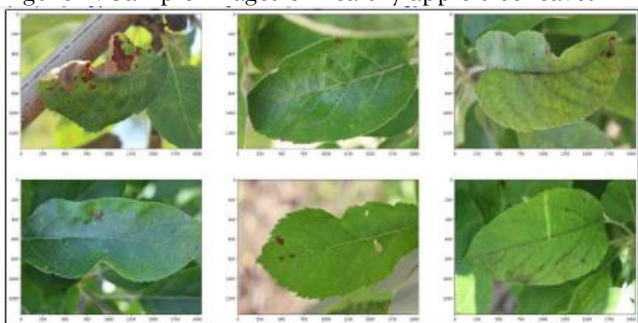
Figure 3. Samples of Images of Apple leaf disease like scab, rust, and powdery mildew.

Figure 4. Single Image of Infected Apple Leaf
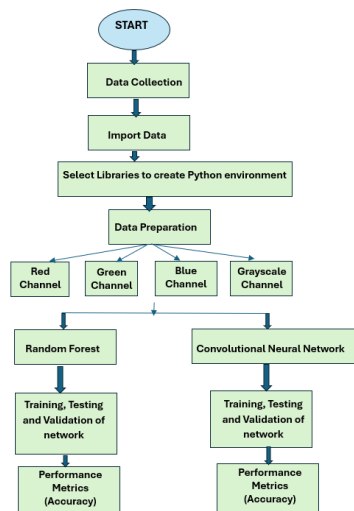
Figure 5. Systematic flowchart architecture for apple disease identification by channel response analysis with Random Forest (RF) and Convolutional Neural Network (CNN) model evaluation and implementation.

Detecting illnesses in apple plant leaves is a challenging task due to the many diseases that are affecting the apple leaves. Nowadays, due to the recent advancement of many advanced artificial intelligence techniques have powerful devices for automated computer-based illness identification in the Apple plant. The automated detection and locating of apple plant illness through artificial intelligence involves many essential stages. In the first stage, a dataset of apple plant leaf images is collected and compiled into two groups, i.e. healthy and infected illness leaves. Following the data preprocessing stage, the color and grayscale images are further split into individual channels for further deep analysis to recognize which channel is giving more information. The observed analyzed the suggested brightness of the three number one shades (Red, Green, and Blue) inside the leaves of regenerated vegetation to determine their dating with chlorophyll content:

- **Red (R) and Green (G):** The implied brightness of those colorations confirmed a terrible correlation with chlorophyll content material, meaning their brightness reduced as chlorophyll content material improved.
- **Blue (B):** In contrast, the implied brightness of Blue showed a growing fashion with chlorophyll content. However, the correlation between blue brightness and chlorophyll content becomes vulnerable.

Once it is collected then the image from that channel is fed and trained with a machine learning model i.e. Random Forest and a deep learning model i.e. CNN to recognize the depth patterns within the images. Figure 5 shows a systematic flowchart architecture for apple disease identification by channel response analysis with Random Forest (RF) and Convolutional Neural Network (CNN) model evaluation and implementation. Following the entire process, the outperforming bets framework was deployed as an application for real-time world usage.

B. Feature Selection and Extraction

The Feature Selection process includes all necessary steps to identify the most active and effective features to enhance the classification performance. Descriptive color statistics, textures, and shape features can be considered to identify easily the crop disease with proper signatures. A strong courting becomes discovered between chlorophyll content and the mean brightness of Red and Green, indicating that these colors are greater applicable for determining chlorophyll content with the use of color factor in comparison to Blue. When analyzing the ratio of implied brightness (RGB) across the three colorations, an enormous correlation is observed with chlorophyll content material:
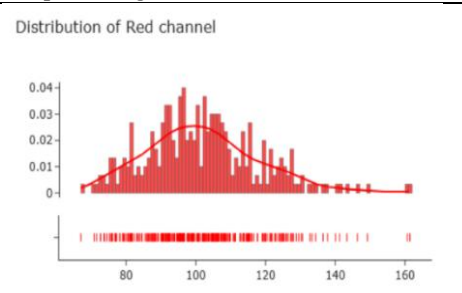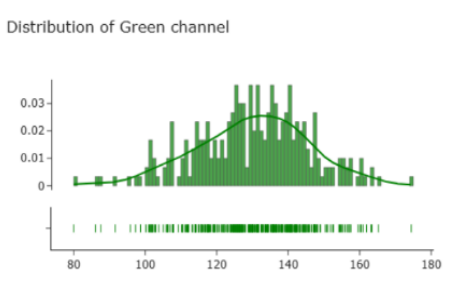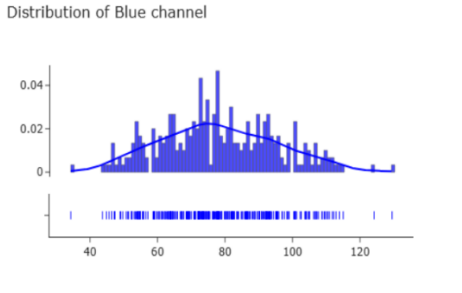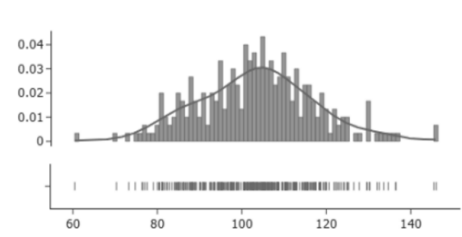
- Red (R) and Green (G) maintained a bad correlation with chlorophyll content.
- Blue (B) showed an advantageous correlation with chlorophyll content material.

This indicates that even as all 3 number one shades are worried, Red and Green are more dependable indicators of chlorophyll content material in regenerated plants.

We observed that the blue channel is the most varied variety in channels and changes diversely for the solid example than other infection tests. Along these lines, we attempt to involve this in a machine-learning model. Then, at that point, we utilize a profound learning model and look at the exhibition of the two models. Many methods overlook

the histogram-combined RGB channel distribution disparities in apple plant images (as described in Figure 6 and Table I) when creating their framework architecture. But here we have considered the spatial properties in the frequency domain of the original image, allowing us to learn from the non-linear restoration information mappings. Table III summarizes the study to obtain detailed plant image sample data by depth, with sample selection guided by the classification of different species. Table II shows the histogram and boxplot analysis of different samples Images of healthy Apple leaves and diseased leaf-like scab, rust, powdery mildew.

Table 1. Histogram distribution of original image samples individual RGB channel distribution analysis

Color Model	Band	Equation	Output Image of Color Bands
RGB Sample	Red	$$\dfrac{R}{R+G+B}$$	Distribution of Red channel
	Green	$$\dfrac{G}{R+G+B}$$	Distribution of Green channel
	Blue	$$\dfrac{B}{R+G+B}$$	Distribution of Blue channel
Grayscale Sample	Grayscale	$$0.299\,R + 0.587G + 0.114\,B$$	Distribution of RGB channels

Distribution of RGB Channels combined
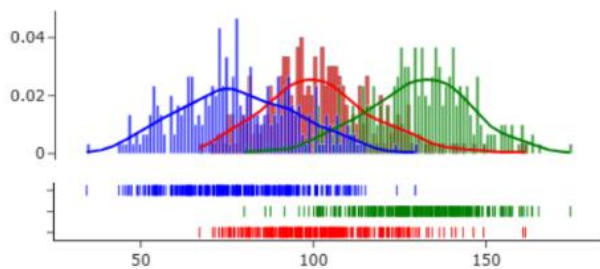
Figure 6. Histogram distribution of original image samples combined with RGB channel distribution analysis

DISCUSSION

C. Evaluation Analysis

This section outlines the performance of the systematic flowchart architecture for apple disease identification by channel response analysis with Random Forest (RF) and Convolutional Neural Network (CNN) model evaluation and implementation in terms of accuracy, precision, recall, and F1-score. The collected data is in the form of color image spaces and further they were converted to grayscale space. The experimental system was set on a Windows PC with an i7 7700HQ processor with 16GB RAM processor running at 2.8GHz. We have focused on training around 70% of the data and 30% rest for testing. Here we had considered a 0.001 learning rate, 128 batch size with 0.9 momentum. For this work, the performance of the framework following parameters such as accuracy, precision, recall, and F1-score must be considered for checking the effectiveness of the model depicted in Eq. (1) – (4).

$$\text{Accuracy} = \frac{\text{True Positive+True Negative}}{\text{True Positive+True Negative+False Negative+False Positive}} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive+False Positive}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive+False Negative}} \quad (3)$$

$$\text{F1} - \text{Score} = 2 \times \frac{\frac{\text{True Postive}}{\text{True Positive+False Positive}} \times \frac{\text{True Positive}}{\text{True Positive+False Negative}}}{\frac{\text{True Positive}}{\text{True Positive+False Positive}} + \frac{\text{True Positive}}{\text{True Positive+False Negative}}} \quad (4)$$

Where True Positive = correctly classified positive apple leaves, True Negative = classified negative apple leaves, False Negative = incorrectly classified apple leaves, and False Negative = incorrectly classified positive apple leaves.

Key insights are exposed by means of this comparative evaluation between the Random Forest-based totally machine learning model and the Convolutional Neural Network (CNN) deep learning model to identify disease spots and classify healthy and diseased plant leaf samples.The classical gadget gaining knowledge of the model, based on Random Forest, was done with 69.59% of accuracy, 83.66% of precision, 70.14% of recall, and an F1 score of 76.30%. The metrics endorse that the Random Forest model discriminates and classifies diseases like apple scab, rust, and multiple diseases exceptionally well, especially with the accuracy metrics. On the other hand, the CNN model, which uses advanced deep gaining knowledge of practices, showed a few vast performances; accuracy of 89.05%, precision of 90.71%, recall of 89.05%, and an F1 score of 89.87%. These consequences prove that this model has awesome competencies to hit upon and classify disease samples with excessive precision and recall, maintaining a low number of false positives and negatives. The performance variations between these models are presented conspicuously excessive in the performance metrics between the Random Forest and CNN models, which simply emphasize the sizeable benefit of the usage of the deep mastering technique in solving tough issues of image classifications—particularly in plant-associated pathology—at ease. The expanded CNN accuracy and F1 rating in comparison to the Random Forest underline that the CNN model needs to be more capable of seizing and interpreting the complicated styles and capabilities of the leaves suffering from sicknesses, considering greater accuracy and dependable disease detection.

Table II. Boxplot analysis of different channels R, G, and B and showcasing the mean values of the individual channels.
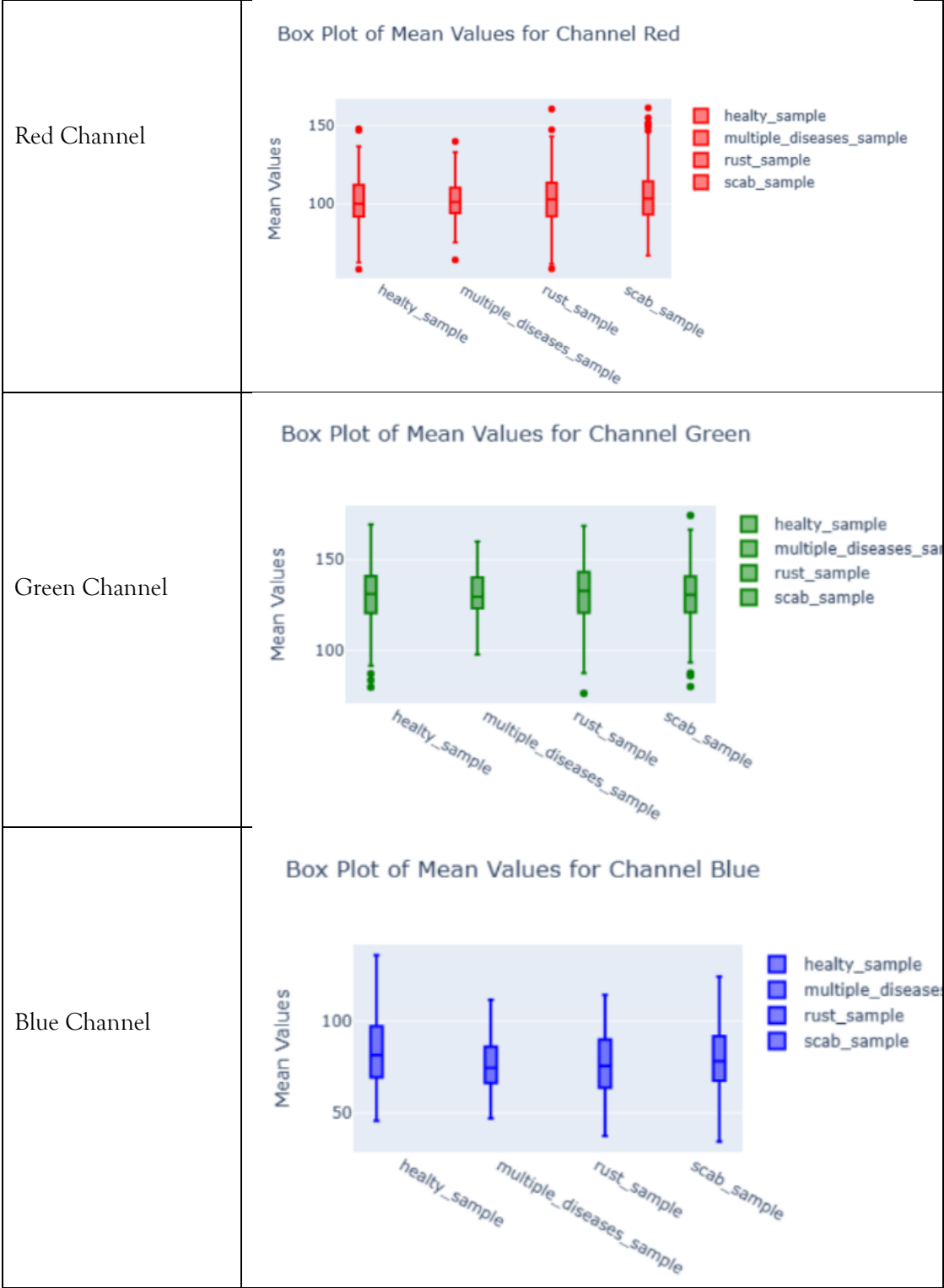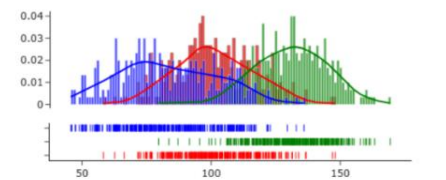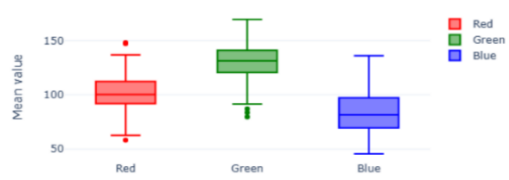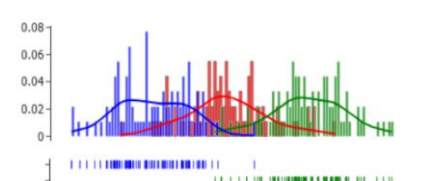
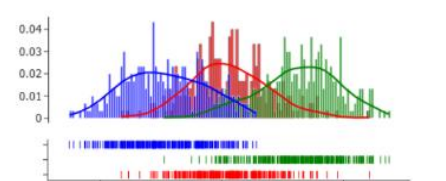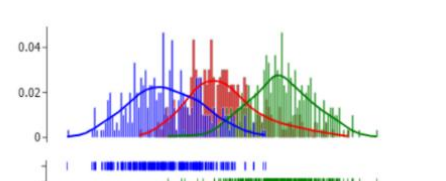|	
Red Channel	
Green Channel	
Blue Channel	

Table III. Histogram and boxplot analysis of different samples Images of healthy Apple leaf and diseased leaf-like scab, rust, powdery mildew

Sample Type	Histogram Analysis	Boxplot Analysis
Healthy Sample		
Multiple Disease Sample		
Rust Disease Sample		
Scab Disease Sample		

Figure 7. Chart showcasing the different performance metrics i.e. accuracy, precision, recall, and F1- Score of two different algorithms i.e. random forest and CNN.
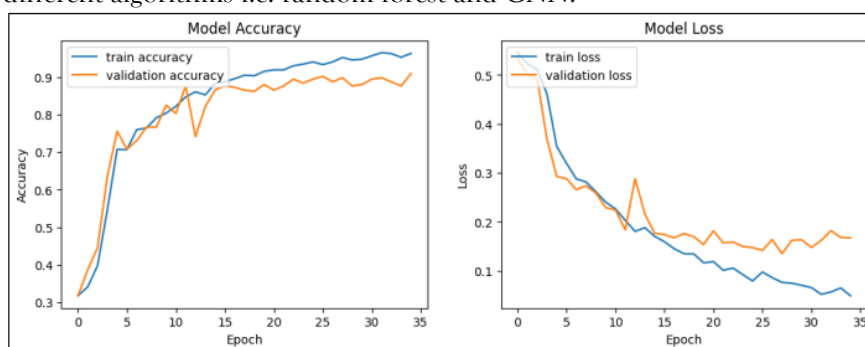
Figure 8. The ROC Curve of the CNN model outperforms the random forest algorithm

D. Execution Time Analysis

Figure 9 shows the details plot for the mean training time in seconds for different color spaces mainly for RGB and grayscale. In this work, we have mainly focused on the amount of time required for training the train data. We have employed two AI techniques, i.e. Random Forest and CNN for this work. After the completion of the entire task, we found that the time required for training is much longer compared to the testing duration, just the same way a classifier works. The agricultural workers are the end users who can use the entire tool for testing and summarizing the tool can be used to work in the real-time field.

Figure 9: Showcases the mean execution time (seconds) of the classifiers for the training data. The blue channel performs the best compared to other channels and requires the least training time.

Figure 10. Confusion Matrix of the random forest and CNN model respectively

E. Confusion Matrices Analysis

Figure 10 illustrates the confusion matrices of 70:30 training: testing samples data for the apple plant. The confusion matrices showcase the statistics of the RGB space where for example row 2 of the predicted class o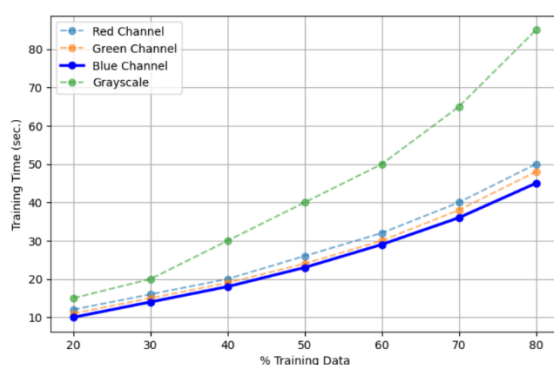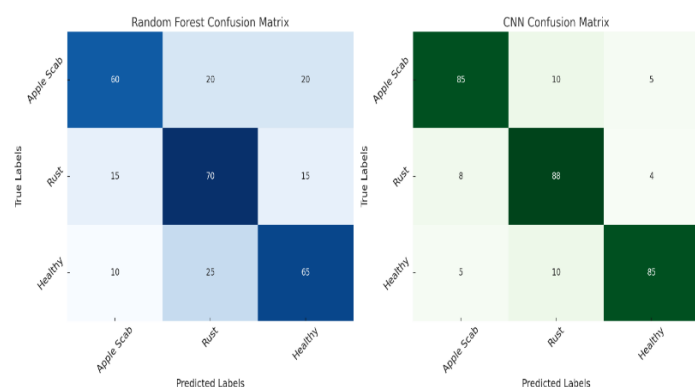f random forest out of 100 apple sample images (15 + 70 + 15), the framework technique classified 15 apple scabs, 70 rust, 15 healthy whereas the predicted class of CNN with 100 apple sample images (8 + 88 + 4), 8 apple scabs, 88 rust, 4 healthy respectively. Therefore, out of 100 apple samples, 70 samples were correctly classified as rust whereas by CNN, 88 samples were correctly classified as rust. If we focus on the two confusion matrices, cell 2 gives the highest value information with the off-diagonal cells.The new comparative frame detects three common apple leaf illnesses i.e. scab, rust, and fine mold or powdery mildew. Further, we analyzed apple plant illness, which also affects apple fruits and creates future enhanced diseases to the fruits too. In the future, a precise experiment should be performed to develop an automated tool that can help to identify and detect a wide range of ailments that may harm plant leaves, stems, and fruits. This type of harmful challenge can be covered in future work.

CONCLUSION

Apple is one of the most popular and important fruits that has multiple antioxidant benefits by reducing chronic illness. Well, this popular fruit is a risk in production due to various organisms like bacterial, fungi, and viruses' infections. Applying pesticides and fertilizers can help to reduce the illness but it will also degrade the quality and quantity of the apple plants. Therefore, early identification and treating the apple plant at the right time is very crucial to reduce productivity and financial loss. In summary, traditional machine learning techniques like Random Forest offer a foundational stage of infection detection, whilst deep learning techniques such as CNNs provide great enhancements in early detection with normal effectiveness, accuracy, precision, and consideration of morphological characteristics. In this paper, RGB images and grayscale images were utilized and analysis of each channel with information transmission is recorded. The integration of CNNs in the detection of plant sicknesses makes early diagnoses and intervention activities extra effective, and therefore, it improves the control of crops, accordingly, shielding the yield and improving agricultural sustainability. The outcomes of this work show that compared with grayscale and RGB image processing procedures, the blue channel has a relevant and reliable result in classifying apple scab illness in early development stages.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest in this work.

DATA AVAILABILITY STATEMENT

Publicly available dataset.

REFERENCES

[1] World Health Organization, 2015. Trends in maternal mortality: 1990-2015: estimates from WHO, UNICEF, UNFPA, World Bank Group and the United Nations population division. World Health Organization.

[2] Oerke, E.C., 2006. Crop losses to pests. The Journal of Agricultural Science, 144(1), pp.31-43.

[3] Chandler, G.N., DeTienne, D.R., McKelvie, A. and Mumford, T.V., 2011. Causation and effectuation processes: A validation study. Journal of Business Venturing, 26(3), pp.375-390.

[4] Lamichhane, J.R., Dachbrodt-Saaydeh, S., Kudsk, P. and Messéan, A., 2016. Toward a reduced reliance on conventional pesticides in European agriculture. Plant Disease, 100(1), pp.10-24.

[5] Mrisho, L.M., Mbilinyi, N.A., Ndalahwa, M., Ramcharan, A.M., Kehs, A.K., McCloskey, P.C., Murithi, H., Hughes, D.P. and Legg, J.P., 2020. Accuracy of a smartphone-based object detection model, PlantVillage Nuru, in identifying the foliar symptoms of the viral diseases of cassava–CMD and CBSD. Frontiers in plant science, 11, p.590889.

[6] Shin, S.Y., Lee, C.M., Kim, H.S., Kim, C., Jeon, J.H. and Lee, H.J., 2023. Ethylene signals modulate the survival of Arabidopsis leaf explants. BMC Plant Biology, 23(1), p.281.

[7] Pattanaik, P.A., Khan, M.Z. and Patnaik, P.K., 2022. ILCAN: a new vision attention-based late blight disease localization and classification. Arabian Journal for Science and Engineering, 47(2), pp.2305-2314.

[8] Tian, F., Yang, D.C., Meng, Y.Q., Jin, J. and Gao, G., 2020. PlantRegMap: charting functional regulatory maps in plants. Nucleic acids research, 48(D1), pp. D1104-D1113.

[9] Sena Jr, D.G., Pinto, F.A.C., Queiroz, D.M. and Viana, P.A., 2003. Fall armyworm-damaged maize plant identification using digital images. Biosystems Engineering, 85(4), pp.449-454.

[10] Karthik, K., Nandiganti, M., Thangaraj, A., Singh, S., Mishra, P., Rathinam, M., Sharma, M., Singh, N.K., Dash, P.K. and Sreevathsa, R., 2020. Transgenic cotton (Gossypium hirsutum L.) to combat weed vagaries: utility of an apical meristem-targeted in planta transformation strategy to introgress a modified CP4-EPSPS gene for glyphosate tolerance. Frontiers in Plant Science, 11, p.768.

[11] Sethy, P.K., Barpanda, N.K., Rath, A.K. and Behera, S.K., 2020. Image processing techniques for diagnosing rice plant disease: a survey. Procedia Computer Science, 167, pp.516-530.

[12] Kacira, M., Ling, P.P. and Short, T.H., 2002. Machine vision extracted plant movement for early detection of plant water stress. Transactions of the ASAE, 45(4), p.1147.

[13] Khotimah, N.N., Rozirwan, R., Putri, W.A.E., Fauziyah, F., Aryawati, R., Isnaini, I. and Nugroho, R.Y., 2024. Bioaccumulation and ecological risk assessment of heavy metal contamination (Lead and Copper) build up in the roots of Avicennia alba and Excoecaria agallocha. Journal of Ecological Engineering, 25(5).

[14] Nhamo, L., Paterson, G., van der Walt, M., Moeletsi, M., Modi, A., Kunz, R., Chimonyo, V., Masupha, T., Mpandeli, S., Liphadzi, S. and Molwantwa, J., 2022. Optimal production areas of underutilized indigenous crops and their role under climate change: Focus on Bambara groundnut. Frontiers in Sustainable Food Systems, 6, p.990213.

[15] Hafeez, M.B., Ghaffar, A., Zahra, N., Ahmad, N., Hussain, S. and Li, J., 2024. Plant growth promoters boost the photosynthesis-related mechanisms and secondary metabolism of late-sown wheat under contrasting saline regimes. Plant Stress, 12, p.100480.

[16] Ding, Y., Shi, Y. and Yang, S., 2020. Molecular regulation of plant responses to environmental temperatures. Molecular plant, 13(4), pp.544-564.

[17] Pattanaik, P.A., Swarnkar, T. and Swain, D., 2021. Deep filter bridge for malaria identification and classification in microscopic blood smear images. International Journal of Advanced Intelligence Paradigms, 20(1-2), pp.126-137.

[18] Pattanaik, P., Ibeh, L. and Cuong, N.M., 2024, November. Identification and Classification of Types of Malarial Parasites Using Customized CNN Model. In 2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) (pp. 1-5). IEEE.

[19] Mohanty, S.P., Hughes, D.P. and Salathé, M., 2016. Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, p.1419.

[20] Li, H., Yan, X., Su, P., Su, Y., Li, J., Xu, Z., Gao, C., Zhao, Y., Feng, M., Shafiq, F. and Xiao, L., 2025. Estimation of winter wheat LAI based on color indices and texture features of RGB images taken by UAV. Journal of the Science of Food and Agriculture, 105(1), pp.189-200.

[21] Pérez-Rodríguez, F. and Gómez-García, E., 2019. Codelplant: Regression-based processing of RGB images for color models in plant image segmentation. Computers and Electronics in Agriculture, 163, p.104880.

[22] Banerjee, A., Tripathi, D.K. and Roychoudhury, A., 2018. Hydrogen sulfide trapeze: environmental stress amelioration and phytohormone crosstalk. Plant Physiology and Biochemistry, 132, pp.46-53.

[23] Gracia-Romero, A., Kefauver, S.C., Vergara-Díaz, O., Zaman-Allah, M.A., Prasanna, B.M., Cairns, J.E. and Araus, J.L., 2017. Comparative performance of ground vs. aerially assessed RGB and multispectral indices for early-growth evaluation of maize performance under phosphorus fertilization. Frontiers in Plant Science, 8, p.2004.

[24] Zhao, C., Zhang, H., Song, C., Zhu, J.K. and Shabala, S., 2020. Mechanisms of plant responses and adaptation to soil salinity. The innovation, 1(1).

[25] Ahmed, A.A. and Reddy, G.H., 2021. A mobile-based system for detecting plant leaf diseases using deep learning. AgriEngineering, 3(3), pp.478-493.

[26] Vigni, M.L., Durante, C. and Cocchi, M., 2013. Exploratory data analysis. In Data handling in science and technology (Vol. 28, pp. 55-126). Elsevier.

[27] Hatuwal, B.K., Shakya, A. and Joshi, B., 2020. Plant Leaf Disease Recognition Using Random Forest, KNN, SVM, and CNN. Polibits, 62, pp.13-19.

[28] Thapa, R., Snavely, N., Belongie, S. and Khan, A., 2020. The Plant Pathology 2020 challenge dataset to classify foliar disease of apples. arXiv preprint arXiv:2004.11958.

[29] Agarwal, A., de Jesus Colwell, F., Bello Rodriguez, J., Sommer, S., Correa Galvis, V.A., Hill, T., Boonham, N. and Prashar, A., 2024. Monitoring root rot in flat-leaf parsley via machine vision by unsupervised multivariate analysis of morphometric and spectral parameters. European Journal of Plant Pathology, pp.1-19.