IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12832

Vision Transformer-Assisted IoT system for smart Agriculture and Multi crop Disease Detection

Geethanjali S G[1], Karan N[2]

Department of CSE, Don Bosco Institute of Technology, Bengaluru, Karnataka, India[1]

Don Bosco Institute of Technology, Bengaluru, Karnataka, India[2]

Abstract: Through the combination of artificial intelligence (AI) and the Internet of Things (IoT), smart agriculture has emerged as a revolutionary step to increase crop yield and sustainability in recent years. This integration has made it possible to continuously monitor farms and automatically assess the health of crops. However, a number of issues plagued the current Convolutional Neural Network (CNN)- based smart agriculture system, including limited applicability in remote agricultural regions, inconsistent data collecting in a variety of field conditions, and poor generalization on single crop diseases. A smart agriculture system based on Vision Transformers (ViT) is suggested as a solution to these problems. The first of four layers in this system architecture is the data acquisition layer, which is equipped with sensor nodes and cameras to gather environmental data and photos of plant leaves. The communication layer follows, which is in charge of carrying out data transfer to the following layer, The processing layer comes next, when threshold evaluation and Simple Moving Average (SMA) filtering are used to preprocess environmental data. Additionally, bilinear interpolation is used to scale and normalize the image data. The pretrained ViT model is then fed this pre processed data in order to classify plants with multiple crops and diseases. Finally, the user receives an interactive web-based dashboard with the output, which comprises the illness kinds that were identified and their confidence levels. The suggested ViT-based system beat the current method and achieved greater accuracy, according to experimental results.

Keywords: Vision Transformer, CNN, IoT, KNN, Simple Moving Average

I. INTRODUCTION

Smart agriculture has developed as a revolutionary strategy to address global concerns like resource efficiency and food security in the modern era. Real-time data collection and analysis of soil moisture, humidity, temperature, nutrient levels, and plant health were made possible by the integration of the Internet of Things (IoT) with crop monitoring and management systems [1]. Precision agricultural practices that seek to increase production, lower input costs, and lessen environmental effect are supported by these IoT-based systems, which use sensor networks, wireless communication, and cloud computing to provide useful information [2]. Despite the integration of many cloud and Internet of Things-based technologies, these systems encountered a number of difficulties when put into practice in the real world. One of the main drawbacks of sensor networks in broad or geographically diverse farm regions is their scalability and reliability [3]. It's still challenging to ensure data integrity and reliable connectivity in these kinds of places. The reliance on continuous and high-quality sensor data for precise forecasts is another problem [4].

In addition, environmental impediments, signal interference, and inconsistent data from node failures all impair system performance [5]. These drawbacks reinforce the demand for scalable, reliable, and user- friendly smart agriculture solutions. The need for smart agriculture and crop monitoring systems, which facilitate early threat identification such as pests or diseases, precise irrigation management, and fertilizer user optimization based on crop needs, is the driving force behind the motivation [6]. Better quality output, sustainable agricultural methods, and lower functional expenses are the outcomes of these strategies and early diagnosis. In order to improve disease prediction, a number of cutting-edge techniques are suggested, such as a hybrid Machine Learning (ML) model coupled with an Internet of Things (IoT)-based soil monitoring system that examines several soil and environmental characteristics [8].

The Digital Twin architecture for smart agriculture is an additional strategy that produced a virtual representation of the actual farm environment. This approach integrates real-time data from drones, weather systems, and IoT devices to simulate farming scenarios and facilitate proactive decision making [9]. Nevertheless, this approach necessitates a robust computational infrastructure and adds complexity to deployment, which is not feasible in regions with limited resources [10].

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12832

Bilinear interpolation for scaling the input images and ImageNet-based normalization in the processing layer to guarantee feature consistency, faster convergence, and better classification even in noisy data are the primary contributions of this study. Because a pretrained Vision Transformer (ViT) can gather global picture features from a variety of visually complex plant images, it is used to improve the accuracy and robustness of crop disease classification.

The sensor nodes delegate all computational work related to illness detection to a central server, allowing them to concentrate only on lightweight data collecting and transmission. The literature reviews are shown in section 2, the entire technique is explained in section 3, the experimental results and discussion are shown in section 4, and the research conclusion is presented in section 5.

II. LITERATURE SURVEY

A platform called SAgric-IoT, which was built on IoT and Convolutional Neural Networks (CNNs), was shown by Juan Contreras-Castillo et al.[11] for greenhouse monitoring. This platform begins by collecting information on greenhouse conditions, including temperature and humidity, using wireless sensor nodes. IEEE 802.15.4, a communication standard, was then used to send the collected data to a central coordinator. Once this data was submitted to the server for further analysis, CNNs were used to detect plant diseases from the collected images. Based on this information, decisions were also made to adjust irrigation and fertilizer. But the SAgric-IoT only dealt with issues related to tomato leaves, which restricted the model's applicability and generalizability.

Gurujukota Ramesh Babu et al. [12] proposes a model for soil monitoring and tomato crop disease prediction at a typical south Indian station. Initially, real-time soil data from the tomato fields, such as humidity, potential hydrogen (pH), and nitrogen phosphorus potassium (NPK) levels, were collected using Internet of Things (IoT) sensors.

The significance of the soil parameters was then ranked using Kendall's correlation after this data was gathered and kept at a rate of one minute. Additionally, a hybrid model that included K-Nearest Neighbor (KNN) and Bayesian Optimization was used to predict disease. Nevertheless, the model is unable to adjust to the various environmental circumstances, and this method relies on localized soil sensor data.

A platform with digital twin deployment for smart agriculture on Cloud Fog-Edge infrastructure was demonstrated by Yogeswaranathan Kalyani et al. [13]. This technology used information from weather stations, drones, and Internet of Things devices to create a Digital Twin (DT) of actual farms. For multi-agent tasks including data gathering, processing, and decision-making, data in this system travels from edge sensors to fog nodes before arriving at the cloud for further analysis and storage. This DT creates the remote conditions needed to carry out and maximize chores like fertilization and irrigation modifications. However, the integration of Cloud-Fog-Edge layers in this system led to a loss of data standardization and an increase in system complexity.

A scalable, sustainable, and peer-to-peer chord-based ecosystem for smart agriculture was presented by Christos-Panagiotis Balatsouras et al. [14]. Initially, LoRa sensor nodes were used to create a peer-to-peer Wireless Sensor Network (WSN). The WiCHORD+ protocol arranged the nodes into a Distributed Hash Table (DHT) structure for effective routing and data search, with each node monitoring environmental variables like humidity and temperature. The model then used the sensor node data that was gathered to predict grapevine disease. However, the real-time data delivery in expansive agricultural contexts was impacted by the overhead created by this chord-based approach during frequent node failures. For smart agriculture, Adeel Ahmed et al. [15] introduced a solution that blends fuzzy data fusion and blockchain technology. The data gathered by sensors, including soil and pest activity, is first grouped into clusters. The data was then detected using a fuzzy similarity matrix, and only the distinct information was transmitted to the edge server. In order to save energy, redundant nodes also switch to sleep mode. Following data gathering, the information is examined and saved on blockchain for future usage and security.

By giving farmers real-time alerts, this method assisted in averting insect outbreaks. Additionally, this strategy reduced energy usage and improved data security through the use of blockchain technology. However, because sleep scheduling and fuzzy clustering need real-time decision making, this system experienced delays in crucial data transfer.

III. METHODOLOGY

This paper suggests a ViT-based smart agriculture and monitoring system that integrates ViT-based crop disease classification, wireless communication, edge-based data processing, environmental sensing, and real-time image capture into a multi-layer Internet of Things architecture. The proposed architecture consists of four layers: data collecting,

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12832

processing, communication, and user interaction. This tiered approach provides real-time crop monitoring and decision support in a range of agricultural scenarios. Important climatic factors including wind, soil moisture, and humidity are also tracked by sensor nodes.

At regular intervals, parallel high-resolution photographs of plant leaves are also taken. After that, the collected data is wirelessly sent to a processing layer edge device for system-level decision-making, ViT- based disease classification, and data preparation. Lastly, the Vit-B model is applied for scalable and accurate plant disease classification. The whole multi-crop Plant Village dataset was used to refine it after it was pretrained on ImageNet-21k. These final results, together with the types of diseases that were found and their confidence ratings, are then shown to the user through an interactive web-based dashboard. The proposed ViT-based smart agricultural and monitoring system's workflow is depicted in Figure 1.

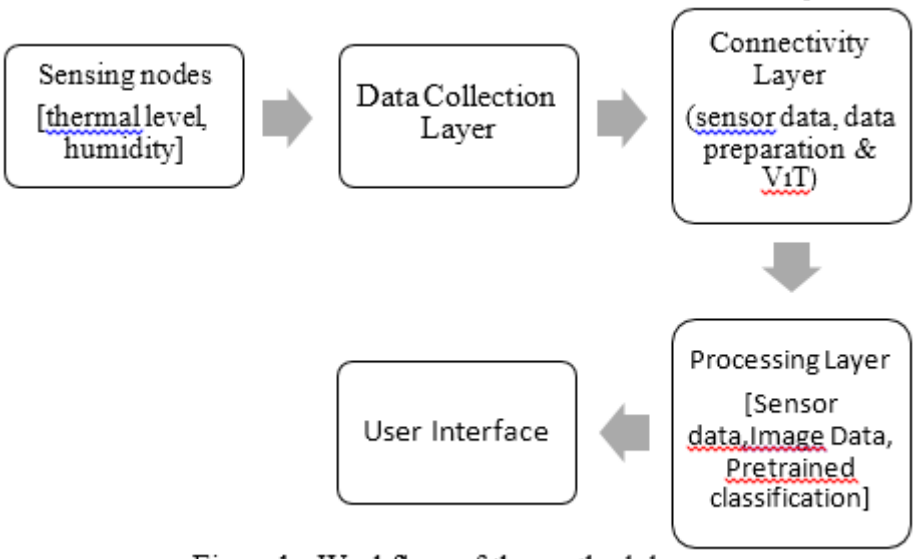


Figure1 : Workflow of the methodology

3.1Data Acquisition Layer

The first tier of the proposed ViT-based smart agriculture and monitoring architecture is the data collection layer. It is the main part of the proposed smart agricultural and monitoring system and is responsible for collecting environmental and visual field data. This layer enables continuous monitoring of crop health and environmental conditions by generating a rich dataset for further analysis and decision-making [11]. This layer collects two types of information: images of plant leaves and environmental features. Wind direction, speed, soil moisture, humidity, and air temperature were among the environmental variables gathered. Identifying hazardous or malevolent growing environments and evaluating the plant's health in relation to its environment depend on these facts.

Concurrently, visual data is also collected, such as high- resolution photos of plant leaves. The ESP32-CAM gadget, which has OV2640 2MP image sensors, was used to take these pictures. To take pictures in real time at specific intervals, these camera nodes are positioned close to specific plants or crops.

A central edge processing unit receives the gathered picture and environmental data, and the ESP32-CAM module sends sensor data via the ZigBee protocol and photos via Wi-Fi. Following collection, all of the data is sent to the processing layer for preprocessing, analysis, and categorization.

Concurrently, visual data is also collected, such as high- resolution photos of plant leaves. The ESP32-CAM gadget, which has OV2640 2MP image sensors, was used to take these pictures. These camera nodes are positioned close to specific crops or plants to take pictures in real time at predetermined intervals. A central edge processing unit receives the gathered picture and environmental data, and the ESP32-CAM module sends sensor data via the ZigBee protocol and photos via Wi-Fi. Following collection, all of the data is sent to the processing layer for preprocessing, analysis, and categorization.

3.2Communication Layer

The second layer in the suggested architecture for smart agriculture and monitoring is the communication layer. This layer safely and efficiently transmits data to the central edge processing unit from distributed field nodes, including sensor and camera modules. This layer ensures a seamless transition between the Data Acquisition Layer and the Processing Layer by permitting bi-directional communication and fault- tolerant data transfer techniques.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12832

Two different wireless communication protocols are utilized, depending on the type of data and hardware limitations: Wi-Fi for camera nodes that take photographs and ZigBee for environmental sensors. Sensor nodes like AM3215, SHT-10, and SEN-08942 use the ZigBee protocol to transmit data in real time. To increase network coverage and connectivity across wide field areas, with intermediary nodes acting as routers, these sensor nodes are set up in a multi-hop topology. In order to establish network paths and synchronize transmission time slots, the node registration and dissemination protocol is utilized to assign end node or routing node roles during commencement.

Each sensor node follows a sleep-wake schedule, recording camera nodes, going into sleep mode when not in use, and waking up at specific times or when an unusual reading is noticed. This strategy results in a longer node lifespan and lower power consumption, particularly in remote locations. EPS32-CAM devices are used to send high-resolution leaf photos over Wi-Fi to the edge device. Wi-Fi enables fast and dependable data transmission because visual data requires more capacity.

Additionally, an NVIDIA Jetson Nano serves as a central gateway and edge computing platform, receiving all transmitted data from sensor nodes or camera devices. After gathering incoming data, including sensor values and photographs, this device serves as an interface to the processing layer, the following step, where data is pre- processed and examined for environmental anomalies and disease detection.

3.3 Processing Layer

The processing layer, the computing centre of the suggested system, now receives the gathered data. Its job is to analyse incoming sensor and picture data and produce crop health status as an output. In order to detect anomalous situations like low soil moisture or high ambient temperature, the environmental sensor data obtained from the communication layer is first processed using Simple Moving Average (SMA) to eliminate noise and then compared to predetermined criteria. Furthermore, the gathered photos undergo preprocessing using ImageNet for normalization and bilinear interpolation for image scaling. Finally, a ViT model is used to classify agricultural diseases using the four pre-processed data.

3.3.1 Simple Moving Average Filtering for Sensor Data

The SMA filter is applied as the initial stage in sensor data preprocessing to eliminate noise and fluctuations from the sensor values. This method is used as a temporal smoothing technique, substituting the average of a predetermined number of the sensor's most recent values for each sensor value. Equation (1) provides the mathematical expression for SMA:

$$SM_t = \frac{1}{n} \sum_{i=1}^{n-1} x_{t-i}$$

Where, SMA_t denotes SMA at time step t, x_{t-i} denotes the individual sensor reading at time $t - i$ and n denotes the chosen window size. This technique facilitates as an effective first stage filter that reduces noise in the variables.

3.3.2 Threshold Evaluation for Sensor Data

Each smoothed sensor value is compared to predetermined thresholds that are specific to the environmental requirements of typical crops after noise reduction via SMA filtering. This method makes sure that abnormalities like heat stress, drought, or humidity that is prone to illness are noted for additional care. Equation (2) states that conditional logic is used to compare each parameter:

$$if\ x_t < T_{min}\ or x_t > T_{max} \Rightarrow Trigger\ Alert$$

Where, x_t denotes the current smoothed sensor value, and T_{min}, T_{max} denotes predefined lower and upper bounds for that sensor. Further, table 1 demonstrates the threshold table which includes the sensor parameters with the threshold conditions.

Table 1. Threshold Table

Sensor Parameter	Threshold Condition
Air temperature	$< 15°C\ or\ > 35°C$
Relative Humidity	$< 40\%\ or\ > 90\%$
Soil Temperature	$< 10°C\ or > 40°C$
Soil Moisture	$< 30\%$
Wind Speed	$> 10\ m/s$

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12832

3.3.3Image Resizing using Bilinear Interpolation

To guarantee conformity with the input requirements of the ViT model, the images obtained using ESP32-CAM modules are pre-processed in this section using the bilinear interpolation approach. All of the raw leaf photos are resized using this method to a set 224 x 224 pixel size. This method maintains the edge smoothness and structural continuity of the input plant leaf images, where visual textures are essential, by determining the intensity value of a new pixel based on a weighted average of the four closest adjoining pixels in the original image grid.First, the input image of size H×W is taken, and the objecive is to resize it to H^'×W^'=224×224. Equation (3) uses the scaling factors to calculate the input image's floating-point location:

$$x = \frac{x'}{W'}.(W - 1), y = \frac{y'}{H'}.(H - 1)$$

where x and y are the pixel values of the input image, and x^' and y^' represent the pixel of the output image. The intensity value at point I(x,y) is then estimated using bilinear interpolation, which combines the values of four nearby pixels as shown in equation (4):

$$(x, y) = (1 - a)(1 - b)Q_{11} + a(1 - b)Q_{21} + (1 - a)bQ_{12} + abQ_{22}$$

Where, Q_{11}, Q_{21}, Q_{12}, Q_{22} are denoted as four known pixel intensities at the top-left, top-right, bottom-left, and bottom-right neighbors respectively and, a and b are the fractional parts of the interpolated location. Each pixel location in the output image undergoes continuous weighted interpolation, producing a smoothly rescaled image that retains important structural elements.

3.3.4Image Normalization using ImageNet based Statistical Scaling

The purpose of this part is to standardize the pixel intensity distribution across all resized photos in order to guarantee that all of the input images are uniform and standardized. ImageNet-based mean and standard deviation scaling, a method employed when working with pretrained models like ViT-B/16, is used in this suggested architecture to accomplish normalization. Red, Green, and Blue are the three color channels that make up each gathered image, with pixel values normally falling between 0 and 255. The mean and standard deviation values, which are taken from the ImageNet dataset that the ViT model is pretrained on, are then used to independently normalize each channel after these raw images have first been scaled to the [0,1] range.Equation (5)

$$I_c^{norm} = \frac{I_c - \mu_c}{\sigma_c}$$

expresses the normalization for every pixel channel: In this case, represents the channel's pixel intensity, mean value, and standard deviation, respectively. Additionally, equation (6) expresses the precise values utilized for ImageNet-based normalization:

$$\mu = [0.485, 0.456.0.406], \sigma = [0.229, 0.224, 0.225] \quad (6)$$

In order to efficiently extract pertinent features from the preprocessed leaf images using a pretrained ViT model, this stage of ImageNet-based normalization guarantees statistical consistency between the training and classification phases

.

3.3.5 ViT model Training

1. The ViT used in the proposed smart agricultural and monitoring system uses a two-phase training technique. The vast ImageNet-21k dataset is used to train the ViT model, while the domain-specific PlantVillage dataset is used to further hone it. This two-phase training method improves the ViT model's capacity to generalize from visual features to crop-specific disease patterns. The selected ViT-B/16 uses a 1616 patch size and processes input images of size 224224. First, this model is pretreated using the ImageNet 211 dataset, which consists of over 14 million natural pictures categorized into 21,841 categories. During this training phase, the model gains the ability to capture the global dependencies. First, as shown in equation 7, each input image is divided into a set of flattened patches called {x_1,x_2,……,x_n}.

2. $$N = \frac{HW}{p^2}$$

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12832

3. Where, and. Additionally, each patch is projected into a fixed length embedding using linear projection, yielding the transformer encoder's initial input, which is represented by equation (8):

4. $z_0 = [x_{cls}; x_1E; x_2E; \dots\dots; x_NE] + E_{pos}$

5. where represents the patch embedding matrix, a learnable classification token, and the positional encoding that was added to preserve spatial links. These encoded representations are then transferred to a series of feed forward layers and multi-head self-attention for additional processing. Lastly, a linear layer and softmax activation are applied to the final hidden state of the classification token in order to provide the output class probabilities. Additionally, the PlantVillage dataset is used to fine-tune the suggested ViT model [11]. More than 50,000 tagged photos of both healthy and diseased plant leaves make up this dataset, which covers 14 different crop kinds and around 38 disease classes.

6. This study specifically focuses on six widely grown crops: tomato, potato, pepper, corn, apple, and grape. Additionally, every image in the collection has a label that combines the names of the crop and the disease, allowing for the classification of several crops and diseases. Three sets are created from the dataset: 70% are used for training, 15% are used for validation, and 15% are used for testing. The new fully connected layer in the ViT-B/16 model, which is correlated with the number of target classes, has taken the role of the previous classification head. Equation (9) states that the model is trained with the goal of minimizing the categorical cross entropy loss:

7. where represents the projected softmax probability, the ground truth label, and the number of illness classes. The ViT model can efficiently learn visual patterns associated with the many leaf diseases across a variety of crops because to the combination of pretraining on large-scale datasets and fine-tuning based on domain-specific datasets.

3.4 based Classification

The ViT model is placed on the edge device to provide real-time sickness classification after pretraining on the ImageNet 21k dataset and fine- tuning on the Plant Village dataset. Based on the visual features, the ViT model correctly determines the crop type and related disease class in this stage of the analysis of the field-collected leaf images. Following preprocessing of the collected photos, the standardized images are sent to the edge device's ViT model. Next, the image is internally partitioned into 1616-sized, non-overlapping segments.

After that, these patches are included into a series of vectors with predefined lengths that are linked to positional encodings. From there, these vectors are processed by a series of transformer encoder layers. All of the patches' global information is combined using the categorization token ([CLS]). A linear classification head receives this class token and uses it to transfer the final representation to a probability distribution over all possible illness classes.

For instance, let denotes the final output of the [CLS] token, and let and denotes the weights and biases of the classification head, where is the number of output classes. Equation (10) is also used to compute the class logits: $\hat{y} = W_{head}. z_{cls} + b$
Finally, a softmax function is applied to these logits, producing a probability distribution across the disease categories, as shown in equation (11):

$$P(c_i) = \frac{e^{\hat{y}_i}}{\sum_{j=1}^{C} e^{\hat{y}_i}}, \forall i \in \{1,2, \dots\dots, C\} \qquad (11)$$

Finally, the output of this classification is sent to the User Interaction Layer, where it is shown to end users with the ambient sensor values via a dashboard interface. This enables a thorough and up-to-date crop health status. The User Interaction Layer receives the categorization output and displays it with ambient sensor values via a dashboard interface. This offers a thorough and up-to-date picture of crop health.

3.5.User Interaction Layer

A lightweight web-based dashboard that exhibits the results of environmental sensor analysis and ViT-based classification is part of the suggested ViT smart agriculture and monitoring system, which makes it easier for users to obtain information and make decisions in real time. Important data is shown on this dashboard, including the label of the crop disease that was found, the classification's confidence score, and the temperature, soil, and humidity sensor readings in real time. The edge device hosts the system's local intranet environment, which enables farmers to use a smartphone to directly monitor crop conditions without the need for an internet connection. By offering timely feedback and visible alarms, this interface also improves utilisation and facilitates data-driven agricultural management.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12832

IV. EXPERIMENTAL RESULTS

A mix of hardware and software components is used to build the suggested ViT-based smart agricultural and monitoring system. These components are made to support intelligent categorisation, local processing, image acquisition, and real-time sensing. Sensor nodes with AM2315 set for air, temperature, and humidity measurements are part of the hardware components. SHT-10 is utilised for temperature and moisture sensing of the soil, and SEN-08942 is used for wind direction and speed monitoring. Images of crops or plants are among the visual data that is acquired by ESP32-CAM modules, which are equipped with 2MP OV2640 sensors. Additionally, the software consists of the PyTorch 2.0 Deep Learning (DL) framework and the Python programming language. A workstation with an NVIDIA RTX 3060 GPU and 12 GB of RAM is used to train the model, and OpenCV and pillow libraries are used for picture preprocessing. Additionally, nodes communicate with one another for sensor data and Wi-Fi to provide photos of leaves via the ZigBee protocol. Last but not least, the farmer dashboard interface is hosted on a webserver built on a lightweight Flask weight and augmented with Bootstrap for responsive design. Equations (12) through (18) provide the measures used to assess the effectiveness of the suggested ViT-based smart agriculture and monitoring system:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (13)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (14)$$

$$F1-\text{Score} = \frac{2TP}{2TP+FP+FN} \quad (15)$$

$$\text{Packet Loss Rate} = \left(\frac{\text{Packets Sent}-\text{Packets Received}}{\text{Packets Sent}}\right) \times 100\% \quad (16)$$

$$\text{Packet Reception Rate} = \left(\frac{\text{Packets Received}}{\text{Packets Sent}}\right) \times 100\% \quad (17)$$

$$\text{Energy Consumed} = \text{Initial Energy} - \text{Remaining Energy} \quad (18)$$

True Positives and True Negatives are indicated by TP and TN, while False Positives and False Negatives are indicated by FP and FN, respectively.

4.1. Performance Analysis

The performance of the suggested ViT-based smart agriculture monitoring system is compared to more conventional methods like digital twin + cloud-fog-edge and hybrid ML + IoT. The metrics of packet loss rate, packet reception rate, and energy usage are used in this assessment. The suggested ViT-based smart agriculture monitoring system's performance analysis is shown in table 2.

Models	Packet Loss Rate (%)	Packet Reception Rate (%)	Energy Consumed (%)
Hybrid ML + IoT	6	92	32
Digital twin + cloud-fog-edge	4	94	26
ViT based smart agriculture system	1	97	12

Table 2 makes it abundantly evident that the suggested ViT-based smart agriculture and monitoring system performed better than the others by achieving a lower packet loss rate (2%). This allowed the system to transfer burdensome

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

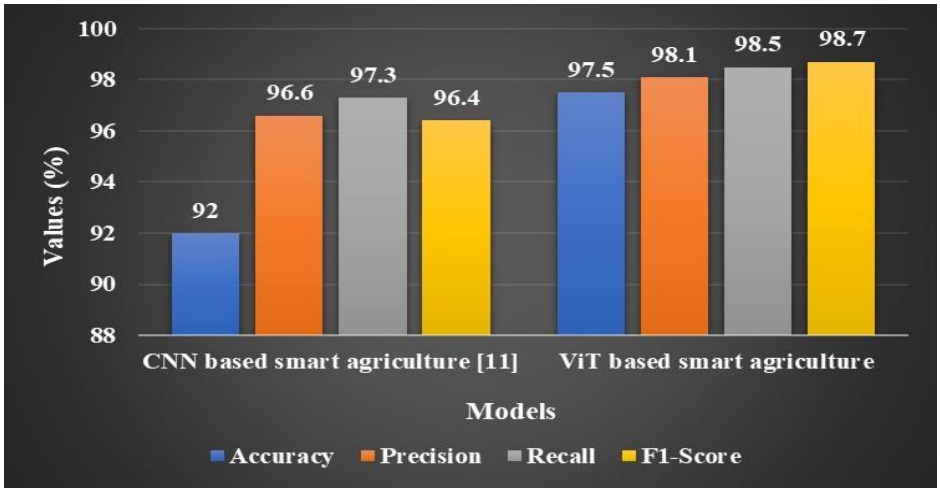
DOI: 10.17148/IARJSET.2025.12832

computations, like disease classification, to a centralised server, avoiding congestion, retransmissions, and packet collisions. Additionally, by minimising pointless data packet exchanges and utilising reliable, low-interference communication protocols, the suggested system was able to obtain a higher packet reception rate (99%). Additionally, the suggested solution only used 11% of the node battery. Thus, the suggested smart agriculture system based on ViT performed better than the conventional methods.

4.2. Comparative Analysis

Using key performance measures like accuracy, precision, recall, and f1- score, a comparative study of the proposed model is conducted against the current CNN-based smart agriculture model [11] in order to assess the efficacy of the proposed ViT-based smart agriculture monitoring system. The suggested ViT-based smart agriculture system outperforms the current CNN-based smart agriculture system, according to a comparison analysis. While the current CNN-based model obtained 92% accuracy, 96.6% precision, 97.3% recall, and 96.4% f1-score, the suggested ViT-based system achieved 97.5% accuracy, 98.1% precision, 98.5% recall, and 98.7% f1-score.

According to the results, the suggested ViT-based smart agriculture and monitoring system performs better in classification, which makes it more efficient and precise for identifying plant diseases. In figure 2, the comparative analysis of the suggested ViT-based smart agricultural and monitoring system is shown.



4.3 Discussion

The main goal of this research is to use a pretrained ViT to achieve a more accurate, efficient, and scalable smart agriculture monitoring system while maintaining the IoT-based framework's communication reliability and energy economy. The current CNN-based approach, although its efficient performance, relied largely on local features and had trouble making important visual differences, especially in complicated plant diseases. Furthermore, the current system's generalisability across various crops and environmental conditions was limited. Traditional IoT-based systems, on the other hand, suffered from shifting soil and climate conditions and used a lot of energy because they had to transmit data often and compute on-device.

The lack of synchronisation between edge, fog, and cloud layers resulted in packet loss and increased energy usage for another conventional Digital Twin solution. These issues are resolved by the suggested ViT- based smart agriculture system, which makes use of the ViT model that was refined on the Plant Village dataset after being trained on ImageNet. The global self-attention mechanism is used by this model to extract intricate patterns. The burden is moved to a centralised server distant from the sensor nodes in this suggested system, which lowers energy consumption. Additionally, by keeping the network lightweight, the suggested solution optimises system packet reception and removes overhead. Thus, in every way with possible solutions, the suggested ViT- based smart agriculture system performed better than the current and conventional methods.

V. CONCLUSION

This paper proposes a ViT-based smart agricultural system to increase the accuracy, scalability, and efficacy of crop disease monitoring. The first layer is the data acquisition layer, which collects data by deploying sensor nodes with Internet of Things capabilities in greenhouse or field environments to capture crop images and environmental parameters.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12832

These nodes perform lightweight sensing and data transfer to ensure minimal energy consumption and reliable data flow free from computational strain.

The communication layer then manages data flow between sensor nodes and the processor layer, ensuring reliable and low latency connectivity while preserving network energy efficiency. The collected images are then pre-processed at the processing layer using bilinear interpolation to resize the images and ImageNet normalization to match the input image data with the pretrained ViT model. Additionally, these pre-processed images are fed into the ViT model for plant disease classification, improving the model's accuracy and generalizability.

Finally, the user interaction layer provides farmers with a real- time interface to monitor crop conditions and get alerts. The predictive capabilities of the proposed ViT-based smart agriculture system will be enhanced in the future by including more data sources and actuation and feedback layers to offer real-time and dynamic system controls.

REFERENCES

[1]. Quy, V.K., Hau, N.V., Anh, D.V., Quy, N.M., Ban, N.T., Lanza, S.,Randazzo, G. and Muzirafuti, A., 2022. IoT-enabled smart agriculture: architecture, applications, and challenges. *Applied Sciences*, *12*(7), p.3396.

[2]. Shaikh, F.K., Karim, S., Zeadally, S. and Nebhen, J., 2022. Recent trends in internet-of-things-enabled sensor technologies for smart agriculture. IEEE Internet of Things Journal, 9(23), pp.23583-23598.

[3]. Sinha, B.B. and Dhanalakshmi, R., 2022. Recent advancements and challenges of Internet of Things in smart agriculture: A survey. Future Generation Computer Systems, 126, pp.169-184.

[4]. Avşar, E. and Mowla, M.N., 2022. Wireless communication protocols in smart agriculture: A review on applications, challenges and future trends. Ad Hoc Networks, 136, p.102982.

[5]. Pagano, A., Croce, D., Tinnirello, I. and Vitale, G., 2022. A survey on LoRa for smart agriculture: Current trends and future perspectives. IEEE Internet of Things Journal, 10(4), pp.3664-3679.

[6]. AlZubi, A.A. and Galyna, K., 2023. Artificial intelligence and internet of things for sustainable farming and smart agriculture. Ieee Access, 11, pp.78686-78692.

[7]. Mowla, M.N., Mowla, N., Shah, A.S., Rabie, K.M. and Shongwe, T., 2023. Internet of Things and wireless sensor networks for smart agriculture applications: A survey. IEEe Access, 11, pp.145813-145852.

[8]. Ghazal, S., Munir, A. and Qureshi, W.S., 2024. Computer vision in smart agriculture and precision farming: Techniques and applications. Artificial Intelligence in Agriculture.

[9]. Cesco, S., Sambo, P., Borin, M., Basso, B., Orzes, G. and Mazzetto, F., 2023. Smart agriculture and digital twins: Applications and challenges in a vision of sustainability. European Journal of Agronomy, 146, p.126809.

[10]. Kabato, W., Getnet, G.T., Sinore, T., Nemeth, A. and Molnár, Z., 2025. Towards climate-smart agriculture: Strategies for sustainable agricultural production, food security, and greenhouse gas reduction. Agronomy, 15(3), p.565.

[11]. Contreras-Castillo, J., Guerrero-Ibañez, J.A., Santana-Mancilla, P.C. and Anido-Rifon, L., 2023. SAgric-IoT: An IoT-based platform and deep learning for greenhouse monitoring. Applied Sciences, 13(3), p.1961.

[12]. Babu, G.R., Gokuldhev, M. and Brahmanandam, P.S., 2024. Integrating IoT for Soil Monitoring and Hybrid Machine Learning in Predicting Tomato Crop Disease in a Typical South India Station. Sensors, 24(19), p.6177.

[13]. Kalyani, Y., Bermeo, N.V. and Collier, R., 2023. Digital twin deployment for smart agriculture in Cloud-Fog-Edge infrastructure. International Journal of Parallel, Emergent and Distributed Systems, 38(6), pp.461-476.

[14]. Balatsouras, C.P., Karras, A., Karras, C., Karydis, I. and Sioutas, S., 2023. WiCHORD+: a scalable, sustainable, and P2P chord-based ecosystem for smart agriculture applications. Sensors, 23(23), p.9486.

