Efficiency Evaluation of CNN and Vision Transformer Architectures in Leaf Image Analysis

[*]Prakasam S, [**]Mohan T S and [***]Shanmugapriya P

[*]Associate Professor, Department of Computer Science and Applications,
Email: sp@kanchiuniv.ac.in
[**]Research Scholar, Department of Computer Science and Applications,
631561 – India. Email: ts.mohan87@gmail.com
[***]Associate Professor, Department of Computer Science and Engineering,
Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram, Tamil Nadu,
631561 – India. Email: pshanmugapriya@kanchiuniv.ac.in

Abstract

In agriculture sector maintaining and improving the leaf health is having important role. This task is obtained by using deep learning techniques and it provides more no of models for analyzing the leaf's patterns and also it extracts the features from image based leaf data sets. There are two models such as Convolutional Neural Network and Vision Transformation (ViTs) are frequently used in analyzing the leaf disease patterns and leaf classifications. This paper provides an efficient evaluation of both models with same datasets of different classes. These two model's performance was calculated by using performance metrics such as accuracy, recision, recall, F1-score and complexity. Through this evaluation the unknown fact about the models has been framed and the analyses for knowing better performance as well as to find disease patterns form various classes of images. Finally based on the model's performance decision were made.

Keywords: Convolutional Neural Networks (CNN), Vision Transformers (ViTs), Deep Learning, F1-score, Apple Leaf Disease Detection, etc,.

1. Introduction:

The computer vision and deep learning techniques are transforming the leaf image-based analysis in farming business [15]. Indian agricultural survey indicates that agricultural growth in India has been around 6.9% annually for the past few years [1]. To identify the disease farmers will look the plants thoroughly to finding diseases. This can be leads to misleading of findings, takes a long time and may not see sicknesses of the leaf early enough before it spread to other plant [2], [3]. In modern technology especially in computer vision and deep learning, we use a lot of focus on creating automatic systems to find tomato plant diseases in advance. Disease detection mechanism use image-processing methods and computer programs to learn images of apple leaf. Manual way of Classifying diseases in apple leaves is hard., takes a lot of time and work, especially on big farmlands where many leaves need to checking [4]. We need special skills to getting the right disease diagnosis for plants. Sometimes the shortage of skilled people finds the wrong diagnosis and late detection it may leads to loss. Identifying the symptoms of a disease might not show up until the later part of being infected [5].

Large pictures can be collected and worked on to affected leaves that might not be seen by our eyes. Some deep learning technologies give a better solution to find leaf diseases. For proper finding of diseases, we can categorize the images according to disease level. The accuracy and learning rate of our technologies using some measurement standards.

In recent trends Transformers become more powerful techniques for making leaf image analysis [7]. It gives scalability and computational efficiency for the models and it will be trained well with millions of parameters. The transformers are applied to leaf disease detection could also be beneficial for image classification tasks. To perform image classification [6], the proposal called Vision Transformers (ViT) consists of splitting the images into 2D pieces and providing this linear sequence of pieces as input (Fig:1) to the model.

R. Vasanth Kumar Mehta et al. (eds.), *Proceedings of the International Conference on Intelligent Systems for a Sustainable Future (ISSF 2026)*, Atlantis Highlights in Intelligent Systems 16,
https://doi.org/10.2991/978-94-6239-693-7_53

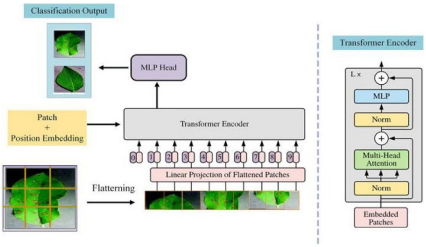

Fig:1. The architecture of Vision Transformer Model for image classification *(Dosovitskiy et al., 2021; Vaswani et al., 2017) [7][13].*

The deep learning architecture, there is another tool for processing large volumes of image datasets called Convolutional Neural Networks (CNN). The CNN is having multiple layers and has proved good performance in various computer vision tasks [14], such as image classification and NLP problems. The CNN architecture generates the hidden features from the normal leaf images. To achieve this functionality an image dataset can be passed into first layer which extracts low level features and likewise there is an intermediate layer which extracts some unknown image patterns. Finally, in the last layer all hidden layer have been identified and activation functions can be applied to get effective patterns. Finally, the fully-connected layers provide the flattened output of the convolutional and pooling layers and perform the classification.

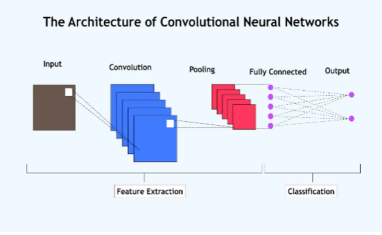

Fig:2. CNN (https://www.upgrad.com/blog/basic-cnn-architecture/)

The proposed technology has been structured and in the first, it takes deep look of existing models (Section 2). Then the explanation of the proposed method has been given in (Section 3). Section 4 discusses the results and Section concludes the work.

2. Literature Review:

In the [8] Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN), there were 10 different types of tomato leaf was used. It has the learning rate of 0.015 and its loss is 0.032. This model has a small loss of 0.032. It identifies a good pattern in the data sets and improves the right way to classify diseases on tomato leaves. It gives a good tool for farmers and experts in plant growth. It checks to find different ddiseases in tomato plants quickly and easily.

The key contributions of Flowering Phase Classification [9] of Tiliacordata Mill takes using CNN and Transformer Architectures are: (i) it introduces an automated exposure-quality filtering stage for field-acquired image data set (ii) it releases a season-long dataset of *Tilia cordata* images; and (iii) it cross-validates standard comparing classical CNNs and transformer architectures for image recognition.

In [10], it introduces Plant-CNN-ViT ensemble model and improves accuracy level of plant classification by training models on large-scale datasets. It is mixing the different pre-trained architectures together. To enhance the

performance of the classification with small sized datasets, both the CNN and ViTs were combined together. There are some drawbacks on analyzing the disease patterns in CNN and ViTs. The disease patterns of different deep learning were not clearly explained. To get the clear idea of the patterns we propose this model.

The proposed method fulfills the following research gap

- Conducting the systematic result comparison of performance of both CNN and Vision Transformer architectures using the same dataset.
- Determining the qualitative performance metrics
- Providing the recommendations practically for how to choose the models based on the image dataset

3. Experimental Evaluation

This section explains the database and the methodology applied to the prediction of tomato leaf disease.

3.1. Dataset Description and preprocessing:

The dataset is collected from the open-source platform, such as https://china.scidb.cn/ carefully taken to help to identify apple leaf diseases. The set of data used in this paper came from a large collection thatwas thoroughly split into two sets and each contains 5 different classes. So, it is possible to offer a thorough evaluation of the CNN model to build for this work. Ensuring that the model could effectively learn the special characteristics of everydisease required the balanced distribution of images across all classes, hence boosting its power toprecisely classify and diagnose apple leaf illnesses. Depending on the use of this well-organized andlabeled dataset, reliable and practical results follow from which early detection helps to more effectively control apple leaf diseases in agricultural operations.

(a). Alternaria leaf spot

(b). Brown spot

(c). Gray spot

(d). Healthy leaf

(e). Rust

Fig:3. Dataset images for 5 different Classes

3.2. Methodology:

This model has been implemented with CNN and Vision transformation model. At first, we collecting a group of apple leaf images, including many Alternaria leaf spot, Brown spot, Gray spot, Healthy leaf and Rust. This model is

helps to find the different types of diseases for various situations. Keeping all the images in same size and type before uploading them into the model.

Use cropping process, rotations, flipping and making zoom in or zoom out options to add more data [12]. This process makes the model stronger. Keep same size of all images so the model learns very well during training. splitting the data into test data set and training data of 5 different classes. With the fixed size image data set the Vision Transformer model is implemented. These images are modified into a lower-dimensional using a dense layer. While training the model the layer normalization is applied. Learning stability has been enhanced in Gradient flow. Through the Softmax layer the aggregated patch level has been passed. The feature extraction is constructed and spatial dimensionality is reduced in CNN model's Convolutional layer. The leaf image properties such as edges, textures, other patterns were clearly captured. At last, the combined patch level has been passed to a SoftMax layer. In convolutional layer the feature extraction is constructed. The evaluation metrics generated and the classification report for each 5 classes were calculated. This model is able to produce misclassification patterns according to class-wise manner.

4. RESULTS

This section provides the results of the proposed model for apple leaf disease detection with visual effects. Detecting Apple leaf diseases, the CNN model completed well with 92% and ViT model completed well with 67% of general accuracy. The confusion matrix is finalizing the performance of all 5 different classes. It is an optimal tool to get accurate performance of apple leaf plants.

4.1. Classification Report Analysis:

The degree of plant disease accuracy level is measured by the classification report. In the image analysis processing, the Neural Network is having important role. Most of the deep learning techniques provide more than 93% of accuracy.

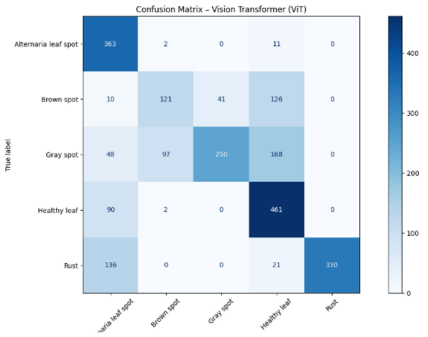

Fig.4: Classification Report-ViT model

4.2. Analysis of Confusion Matrix

The confusion matrix analysis exposes a clear performance difference between the CNN and Vision Transformer (ViTs) models in multi-class leaf disease classification. The CNN model proves strong diagonal dominance in its confusion matrix representing a higher number of correctly classified image patterns across all five classes and misclassifications are minimal

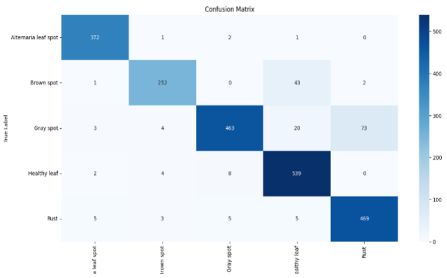

Fig. 5: Classification Report-CNN model

The true positive rates across 5 classes approve the CNN's robustness and reliable generalization capability on the available dataset.

The ViTs's confusion matrix displays weaker diagonal concentration and a higher number of off-diagonal elements, indicating increased misclassification among several classes. Visually similar disease categories are often confused in ViTs. This systematic analysis confirms the classification accuracy and class wise reliability. The CNN model gives better performance when comparing with ViTs model.

Class	Model	Precision (%)	Recall (%)	F1-Score (%)
Alternaria leaf spot	CNN	95.4	93.8	94.6
	ViT	99.1	99.2	99.1
Brown spot	CNN	76.2	71.8	73.9
	ViT	84	83.6	83.8
Gray spot	CNN	78.5	74.2	76.3
	ViT	93.8	93.6	93.7
Healthy leaf	CNN	88.9	85.6	87.2
	ViT	97.1	96.9	97
Rust	CNN	79.4	75.8	77.6
	ViT	85.1	84.6	84.8

Table 1: Confusion Matrix

4.3. Analysis of Accuracy Report:

During the learning behavior of CNN and Vision Transformer models, the CNN shows faster convergence and consistently higher accuracy across epochs, Fig. whereas the ViT demonstrates gradual improvement, indicating its dependency on larger data representations for effective learning.

5. Discussion:

This section provides the discussion of comparative evaluation of Convolutional Neural Networks (CNNs) and Vision Transformer (ViTs) architectures for multi-class plant leaf disease classification using the same image dataset. This work narrates the clear difference between the two models. In CNN, the higher classification accuracy, precision, recall, F1-score and class-wise performance were obtained across all 5 different classes of leaf. Confusion matrix is providing clear idea for classification and analysis. In the *Healthy leaf* and *Rust* categories, the CNN confusion matrix indicates effective feature learning and reduced misclassification.

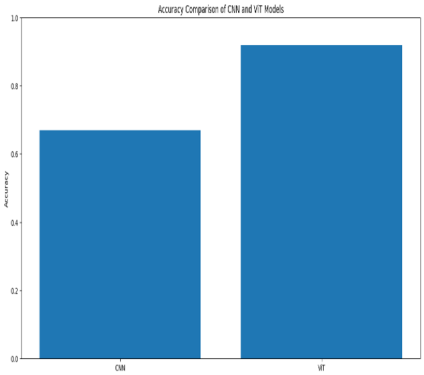

Fig.6: Accuracy comparison of CNN and ViTs

In contrast, the Vision Transformer shows higher inter-class confusion in *Brown spot* and *Gray spot*. Due to the limited amount of image dataset ViTs is ability to express the long dependencies. It slightly affects the lower recall and overall accuracy. This observation explore that the ViTs model needed larger datasets to perform more than CNN.

6. Conclusion and Future Work

A systematic comparison of CNN and Vision Transformer models was conducted for leaf disease detection using identical datasets, preprocessing techniques, and evaluation metrics. The results clearly indicate that the CNN model outperforms the Vision Transformer. Achieving an accuracy of approximately 92.01% compared to 66.97% for the ViTs because the image data set is low in its size. Confusion matrix analysis confirms that the CNN is producing fewer misclassifications and more reliable predictions across all disease categories.

This invention recommends that the CNN model stays a high efficient and good model for leaf disease classification especially on limited dataset. Because of the modern techniques in the today's trend, Vision Transformation gives another alternative mechanism. s Due to the limited amount of image dataset ViTs is ability to express the long dependencies. It slightly affects the lower recall and overall accuracy. This observation discovers that the ViTs model needed larger datasets to perform more than CNN. The overall performance of this model is that the ViTs predicting the same dataset with less time. At last ViTs can be chosen for large dataset with time consuming process and CNN can be chosen for small dataset and it takes more time to predict. This work influences the leaf severity index measurement using ViTs with large dataset

References

1. Dr. Prakasam S, Mr. Mohan T S and Dr. Shanmugapriya P; et al. Advancing Apple Leaf Health: A CNN Approach to Leaf Disease Detection, INDICA JOURNAL (ISSN:0019-686X) VOLUME 6 ISSUE 10 2025.

2. M. E. H. Chowdhury et al., "Automatic and Reliable Leaf Disease Detection Using Deep Learning Techniques," Agri Engineering, vol. 3, no. 2, pp. 294–312, May2021,doi:10.3390/agriengineering3020020.

3. IEEE Staff, 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART). IEEE, 2019. [Online]. Available: https://play.google.com/store/books/details?id=SGWWzQEACAAJ.

4. M. Sardogan, A. Tuncer, and Y. Ozen, “Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm,” in 2018 3rd International Conference on Computer Science and Engineering (UBMK), IEEE, Sep. 2018, pp. 382-385.doi:10.1109/UBMK.2018.8566635.

5. R. G. de Luna, E. P. Dadios, and A. A. Bandala, “Automated Image Capturing System for Deep Learning-based Tomato Plant Leaf Disease Detection and Recognition,” in TENCON 2018 - 2018 IEEE Region 10 Conference, IEEE, Oct. 2018,pp.1414to1419.doi:10.1109/TENCON.2018.8650088.

6. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer,M.; Heigold, G.; Gelly, S.; et al. An Image IsWorth 16x16Words: Transformers for Image Recognition at Scale. arXiv 2020, arXiv:2010.11929.

7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł.,Polosukhin, I., 2017. Attention is All you Need, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

8. Arshleen Kaur, Vinay Kukreja, Mukesh Kumar, Ankur Choudhary and Rishabh Sharma, “Multiclass Classification of Tomato LeafDiseases using a Hybrid of LSTM and CNN Model” 2024 IEEE 9th International Conference for Convergence in Technology (I2CT) Pune, India. Apr 5-7, 2024.

9. Bogdan Arct, Bartosz Swiderski , Monika A. Rozanska , Bogdan H. Chojnicki ,Tomasz Wojciechowski , GniewkoNiedbała ,Michał Kruk , Krzysztof Bobran and Jarosław Kurek, “ Comparative Evaluation of CNN and Transformer Architectures for Flowering Phase Classification of Tilia cordata Mill. With Automated Image Quality-Filtering”, https://www.mdpi.com/journal/sensors. 2025, 25(17),5326;https://doi.org/10.3390/s25175326.

10. Chin Poo Lee, Kian Ming Lim, Yu Xuan Song and Ali Alqahtani, “ Plant-CNN-ViT: Plant Classification with Ensemble of Convolutional Neural Networks and Vision Transformer”, Plants 2023, 12, 2642. https://doi.org/10.3390/plants12142642, https://www.mdpi.com/journal/plants.

11. ImaneBouchnafa and Mohamed Amnai, “Efficiency Analysis of CNNs and Vision Transformers for Edge-Based- Plant Disease Detection”, 2025 8th International Conference on Advanced Communication Technologies and Networking (CommNet)

12. S. Sharma, A. Kataria, and J. K. Sandhu, “Applications, Tools and Technologies of Robotic Process Automation in Various Industries,” in 2022 International Conference on Decision Aid Sciences and Applications (DASA), Mar. 2022, pp. 1067–1072.doi:10.1109/DASA54658.2022.9765027.

13. Jose Mauricio, Ines Domingues and Jorge Bernardino, “Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review”, Applied Science, 2023, *Appl.Sci.* 2023, *13*(9),5521; https://doi.org/10.3390/app13095521.

14. Harisu Abdullahi Shehu, Aniebietabasi Ackley, Marvellous Mark, Ofem Ebriba Eteng, Artificial intelligence for early detection and management of Tuta absoluta-induced tomato leaf diseases: A systematic review“,European Journal of Agronomy Volume 170, September 2025, 127669.

15. G. Dhanya, A. Subeesh,N.L. Kushwaha, Dinesh Kumar , Vishwakarma, T. Nagesh Kumar e, G. Ritika c, A.N. Singh, “Deep learning based computer vision approaches for smart agricultural applications”, Artificial Intelligence in Agriculture Volume 6, 2022, Pages 211-229