Deep Learning Based Techniques for Corn Plant Disease Detection Using UAV Imagery

Sai Varun Nimmagadda, Tej Mahanth Jammula, Chandra Naga Sai Manikanta Kona, Sasank Sonti, V Sateeshkrishna Dhuli, *Member, IEEE*, and Vineela Chandra Dodda, *Member, IEEE.*

***Abstract*—Plant diseases have become a global concern as they pose a significant threat to food security. These diseases have the potential to cause damage to crops, reduce yields, and compromise food quality. Moreover, the rapid spread of fungi, bacteria, and viruses can lead to widespread crop failures, creating food shortages, price increases, and ultimately, the risk of hunger. However, identifying plant diseases at the early stages remains challenging and time-consuming. Traditional methods of disease identification are often time-consuming and require manual inspection, which slows down timely actions and interventions. In recent years, deep learning-based techniques have shown promising outcomes in the realm of plant disease detection. In this paper, we develop three deep-learning models for the identification of healthy and unhealthy corn plant leaves. The models include a fully convolutional auto-encoder model, and a vision transformer, along with a baseline CNN (Convolutional Neural Network) model. We acquired the data set of corn plant images from UAV drones which were used to train and validate the proposed models. The results indicate that the proposed fully convolutional auto-encoder model achieved an accuracy of 95.09% outperforming both the Vision Transformer and CNN models. By leveraging deep learning-based techniques, such as the fully convolutional auto-encoder model, the agricultural industry can benefit from improved disease detection and prevention. Early and accurate identification of plant diseases allows for timely interventions, such as crop protection measures, which can ultimately safeguard food security.**

***Index Terms*—Deep learning, Plant disease, Convolutional neural networks, Vision Transformer, fully convolutional auto-encoder, UAV.**

I. Introduction

Corn is an important crop in India, serving food and raw materials for various companies in the food industry. However, corn fields are prone to various diseases, which can lead to a decrease in the outcome of the crop. Early detection of these diseases can increase our chances for the effective management and prevention of their spread. A fungal disease caused by Bipolaris maydis called Southern corn leaf blight affects corn plants [1]. The telltale signs of this disease are lesions on leaves, which hamper photosynthesis and hinder growth, ultimately reducing crop yield [2]. Southern corn leaf blight is a significant concern for corn growers worldwide, including India, where it caused a devastating epidemic in the 1970s. The precise impact of Southern corn leaf blight on corn plants remains unclear, and further research in this area could improve our understanding of its effect on both yield and quality. Recent advancements in deep learning-based methods show great potential for automating plant disease identification [3]. Convolutional Neural Networks (CNNs) represent a burgeoning field of research that has garnered significant attention in recent years. The study of CNNs in plant disease detection focuses on developing and implementing various deep learning algorithms to automate disease detection. Unlike traditional methods, these deep learning techniques can automate the detection process and achieve high accuracy in identifying diverse plant ailments, encompassing fungal infections, viral diseases, and nutrient deficiencies. The reviewed literature delves into topics like the creation of CNN architectures for plant disease detection, the development and annotation of image datasets for training and validation purposes, and the evaluation of CNN performance metrics such as accuracy, precision, recall, sensitivity, and specificity. With technological advancements, the integration of deep learning techniques with Unmanned Aerial Vehicles (UAVs) has garnered considerable interest for plant disease detection [4]..

Several studies [5]–[7] emphasize the potential of deep learning models, particularly those utilizing Convolutional Neural Network (CNN) architectures, for achieving accurate plant disease detection. The marriage of UAVs and deep learning techniques offers a promising approach for non-invasive and efficient disease monitoring in corn crops, leading to improved agricultural practices and crop management, as highlighted in [8]. The application of neural networks on UAV imagery [9] showcases the efficacy of deep learning techniques in plant disease detection. This research underscores the significance of dataset creation, robust feature extraction, and classification methods for accurate plant identification. Although it does not directly address disease detection, the study underscores the significance of these techniques in accurately identifying plants using UAV imagery. As mentioned in [8], [10] with the aid of UAV photos, it proposes a deep learning-based method for automatically identifying soybean leaf illnesses, illustrating the potency and utility of deep convolutional neural networks in disease identification. In order to achieve high accuracy in illness classification, it emphasizes the importance of large-scale datasets, dataset quality, transfer learning, and

Sai Varun Nimmagadda, Tej Mahanth Jammula, Chandra Naga Sai Manikanta Kona, Sasank Sonti, Department of Electronics and Communication Engineering, SRM University, Amaravathi, Andhra Pradesh-522502, India.

V Sateeshkrishna Dhuli, Assistant Professor, Department of Electronics and Communication Engineering, SRM University, Amaravathi, Andhra Pradesh-522502, India. e-mail: (sateeshkrishna.d@srmap.edu.in).

Vineela Chandra Dodda, Assistant Professor, Department of Electronics and Communication Engineering, Amrita School of Engineering Amaravati, Amrita Vishwa Vidyapeetham, Andhra Pradesh, India.
e-mail: (c_vineela@av.amrita.edu).

model evaluation metrics.

Later, the authors in [11] proposed a methodology for calculating the winter wheat's leaf area index (LAI) using UAV data and CNNs. While it does not specifically focus on disease detection as mentioned in [6], it highlights the potential of CNNs and UAV imagery in quantifying crop growth parameters. This passage highlights the importance of image preprocessing and CNN architecture design for accurate Leaf Area Index (LAI) estimation. While not directly related to disease detection, the study in [12] proposes a fully convolutional auto-encoder for change detection in UAV imagery, demonstrating the effectiveness of deep learning in identifying variations in crop growth, which can be indicative of disease presence. Furthermore, the successful application of convolutional auto-encoders on hyperspectral data for predicting canopy-averaged chlorophyll content in pear trees [13] offers valuable insights for optimizing feature extraction techniques in corn disease detection using drone images. Even the effectiveness of using deep convolutional auto-encoders to learn representative features [7], can contribute to improved accuracy and robustness in corn plant disease detection. A CNN-based approach and methodology for accurately identifying soybean foliar diseases by using UAV images similar to [10] was mentioned in [8]. By taking insights into multi-scale image analysis from [6] this study emphasizes fine-tuning pre-trained models, employing multi-scale image analysis, and ensuring accurate labeling and dataset quality to improve disease recognition. Even the CNN-based methodology in [14] emphasizes the significance of large-scale datasets, transfer learning [9], and model optimization for improved disease classification. Unlike limiting for analysis on one particular disease as in [8] authors of [10] researched and worked on multiple diseases in the case of soybean leaf disease detection.

Deploying drones, collecting high-resolution images, and using UAV imagery along with deep learning techniques gave better results in counting the corn plants [5] in the field, identifying the plant diseases [15], and monitoring fields. The application of deep learning and UAV imagery for crop assessment can aid in accurate plant disease detection with better outcomes. However, challenges like occlusions from overlapping plants, variations in growth stages, and optimizing image analysis algorithms need to be addressed [5], [15]. Existing research on general plant diseases provides valuable insights for corn disease detection [10]. A multi-condition training approach for deep convolutional neural networks (CNNs) enhances disease detection robustness [6]. Techniques for handling variations in lighting and background conditions can also be adapted for corn disease detection [16]. Studies exploring deep convolutional autoencoders for feature extraction in SAR images offer valuable insights for corn disease detection, even though the data type is different [17]. Learning discriminative features can further improve detection capabilities [16]. The use of hyperspectral data combined with CNN processing has proven effective in enhancing disease detection precision [18]. This synergy holds promise for corn disease detection as well. Deep CNN architectures trained on extensive datasets showcase the remarkable potential for achieving exceptional disease detection accuracy [19]. This knowledge can inform the development of similar CNN-based systems for corn. Our research leverages deep learning and UAV technology to develop an accurate and effective system for corn disease detection [9], [10]. This aligns with precision agriculture goals, empowering farmers with proactive disease management strategies to minimize yield losses and promote sustainable practices [11]. We focused on detecting Indian maize diseases. We created a dataset of healthy and diseased corn plant images captured by a drone and trained a deep learning model on it. This model achieved high disease detection accuracy and demonstrates potential for real-world application [20]. Our work contributes valuable information on deep learning for crop disease detection in India, highlighting its potential for improved crop productivity and sustainability. Our approach offers several advantages over traditional, human-reliant methods, which can be time-consuming, subjective, and error-prone [21]. UAVs provide a unique aerial perspective, enabling the capture of high-resolution images from various angles for greater crop coverage [22]. They can access hard-to-reach areas, facilitating more comprehensive crop health assessments. The adaptability and maneuverability of UAVs allow for on-demand data acquisition, reducing reliance on ground surveys and enabling real-time disease monitoring [23] Overall, UAV imagery improves the accuracy, efficiency, and scalability of plant disease detection, thereby enhancing crop management and yield. They offer valuable aerial perspectives for training models and analyzing data in various industries. Overall, the background study explains the potential applications and benefits of deep learning techniques in the area of plant disease detection. That also includes discussions of how automated detection can improve crop production, reduce chemicals and pesticide usage, and contribute to more sustainable agriculture practices [24]. In our study, we experimented with three different techniques for corn plant disease detection namely Convolutional Auto-encoder, Vision Transformer, and CNN baseline model. The Convolutional Auto-encoder demonstrated the highest accuracy and efficiency among the three models, making it the primary focus of our research. The Vision Transformer model exhibited moderate performance, while the CNN baseline model achieved the least accuracy among the three models.

The rest of the paper is structured as follows: Section II discusses the concepts and methodology, whereas Section III explains the proposed models' architecture, experimental results, and a comparative analysis of the proposed models' performance. Finally, the conclusion is given in Section IV.

II. Methodology

In this section, we discuss the pre-processing steps and methodology for disease detection in corn plants. Data pre-processing constitutes a crucial stage in the examination of the dataset acquired through drone technology to identify plant diseases. Our initial approach involves the elimination

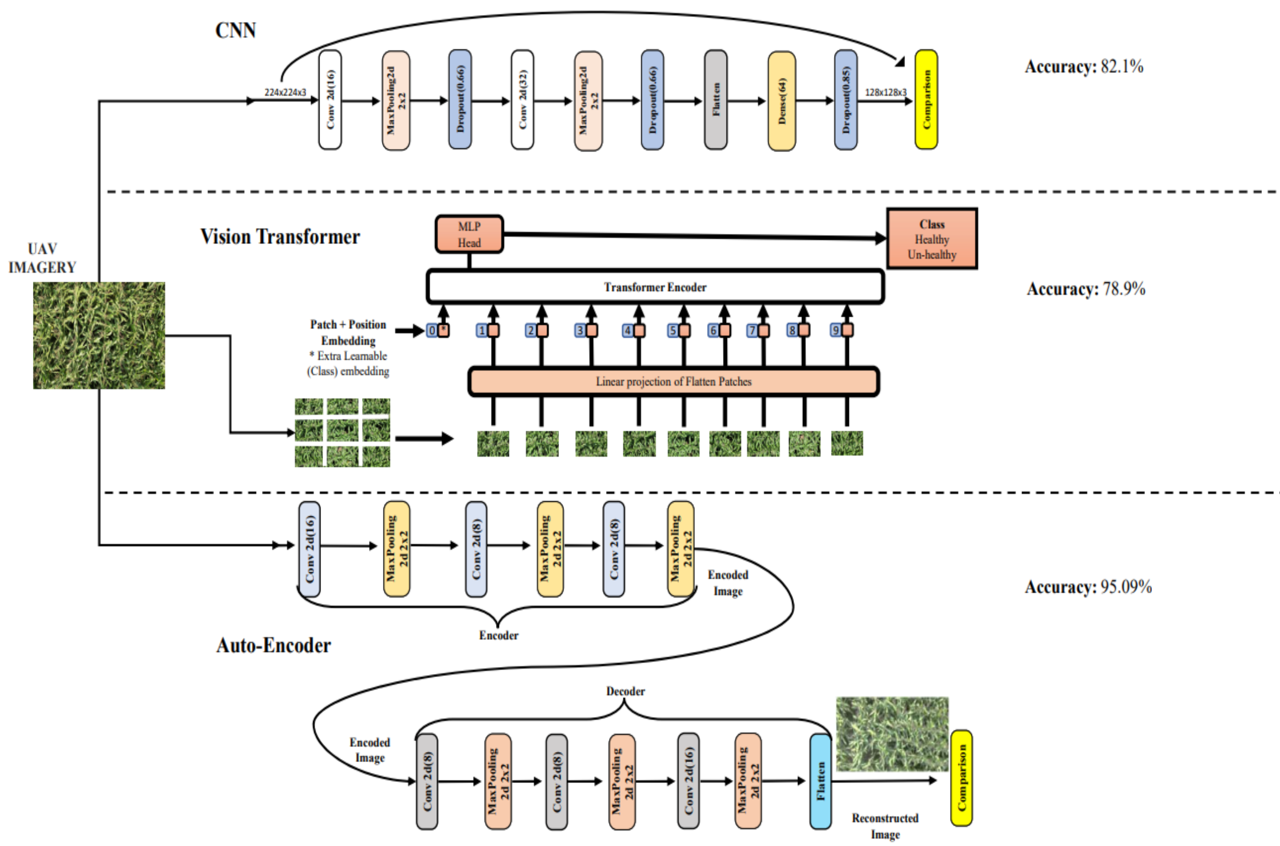


Fig. 1: Network Architecture

of redundant images marred by either being out of focus or exhibiting motion blur, as these instances have the potential to detrimentally influence the outcomes. Subsequently, the images transform to achieve the necessary format, involving a reduction in pixel density to the designated specifications. It's worth noting that all of these methods were applied to full-color images, each described by RGB metrics.

We used a DJI Mavic 3 Pro drone to take real-time pictures of wheat fields near SRM University Amaravati, India. Here are some key features of this drone:

- Lightweight: It weighs only 895 grams, making it easy to carry and fly.
- High-quality images: It has a good camera sensor and takes 20-megapixel photos in both JPEG and DNG formats.
- Great for videos too: It can record high-resolution videos in 5.1K or 4K, perfect for capturing details or slow-motion analysis.
- Long flights: A single battery charge lets it fly for up to 46 minutes, covering a large area.
- Stays connected: It can transmit data back to the controller from up to 15 kilometers away.
- Safe flyer: Sensors around the drone help it avoid obstacles in all directions.
- Storage and control: It has 8GB of storage for pictures and videos, and comes with a standard remote control.

This advanced drone allowed us to collect valuable real-time data on the wheat crops near the university. Its features make it a great tool for agricultural research and monitoring.

A. CNN Architecture

Imagine a powerful image analysis tool called a Convolutional Neural Network, or CNN for short. These CNNs are like assembly lines specifically designed to understand pictures. The first stop is the input layer, where the image gets fed in. Then comes the key step: the convolutional layer. Here, the image is scanned by a series of filters, like detectives looking for clues. These filters, also called kernels, search for important details like edges, shapes, and textures. The result is a "feature map" highlighting these visual fingerprints. Next, the pooling layer acts like a summarizer, condensing this information and keeping only the most critical parts. This helps the CNN focus on what truly matters. After that, the fully connected layer takes over, putting all the gathered information together. It's like a team of analysts working collaboratively to make sense of the clues. Finally, the output layer delivers the CNN's verdict, such as classifying a plant in the image as healthy or diseased. But the real magic lies in how CNNs learn. By analyzing massive datasets of labeled images (think healthy vs. diseased plants), CNNs can fine-tune their filters, becoming

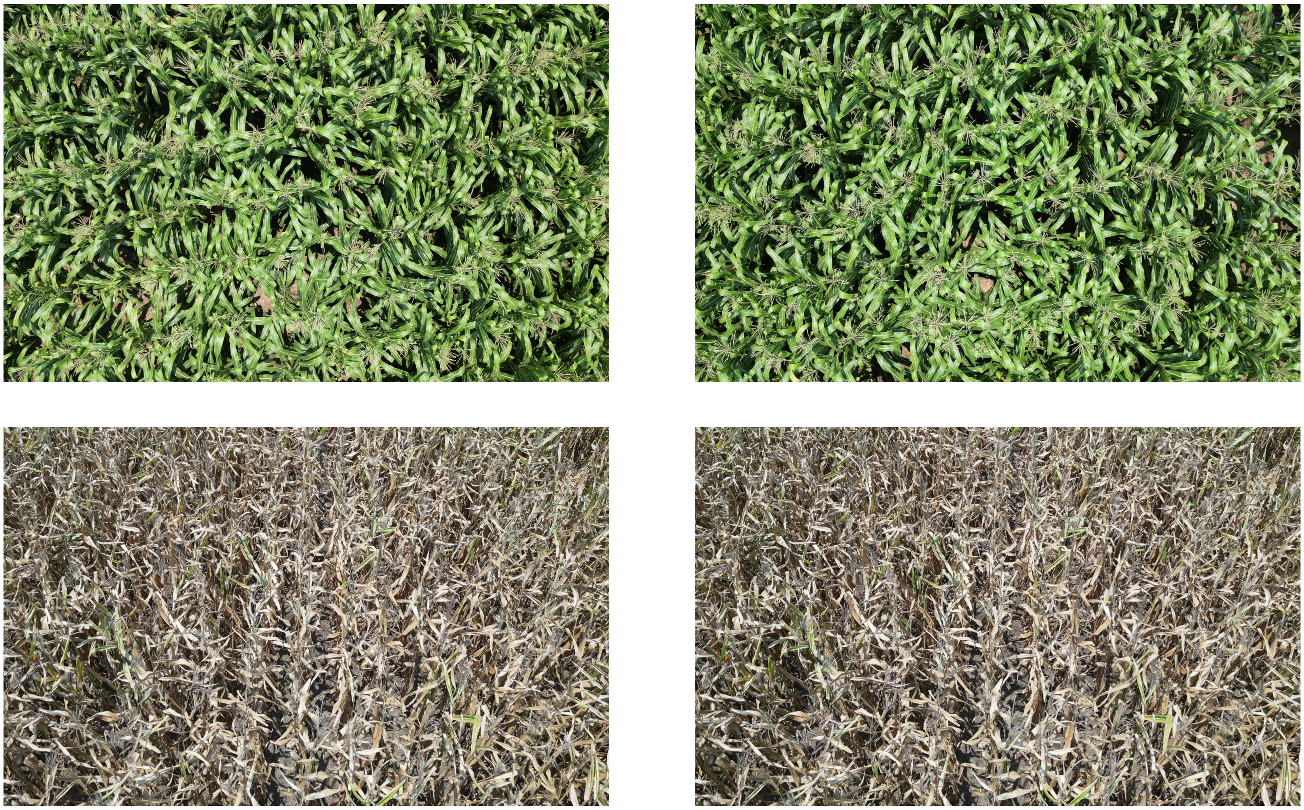


Fig. 2: Healthy and Unhealthy Plant images from Dataset

experts at recognizing the most distinguishing features. This learning ability allows CNNs to achieve impressive accuracy in tasks like identifying plant diseases.

$$H(I,J,K) = Activation(\sum(M,N,O)X(I+M,J+N,O) * W(M,N,O,K) + b(K)) \quad (1)$$

The convolutional layer is subsequently linked to a pooling layer, which operates to decrease the spatial dimensions of the feature map derived from the convolutional layer. This reduction preserves the most essential information, leading to the creation of a pooled feature map tensor. Mathematically, the pooling layer can be described as follows:

$$P(i,j,k) = max(H(i * s, j * s, k)) \quad (2)$$

Further, the final layer i.e., the fully connected layer consists of neurons, in which each neuron has its own weight and bias term. These neurons are utilized to extract a vector of logits representing the class probabilities from a flattened vector representing the features from its previous layers.

$$Z = F \cdot W + b \quad (3)$$

The heart of our image analysis system is a Convolutional Neural Network (CNN) model. Imagine this CNN as a series of processing steps. First, it receives an image as input, with specific dimensions for width, height, and color channels. Then comes the critical part: convolutional layers. These act like filters, scanning the image for important details like edges or shapes. Each filter calculates a score based on how well a specific pattern matches a small area of the image. An activation function, like a switch, then decides if this score is significant enough to keep. To avoid getting overwhelmed with too much data, pooling layers step in. They shrink down the information, keeping only the most important features. Our model uses two sets of these filtering and pooling steps, with a special technique called dropout added to prevent overfitting. Dropout randomly removes some information, forcing the model to learn more robust features. Finally, the information is flattened and fed into a fully connected layer. This layer acts like a team of analysts, putting all the clues together to make a final decision. In our case, this decision might be classifying a plant in the image as healthy or diseased. The key to the CNN's power lies in its ability to learn. By analyzing massive datasets of labeled images, the model fine-tunes its filters, becoming an expert at recognizing the most distinctive features of healthy and diseased plants. Finally, the output layer with softmax activation is added to perform the classification into two classes.

Fig.1 shows all the layers in the CNN architecture. The input layer takes the input image in size 224x224x3, which then undergoes multiple Convolution layers, Max pooling layers, and Dropout layers to extract necessary features for classification along with flattened and Dense Layers to avoid over-fitting. At last, the model will be compiled with input images as the input given to the model and output labeled as the output of the model with 'categorical-cross entropy' as a loss function.

B. Vision Transformer

Vision Transformer (ViT) is a deep-learning architecture. Unlike traditional convolution neural network(CNN) architectures, the vision transformer is a transfer-based model which is originally introduced for natural language processing tasks. ViT divides the input images into fixed-size patches and treats them like tokens, just like words in a sentence. These tokens are then fed into the transform encoder in a linear arrangement. The main idea behind the Vision Transformer is to leverage the power of the self-attention mechanisms to capture long-range dependencies and relation patterns between these image patches or tokens, as shown in Fig.1.

At the core of this system lies a powerful processor called a transformer encoder. This encoder is built with two key components working together. The first is self-attention, which allows the system to analyze the image not just piece by piece, but also how different parts of the image relate to each other, like looking at the big picture. The second component is a feed-forward network that takes the information gathered and refines it, making it more nuanced. The result of this combined effort is a detailed breakdown of the image, where each section is represented by a unique set of features. All these steps of the vision transformer are shown in Fig.1 for better understanding. In order to make predictions, a classification head is added on top of the transformer encoder. This head takes the final feature vector as input and produces class probabilities or regression outputs depending on the task at hand. During training, the ViT model is optimized using supervised learning techniques, such as cross-entropy loss. In our model, we utilized the pre-trained VGG-16 architecture to design the vision transformer and initialized the model with the weights of Image-net. We obtain a 4D tensor as output from the vision transformer and to convert it into a 2D tensor, a flattened layer is added after the VGG-16 layer along with a fully connected dense layer which consists of 28 units and a ReLU activation function. After all the image analysis steps, we're left with a simplified version of the image with its key features identified. But there's more to the story. To squeeze even more meaning out of this data, we use fully connected dense layers. These layers act like analysts, searching for deeper connections and patterns within the features. To prevent the model from simply memorizing specific examples instead of learning true patterns, we introduce a dropout layer. This layer randomly throws away some information, forcing the model to be more flexible and adaptable. Finally, the model needs to make a decision: is the plant in the image healthy or diseased? An output layer with two options and a special function called softmax helps translate the analyzed features into these two probabilities. To train the entire model effectively, we use a specific optimizer and a way to measure its performance, like checking its accuracy.

C. Auto Encoder

Imagine a special kind of CNN called an autoencoder. Unlike regular CNNs that focus on classifying things, autoencoders are more like artistic copiers. They have two parts: an encoder and a decoder. The encoder takes an image as input and shrinks it down to a smaller version, capturing only the most important details. Then the decoder takes this mini-image and tries to rebuild the original picture. The goal is for the decoder to be as accurate as possible, with minimal errors. By trying to perfectly copy the image, the autoencoder actually learns to recognize the key features and patterns that make the image what it is. This makes autoencoders great for tasks like feature extraction, where you want to identify the most important aspects of something, like a healthy leaf versus a diseased one.

We built a special kind of CNN called an autoencoder to tackle this task. Imagine it as a two-part machine for learning image features. The first part, the encoder, takes an image and shrinks it down, capturing only the crucial details. It uses a series of filters (convolutional layers) and pooling layers to achieve this. These filters act like spotlights, highlighting the important aspects of the image, while pooling layers condense the information. This compressed version, essentially the key features of the image, then gets fed into the second part, the decoder. The decoder's job is to use this information to rebuild the original image as accurately as possible. By trying to perfectly copy the image, the autoencoder actually learns to recognize the most important features and patterns. This makes it a powerful tool for tasks like feature extraction, where we want to identify the essence of something, like the difference between a healthy and diseased leaf. The beauty of this design is its symmetry. The encoder and decoder mirror each other, with the compressed feature layer acting as the bridge between them.

All of these steps explained above are represented in detail in Fig.1, which depicts that the encoder is designed using a set of Convolutional layers in which two layers consist of 8 neurons and one layer consists of 16 neurons. Along with a set of Max-pooling layers with (2X2) dimensions. The output of this encoder is a low dimensional encoded image which is further reconstructed by the decoder, whose design is similar to that of the encoder with a flattened layer added at the end of it. After the flattened layer we obtain the reconstructed image, which will be compared with the input image to calculate the reconstruction error of the model. The optimizer is set to Adam, and categorical cross-entropy is used as a loss function to measure the reconstruction loss, which should be as minimal as possible for the model to be efficient. Additionally, accuracy is chosen as a metric to evaluate the models' efficiency. Finally, the convolutional auto-encoder is trained using the fit function. The training data set is passed to the function, and the training of the model is carried out for a specified number of epochs (in this case, 4).

D. Metrics

1) ***Accuracy:*** One way to measure how well a classification model performs is by looking at its accuracy. This simply means how often the model makes the right prediction. Imagine you ask the model to identify 100 images of cats and dogs. If it correctly classifies 85 of them, its accuracy would be 85%. The higher the accuracy, the better the model is at distinguishing between different categories, like healthy and diseased plants in your case.:

$$Accuracy = \frac{T.P + T.N}{T.P + T.N + F.P + F.N} \tag{4}$$

*2) **Loss**:* Imagine you're training a model to tell the difference between healthy and diseased plants. The model makes predictions, but how do you know how well it's learning? This is where loss comes in. Loss is a way to measure the gap between the model's guesses (predictions) and the actual truth (healthy or diseased). It's like a score that tells you how much improvement is needed. There are different ways to calculate loss, depending on the task. In this case, we used two specific methods: binary cross-entropy (for situations with two options like healthy/diseased) and categorical cross-entropy (for problems with more than two categories). By minimizing this loss during training, we help the model learn from its mistakes and become better at making accurate predictions.

*3) **Binary Cross-entropy**:* When dealing with two categories, like healthy versus diseased plants, a specific loss function called Binary Cross Entropy helps us understand how well our model is learning. This function calculates the difference between the model's predictions (how likely it thinks something is healthy or diseased) and the actual truth (healthy or diseased). The lower this difference, or loss, the better the model is performing. We can think of it as a score that guides the model's training, helping it minimize its mistakes and improve its accuracy in distinguishing between the two categories.:

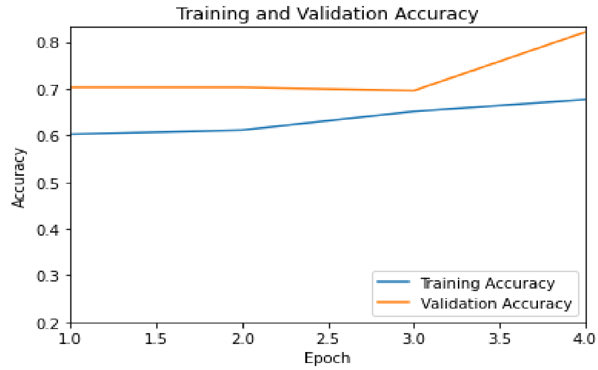
$$L_{\text{Binary Cross-entropy}} = -(y \log(p) + (1 + y) \log(1 - p)) \tag{5}$$

*4) **Categorical Cross-Entropy**:* Imagine you're training a model to classify not just healthy and diseased plants, but also maybe stressed or nutrient-deficient ones. In this case, with multiple categories, we use a different loss function called Categorical Cross Entropy. This function works similarly to Binary Cross Entropy, but instead of comparing two options, it considers all the possibilities. It calculates the difference between the model's predictions (how likely it thinks something is healthy, diseased, stressed, etc.) and the actual truth (the real category of the plant). The lower this difference, the better the model is performing. Categorical Cross Entropy acts like a guide during training, helping the model minimize its mistakes and improve its accuracy in recognizing all the different plant conditions.:

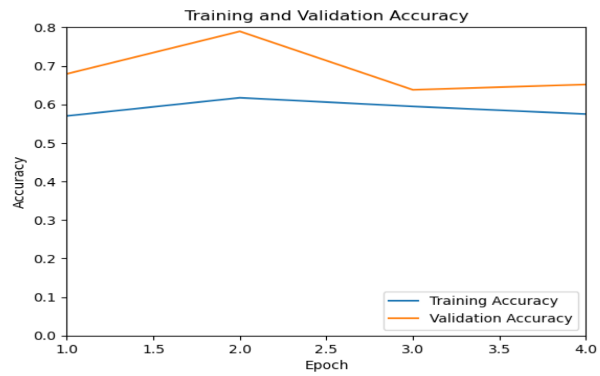
$$L_{\text{Categorical Cross-Entropy}} = -\sum_{i=1}^{N} y_i \log p(i) \tag{6}$$

III. Experimental Results

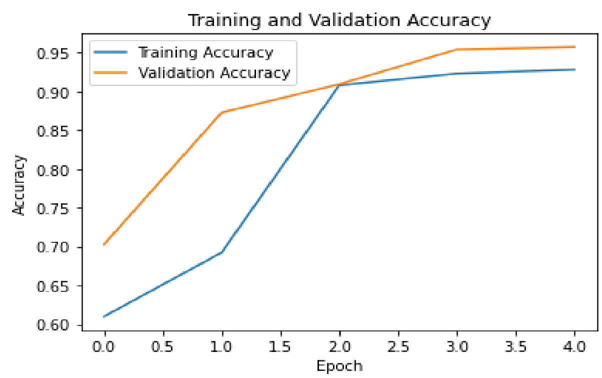
This section provides a comprehensive analysis and discussion of the results achieved using the proposed approach. Table I thoroughly examines and elaborates on the validation



(a) CNN Architecture



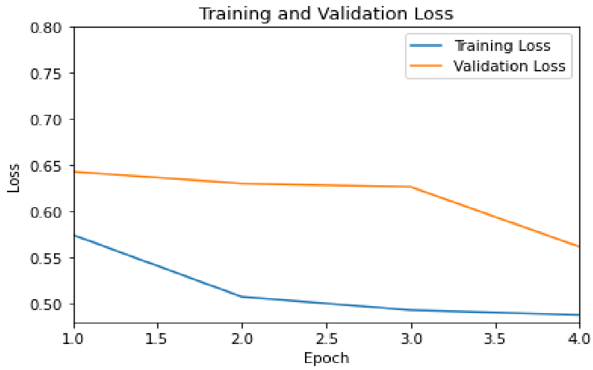
(b) Vision Transformer



(c) Auto Encoder

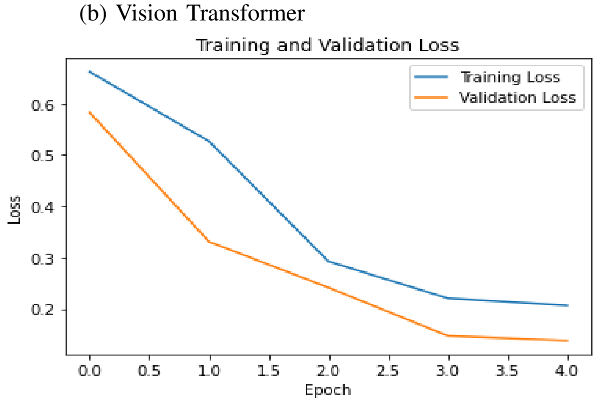
Fig. 3: Accuracy of CNN vs Vision Transformer vs Auto Encoder

outcomes attained through the proposed approaches. The proposed models were trained and validated using a dataset acquired by UAV imagery throughout four epochs. The proposed CNN model has attained its highest validation accuracy of 82.1% at a loss rate of 0.5613 in its ultimate epoch. Fig. 3(a) gives an overview of the training and validation accuracy of the convolution neural network model with the number of epochs on the x-axis and Accuracy on the y-axis as parameters. Fig. 4(a) gives an overview of training loss and validation Loss of the same model with the number of epochs on the x-axis and parameter loss on the y-axis as parameters. In Fig. 4(a) as we pass by the epochs, the decrease in Loss is evident at the 4th epoch which subsequently resembles an increase in accuracy which can be witnessed in Fig. 3(a) at epoch 4 and this is the



(a) CNN Architecture

(b) Vision Transformer



(c) Auto Encoder

Fig. 4: Loss Metric of CNN vs Vision Transformer vs Auto Encoder

highest accuracy attained by the model overall.

Similarly, the proposed Vision transformer has acquired a validation accuracy of 78.97% at a loss rate of 0.5629. Fig. 3(b) presents an overview of training and validation accuracy and Fig. 4(b) presents an overview of the training and validation loss of the vision transformer model.

The proposed fully convolutional auto-encoder model has achieved an accuracy of 95.09% at a Loss rate of 0.1442. Fig. 3(c) depicts the relation between the validation and training accuracy of the model and in the same way Fig. 4(c) depicts the correlation between loss values. Overall, the fully convolutional auto-encoder has achieved the desired results among all the proposed models with an accuracy of 95.09%.

TABLE I: Performance Metrics

	Metrics	epoch1	epoch2	epoch3	epoch4
CNN	Loss	0.6427	0.6299	0.6262	0.5613
	Accuracy	70.3%	70.3%	69.6%	82.1%
Vision Transformer	Loss	0.6133	0.5629	0.6299	0.6093
	Accuracy	67.9%	78.9%	63.7%	65.1%
Auto Encoder	Loss	0.5610	0.3397	0.2310	0.1442
	Accuracy	70.31%	91.52%	94.02%	95.09%

IV. Results & Discussions

Data acquisition plays a pivotal role as obtaining an accurate dataset holds paramount significance in the creation of an effective model. Diverging from the conventional method of gathering plant data through DSLRs or similar approaches, our research employs an Unarmed Aerial Vehicle (UAV) to obtain images of corn plants, unlike the traditional approach. The DJI Mavic 3 UAV was employed, boasting the capability to capture high-resolution 4k videos at a remarkable 120 fps. The drone was piloted at an altitude exceeding 40 meters above the field to record videos of the plants. These videos were captured from various angles, encompassing all possible directions, to ensure an extensive scope for the identification of plant diseases. In order to achieve comprehensive coverage of the field, we operated the drone in a zigzag pattern. Subsequently, a Python script was employed to transform the video footage into individual images (for instance, a few images shown in Fig.2). Each frame within the video was extracted, generating a dedicated image folder containing field images. Figure 2 visually presents the aerial view images of both categories: healthy and unhealthy plants, captured by the UAV.

Table I presents the performance metrics for all models, specifically the CNN, vision transformer, and fully convolutional auto-encoder model. Each of these models underwent training through 4 epochs. Throughout the training process of an auto-encoder, the binary cross-entropy loss function is employed. This function serves to gauge the extent of reconstruction error between the original input image and the generated reconstructed image. Typically, this loss function diminishes progressively with the advancement of each epoch. Initially, the auto-encoder might encounter challenges in accurately reconstructing the data, which consequently results in elevated loss values. However, as the training unfolds, the model refines its internal parameters utilizing methods such as backpropagation and gradient descent. This iterative refinement enhances the model's capability to identify pertinent features and generate more precise reconstructions. Consequently, the loss gradually declines, indicative of an increasingly accurate alignment between the input image and its reconstructed counterpart. The pace at which the loss diminishes can fluctuate, influenced by factors such as data intricacy, architectural configuration, optimization algorithms, and dataset dimensions. The vigilant monitoring of loss is imperative to ensure progression towards a desirable solution. This oversight facilitates the fully convolutional auto-encoder model in attaining a stable and desirable level of accuracy. The proposed fully convolutional auto encoder model has attained 95.09% accuracy at a loss of 0.1442 which can considered as a very low loss rate when compared within its concurrent epochs and also when compared with the CNN and vision transformer models. CNN model attained its best accuracy of 82.19% at a loss of 0.5613 whereas the vision transformer model attained its best accuracy of 78.97% at a loss of 0.5629. Fig. 3 and Fig. 4 give a complete overview of the accuracy and loss attained by the CNN, Vision Transformer, and Auto Encoder models respectively over their training and validation datasets.

Table II provides a representation of the achieved accuracy for each model. Notably, the fully convolutional auto-encoder has achieved an impressive accuracy of 95.09%, while the CNN and Vision Transformer models have achieved accuracy of 82.19% and 78.97%, respectively. The fully convolutional auto-encoder model has exhibited superior performance compared to the conventional CNN and Vision Transformer models.

TABLE II: Accuracy Measurement

Model	Accuracy
CNN	82.19%
Vision transformer	78.97%
Auto Encoder	95.09%

V. Conclusions

Results obtained by experimental analysis allow us to conclude that the implementation of deep Learning techniques in the agriculture sector has given good results in terms of disease detection. Based on our thorough analysis, we can deduce that the fully convolutional auto-encoder methodology presented in this paper has achieved a validation accuracy of 95.09%, surpassing all other deep learning techniques. Our study further suggests that implementing deep learning models for plant disease detection can mitigate the impact of diseases on crop yields, thereby promoting increased production.

References

[1] L. Picek, M. Šulc, J. Matas, J. Heilmann-Clausen, T. S. Jeppesen, and E. Lind, "Automatic fungi recognition: Deep learning meets mycology," *Sensors*, vol. 22, no. 2, p. 633, 2022.

[2] F. Li, Z. Liu, W. Shen, Y. Wang, Y. Wang, C. Ge, F. Sun, and P. Lan, "A remote sensing and airborne edge-computing based detection system for pine wilt disease," *IEEE Access*, vol. 9, pp. 66346–66360, 2021.

[3] K. Sornalakshmi, G. Sujatha, S. Sindhu, and D. Hemavathi, "A technical survey on deep learning and ai solutions for plant quality and health indicators monitoring in agriculture," in *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 984–988, 2022.

[4] F. Sabrina, S. Sohail, S. Thakur, S. Azad, and S. Wasimi, "Use of deep learning approach on uav imagery to detect mistletoe infestation," in *2020 IEEE Region 10 Symposium (TENSYMP)*, pp. 556–559, 2020.

[5] J. Kolli, D. M. Vamsi, and V. M. Manikandan, "Plant disease detection using convolutional neural network," in *2021 IEEE Bombay Section Signature Conference (IBSSC)*, pp. 1–6, 2021.

[6] R. S. Yuwana, E. Suryawati, V. Zilvan, A. Ramdan, H. F. Pardede, and F. Fauziah, "Multi-condition training on deep convolutional neural networks for robust plant diseases detection," in *2019 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, pp. 30–35, 2019.

[7] K. Trang, L. TonThat, and N. G. Minh Thao, "Plant leaf disease identification by deep convolutional autoencoder as a feature extraction approach," in *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pp. 522–526, 2020.

[8] E. Castelão Tetila, B. Brandoli Machado, N. A. Belete, D. A. Guimarães, and H. Pistori, "Identification of soybean foliar diseases using unmanned aerial vehicle images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2190–2194, 2017.

[9] Z. Fan, J. Lu, M. Gong, H. Xie, and E. D. Goodman, "Automatic tobacco plant detection in uav images via deep neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 876–887, 2018.

[10] E. C. Tetila, B. B. Machado, G. K. Menezes, A. Da Silva Oliveira, M. Alvarez, W. P. Amorim, N. A. De Souza Belete, G. G. Da Silva, and H. Pistori, "Automatic recognition of soybean leaf diseases using uav images and deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 5, pp. 903–907, 2020.

[11] L. Wittstruck, T. Jarmer, D. Trautz, and B. Waske, "Estimating lai from winter wheat using uav data and cnns," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[12] D. B. Mesquita, R. F. d. Santos, D. G. Macharet, M. F. M. Campos, and E. R. Nascimento, "Fully convolutional siamese autoencoder for change detection in uav aerial images," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 8, pp. 1455–1459, 2020.

[13] S. Paul, V. Poliyapram, N. İmamoğlu, K. Uto, R. Nakamura, and D. N. Kumar, "Canopy averaged chlorophyll content prediction of pear trees using convolutional autoencoder on hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1426–1437, 2020.

[14] B. T. Kitano, C. C. T. Mendes, A. R. Geus, H. C. Oliveira, and J. R. Souza, "Corn plant counting using deep learning and uav images," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2019.

[15] S. Rachmawati, A. Syah Putra, A. Priyatama, D. Parulian, D. Katarina, M. Tri Habibie, M. Siahaan, E. Prawesti Ningrum, A. Medikano, and V. Valentino, "Application of drone technology for mapping and monitoring of corn agricultural land," in *2021 International Conference on ICT for Smart Society (ICISS)*, pp. 1–5, 2021.

[16] A. Zeggada, F. Melgani, and Y. Bazi, "A deep learning approach to uav image multilabeling," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 694–698, 2017.

[17] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution sar image classification via deep convolutional autoencoders," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2351–2355, 2015.

[18] H. Amin, A. Darwish, A. E. Hassanien, and M. Soliman, "End-to-end deep learning model for corn leaf disease classification," *IEEE Access*, vol. 10, pp. 31103–31115, 2022.

[19] M. Agarwal, V. K. Bohat, M. D. Ansari, A. Sinha, S. K. Gupta, and D. Garg, "A convolution neural network based approach to detect the disease in corn crop," in *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, pp. 176–181, 2019.

[20] T. Zhang, Z. Yang, Z. Xu, and J. Li, "Wheat yellow rust severity detection by efficient df-unet and uav multispectral imagery," *IEEE Sensors Journal*, vol. 22, no. 9, pp. 9057–9068, 2022.

[21] S. I. Moazzam, U. S. Khan, T. Nawaz, and W. S. Qureshi, "Crop and weeds classification in aerial imagery of sesame crop fields using a patch-based deep learning model-ensembling method," in *2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, pp. 1–7, 2022.

[22] H. G. Park, J. P. Yun, M. Y. Kim, and S. H. Jeong, "Multichannel object detection for detecting suspected trees with pine wilt disease using multispectral drone imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 8350–8358, 2021.

[23] M. Ouhami, Y. Es-saady, M. E. Hajj, R. Canals, and A. Hafiane, "Meteorological data and uav images for the detection and identification of grapevine disease using deep learning," in *2022 E-Health and Bioengineering Conference (EHB)*, pp. 1–4, 2022.

[24] C.-J. Chen, Y.-Y. Huang, Y.-S. Li, C.-Y. Chang, and Y.-M. Huang, "An aiot based smart agricultural system for pests detection," *IEEE Access*, vol. 8, pp. 180750–180761, 2020.