Revolutionizing Citrus Health: AI-Based Detection of Greening Disease and Nutrient Deficiencies in Sweet Orange

Abstract
This paper explores the application of Artificial Intelligence (AI) techniques for detecting plant diseases, focusing on citrus greening disease and various foliar nutrient deficiencies in sweet orange. Agriculture faces numerous challenges from cultivation to harvesting, including significant yield losses due to disease infections and environmental hazards from excessive use of insecticides and fungicides. As the global population grows, the demand for food is surging, and traditional farming methods fall short in meeting this demand, often degrading soil health through intensive pesticide use. AI offers significant advantages over conventional techniques in disease detection. This study employs AI for visual detection of citrus greening disease and foliar nutrient deficiencies, achieving 87% accuracy on a test dataset of infected, nutrient-deficient, and healthy sweet orange leaves. Performance is evaluated using metrics such as Accuracy, Recall, Precision, and F1-Score. Specifically, Convolutional Neural Network (CNN) architectures, including Visual Geometry Group (VGG-16), are utilized for image-based detection and classification, demonstrating the potential of AI in enhancing plant health monitoring.

Keywords: Nutrient deficiencies, Convolutional Neural Network (CNN), Visual Geometry Group (VGG), Accuracy, Recall, Precision and F1-Score.

Introduction

Citrus and its families have horticultural significant because they include important fruit crops such as mandarins, oranges, lemons, limes and grapefruit, among others. It is belonging to Rutaceae family and native to South East Asia. It is restricted to 0-400 latitudes from North to South of the equator, covering different regions with different soil and climatic conditions. Citrus is one of the world most important fruit crop and more than 100 countries growing it. Citrus productivity is influenced by a variety of biotic and abiotic factors. Bacteria, fungi, viruses and nematodes are among the biotic factors that are causing significant reductions in citrus production worldwide. Citrus greening is the most destructive bacterial disease and has played a significant role in the decline of citrus in India and other Asian, Pacific and African countries Ahlawat [1]. The disease was first reported in Southern China in 1919 Reinking [13] and it is now known to be present in Asian, African, South and North American countries. Dr. Lilian R. Fraser and his co-workers first reported citrus greening in India during the year 1966. The disease is caused by the Gram-negative fastidious bacterium *Candidatus liberibacter asiaticus* in Asia, *Candidatus liberibacter africanus* in Africa Jagoueix [8] and *Candidatus liberibacter americanus* in South America Garnier [6].

Citrus crop yield is currently declining due to nutrient deficiencies such as Nitrogen (N), Potassium (K), Zinc (Zn), Manganese (Mn), Magnesium (Mg), Ferrous (Fe), Boron (B) and others. All these nutrient deficiencies are affecting citrus growth, yield and quality. Micronutrient deficiencies reduce crop productivity as well as the efficiency of applied major nutrients in crop. Micronutrient deficiencies are on the rise and are likely to continue with their negative impact. Identifications of these nutrient deficiencies in crop are confused with diseases too. Sweet orange productivity in India is significantly lower than in some of the frontline citrus growing countries such as Brazil, the United States, Spain, and Italy. Similarly, the average productivity of sweet orange orchards varies according to sweet orange variety [10-12]. Nutrient deficiencies are one of the primary causes of low sweet orange orchard productivity in the Marathwada region. Because the soils in this region are mostly derived from basaltic parent material and are commonly deficient in multiple nutrients such as N, P, K, Fe, Mn, Mg, Bo and Zn the traditional nutrient management strategy based primarily on macronutrient application in citrus orchards has not been very successful in increasing productivity Srivastava [16]. Plant disease identification is a major concern that has been studied over all the years, and it is motivated by a desire to grow healthy food. However, budget, user-friendliness, sensitivity, and reliability are some desirable characteristics to remember Irudayaraj [7]. Convenient and precise disease detection in plants could help in the creation of an earlier treatment method while greatly reducing economic losses Fuentes [5].

A series of new methods for successfully recognizing plant diseases particularly citrus greening and its symptoms in real-world situations compare to nutrient deficiencies, while also provide some recommendations for developing more reliable identification methods that can be accommodated depending on the circumstances. Approaching deep learning (DL) by proposing the following two structures: (1) A comprehensive meta-architecture for detecting plant diseases and pests in real-world situations. (2) An effective method for dealing with issues such as imbalance of class and assumed positives, particularly in applicability with an inadequate number of data are now need of present era. Hence, 'Detection through artificial intelligence (AI) of citrus greening disease (CGD) and various foliar nutrient deficiencies (FND) of sweet orange.' study was being undertaken.

Materials and methods

Related to detection through AI of CGDe of sweet orange and nutrient deficiency, different experiments were carried out at the ARS, Badnapur, SORS, Badnapur, Department of Plant Pathology, COA, Badnapur, District. Jalna (MS) and Shri Guru Gobind Singhji Institute of Engineering and Technology (GGSIET), Vishnupuri, Nanded for classification work as data splitting, training, validation and testing dataset related to AI during the year 2020-2021 for fulfilment of the defined objectives. The materials and techniques utilized in the current research are mentioned below.

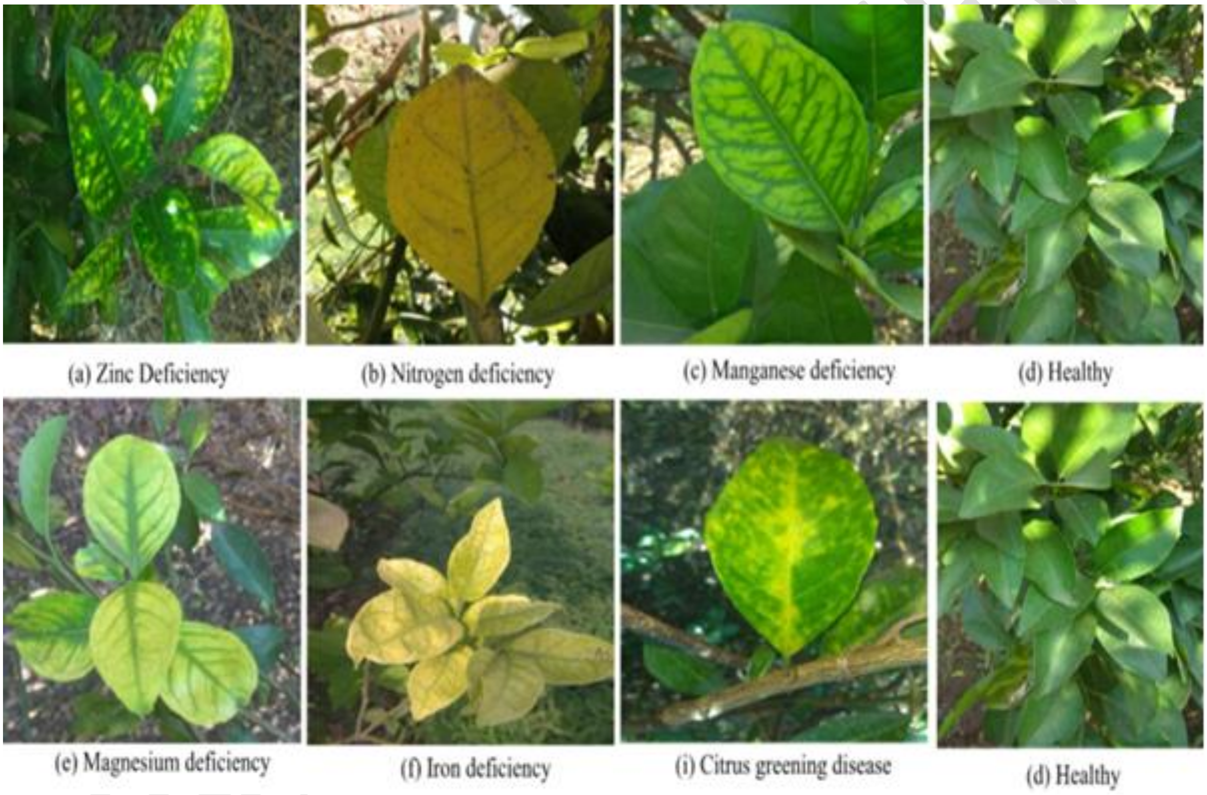

Fig. 1 Visual detection of CGD and various FND of sweet orange through AI.

Table 1. Sweet orange datasheet statistics

Classes	Images collection	Resolution	Annotation type	Number of images
CGD				1824
Healthy leaf				2542
Fe deficiency				1956
Mn deficiency	Real field images	1080 X 1080	Image level	1538
Zn deficiency				2270
N deficiency				951
Mg deficiency				1269
Grand total: Twelve thousand three hundred and fifty photographs				12350

Data collection and annotation: dataset contains approximately 12,350 sweet orange field images taken by Sony DSLR and Smartphone. These field images were taken under varying environmental conditions to create a comprehensive model. Visited many sweet orange orchards to capture images of citrus greening and nutritional deficiencies affected leaves for this purpose. Encountered certain difficulties when capturing an image of leaves, such as the difficulty in distinguishing between citrus greening and various nutritional deficiencies when capturing images. Proposed dataset contains a variety of images, including photographs of varying resolutions captured by smartphone in varying light conditions based on the time of year, *e.g.*, temperature, humidity and different environmental locations (Maharashtra, India). Gathered the photographs from various perspectives. To avoid confusion between citrus greening and nutritional deficits in model, photos were obtained from a variety of citrus species, orchards and environments (Maharashtra, India). However, in this study, used datasets that included healthy leaves, citrus greening and nutrient deficiency leaves on the base of visual detection through practice.

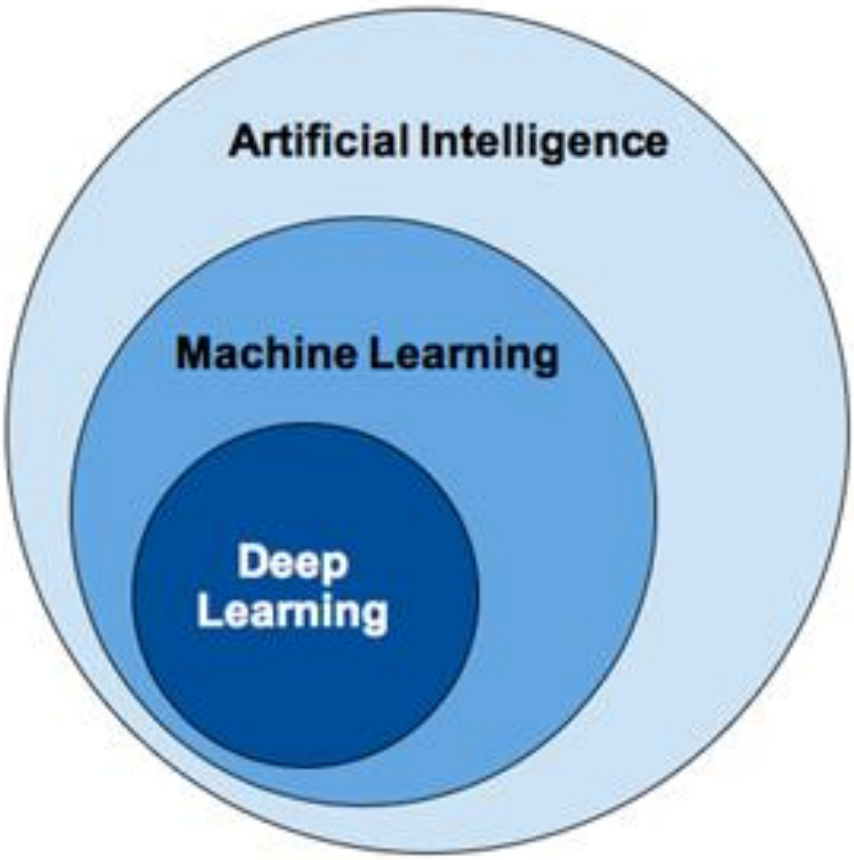

Fig 2. The overall diagram of Artificial Intelligence

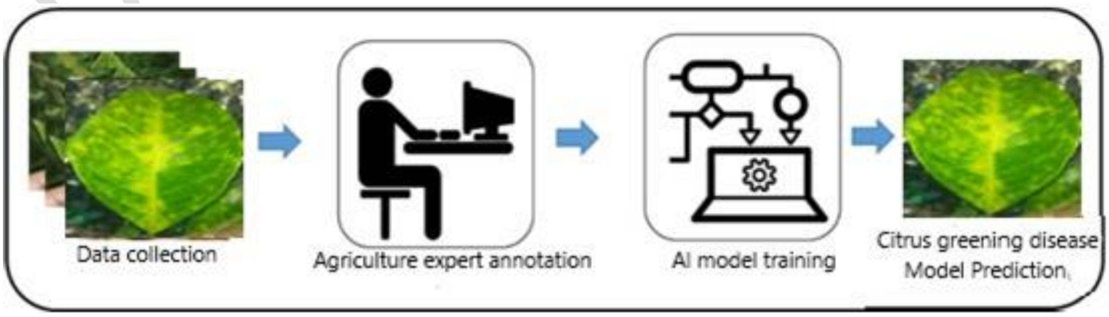

Fig. 3. The sequential procedure of proposed method.

Convolutional Neural Networks (CNN):

Recent development in AI promotes use in the Industrial sector as well in the agriculture sector. AI is broadly classified in Machine learning (ML) and DL fields. Whereas DL is a subset of ML and ML is a subset of AI as shown in fig.3 CNN is a subfield of the DL field which is very popular for computer vision tasks. CNN takes advantage of the spatial structure of the input image. The first CNN network structure was proposed by Fukushima in 1988. It was not widely used due to limitations of computation hardware for training the network. Le Cun et al. (1990) used a CNN based algorithm and obtained accurate results for the hand-written digit classification task. CNNs have been extremely successful and widely used in computer vision applications, such as Image classification, segmentation and object detection. Standard CNN structure comprises three main types of neural layers, namely, convolutional layers, pooling layers, and fully connected layers as shown in Fig 3. Each layer has a specific role. Purpose of the convolutional layer is to extract meaningful features from the image, the pooling layer is used to reduce dimension of image, specifically max pool applied on neighbourhood values to reduce dimension. The ReLu activation function is used to add non linearity, a further fully connected layer is used to flatten the output. After these steps, a softmax function is applied to classify the input into one of predetermined classes.

In such work, evaluated CNN-based architectures, namely, VGG-16 for image-based detection and classification of sweet orange citrus greening diseases and nutritional deficiency. In order to solve data scarcity problem, collected real field images data from a citrus field in an uncontrolled environment detail found in Table 3. Further utilized data augmentation techniques to increase the number of instances. Also collected images from different geographic locations to add diversity in the dataset. Collected dataset also covers challenges such as image blur due to wind, brightness variation, overlapping of leaves and shadow problem. These techniques of evaluation are more reliable and representative of a real-world situation. The findings were reported using a variety of performance indicators, including accuracy, recall, precision, and F1- score. (The VGG Net.) The Visual Geometric Group (VGG) of the University of Oxford was the first to introduce a pertained model. VGG Net's main operating idea is to employ deeper layers with smaller filters. (The VGG architecture's input layer dimension is configured for a picture size. After pre-processing, there was a stack of five convolutional layers, each of which was followed by a MaxPool layer, thus each set of convolutional layers was followed by a MaxPool layer. (Three fully connected (FC) levels come before the final MaxPool layer. The first two FC layers each contain 64 64 (4096) channels, whereas the final FC layer has 1000 channels and was followed by a softmax activation function. VGG network comes in a variety of varieties, including VGG-16 and VGG-19. The design of VGG-16 and VGG-19 were the same, however the number of layers differs. VGG-16 has sixteen layers, whereas VGG-19 has nineteen. The number of convolution layers in the 3rd, 4th, and 5th layers of convolutional layers stacks is a distinguishing element.

Performance Evaluation Metrics:

As performance assessment measures, they employed accuracy, precision, recall, and F1-score. Note that the basic confusion matrix was deceiving, used the performance evaluation criteria mentioned earlier.

- Accuracy:

  Accuracy (A) was the percentage of presently categorized forecasts that are correct. It is calculated as follows:

$$A = \frac{\mathbf{TP} + \mathbf{TN}}{\mathbf{TP} + \mathbf{TN} + \mathbf{FP} + \mathbf{FN}}$$

(True positive, true negative, false positive, and false negative are represented as TP, TN, FP and FN, respectively.) Precision (P) was the percentage of positive outcomes that were truly right, and it's computed this way:

$$P = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FP}}$$

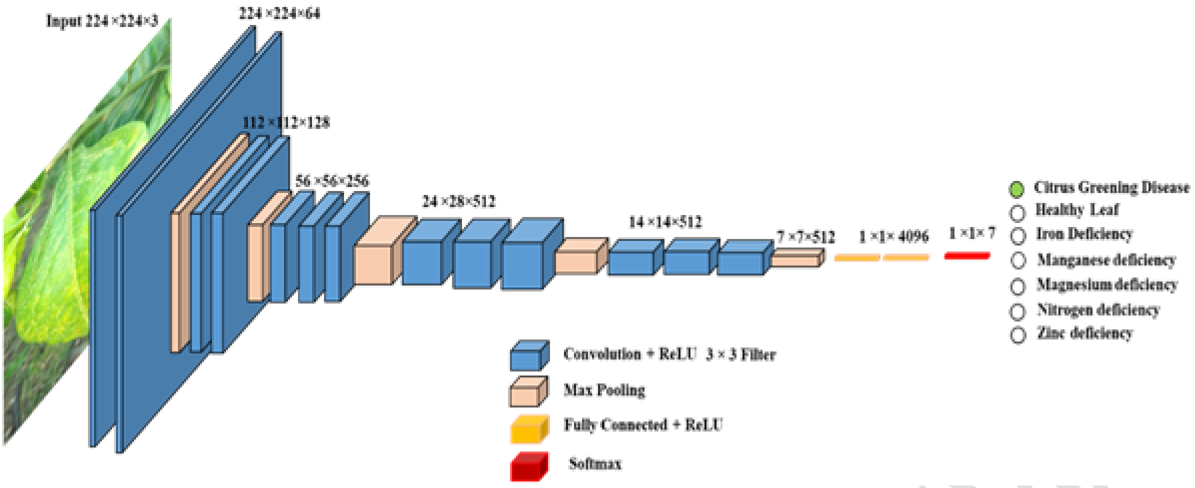

Fig. 4. Structure CNN model including convolutional, pooling, and fully connected layers.

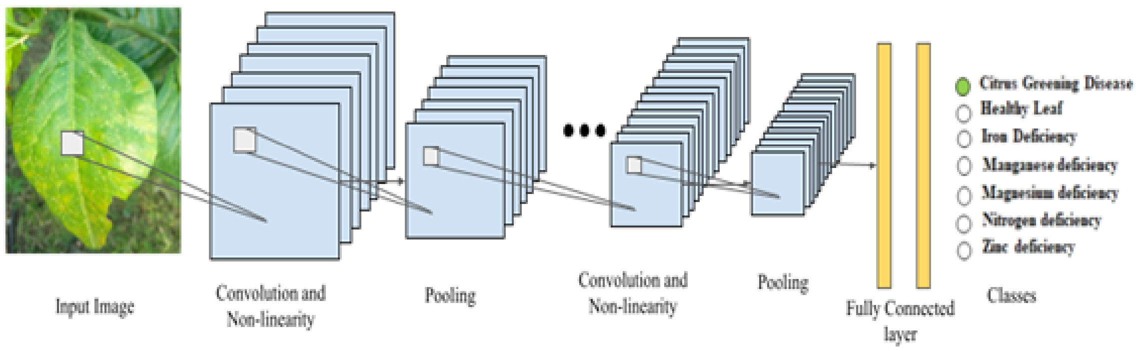

Fig. 5. Basic Architecture of CNN of model.

- **Recall :**
  The proportion of real positives that were properly recognized is measured by recall (R), which was computed as follows:

$$R = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FN}}$$

- **F1-Score :**
  The F1-score was derived by taking the harmonic mean of accuracy and recall:

$$F1 = 2\frac{\mathbf{P} - \mathbf{R}}{\mathbf{P} + \mathbf{R}}$$

Result and discussion

The current investigations on 'CGD as well as various foliar nutrient deficiencies of sweet orange detection through AI and its management' were carried out at ARS, Badnapur, SORS, Badnapur, Department of Plant Pathology, College of Agriculture, Badnapur, Dist. Jalna (MS) and SGGSIET, Vishnupuri, Nanded (for classification work as Data Splitting, Training, Validation and Testing Dataset related to AI) during the year 2020-2021 to fulfil the defined objectives. The results obtained during the research work of these investigations are being presented statistically in the form of following tables.

Visual detection through AI of CGD and FND of sweet orange

To solve the data scarcity problem, real field images collected from sweet orange fields in an uncontrolled environment detail given in (Table 1). Further utilized data augmentation techniques to increase the number of instances. Images were collected from different geographic locations to add diversity to the dataset. The collected dataset also covers challenges such as image blur due to wind, brightness variation, overlapping of leaves and shadow problems. Our techniques of evaluation are more robust and representative of a real-world scenario. The results are based on different performance criteria, such as Accuracy, Recall, Precision, and F1-Score. This model had got 87 % accuracy on a test sample of CGD, FND and healthy leaves of sweet orange dataset, although in this work, CNN architectures, namely, VGG-16 for image-based detection and classification of CGD, FND and Healthy leaf.

Accuracy using proposed method architecture on sweet orange dataset (pretrained model evaluation metrics accuracy, precision, recall, F1–score)

Evaluation metrics of Precision, Recall and F1- score on each dataset of CNN architecture as summarized in the (Table 2 and Fig. 7, 8, 9, 10). The main goal of this research work was to detect CGD and FND's including healthy leaves of sweet orange. CGD accuracy using proposed method architecture on sweet orange dataset included Precision (75 %), Recall (87 %) and F1- Score (80 %). Healthy leaves accuracy using proposed method architecture on sweet orange dataset included Precision (99 %), Recall (99 %) and F1- Score (99 %). Zinc deficiency accuracy using proposed method architecture on sweet orange dataset included Precision (78 %), Recall (61 %) and F1- Score (68 %). Manganese deficiency accuracy using proposed method architecture on sweet orange dataset included Precision (89 %), Recall (89 %) and F1- Score (89 %). Magnesium deficiency accuracy using proposed method architecture on sweet orange dataset included Precision (83 %), Recall (88 %) and F1- Score (85 %). Iron deficiency accuracy using proposed method architecture on sweet orange dataset included Precision (99 %), Recall (95 %) and F1- Score (97 %). Nitrogen deficiency accuracy using proposed method architecture on sweet orange dataset included Precision (92 %), Recall (84 %) and F1- Score (88 %). Among the pretrained models, VGG-16 model performed best, achieving an averaged accuracy of 87.00 % on the test set compared to other models.

Classes wise accuracy using proposed method architecture on sweet orange dataset.

Prediction of the correct class between 07 (Seven) possible classes. *i. e.*, CGD, healthy leaves and FND's (Zn, Mn, Mg, Fe and N) are shown in (Table 2, 3 and Fig 10, 11). Total test image samples included in the healthy leaves class were 508 (Five hundred and eight). Among them right predicted images by the model were 505 (Five hundred and five) images. Wrong predicted images by the model were 03 (three) images and the sensitivity of the model for healthy leaves class was 99 per cent. Total test image samples included in the citrus greening disease class were 454 (Four hundred and fifty-four). Among them right predicted image by the model were 393 (Three hundred and ninety-three) images. Wrong predicted images by the model were 61 (sixty-one) images and the sensitivity of the model for CGD class was 87%.

Total test image samples included in the nutrient deficiency of zinc class were 254 (Two hundred and fifty-four). Among them right predicted image by the model were 154 (One hundred and fifty-four) images. Wrong predicted images by the model were 100 (Hundred) images and the sensitivity of the model for zinc class was 61 %. Total test image samples included in the nutrient deficiency of manganese class were 391 (Three hundred and ninety-one). Among them right predicted image by the model were 347 (Three hundred and fourty seven) images. Wrong predicted images by the model were 44 (Forty-four) images and the sensitivity of the model for manganese class was 89 %. Total test image samples included in the nutrient deficiency of magnesium class were 308 (Three hundred and eight). Among them right predicted image by the model were 271 (Two hundred and seventy-one) images. Wrong predicted images by the model were 67 (Sixty-seven) images and the sensitivity of the model for magnesium class was 88 %.

Total test image samples included in the nutrient deficiency of iron class were 190 (One hundred and ninety). Among them right predicted image by the model was 181 (One hundred and eighty-one) images. Wrong predicted images by the model were 09 (Nine) images and the sensitivity of the model for iron class was 95 %. Total test image samples included in the nutrient deficiency of Nitrogen class were 365 (Three hundred and sixty-five). Among them right predicted image by the model were 307 (Three hundred and seven) images. Wrong predicted images by the model were 58 (Fifty-eight) images and the sensitivity of the model for Nitrogen class was 84 %.

Such similar results obtained in the present investigation are in accordance with earlier works *viz*., Denis et al. (2020) [2] reported that a DL approach for determining affects Tomato leaf miner (*Tuta absoluta*) in tomato plants. Among the pretrained architectures, experimental results showed that Inception - V3 yielded the best results with an average accuracy of 87.2 percent in estimating the severity status of *T. absoluta* in Tomato plants. Selvaraj et al. (2019) [15] reported AI powered banana diseases and pest detection. These architectures represent the state-of-the-art results of banana diseases and pest detection with an accuracy of more than 90% in most of the CNN models tested.

Confusion matrix:
In the field of DL, specifically the problem of statistical classification, the confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm. It considers different metrics: The True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) etc. Based on the results obtained on the test dataset, we generated a confusion matrix for each of the best architectures. Due to the complexity of the pattern shown in each class from different plant parts, the system tends to be confused on several classes has results in lower performance. Based on the results, we can visually evaluate the performance of the classifier and determine which classes and features are more prone to confusion. If the number of misclassifications between two particular classes becomes high, it indicates that we need to collect more data on those classes to properly train the convolutional architecture so that it can differentiate between those two classes. On comparing among the models, leaves produced a lot of confusion and low accuracy (87 %) especially citrus greening disease class and zinc class, followed by magnesium and manganese. They produced similar symptom and they are difficult to identify for human that's why this model also confused in this class. This confusion of model result in low accuracy for prediction of disease. Shown in (Fig 4). Based on a result, visually analysed classifier performance and decide which classes and characteristics are more highlighted by the neurons in the network. It also helps us in analysing a subsequent process to avoid inter-class confusions. For instance, the CGD class shows to be confused in more intensity with zinc class.

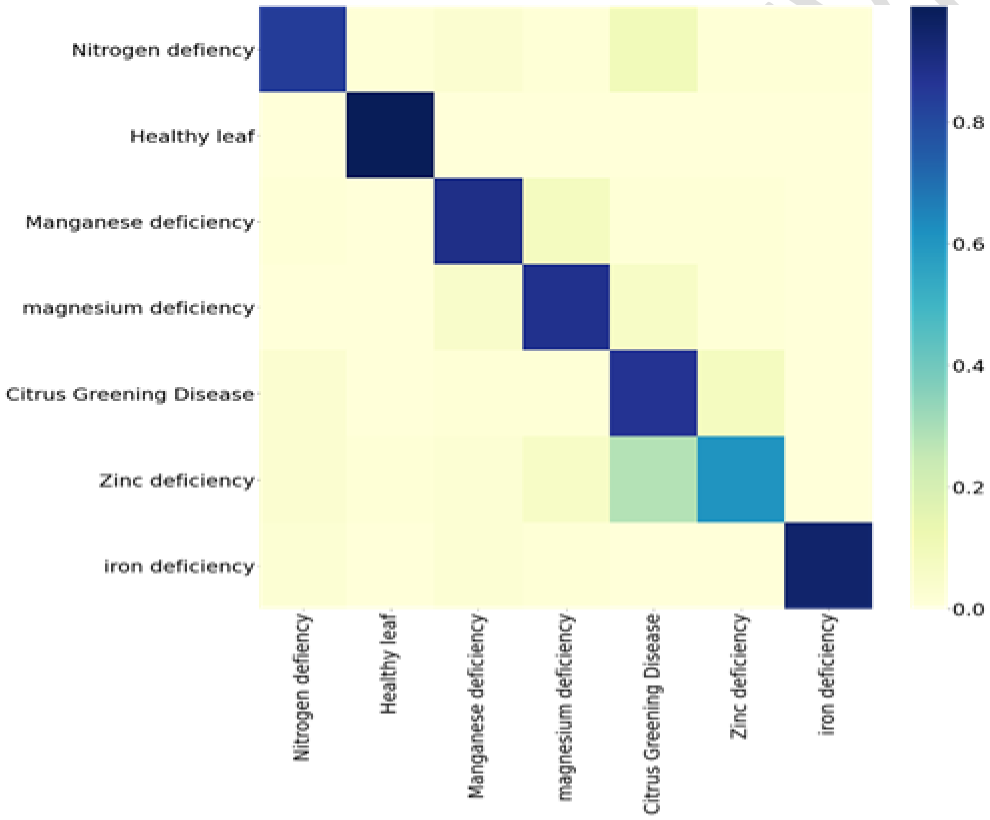

Fig 6. Confusion matrix of CGD and nutrient deficiencies of sweet orange.

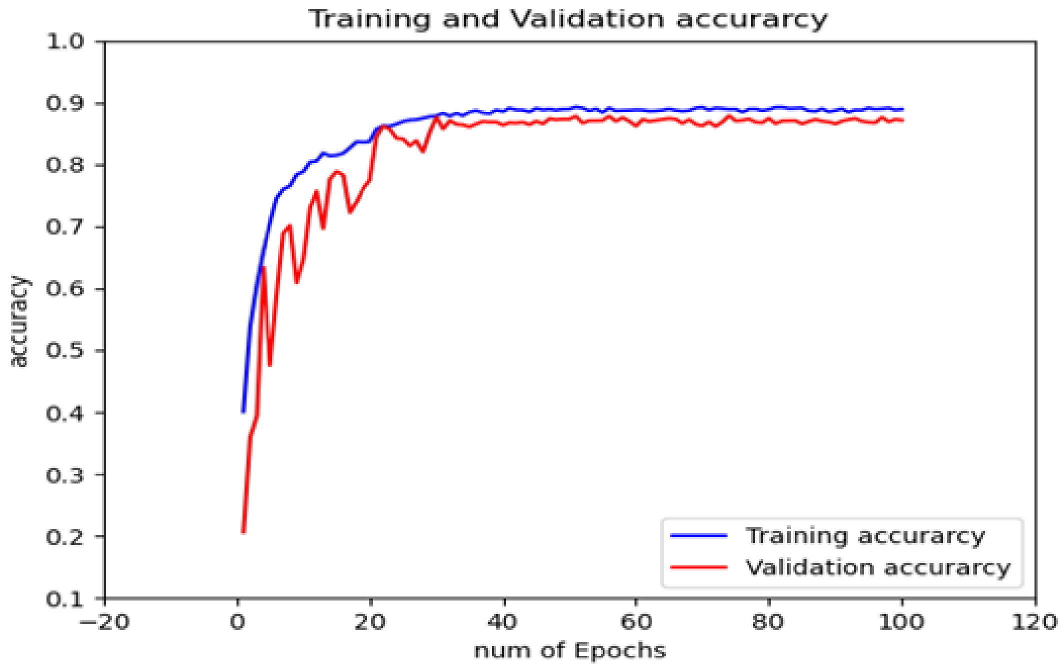

Fig 7. Training and validation loss of the model

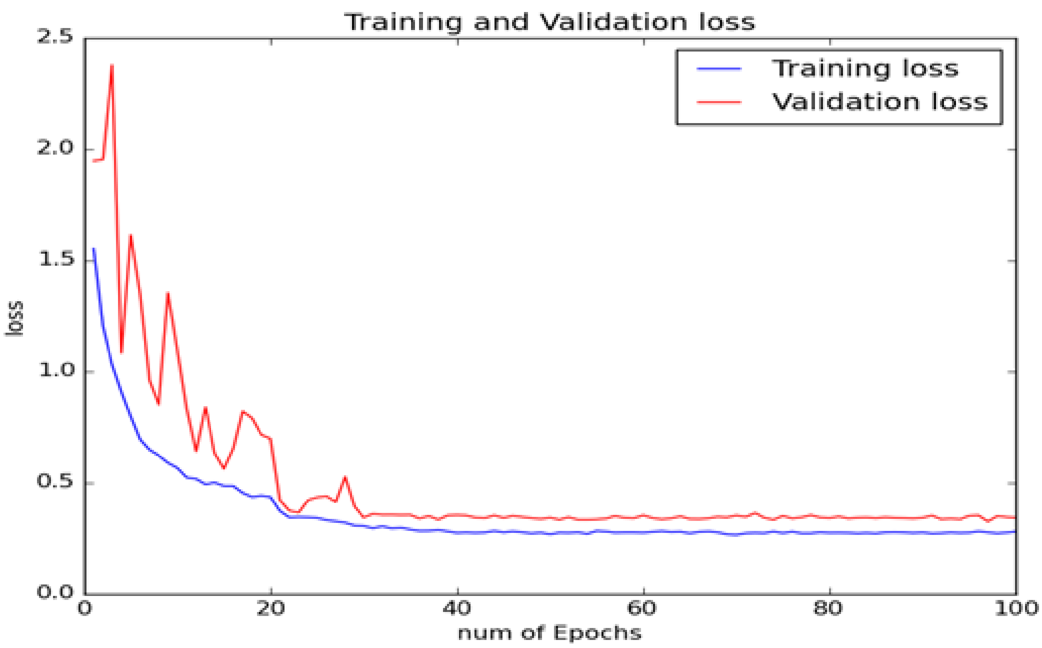

Fig 8. Training and validation accuracy of the model

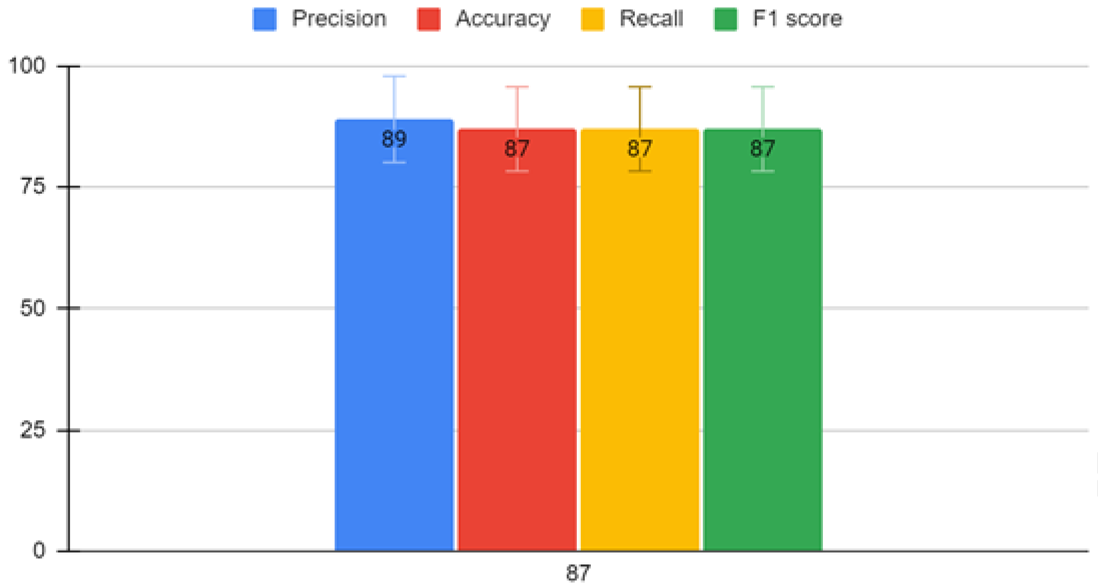

Fig 9. Pretrained model evaluation metrics accuracy, precision and F1-score

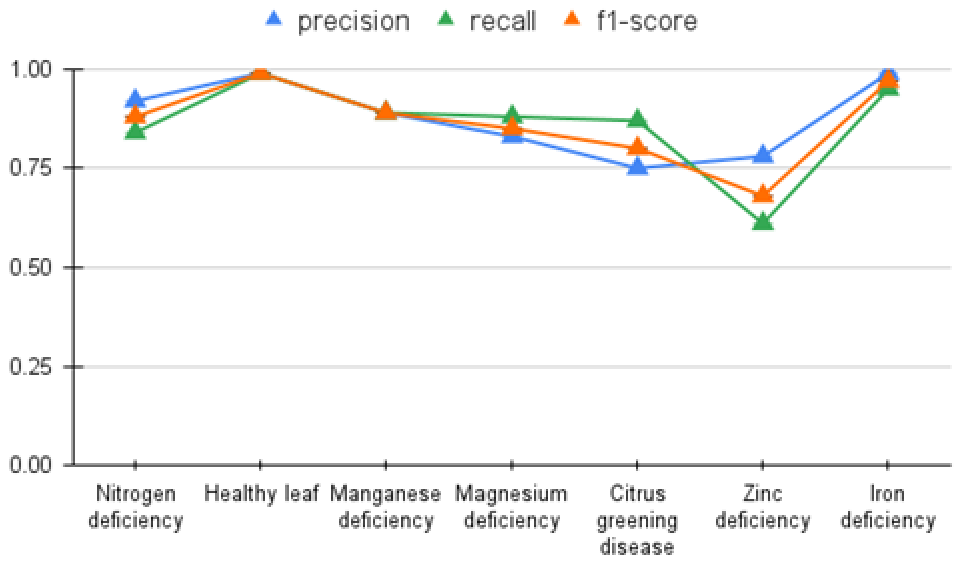

Fig 10. Classes wise accuracy for precision, recall and F1-score by using proposed method architecture on deep sweet orange dataset

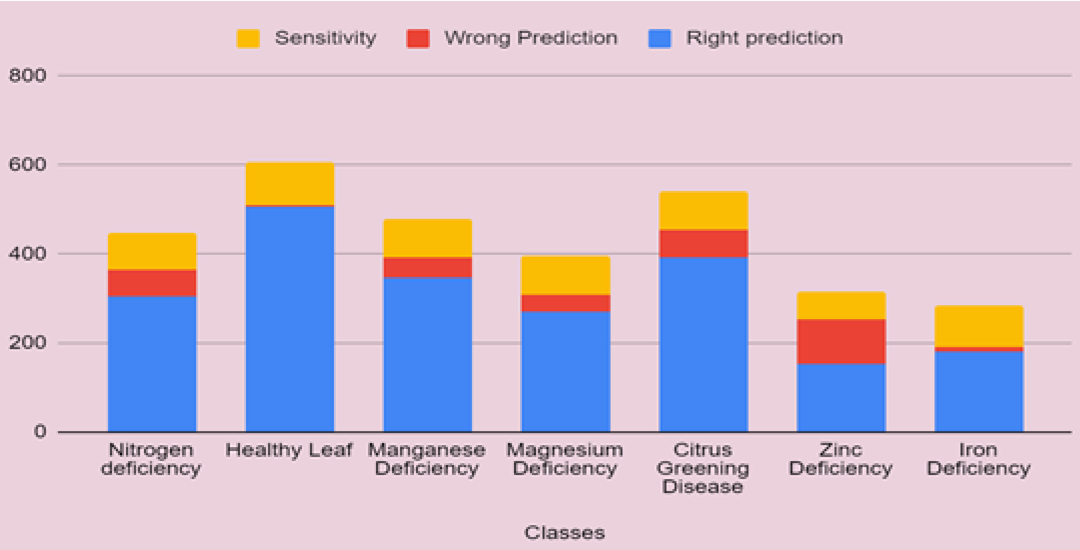

Fig 11. Classes wise accuracy for sensitivity, wrong prediction and right prediction by using proposed method architecture on deep sweet orange dataset

Table 2. Classes wise accuracy for precision, recall and F1-score by using proposed method architecture on CGD and nutritional deficiencies of deep sweet orange dataset

Classes	Precision	Recall	F1-score	Support
N deficiency	92	84	88	365
Healthy Leaf	99	99	99	508
Mn Deficiency	89	89	89	391
Mg Deficiency	83	88	85	308
CGD	75	87	80	454
Zn Deficiency	78	61	68	254
Fe Deficiency	99	95	97	190
Accuracy	-	-	**87**	**2470**
Macro Average	**88**	**86**	**87**	**2470**

Where, precision= correct classify image, recall= false positive image, F1 score= harmonic mean (average) of the precision and recall, support= total sample

Table 3. classes wise accuracy for sensitivity, wrong prediction (WP) and right prediction (RP) by using proposal method architecture on CGD and nutritional deficiencies of deep sweet orange dataset.

Classes	Total sample	RP	WP	Sensitivity
N deficiency	365	307	58	84
Healthy Leaf	508	505	03	99
Mn Deficiency	391	347	44	89
Mg Deficiency	308	271	37	88
CGD	454	393	61	87
Zn Deficiency	254	154	100	61
Fe Deficiency	190	181	09	95

Conclusion

The conclusions from studies on various aspects during research work entitled as ‘Detection through AI in CGD and various FND’s of sweet orange’ are presented below in brief. CGD of sweet orange is one of the most destructive diseases caused by *Candidatus liberibacter* and responsible for quality and quantity yield losses in almost orchards belongs to Jalna district of Marathwada region. FND’s are on the rise and are likely to continue with their negative impact affecting citrus growth, yield, quality and identifications of these nutrient deficiencies in crop are confused with biotic diseases of sweet orange crop.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

References

1. Ahlawat, Y. S. (1997). Viruses, greening, bacterium and viroids associated with citrus (*Citrus* spp.) decline in India. *Indian J. Agril. Sci*., 67: 51-57.
2. Denis, P. R., Loyani, K. L., Mgaya, R., Sawahiko, S. (2020). A Deep Learning Approach for Determining Effects of *Tuta Absoluta* in Tomato Plants. ArXiv preprintarXiv:2004.04023.
3. Fuentes, A., Yoon, D. H. and Park, D. S. (2017). Spectral Analysis of CNN for Tomato Disease Identification. In: Rutkowski L., Korytkowski M., Scherer R., Tadeusiewicz R., Zadeh L., Zurada J. (eds) Artificial Intelligence and Soft Computing. ICAISC 2017. Lecture Notes in Computer Science, vol 10245. Springer, Cham. https://doi.org/10.1007/978-3319 59063-94.
4. Garnier, M. and Bove J. M. (1984). The greening organism is a gram-negative bacterium, *Proc. Con. Int Org. Citrus Virol*, 9th, pp, 115-124.
5. Irudayaraj, J. (2009). *Pathogen Sensors*. 9, 8610–8612.
6. Jagoueix, S., Bove, J. M. and Garnier, M. (1994). The phloem limited bacterium of greening disease of citrus is a member of a subdivision of the Proteobacteria. *Current Microbiology.* 44:379-386.
7. Reinking, O. A. (1919). Diseases of economic plants in southern China. *Philipp Agric* 8: 10935.
8. Selvaraj, M. G., Vergara, A. and Ruiz, H. (2019). AI-powered banana diseases and pest detection. *Plant Methods* 15, 92 (2019). https://doi.org/10.1186/s1 3007-019 0475z.
9. Srivastava, A. K., Shyam, S., Diware, V. S. and Hamandeep, S. (2009). Site-specific nutrient management in 'Mosambi' sweet orange, Better Crops- India: pp. 10-11
10. Pezantes-Orellana C, German Bermúdez F, Matías De la Cruz C, Montalvo JL, Orellana-Manzano A. Essential oils: A systematic review on revolutionizing health, nutrition, and omics for optimal well-being. Frontiers in Medicine. 2024 Feb 16;11:1337785.
11. Munir S, Ahmed A, Li Y, He P, Singh BK, He P, Li X, Asad S, Wu Y, He Y. The hidden treasures of citrus: finding Huanglongbing cure where it was lost. Critical Reviews in Biotechnology. 2022 May 19;42(4):634-49.
12. Riaz S, Ahmad A, Farooq R, Hussain N, Riaz T, Hussain K, Mazahir M. Citrus: an overview of food uses and health benefits. Citrus Research-Horticultural and Human Health Aspects. 2022 Nov 14.