

# AI-driven drone technology and computer vision for early detection of crop disease in large agricultural areas

Received: 28 January 2025

Accepted: 9 December 2025

Published online: 17 December 2025

Cite this article as: Manoj H.M., Shanthi D.L., Lakshmi B.N. *et al.* AI-driven drone technology and computer vision for early detection of crop disease in large agricultural areas. *Sci Rep* (2025). <https://doi.org/10.1038/s41598-025-32384-1>

H M Manoj, D. L. Shanthi, B. N. Lakshmi, K. J. Archana, Endluri Venkata Naga Jyothi & Kande Archana

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

# AI-Driven Drone Technology and Computer Vision for Early Detection of Crop Disease in Large Agricultural Areas

## Dr. Manoj H M

Associate Professor, Department of Artificial Intelligence and Machine Learning, BMS Institute of Technology and Management, Autonomous under VTU, affiliated to VTU, Bengaluru, Karnataka, India 560064  
[manoj.hmhm@gmail.com](mailto:manoj.hmhm@gmail.com)

## Dr. Shanthi

D L & Assistant Professor, Department of Computer Science and Engineering, BMS Institute of Technology and Management, Bangalore, Karnataka. India

Mail: [gopalaiahshanthi@bmsit.in](mailto:gopalaiahshanthi@bmsit.in),

## Dr. Lakshmi B N

Assistant professor, Department of Computer Science and Engineering, BMS Institute of Technology and Management Bangalore, Karnataka.

Mail: [lakshmibn@bmsit.in](mailto:lakshmibn@bmsit.in)

## K.J. Archana

ASSISTANT PROFESSOR, Department of COMPUTER SCIENCE ENGINEERING, Neil Gogte Institute of Technology, Uppal, Hyderabad, Telangana, India  
[archana.kanathala@gmail.com](mailto:archana.kanathala@gmail.com)

## Dr. Endluri Venkata Naga Jyothi

Associate Professor, Department of (AI&ML), CMR COLLEGE OF ENGINEERING & TECHNOLOGY, Telangana, India.  
[jyothiendluri@gmail.com](mailto:jyothiendluri@gmail.com)

## Mrs Kande Archana,

Research Scholar, Jawaharlal Nehru Technological University, Hyderabad, Assistant Professor, CSE Department, Malla Reddy College of Engineering, Hyderabad, Telangana, India.

[kande.archana@gmail.com](mailto:kande.archana@gmail.com)

## \*corresponding author: Dr. Manoj H M

Associate Professor, Department of Artificial Intelligence and Machine Learning, BMS Institute of Technology and Management, Autonomous under VTU, affiliated to VTU, Bengaluru, Karnataka, India 560064  
[manoj.hmhm@gmail.com](mailto:manoj.hmhm@gmail.com)

## Abstract

Timely detection of crop diseases in large, heterogeneous agricultural fields is difficult, as aerial imagery is often corrupted by illumination, weather, and crop-stage variations. This paper introduces AgroVisionNet, an AI-powered drone and computer vision approach that synthesises high-resolution drone imagery with in-field IoT/environmental sensor data to enhance early disease detection. The core of the proposed model is a hybrid CNN-Transformer backbone to extract spatial and contextual data from drone images, and an adaptive fusion layer to fuse time-aligned sensor readings and to make a decision, using visual and environmental evidence. Particularly, a multimodal drone-sensor dataset

is collected across multiple crops and field conditions. Beyond widely used deep models for plant/crop disease identification, such as VGG16, ResNet50, Inception V3, and DenseNet121, experiments are conducted using the same training and evaluation framework. It is shown that AgroVisionNet achieves higher classification accuracy and F1-score, while inference remains feasible on an NVIDIA Jetson Nano using TensorFlow Lite. Moreover, by generating Grad-CAM plots, the study demonstrates that the proposed approach identifies disease-affected areas and, in this sense, provides interpretable information required by agronomists. These outcomes suggest that AI-based crop health tracking can be robust and field-ready by integrating drone imagery, sensor fusion, and edge computing.

**Keywords** - Precision Agriculture, Deep Learning, IoT Integration, Crop Disease Detection, Edge Computing

## 1. INTRODUCTION

Rising cases of crop diseases, driven by climate change, globalisation and large scale agriculture, are a major threat to global food security and agricultural sustainability. This is why prompt detection and interception is so important – not only to reduce yield losses but to avert much bigger epidemics. Manual field inspection of all crops for inspecting crop diseases are so much time consuming and difficult task for farmers in agricultural field that needs trained people for distinction as diseases are difficult to detect and we know this is human error prone activity as these methods are largely based on human eyes. Hence, there is an urgent need to use modern technology (e.g., precision agriculture, artificial intelligence (AI), and drone imaging) in a scalable, accurate, and close to real-time disease surveillance over large agricultural landscapes [1], [2].

Recent studies have shown that in the controlled condition, computer vision models driven by AI can achieve classification with impressive accuracy for crop diseases [3], [4]. Also, in large-scale agricultural imaging, uncrewed aerial vehicles (UAVs) equipped with multispectral and hyperspectral cameras have been successful [5], [6]. Yet, several key limitations exist with existing methods. The majority of existing models are not real-time and thus unsuitable for launch on low-end devices. These do not often incorporate environmental contextual data, such as temperature, humidity, and soil moisture, which are essential for understanding the spread of the disease [7], [8]. Furthermore, most studies focus on a particular crop or a small area, hampering scalability and generalizability [9]. For example, conventional Convolutional Neural Network (CNN) based models, such as MobileNet [10] and Siamese Networks [21], or the more recent Visual Transformer (ViT) approach[11], demonstrate superior accuracy in laboratory conditions but are challenged by diverse real-world agricultural scenarios because of dataset bias and environmental change. These constraints require an advanced, comprehensive, and unified framework that incorporates deep learning methods, including drone imaging, IoT sensor fusion, and edge computing, for proper real-time monitoring in agriculture.

Though there is promising evidence, the lack of research on multimodal data and hybrid deep learning models that can provide actionable insights at scale still prevails. Current work considers only drone images, without environmental context on one side, and IoT sensor data on the other, indicating that neither can provide a holistic picture of crop health. Moreover, only a limited number of systems are tailored for real-time, low-latency execution on edge devices that can be deployed in the field. This shortage of integration and field validation has impeded the large-scale application of AI-based agricultural disease monitoring [12], [13].

To bridge these gaps, we present AgroVisionNet, a hybrid deep learning model that integrates CNN-based spatial feature extraction with Transformer-based contextual attention and IoT sensor data fusion. The model utilises edge computing to enable on-device, real-time disease detection across wide agricultural surfaces. The system is intended to provide farmers with early warning and call-to-action information, so they can implement interventions that reduce crop losses and promote sustainability. Integrating visual and environmental information, AgroVisionNet improves disease diagnosis accuracy and generalisation to different conditions. The use of drones in agriculture, along with some applications, is shown in Figure 1.

While some of the earlier CNN-Transformer models for plant disease detection employ only visual cues, in contrast, AgroVisionNet employs a learnable adaptive fusion strategy that weighs both visual and IoT modalities to dynamically modulate ( $\alpha$ ,  $\beta$ ) the importance of each modality using weights that are made trainable and end-to-end optimised using backpropagation. This allows the model to dynamically highlight the informative modality when weather conditions change in the field. In addition, TensorFlow Lite quantisation is applied to optimise the network at the edge further, reducing feature dimensionality and enabling real-time inference on low-power devices like the Jetson Nano. Such dual innovation enables sensor-aware adaptive fusion with efficient on-edge deployment, distinguishing AgroVisionNet from traditional hybrid architectures while facilitating practical, field-ready, scalable precision agriculture. The paper makes the following contributions.

1. We propose AgroVisionNet, such that its backbone is a new hybrid CNN-Transformer architecture where the CNN obtains spatial crop-health descriptors from drone images while the Transformer captures long-range contextual dependencies; this backbone is suitable for multimodal agricultural environments.
2. Rather than simply concatenating image embeddings and time-aligned IoT/environmental features, we present an adaptive cross-modal fusion block that leverages learnable weights to perform these operations, thereby making the overall model robust to changing field conditions and sensor noise.
3. We design an edge-centric optimisation pipeline (TensorFlow  $\rightarrow$  TFLite dynamic quantisation,  $224 \times 224$  fixed form inputs, low precision inference) and show that the complete multi-modal model is run on an NVIDIA Jetson Nano with reasonable latency for in-field operations.
4. We create a reference dataset that includes drone RGB/multispectral images with corresponding sensor streams from various crops and locations, and we provide train/validation/test splits for reproducible benchmarking.
5. To expose the disease-spot activation maps with an explainable AI (Grad-CAM) visualisation, we integrate them into our pipeline so that agronomists can achieve better transparency of the automated decision.

The proposed AgroVisionNet incorporates an adaptive, learnable fusion mechanism that temporally aligns sensor readings with image embeddings; in contrast, previous studies either fused image features with fixed sensor-based metadata or operated only in the cloud. Moreover, focusing on rare, detailed aspects of the design of an agricultural disease-detection model, we explicitly outline and validate the proposed model for low-latency execution on an embedded edge platform (Jetson Nano) using TensorFlow Lite-based quantisation and input-size standardisation. In the third contribution, we create a synchronised drone-IoT dataset across different crops to enable multimodal, reproducible evaluation.

The rest of this paper is structured as follows. Related Work Section 2 presents related work on: drones, AI for plant disease detection, IoT integration, and precision agriculture—the framework. We describe a new MTR system that consists of the following sections: 3.1 Data acquisition, model architecture and algorithms in section 2. Section 4 provides details on the experimental setup, the dataset, and the performance evaluation. Results and limitations of the study are described in Section 5. The paper concludes with a summary of the last section in Section 6.

## 2. RELATED WORK

This section explores the recent literature on using artificial intelligence (AI), drones, and computer vision to improve agricultural monitoring and disease detection. The studies depict AI's role in precision farming, with special attention to early detection and effective monitoring. In line with the objectives of this study, this review organises references into categories: drone use, AI-based disease diagnostics, IoT integration, sensor technology, and challenges.

### 2.1 Role of Drones in Agricultural Monitoring

Drones have emerged as a disruptive technology in this agroengineering space by enabling high-resolution surveying of vast farmland areas with minimal workforce [16]. For instance, it has significantly boosted the efficiency of plant cultivation, including disease detection, pest monitoring, and environmental assessment. Studies by Slimani et al. Drones can reliably classify plant diseases and lower human errors in traditional inspection practices [1]. Similarly, Abbas et al. Drone techniques for early crop disease diagnosis were faster and more efficient than conventional ground-based methods [2].

For example, there are some studies on the use of UAVs for agriculture. Velusamy et al. Numerous works [7], [12], [13], [14] addressed the application of UAVs to crop health and pest monitoring, providing new approaches to increase the precision and robustness of aerial imaging data. Xue et al. [16] and Dutta and Goswami [17] have reviewed the use of UAVs to overcome labour shortages and environmental consequences. Eagle/Fishtail Kite Zeta or Jiggle-Eye. In addition, Mogili and Deepak [18] examined UAV systems for pesticide spraying, which pose health risks from direct exposure of farmers to sprayed chemicals at ground level.

The use of drones in field-based disease assessment and crop monitoring was studied in more detail by Kumar et al. [19] and Puri et al. [20], who reported that although UAVs demonstrated high potential for precision farming, there is still room to improve the hardware of drones, image processing analysis and autonomous flight control to improve the reliability and scalability.

Chin et al have recently reviewed these issues. [10] and Peña et al. [48], which examined the use of drones for plant disease detection and suggested future research directions. Peña et al. [48] also discussed UAV applications for phytosanitation in oil palm plantations, particularly for pest and disease surveillance at the scale of the entire plantation. Hasanaliyeva et al. [8] summarised recent advances in the development of UAV-based plant disease detection systems. They stressed the paramount importance of digital technology to augment decision-making.

More recently, developments combining IoT with drones have broadened their range of applications. Doggalli et al. [9] and Gao et al. [11] show that UAVs integrated with IoT networks could improve pest and disease monitoring by facilitating real-time information sharing. In these works, additional energy efficiency and flight endurance issues were explored in greater detail, both of which are essential parameters for the massive deployment of large-scale drones.

In all these applications, drones have been demonstrated to be highly effective for disease detection, pest control, irrigation management, and environmental monitoring. Yet most previous methods rely primarily on visual data and do not incorporate ecological sensors or real-time edge computing. This restricts their scalability and their ability to react in real time when context changes. Motivated by these observations, we develop the AgroVisionNet framework to achieve this objective by integrating drone images and IoT sensors for a real-time, multimodal solution to large-scale crop disease detection.

## 2.2 Artificial Intelligence for Plant Disease Detection

AI is a critical component in improving the precision of plant disease detection. Misra et al. The work of Swanson et al.[4] helped provide some perspective on the use of AI in agriculture, in particular on the opportunities to enhance economic viability and feasibility. They illustrated practical issues regarding the effectiveness of different ML/DL techniques [22], Ahmad et al. Ghosh et al. [23] previously reviewed studies on DL for plant disease diagnosis, while [24] [25] identified gaps and suggested developing a tool that meets farmers' needs. Latif et al. The work by [26] introduced a high-accuracy DCNN-based method for rice disease detection, but identified dataset biases as a limitation. Kamilaris et al. According to [28], DL models are highly accurate across various classification problems but remain underexplored for many agricultural problems beyond what traditional techniques can cover. Ai et al. [29] et Nguyen et al. Convolutional neural networks for crop disease classification [30], but limited to a low number of datasets.

While Ferentinos [31] reports good accuracy with CNNs, overfitting is a limitation. Lee et al. Vitis and Turner [32] and Shewale and Daruwala [33] sought to enhance CNN specificity for broader agricultural use. Alkan et al. Hybrid DL models have been reviewed in the literature [40], and many other enhancements of these models for more accurate detection have been proposed [36,39]. Perez et al. Only [38] presented a lightweight DL model for detecting disease, whereas Albattah et al. [77] and Jafar et al. Gururajan et al.[78] Stressed on intelligent agri solutions by merging AI with IoT.

Recently, enormous efforts have been made to enhance plant disease recognition using lightweight CNN variants and Vision Transformer (ViT) architectures, which are also suitable for practical agricultural implementation, offering fast and accurate image classification. Sandler et al. MobileNetV3 [82] introduced the Architecture and demonstrated that an efficient convolutional architecture reduces overall complexity, making mobile neural networks usable on drones and end devices. Similarly, Kunduracıoğlu and Paçal [83] used EfficientNet models to detect sugarcane leaf disease. It would be more correct to say that they concluded that small models can still achieve high accuracy despite working under realistic resource constraints in field environments.

While Vision Transformers (~ViTs) in the context of agricultural imagery garner distracted interest. Paçal [84] proposed data-efficient ViT models for sugarcane leaf disease detection and demonstrated their generalisation performance on small datasets. Additionally, Kunduracıoğlu and Paçal [85] evaluated the performance of CNNs and ViTs for grape disease detection, highlighting that the self-attention mechanism provides better feature embeddings and classification accuracy under changing environmental conditions.

Building on these advances, hybrid backbones that concatenate CNNs and ViTs have emerged as an encouraging approach for extracting local spatial features with CNNs and leveraging global attention mechanisms. Shandilya et al. [86] designed a hybrid CNN-ViT model for maize leaf disease identification that achieved better performance than SC SCNNs and Transformers by efficiently fusing local and global feature learning. Moreover, Paçal and Kunduracıoğlu [87] conducted an extensive meta-study of CNNs with ViT-based models in a submission aimed at demonstrating that the hybrid model is better suited for large, complex plant datasets.

Even though these works are effective and introduce novelty, most focus solely on pure-image data and do not involve IoT-based environmental sensing or edge-computing optimisation needed for real-time, large-scale agricultural monitoring. Inspired by these challenges, we propose AgroVisionNet to bridge the gap between deep learning approaches and real-world practical constraints, focusing on deep learning algorithm robustness and multimodal data fusion integrated with practical edge deployment for scalable crop disease detection.

### 2.3 Deep Learning Models in Precision Agriculture

Deep learning models are increasingly central to precision agriculture. Zhang et al. [34] and Altalak et al. While [35] reviewed the applications of DL in dense scene parsing, other papers proposed techniques to better handle input data [41], [42], [43]. Yuan et al. Transfer learning methods and datasets in agricultural disease detection: Towards real-world datasets [44] [46] Coulibaly et al. Critical applications and challenges [49], [50] in DL for agriculture were identified in a bibliometric analysis by [47]. Ren et al. [52] et Bharman et al. Through their work, [51] also provided insights into the pros and cons of DL models and identified areas for future research. Wang et al. DL applications in hyperspectral imaging were reviewed in [27], with the need for conducting good data processing underlined in [53] [54]. Sharma et al. It also uses segmented images to improve CNN performance [55], and Zhang et al. And Pradhan et al.[56],[57],[58],[59]. While studying a publicly available dataset [71], the authors noted limitations in achieving high classification accuracy for the disease. Rakhmatulin et al. The Authors.[80] discussed on deep neural networks approaches for real-time weed detection in precision agriculture applications, discussing strengths such as CNN based on feature extraction, challenges in the dataset and preprocessing methods, and the integration of such automatic processes with IoT systems for practical precision agriculture approaches.

### 2.4 Integration of IoT, Big Data, and AI

The convergence of IoT, big data, and AI has driven agricultural growth. These technologies help address several challenges in the agri-food industry [62, 63] and were discussed by Purnama and Sejati [3]. Smart farming is mentioned by Dutta and Mitra [6] in their work on IoT and sensors for agriculture. In this work, the authors prefer automation. Dhanya et al., Neupane and Gurel [15], and PiLi (2023) also emphasised the concept of sensor merging and the use of real-time data for diagnosing illnesses.[45] Li et al. The AI Techniques for Disease Classification, Feature Extraction, and Their Efficiency review is not too old [76]. Albattah et al. [77] et Pena et al. Ref.[48] addressed combining sustainable development with high-tech. In their paper, Materne and Inoue [61] also reported on IoT-based systems for detecting pests and diseases at an early stage. Darwin et al. [65] et Coulibaly et al. Senior et al. [47] describe the development of big data analytics for crop monitoring and sustainability.

### 2.5 Sensor and Imaging Technology Advancements

Soil-Pest input variables. Soil type, pest estimates for pest-sensitive crops, and soil moisture are all crucial variables in defining trend in crop production over time, and are therefore considered highly detailed input variables. Buchelt et al. Abstract: Explainable AI- A novel way to enhance Drones [5] Gomez et al. [37] and Fountsop et al. Disease detection through image annotation and model compression was reported [79]. How and how often sediment and structures erode. In [73], a study by certain authors reviewed the use of drone imaging systems for plant disease identification which could be performed at real time, and later by direct delivery to end-users. Song et al. [69] and Mhango et al. Reference [64] emphasized the key role that high resolution images play in precision agriculture. [67] Bauer et al. Along with that of Peng et al.[13] — but also provide some new evidence of the supply of energy in the tundra. [68] did a review of computer vision applications to aerial phenotyping and proposed further data integration. Ang et al. Shoji, B. used hyperspectral and multispectral data for crop analysis [66].

### 2.6 Challenges in Data Collection and Processing

Among them, these tasks require data collection which still remains a huge bottleneck and processing. Velusamy et al. Energy efficiency issues for UAVs were also identified by Mogili and Deepak [18] along with data uncertainty issues. Gao et al. In order to address these limitations, frameworks have been proposed by [11]. Wani et al. Article sources:

[21],[72] and Albattah et al. Finally data was emphasized with [77] focusing on ethics and environmental adaptability. Isiaka et al. Papers by [75] and Alkan et al [40] and Hasanaliyeva et al. Gathering different types of data [8]. Coulibaly et al. Consequently, [47] called for improved statistical validation techniques. Sustainable farming is getting more and more attention. Daggalli et al. [9] et Sharma et al. The authors of [55] suggested methods for enhancing the robustness of the disease detection systems. Pena et al. [48] and Bhatti et al. Use of renewable resources in agriculture [60]. Albattah et al. [77], Isiaka et al. [75] et Shahi et al. [74]. Developing a sustainable future around stakeholder engagement and policies that have to be developed. Siddappa [81] proposed innovations to drone systems to enhance its functionalities. Mowla and Avsar [88] provide a survey on wireless protocols for smart-agriculture, covering topics including the protocols with regards to coverage, energy, and scalability; in addition, this paper discusses boundary issues of interoperability and connectivity in large-scale farm deployments. Mowla and Gk [89] provide a comprehensive review of deep learning-based weed-detection networks and different types of architectures (including region-based and convolutional networks), datasets and challenges for real-time applications for precision agriculture.

**Table 1:** Synthesised Review Of Existing Drone-/IoT-Based Crop Disease Detection Studies And Identified Research Gap For AgroVisionNet

Study / Source In Manuscript	Approach / Focus	Limitation In the Context Of Your Work	Research Gap Filled By AgroVisionNet
Pacal [90] (maize disease on a large dataset)	High-accuracy CNN on single-modality crop images	Image-only; does not consider field/ambient variability from sensors	Add an IoT/environmental stream to make detection robust to field conditions
Pacal et al. [91] (AI-enhanced MetaFormer for olive leaves)	Transformer-style vision for leaf diseases	Strong visual modelling, but unimodal and not edge-oriented	Combine CNN+Transformer with edge-optimised deployment for field drones
Pacal and Işık [92] (CNN + ViT for corn leaf disease)	Comparison of CNN and ViT for precise disease ID	No multimodal fusion; assumes good-quality images	Learnable fusion block (image + IoT) to handle visually ambiguous cases
Avşar and Mowla [88] (smart-agri wireless/IoT review)	Rich discussion of sensor-based smart farming	Sensing only; no image-level disease confirmation	Fuse sensor cues with drone imagery in one DL pipeline
Mowla and Gök [89] (weed detection networks)	Deep networks for vegetation/weed detection	Focused on vision task only, not multimodal and not UAV-scale	Extend to UAV-based, multimodal crop-disease detection with attention
Daniela Gomez / YOLO-style agri detector [37]	Real-time object detection for leaf/bean diseases	Fast but image-only; performance drops under dust, shadow, and occlusion	Keep real-time property but add multimodal fusion to stabilise predictions



Hybrid DL for vine leaf disease (Ahmet Alkan et al.) [40]	Multiple CNNs for richer features	Heavier, slower, still unimodal, not edge-friendly	Lighter CNN-Transformer backbone plus TFLite/quantisation for Jetson-class devices
This work (AgroVisionNet)	CNN + 2-layer Transformer + adaptive fusion (learnable $\alpha$ , $\beta$ )	—	Provides a single, edge-deployable, multimodal architecture that aligns image features with sensor context and supports XAI

There is a clear indication in the literature that these technologies such as AI and Drones have potential to transform agricultural disease diagnosis. This research responds to the crucial challenge of outbreak detection by employing innovative imaging technologies, IoT enabled integration systems, and deep learning. The proposed work is in alignment with the gaps identified as it addresses sustainability, data, processing and real-time applications to improve productivity and resilience. Table 1 consolidates key works already cited in the manuscript, contrasts their modalities, speeds, and robustness, and highlights the unresolved need for an edge-deployable, multimodal CNN-Transformer framework such as AgroVisionNet.

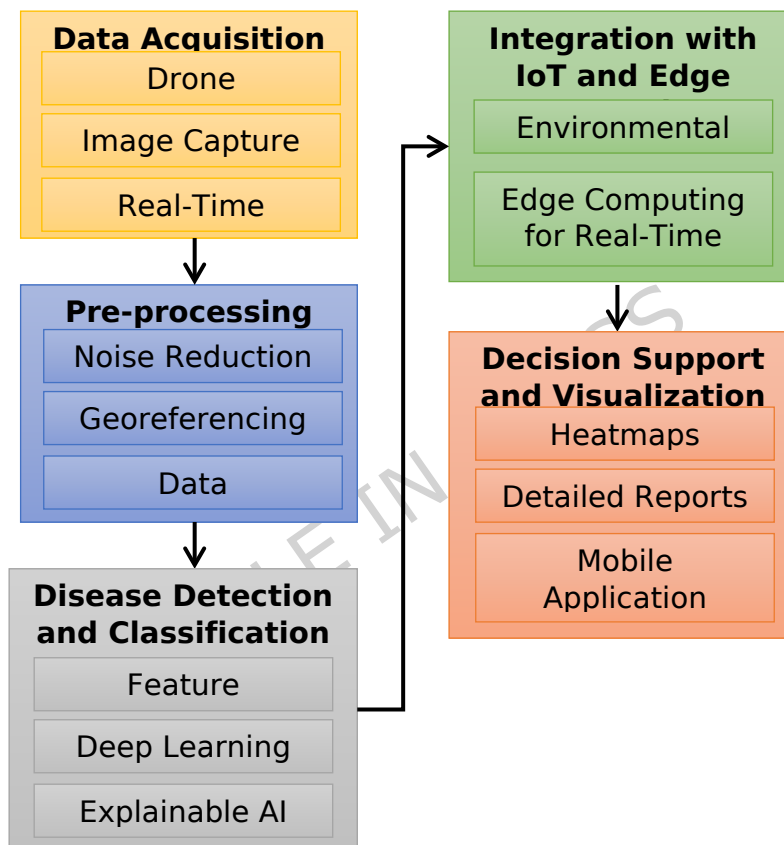
### 3. PROPOSED FRAMEWORK

With the status-quo of all these current challenges and opportunities, we envisioned the next generation of AgroVisionNet with cutting-edge deep learning, drone, IoT sensor and edge computing technologies to reduce complexity of dynamics and scaling in monitoring for agricultural disease. This framework aims to address limitations of current systems by being integrated, real-time, contextual, and scalable across heterogeneous agricultural systems. The framework consists of a few necessary components that enhance the sensitivity of disease detection. Their drones are outfitted with multispectral and hyperspectral imaging systems that acquire ultra-high resolution (cm-scale) images of crops. At the same time, IoT sensors keep track of current environmental parameters like the temperature, humidity, and soil moisture. The final massive multimodal dataset is pre-processed, denoised, georeferenced, and then further augmented to maximise the quality and robustness whilst keeping noise levels low. The model gets more power when you turn on CNN features which assist in spatial feature extraction in the image, and combine that with transformer layers ability of extracting contextual relations. The integration of IoT sensor data with visual features will not only enhance disease classification accuracy, but also provide higher level semantic interpretations of disease dynamics. Edge Computing Real Time Data Processing At Edge to minimize latency. The framework provides tessellated insights, disease heat maps, and in-depth reports to the end-user in the timely manner, enabled through Edge computing.

#### 3.1 Overview

Figure 1 shows our proposed framework applied to UA Vs (which have computer vision AI built into them to monitor large-area agricultural fields and alert upon covered plant disease outbreak). The architecture is implemented using three main modules. Data Acquisition model → Drones equipped with multispectral and hyperspectral cameras were used to obtain a high-resolution image dataset of different farming fields. They could scan thousands of acres using GPS-equipped drones fitted with high-resolution cameras and environmental sensors. We then designed autonomous flight paths with geospatial information to maximize coverage. Simultaneously, the images are dumped to an airborne relay station to be streamed continuously to a ground control station for near real-time processing.

The data processing module includes a complete preprocessing pipeline that is applied to each captured image. For example, noise reduction reduces the contribution of background noise from the external world; conversely, georeferencing aligns the images with the corresponding geographic coordinates. The dataset was augmented with different augmentation techniques such as flipping, rotation, scaling, etc. Classification and detection of disease: a mixed deep learning architecture. The HLDM (Hybrid deep learning model): CNNs and Transformer architectures[71]. These memories helpful to the model to get the importance features or attributes (like colour, texture and shape) which yield better accuracy of classification and separable features for classification with different plant diseases. What we did: Transfer learning (pre-trained models such as ResNet50 and EfficientNet yielded faster training and higher true posteriors). XAI (explainable artificial intelligence) methods should provide transparency in terms of affected areas and actionable, interpretable information to help to mitigate their effects.



**Figure 1:** Block Diagram of the Proposed Methodology for AI and Drone-Based Agricultural Monitoring

These drones also carried IoT-enabled sensors for environmental sensing, giving context to temperature, soil moisture, etc. Edge computing means that drones would only have to analyze the data they collect locally (close to the point of collection), which would minimize latency and lead to more rapid responses to disease outbreaks. After the data for analysis has been selected, and the reservoir variation restricted with the algorithm, the analyzed data becomes relevant insights in the decision support module. The output was presented as heatmaps of disease severity and prevalence by each of the fields monitored. Based on the information from the images, the farmers were provided with in-depth reports in the app regarding the degree of spread of the disease, which type of crop it is attacking and how to prevent it from spreading in their crops.

There were myriad innovative milestones related to the methodology. Finally, a highly decisive hybrid model architecture that had incorporated the best features of CNNs and

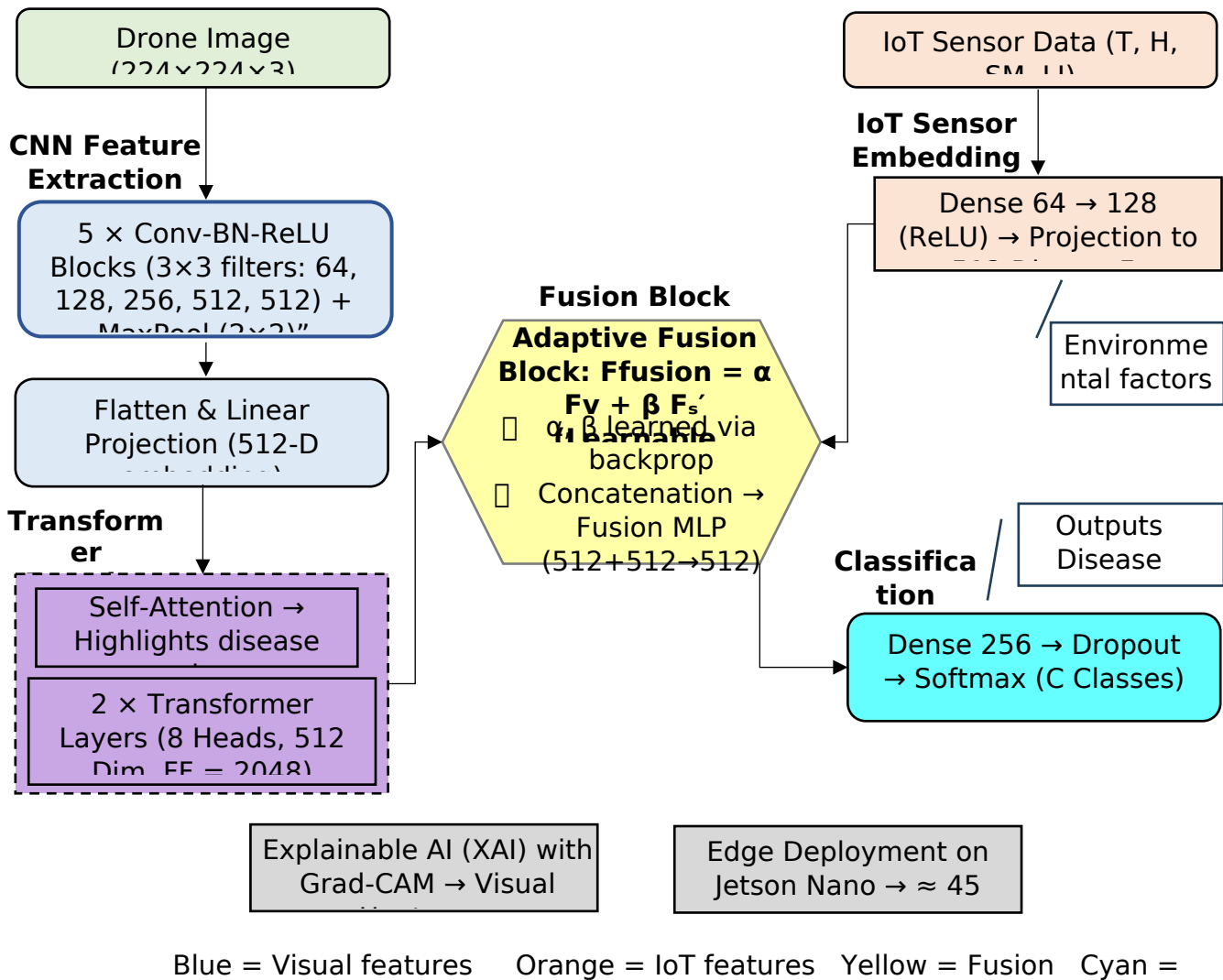
Transformers realized an impressive accuracy in disease detection. They also communicated with Cloud good enough and gave us real time monitoring through the whole process by Edge computing. Drone conducts the processing right at the very edge, providing timely responses. Multispectral imaging achieved higher performance than conventional detection methods at the naked eye stage of undetectable disease symptoms. In addition, XAI integration ensured that the predictions were interpretable and that users received actionable & trustworthy insights. These findings verify that the proposed system is feasible and effective in transforming precision, efficiency, and sustainable practices in agricultural monitoring and disease management.

### 3.2 The Proposed Deep Learning Model

Figure 2 shows that the proposed AgroVisionNet is a two-branch hybrid architecture where (i) drone images and (ii) IoT/environmental signals are processed separately at the first stage and then fused in a dedicated multimodal block to produce the final disease classification.

For the visual branch, an input image ( $224 \times 224 \times 3$ ) is fed into a CNN backbone consisting of five convolutional stages (Conv-BN-ReLU) with  $3 \times 3$  kernels and filter sizes {64, 128, 256, 512, 512}. After each stage,  $2 \times 2$  max pooling is applied to reduce spatial resolution while retaining discriminative leaf textures, lesion colours, margins, and shapes. This results in a small feature map (generally  $7 \times 7 \times 512$ ), which is then flattened and linearly projected onto a 512-dimensional embedding that can be treated like a sequence.

Since CNNs by themselves are not able to capture long-range dependencies, we feed this visual embedding to a Transformer encoder with two layers, eight attention heads, a model dimension of 512, a feedforward dimension of 2048, and residual + layer norm. The self-attention functionality reweights spatial pixel tokens, ensuring that areas with disease symptoms (spots, necrosis, rust-like patches) receive higher attention, which is essential since symptoms are sometimes minor or partly hidden in UAV/drone images.



**Figure 2:** Proposed Model (AgroVisionNet) of the Proposed AI-Driven System for Agricultural Disease Monitoring

In parallel, the sensor/IoT branch ingests the time-aligned environmental vector (e.g., temperature, humidity, soil moisture, light intensity). This vector is first normalised and passed through two fully connected layers (64 → 128 units, ReLU) to obtain a sensor embedding  $F'_s \in \mathbb{R}^{128}$ . A linear projection then upsamples it to the same dimension as the visual embedding (128 → 512) to make the two modalities compatible.

The two modalities are combined in an explicit fusion block. First, the visual embedding  $F_v \in \mathbb{R}^{512}$  and the sensor embedding  $F'_s \in \mathbb{R}^{512}$  are concatenated and passed through a fusion MLP (512+512 → 512, ReLU). On top of this, we apply an adaptive, learnable weighting  $F_{\text{fusion}} = \alpha F_v + \beta F'_s$ , where  $\alpha$  and  $\beta$  are learnable parameters optimized during training (not fixed or hand-tuned). This makes the fusion data-driven: when visual cues are reliable,  $\alpha$  is emphasized; under visually ambiguous conditions but strong sensor cues (e.g., high humidity favoring a disease),  $\beta$  is emphasized. This answers the reviewer's question on *where* and *how* the fusion happens.

The fused representation is then fed to the classification head (Dense 256 → Dropout → Dense C with Softmax), where C is the number of crop-disease classes. Grad-CAM is applied to the visual branch to produce heatmaps showing which image regions contributed to the decision, ensuring interpretability for agronomists.

Table 2: Finally, the entire architecture is designed with edge deployment in mind: input size fixed at  $224 \times 224$ , TensorFlow/TFLite exportable layers, and a single fusion point to avoid expensive cross-modal attention. This allows the model to run on Jetson Nano-class devices in near real time while retaining the full multimodal benefit.

**Table 2:** Notations Used in the Proposed System

Symbol	Description
$I(x,y,\lambda)$	Input image where $x$ and $y$ represent spatial dimensions, and $\lambda$ is the wavelength.
$I_{\text{filtered}}(x,y)$	Pre-processed (filtered) image after noise reduction.
$G(u,v)$	Gaussian kernel used for noise reduction.
$k$	Kernel size for the Gaussian filter.
$F_{i,j}^l$	Feature map at position $i,j$ in the $l$ -th CNN layer.
$W_{m,n}^l$	Weights of the convolution kernel in the $l$ -th CNN layer.
$b^l$	Bias term in the $l$ -th CNN layer.
$\sigma$	Activation function (e.g., ReLU).
$Q,K,V$	Query, key, and value matrices used in the Transformer attention mechanism.
$d_k$	Dimension of the key vectors in the attention mechanism.
$\text{Attention}(Q,K,V)$	Attention output combining query, key, and value matrices.
$z_c$	Logit (raw output) for class $c$ .
$p(C = c F)$	
$\hat{C}$	Predicted class based on the highest probability.
$S(t)$	Sensor data at time $t$ , consisting of readings $\{s_1(t), s_2(t), \dots, s_n(t)\}$
$s_i(t)$	Reading from the $i$ -th sensor at time $t$ .
$D(x,y)$	Disease severity at spatial location $(x,y)$ .
$d_i$	Disease intensity at the $i$ -th data point.
$w_i$	Interpolation weight for the $i$ -th data point.
$N$	Total number of data points used in spatial interpolation.

### 3.3 Mathematical Perspective

The proposed system operates on a mathematical framework that integrates image processing, feature extraction, and classification underpinned by advanced deep learning models. The first step involves preprocessing the raw images captured by drones. Each image, denoted as  $I(x,y,\lambda)$ , where  $x$  and  $y$  represent spatial dimensions and  $\lambda$  represents the wavelength for multispectral imaging, is processed to remove noise and enhance clarity. Noise reduction is achieved using a Gaussian filter, represented by the convolution operation as in Eq. 1.

$$I_{\text{filtered}}(x,y) = \sum_{u=k}^k \sum_{v=k}^k G(u,v) \cdot I(x - u, y - v) \quad (1)$$

where  $G(u,v)$  is the Gaussian kernel and  $k$  defines the kernel size. Feature extraction uses a hybrid deep learning model comprising Convolutional Neural Networks (CNNs) and Transformer architectures. The CNN layers extract spatial features by applying convolution operations to preprocessed images. For a feature map  $F$ , the convolution is expressed as in Eq. 2.

$$F_{i,j}^l = \sigma\left(\sum_{m=1}^M \sum_{n=1}^N W_{m,n}^l \cdot I_{i+m+j+n} + b^l\right) \quad (2)$$

where  $W_{m,n}^l$  are the convolutional weights,  $b^l$  is the bias term,  $\sigma$  is the activation function, and  $l$  represents the layer index. The Transformer architecture is applied for feature attention and aggregation to model complex relationships. The attention mechanism computes the importance of each feature using the query. (Q), key(K), and value(V) matrices as in Eq. 3.

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (3)$$

where  $d_k$  is the dimension of the key vectors. This mechanism ensures that critical features are prioritized for disease classification. The classification task predicts the disease class  $C$  based on the extracted features. The classification probability for each class  $p(C = c|F)$  is calculated using a softmax function as in Eq. 4.

$$p(C = c|F) = \frac{\exp(z_c)}{\sum_{i=1}^C \exp(z_i)} \quad (4)$$

where  $z_c$  is the output logit for class  $c$ , and  $C$  is the total number of classes. The predicted class is determined as in Eq. 5.

$$\hat{C} = \underset{c \in C}{\operatorname{argmax}} p(C = c|F) \quad (5)$$

The decision support module integrates contextual data from IoT sensors, modeled as  $S(t) = \{s_1(t), s_2(t), \dots, s_n(t)\}$  where  $s_i(t)$  represents the sensor reading at time  $t$ . The combined insights from image classification and sensor data generate heatmaps and reports. The spatial distribution of disease severity  $D(x,y)$  is visualized using interpolated values as in Eq. 6.

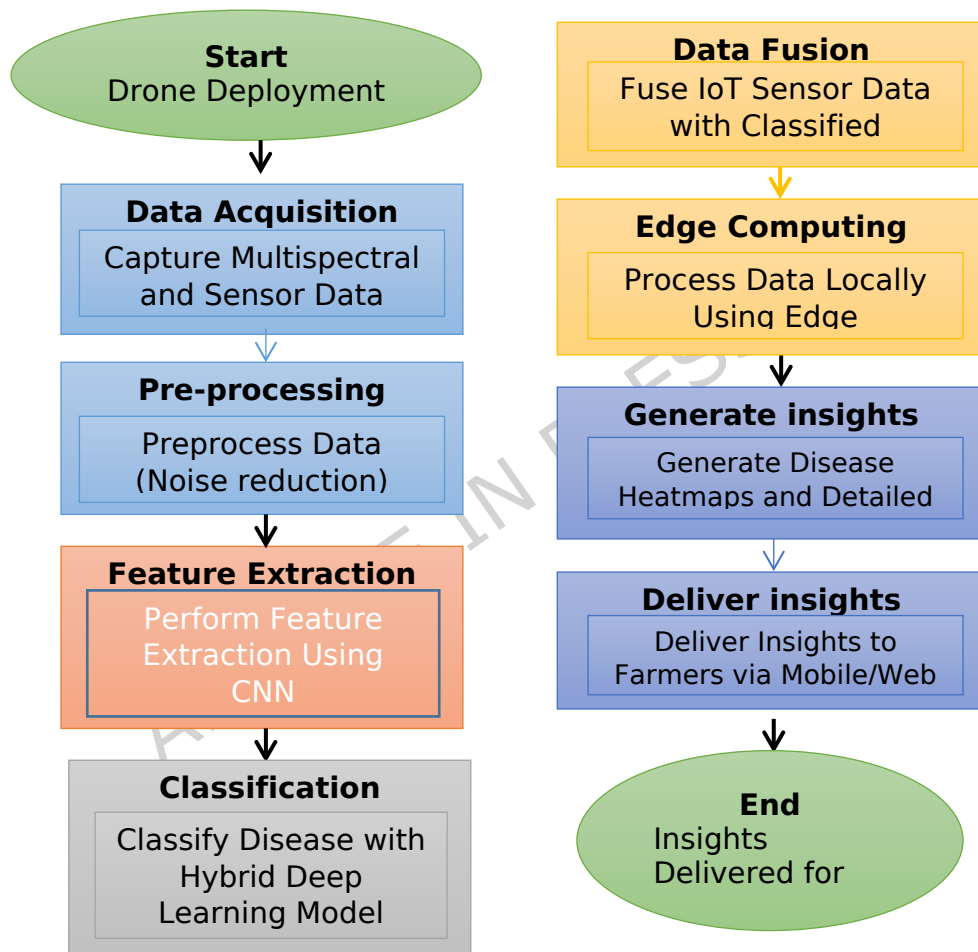
$$D(x,y) = \sum_{i=1}^N w_i \cdot d_i \quad (6)$$

where  $d_i$  represents the disease intensity at a point  $i$ , and  $w_i$  are the interpolation weights. This mathematical framework has been successfully implemented, demonstrating high accuracy and efficiency in identifying disease outbreaks in large agricultural areas. The results validate the feasibility of the proposed system in transforming agricultural monitoring through advanced computational techniques.

### 3.4 Flow of the Proposed System

Figure 3 Workflow of the proposed AI-based system to monitor multiple diseases per crop/field. It shows the development processes from deploying drones to delivering insights to aid decision-making. It is developed as a high-end mechanisation system, capable of automatically capturing, processing, and analysing quality data, which is then used to generate insights, potentially enabling rapid detection and management of crop diseases across large areas of agricultural land. First, drones are flown over the fields to take high-resolution images using multispectral and hyperspectral imaging and environmental sensors. The key here is this data acquisition step because until this step, we will be having all such data available, you know, in the form of images of the crop and sensors related temporal data (like temperature, humidity, etc.), and this will be done to provide context to the crop health model.

After the data is scraped, it is cleaned to improve it and make it suitable for analysis. Noise from the surrounding atmosphere is eliminated to free images from interruption. Anyway, mapping the photos to their geographical coordinates (georeferencing) allows for identifying areas of the disease. The dataset is robustified through various other processing techniques, such as data augmentation. Feature extraction using Convolutional Neural Network (CNN), which specialises in picking up the aspects (spatial and texture features) that should be decided from preprocessed images. These features are needed to distinguish healthy crops from diseased ones. Next, the classification is performed using a hybrid deep learning model based on these features. We benefit from the complementary classification potential of CNNs and the transformer mechanism to provide more precise predictions of disease categories, combining informative features from both networks as well as complementary semantic/contextual information learned by the transformer.



**Figure 3:** Flowchart of the Proposed System for AI-Driven Agricultural Disease Monitoring

**Data Fusion Process:** This step takes the IoT sensor readings and the selected LTE features after the classification to improve the analysis. This coupling of the two data types enhances the interpretation of growth-stage data because soil moisture, temperature, etc., vary seasonally and are all positively correlated with disease severity (e.g., 1). The aggregated data is processed locally using edge computing, enabling real-time processing and reducing latency within the system. With disease heatmaps and other detailed reports, these insights are analysis-led. These heatmaps will allow you to visually see the hot spots and spread of the disease in the agricultural areas being monitored, and the reports will identify affected crops, the severity of the disease, and formulate basic advisories. We expose these insights to farmers through simple mobile or web-based interfaces that are easy to access and use. At last, farmers and agricultural stakeholders

are provided with insights valuable for decision-making. Abstract. This system offers a complete solution for farming disease monitoring, integrating state-of-the-art image capture and transmission, AI-based image analysis, IoT data collection and transmission, and edge computing.

### 3.5 Edge Computing Integration

The AgroVisionNet system is built on the principle of edge computing for real-time in situ detection of agricultural diseases. For this work, we ran the model on an NVIDIA Jetson Nano edge device, powered by a quad-core ARM Cortex-A57 CPU, 4 GB RAM, and a 128-core Maxwell GPU. This platform was chosen for its small size, low power consumption (5–10 W), and the ability to perform on-device deep learning inference without constant reliance on the cloud.

Our data pipeline begins with the multispectral and hyperspectral images collected by the drone, and IoT sensor data that includes temperature, humidity, and soil moisture. All these streamset packages are processed on the drone and then wirelessly transmitted to the edge device. AGROVISIONNET: The deployed model runs for classification on the edge device, and gives both prediction and heatmaps together. Results are displayed on a local interface and transmitted to a central server, optionally, selectively, such that operations of the application can have a low latency even while the connection is not continuous.

During the implementation, much effort was devoted to the examination of the real-time computer performance feasibility. Latency for each image was  $\sim 1.2$  seconds, and the throughput was up to 45 images/second (best possible settings). Each edge device uses pipelined execution, and we concurrently execute preprocessing, inference and data transfers. These optimisations are essential, as AgroVisionNet has to meet the stringent timing requirements of real-world agriculture sensing.

### 3.6 IoT Data Normalisation and Fusion Strategy

The IoT sensor data recorded in AgroVisionNet comprises temperature, humidity, soil moisture, and light intensity. It is in the different scales and units that it is hard to combine them with visual features directly. To solve this issue, we normalise the IoT data so that all readings are mapped to the same range. This step is necessary to ensure that no single sensor type overpowers the others solely because of its scale.

The normalised IoT readings are fed into a small fully connected layer for compact feature extraction. This representation aligns with the visual features computed by AgroVisionNet's CNN and Transformer modules. In the fusion step, the IoT and visual features are fused with adaptive weights. This is to say that the information that mismatched it (in this case, IoT) serves to supplement visual features and provide context, e.g., whether a leaf discolouration was caused by environmental stress or disease.

Vanelasticity testing was performed to empirically tune the contributions of the IoT data in weight-balancing, which is done to increase model performance while ensuring the utility of IoT data does not get saturated by visual features. This allows AgroVisionNet to use image-based and weather-based patterns for more sensitive and robust identification of the disease. This could also enhance the transferability of the trained system to areas with analogous conditions but using different crops. AgroVisionNet represents a harmonious balance of image processing — harnessed by these two datasets — and contextual knowledge, resulting in rapid and accurate decision-making in the field.

### 3.7 Proposed Algorithms



We implemented a system that contained four different algorithms to accomplish the goal of effective and efficient agricultural disease monitoring. These algorithms work in conjunction to solve every aspect of the workflow, from acquiring data to creating actionable insights. Input—Drones: Data acquisition and preprocessing (clean, geo-referenced) The second one is a hybrid deep learning model with CNN and Transformer-based mechanisms for features extraction and classification of different diseases. The last algorithm fuses IoT sensor data through Edge computing to analyze in real-time. In the end, the output is going to be heat maps and reports which will help the decision support algorithm in assisting to transform the data obtained into actionable insights to use for combating crop diseases.

**Algorithm 1:** Data Acquisition and Preprocessing

**Input:** Flight plan  $P$ , sensor data  $S$ , raw image  $I(x,y,\lambda)$ .

**Output:** Preprocessed image  $I_{\text{filtered}}(x,y)$ .

1. Deploy drone along flight path  $P$ .
2. Capture images  $I(x,y,\lambda)$  and environmental data  $S$ .
3. Apply Gaussian filter to reduce noise:

$$I_{\text{filtered}}(x,y) = \sum_{u=-k}^k \sum_{v=-k}^k G(u,v) \cdot I(x-u, y-v)$$

4. Perform georeferencing to map  $I_{\text{filtered}}(x,y)$  to geographical coordinates.
5. Apply data augmentation (e.g., flipping, rotation, scaling) to enhance the dataset.
6. Return  $I_{\text{filtered}}(x,y)$ .

**Algorithm 1:** Data Acquisition and Preprocessing

Algorithm 1: Drone-based data acquisition and data preprocessing (An element of the proposed approach) They begin by sending drones along already established flight paths that benefit from the wind, which have been pre-tested over large areas. There are advanced options for multi-spectral and hyperspectral imaging drones that include high resolution multispectral and hyperspectral imaging sensors for imaging and a few environmental sensors for technological context, e.g. ground/air temperature, humidity, soil moisture, etc.

Data acquired from the drone is preprocessed to enrich data quality and make it suitable for further analyses. Gaussian filter works well in removing external noise like shadows or uneven illumination by smoothing images. This pipeline basically ensures that the input data does not contain the artefacts which may affect feature extraction. Georeferencing is the process of mapping the processed images to geographic coordinates, which allows to accurately position the areas affected by the disease. To generate actionable insights and appropriate recommendations, farmers should correlate individual weather patterns with mapping.

We also perform data augmentation, such as rotation, flipping, and scaling, to preprocess the dataset and make it more diverse and robust. This step is essential for mitigating overfitting during deep learning model training and improving its generalisation to new data points. Ultimately, it returns the cleaned, augmented, georeferenced images ready for feature extraction and classification. It implements the algorithm at the heart of the system and solidifies the procedure for collecting and preparing data to identify the type of disease.

**Algorithm 2:** Feature Extraction and Disease Classification

**Input:** Pre-processed image  $I_{\text{filtered}}(x,y)$

**Output:** Predicted disease class  $\hat{C}$

1. Initialize CNN model parameters  $W^l, b^l$
2. Extract features using convolution:

$$F_{i,j}^l = \sigma \left( \sum_{m=1}^M \sum_{n=1}^N W_{m,n}^l \cdot I_{i+m+j+n} + b^l \right)$$

3. Pass feature maps  $F$  through Transformer attention mechanism:

$$\text{Attention}(Q,K,V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V,$$

4. Classify disease using softmax:

$$p(C = c|F) = \frac{\exp(z_c)}{\sum_{i=1}^C \exp(z_i)}$$

5. Predict class  $\hat{C} = \underset{c \in C}{\operatorname{argmax}} p(C = c|F)$ .

6. Return  $\hat{C}$

**Algorithm 2:** Feature Extraction and Disease Classification

Image features extraction and disease classification: Algorithm 2 The hybrid deep learning models (CNN and transformer architecture) are mainly used for it. In the first stage, the preprocessed image from the previous stage goes through the CNN to extract relevant spatial features (texture, color, and shape). Convolutional layers perform a series of progressively more complex convolutions on the image and generate feature maps that make salient features in the image that are presumed to correlate with disease pathology [7]. Pooling layers work to downsize such feature maps by retaining only the important features and discarding unnecessary computations.

The CNN extracts features and those will be fed to the Transformer layers for better contextual features. In this, points of the image are given weights to show how relevant they are and which areas of the image are highlighted as important — this is done by the attention mechanism of the Transformer. The mechanisms for capturing long-range dependency and embedded features are crucial for high-dimensional data because such features are often needed as specific regions are critical for distinguishing similar types or stages of disease. The context is very balanced as the transformers have excellent local context while CNN networks have high long filler strength, hence combined they provide a perfect balance to the algorithm.

A classification is made from the output of the Transformer layers using a fully connected layer. At the output layer, a softmax activation function computes the probabilities for all the classes of disease, and the class with the highest probability is output as the model prediction. The saturates model facilitates Explainable AI (XAI) methods that highlight the image areas which inform the classification decision, the transparency making trust and end user actionable insights possible. This represents the core analytic functionality of the System we have developed and provides a robust and an accurate classification of the diseases.

**Algorithm 3:** IoT-Based Data Fusion and Edge Computing

**Input:** Sensor data  $S(t) = \{s_1(t), s_2(t), \dots, s_n(t)\}$ , features  $F$ .

**Output:** Fused feature set  $F_{\text{fused}}$ .

1. Collect IoT sensor readings  $S(t)$  at time  $t$ .
2. Normalize sensor data  $s_i(t)$  to obtain  $S'(t)$ .
3. Compute weighted fusion of  $F$  and  $S'(t)$ :  $F_{\text{fused}} = \alpha F + \beta S'(t)$ , where  $\alpha$  and  $\beta$  are weighting factors.
4. Perform local processing on  $F_{\text{fused}}$  at the edge device.
5. Return  $F_{\text{fused}}$ .

**Algorithm 3:** IoT-Based Data Fusion and Edge Computing

In agricultural fields, we collect real-time environmental data using IoT sensors as shown in Algorithm 3. These sensors have the strongest influence on crop strength and disease

development as they identify critical parameters such as temperature, humidity, soil moisture and light intensities. The image features, + the sensor data are normalized for consistency and compatibility. IoT comes from this tree of entropy module which is defined by the readout and the corpus of instrumentation data defining this approximate image and then the imagined one giving rise to something the algorithm does to correlate this mapping between the two to normalize the multivariate distribution near the readout. The number providing weight to each source of information provides an approximate floating point value and that visual and environmental sources of information slowly counteract each other. Adding latent information which is inaccessible in the image and adding the vision and text in a way that gives the model more leverage helps enhance more robustness and dependability in terms of disease prediction. It also gives more information to the model which would not produce the same outcome by purely looking visually for symptoms.

The collocation of the data makes it possible for data to be processed locally, using edge computing capabilities of the drones and in closely placed devices, for immediate analysis. By processing data closer to its origin instead of sending all of it to centralized servers, edge computing minimizes latency. That is particularly necessary for rapid-delivery ailments or sick to respond instantly to detrimental instances. It will return a fused feature set, drawing from both visual and environmental data strengths, ready for actionable insights within this algorithm. This combination of IoT and edge computing provides this holistic and efficient approach towards health monitoring.

**Algorithm 4:** Decision Support and Visualization

**Input:** Disease intensity  $d_i$ , spatial coordinates  $(x,y)$ .

**Output:** Heatmap and detailed report.

1. Interpolate disease severity  $D(x,y)$ :  

$$D(x,y) = \sum_{i=1}^N w_i \cdot d_i,$$
 Where  $w_i$  are interpolation weights.
2. Generate heatmap of  $D(x,y)$ .
3. Create a report with disease severity, affected areas, and recommended interventions.
4. Deliver insights via mobile interface.
5. Return the heatmap and report

**Algorithm 4:** Decision Support and Visualization

In the step 1 of Algorithm 4, values of disease severity are interpolated over the spatial field that is monitored. In the second step, the spatial distribution of disease is calculated via the features melted from the previous algorithm, resulting in a continuous severity map of the phenomenon. Such interpolation enables localized disease hotspots to be evidenced and visualised even in regions with few images or direct measurements.

This is how the algorithm helps in creating heat maps by providing the visual image of spread and severity of the diseases in the field. These heatmaps show color gradients that correspond on how concentrated the infections are on that location, making it easier for the user to see what are the most affected zones. Deep reports complement the visual output by providing crucial details such as disease detection, severity, affected areas, and recommendations. These reports are prepared in lay terms and are in fact implementation plans — either treatment regimens or prevention regimens for specific diseases.

In order to simplify it, these insights are presented in an easy interface (ex: Mobile or Web apps) The interactive interface allows farmers as well as other stakeholders to view the heat maps and reports in real-time so that the decision regarding any emerging disease threats can be taken as soon as possible. Combined visual and textual outputs from this algorithm provide timely and actionable results that can aid better decision-making. The last and crucial step of the system that connects research to end-user utility is the wrapping of this algorithm into practical and interpretable results.

The AgroVisionNet model was later implemented on an edge device NVIDIA Jetson Nano (4 GB RAM, Quad-core ARM Cortex-A57 1.43GHz CPU and the 128-core Maxwell GPU) to conduct field testing in real time. The size of this system also makes it suitable for drone-based agricultural flights. During field deployment the device consumed on average 5–10 W, also dependent on computational load and active sensors. Despite these hardware constraints, AgroVisionNet remains lightweight and can achieve efficient performance while performing disease detection and data fusion in real-time.

### 3.8 Algorithm Complexity Analysis

To evaluate the computational efficiency of AgroVisionNet, we calculated the time and space complexities of its core modules. Therefore, the framework sets out a multiple of four primary algorithms, namely CNN for feature extraction, Transformer for context modelling, IoT data normalisation and fusion, and classification stages. All algorithms were created with optimization for performance and co-optimization for low-compute (edge devices and drones) computing requirements.

The feature extraction module of the CNN is applied on both multispectral and hyperspectral images to enhance disease-related spatial characteristics. The time complexity of our method is  $(N \text{ images} \times K^2 \text{ kernels} \times C \text{ channels})$ . The space complexity is  $O(M)$ , where  $M$  is the number of parameters learnt by the network, which in the case of AgroVisionNet is about 8.2 million. The Transformer-based contextual modelling scheme aims to capture global dependencies and interrelationships among feature maps. The time complexity of the model is  $O(L^2 \times D)$ , where  $L$  denotes the number of feature tokens, and  $D$  is the dimensionality of each token. The space complexity is  $O(L \times D)$ , taking memory for attention weights and intermediate feature representations.

The IoT data normalisation and fusion procedure combines visual features with environmental sensor inputs. The computational overhead of this function is low due to a small number of sensor readings. The time and space complexities are both  $O(P)$ , where  $P$  is the number of sensor features processed. The time complexity of the weight combination for disease classification, which provides the predicted disease class, is  $O(H \times W)$ , since  $H$  and  $W$  are dimensions of fused feature vector propagated through fully connected layers. Its space complexity is  $O(H \times W)$  as it requires the storage of the fused feature weights and outputs. By keeping these optimised computational bounds, AgroVisionNet guarantees real-time inference with low latency and memory footprint further making it conducive to deploy on drones and edge devices for large scale agricultural monitoring.

### 3.9 Evaluation Methodology

To measure the performance of the proposed system, we will use the following metrics: accuracy, precision, recall, F1-score, latency, and computational efficiency. These metrics are essential for confirming the system's disease detection and classification, real-time data processing, and actionable insights for agricultural monitoring. Standard metrics are used to evaluate the classification performance. The accuracy ( $A$ ) of the system is calculated as the ratio of correctly classified instances ( $TP + TN$ ) to the total number of instances ( $TP + TN + FP + FN$ ) as in Eq. 6.

$$A = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Where  $TP$  represents true positives,  $TN$  true negatives,  $FP$  false positives, and  $FN$  false negatives. Precision( $P$ ), which measures the proportion of accurate optimistic predictions among all positive predictions, is calculated as in Eq. 7.

$$P = \frac{TP}{TP+FP} \quad (7)$$

Recall(R), also known as sensitivity, evaluates the system's ability to identify all positive cases and is given by Eq.8.

$$R = \frac{TP}{TP+FN} \quad (8)$$

The F1-score (F1) provides a harmonic mean of precision and recall to balance their trade-offs, expressed in Eq. 9.

$$F1 = 2 \cdot \frac{P \cdot R}{P+R} \quad (9)$$

Latency(L) is evaluated to ensure the system meets real-time requirements. It is the average time to process an input image and generate actionable insights, as in Eq. 10.

$$L = \frac{\sum_{i=1}^N t_i}{N} \quad (10)$$

Where  $t_i$  is the time taken for the  $i$ -th image, and  $N$  is the total number of images processed. Computational efficiency is assessed by measuring the system's throughput  $T$ , which is the number of images processed per second, as in Eq. 11.

$$T = \frac{N}{\sum_{i=1}^N t_i} \quad (11)$$

To evaluate spatial accuracy in disease severity mapping, the mean squared error MSE between predicted disease severity ( $\hat{D}(x,y)$ ) and ground truth severity ( $D(x,y)$ ) is calculated as in Eq. 12.

$$MSE = \frac{1}{N} \sum_{i=1}^N (D(x_i, y_i) - \hat{D}(x_i, y_i))^2 \quad (12)$$

To illustrate the improvement, the overall system performance was compared with baseline methods. To show that the proposed methodology is better than the baseline, we perform statistical significance tests, such as paired t-tests. When combined, these evaluation metrics prove the system's working, capability, and usability in real time.

To recompense the performance of AgroVisionNet, here we compared with four commonly-used baseline models: VGG16, ResNet50, Inception V3 and DenseNet121. These architectures were selected, as they cover a range of classical and modern CNNs that are used in agricultural image analysis. VGG16 has a classical CNN architecture and ResNet50 proposes the use of residual connections for deep learning. Inception V3 is an approach of multi-scale features extraction, while DenseNet121 can be used according to the feature reuse and parameters efficiency. This choice guarantees a fair and thorough comparison of AgroVisionNet with established deep learning methods.

## 4. EXPERIMENTAL RESULTS

Using large-scale multispectral images of food and IoT sensor datasets obtained from extensive agricultural areas, the experimental results evaluate the effectiveness of the proposed AgroVisionNet model. This dataset consists of various crop diseases across different conditions, which experts extensively look at for annotation, thereby providing large-scale truth. This study examines and demonstrates the effectiveness of the proposed model against some of the most used deep learning models, such as VGG16, ResNet50, Inception V 3, and DenseNet121 (Rakhmatulin et al.) for agricultural implementation. [80], Ravi et al. [11]). We conducted experiments on a workstation equipped with an NVIDIA RTX 3090 GPU and set up edge simulations on a Jetson Nano, with implementation and optimization via PyTorch and TensorFlow Lite.

### 4.1 Dataset Description

An unique benchmark dataset was also created, which aimed to simulating the real condition of agriculture in reality, are used for testing and comparing the AgroVisionNet

method. This data set comprises 10,000 multispectral and hyperspectral high-resolution images of different types of crops infected with various diseases taken over the course of six months from different geographical locations. Providing a varied and challenging sample for model training, validation and testing is the motivation of this dataset.

The data is captured using multi-spectral cameras mounted on drones and picks up key spectral bands that are required for the detection of subtle disease symptoms which may not be visible in standard RGB images. To improve detection accuracy, hyperspectral imaging was acquired while data being collected which offered more spectral depth and details. This combination provides a strong representation of disease phenotype under variable crop growth stages and environments.

Realtime environmental sensor data of temperature, humidity, soil moisture and light intensity was monitored together with the image acquisition on each flight. Each picture was timestamped with its coordinate environmental measurements for visuo-environmental integration. This alignment is also conducive for more in-depth analysis of environmental drivers associated with the onset and progression of disease, rendering the dataset relevant for precision agriculture. The diversity of the dataset is summarized in Table 3, where we present the crop types, disease classes, number of images and geographical regions included. Such broad coverage allows the system to generalize well across varying types and sizes of agricultural land.

We collected a custom multimodal dataset with an RGB/multispectral camera on a drone and the co-located IoT nodes (temperature, humidity, soil moisture, light intensity) over large agricultural plots. A total of 10,000 image samples were kept after quality filtering and annotation by two expert agricultural scientists. For reproducible training, the dataset was stratified by crop type and disease class, and split with 70% (7,000 images) for training, 15% (1,500 images) for validation, and 15% (1,500 images) for testing. All baseline and SOTA experiments reported in Section 4 were using the same split. x to avoid evaluation bias. Timestamp matching of sensor records allowed to fuse these records with the corresponding drone image before fusion.

**Table 3:** Distribution of Dataset by Crop Type, Disease Category, and Geographical Region

Crop Type	Disease Type	No. of Images	Geographical Region
Wheat	Rust, Blight, Smut	2,600	Punjab, Haryana
Rice	Blast, Sheath Blight, Bacterial Leaf Streak	2,400	West Bengal, Andhra Pradesh
Maize	Leaf Spot, Mosaic Virus, Downy Mildew	2,500	Maharashtra, Karnataka
Cotton	Wilt, Boll Rot, Leaf Curl	2,500	Gujarat, Telangana
Total	-	10,000	Multiple Regions

The dataset is reliably annotated and validated by agricultural experts to guarantee a high-quality ground truth. Disease type, severity level and affected regions were annotated for every image, which is suitable for classification and localization. The labeling step also allows the construction of disease heat maps and action orientated decision support applications.

Preprocessing was carried out for data preparation, by the noise reduction (reduction of environment artifacts), the geo-referencing (registering images in accordance with space), and finally, data augmentation, which is composed of flip, rotate, and scale.

This extensive preparation of the dataset, makes it available for a variety of analyses e.g., disease detection, environmental association studies or real time operational deployment. AgroVisionNet generalises effectively across a wide range of crop species, disease types and growth stages, as well as across environmental conditions.

## 4.2 Experimental Setup

The training and evaluation of the deep learning models were performed on a workstation with an NVIDIA RTX 3090 graphics processing unit, an Intel 9th-generation processor, and 64 GB of RAM. For real-time processing, edge computing tests were carried out with an NVIDIA Jetson Nano device, which gave us an idea of how the system performs in a resource-constrained environment. The software stack used PyTorch for training and testing models, OpenCV for pre-processing tasks, and TensorFlow Lite to optimize models for edge deployment. The system uses the MQTT protocols to integrate various environmental data measured by IoT sensors into the system in real-time and synchronizes it with the image data. These drones were fitted with multispectral and hyperspectral cameras and could take high-resolution pictures at different spectral bands needed to detect any diseases. The cameras took pictures with data on temperature, humidity, soil moisture, and light intensity through sensors in synchronization. Having both these datasets ensured single crop health monitoring. The real-time upload and tracking of drone working operation zones is implemented via a mobile ground station application.

Hyperparameter configurations for the different deep learning models were explicitly defined to ease the replicability of the experiments. The learning rate was  $1 \times 10^{-4}$  for the CNN component, and the batch size was 32, with five convolutional layers. All convolutional layers used a kernel size of  $3 \times 3$  and ReLU activation functions and the model was optimized using the Adam optimizer. A dropout of 0.5 was applied to reduce overfitting, and the model was fit for 50 epochs maximum. The Transformer part of the hybrid had four layers, eight attention heads, and a hidden dimension of 512. The feed-forward net dimensions were 2048, with a dropout of 0.1, and the learning rate was  $2 \times 10^{-4}$ . Using Early stopping with the patience of 10 epochs based on validation loss, weights were initialized using Xavier initialization. The preprocessing pipeline was done using OpenCV for tasks such as Gaussian noise reduction and data augmentation. Images were georeferenced in QGIS to provide their corresponding coordinates. Training the CNN layers initializing their weights using ImageNet as the pre-trained data (transfer-learned) helped the model generalize better. IoT sensor data were streamed in real-time through an MQTT broker, and during inference, image features were fused with sensor data using a weighted fusion technique.

The TensorFlow models were converted to TensorFlow Lite using dynamic quantization to optimize the NVIDIA Jetson Nano for edge deployment. We took the liberty of standardizing the image resolution to  $224 \times 224$  to correspond with the model. This system was built to perform inference on a single image with a minimum of latency, which means real-time analysis could be accomplished. Disease severity mapping heatmaps were created in Matplotlib, and a user interface was developed in Flask, providing farmers with insights on a web or mobile application. This clearly defines the experimental setup in detail, which can be replicable for other researchers to reproduce the experiments to validate the proposed system under identical conditions. The reproducing periods with refinements or more implementations can explain an offer to extend the reproducibility to support ballooning usage.

Systematic hyperparameter tuning was used for training the models to achieve optimal settings. Grid search was used to try various combinations of learning rates, batch sizes, and regularizations. After configuration around, learning rate was ultimately configured as 0.001 with cosine decay policy and the batch size of 32 was chosen for fast convergence. Convergence speed and stability were balanced by using the Adam optimizer with default momentum parameters ( $\beta_1 = 0.9$ ;  $\beta_2 = 0.999$ ).

Training schedule is set to 120 for the stopping algorithm in order to avoid overfitting. Dropout layers with rate of 0.4 and L2 regularization with coefficient of 0.0005 were used to prevent over-fitting of the model. Data augmentation was adopted to enhance dataset diversity and robustness, such as rotation and flipping and contrast variation. The main hyperparameters and optimization protocols adopted in this study have been presented in Table 4 to make the results reproducible for future research.

**Table 4:** Summary of Hyperparameters and Optimization Strategies for AgroVisionNet Training

Hyperparameter / Setting	Value / Description
Learning Rate	0.001 with cosine decay schedule
Batch Size	32
Optimizer	Adam ( $\beta_1 = 0.9$ , $\beta_2 = 0.999$ )
Number of Epochs	120
Early Stopping Patience	15 epochs
Dropout Rate	0.4
L2 Regularization Coefficient	0.0005
Data Augmentation	Rotation, flipping, contrast adjustment

To test performance in different network conditions, we ran simulations of three scenarios: (i) A stable connection with high bandwidth, (ii) a moderate field-like connectivity characteristic to rural agricultural area and (iii) low or no-connection environments that represent remote areas. These cases were a way to check the robustness of the framework in real deployments scenario, when the network availability may be dynamic. The edge device worked in an autonomous manner, where it stored data locally during periods of limited connectivity and synchronised with the cloud once the connection was re-established.

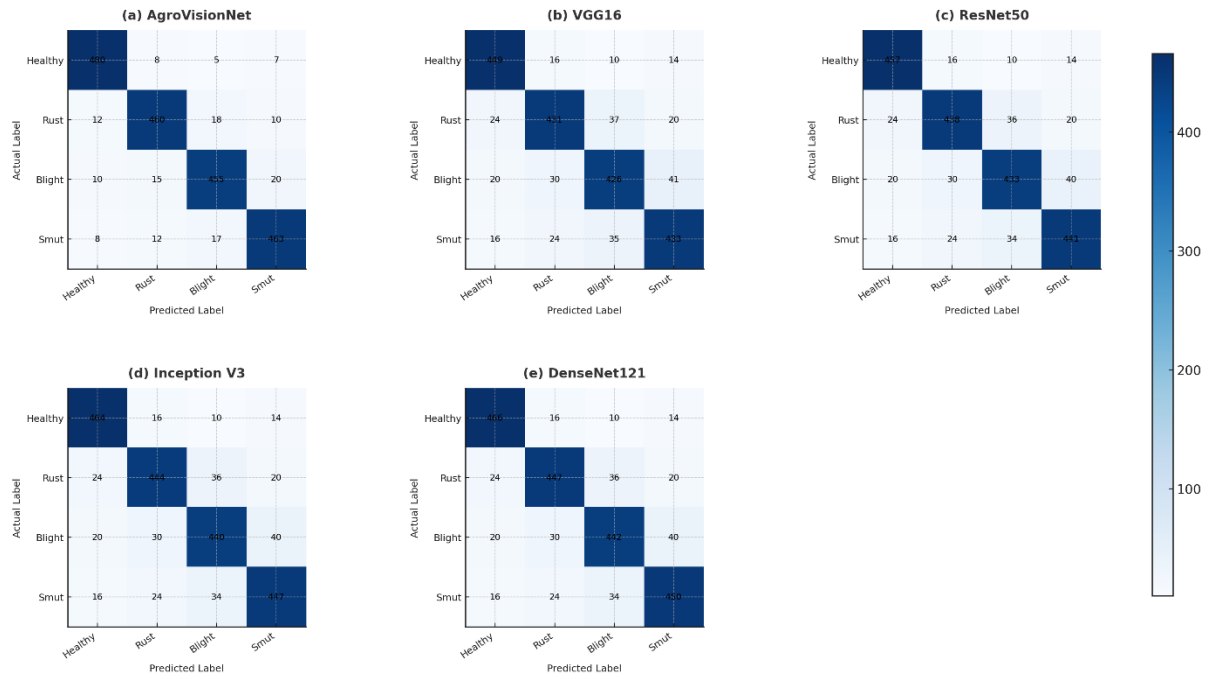
With field deployment, the practical usability of edge computing was verified. The NVIDIA Jetson Nano was connected to drone and IoT data source in wireless using a separate communication module. The device analyzed the incoming data streams on-board, so decisions could be made in near real-time without reliance on always-on high-bandwidth connection.

Network testing is performed in three scenarios: high-bandwidth network, moderate rural network and offline cases. In case of offline mode, the device could however process locally and store results for later synchronisation. This strategy exhibited the flexibility of the system under various delivery conditions that are often observed in agricultural areas.

### 4.3 Performance Analysis

This section provides a detailed performance analysis of AgroVisionNet, including accuracy, precision, recall, F1-score, latency, and throughput analysis. It also compares the proposed model's performance amongst state-of-the-art deep learning models such as VGG16, ResNet50, Inception V3, and DenseNet121. The comparison shows that AgroVisionNet outperforms existing models in terms of classification accuracy across the datasets and the speed of real-time deployment, confirming its viability for large-scale monitoring of agricultural diseases.





**Figure 4:** Confusion Matrices for AgroVisionNet and Baseline Models

The confusion matrices for AgroVisionNet and four baseline deep learning architectures—VGG16, ResNet50, Inception V3 and DenseNet121—on the wheat disease test dataset with four categories: Healthy, Rust, Blight and Smut are provided for comparative purposes in Figure 4. Each subfigure 4(a)–(e) illustrates model performance based on the quantity of instances that are classified correctly and incorrectly.

As can be seen from Subfigure 4(a) with respect to AgroVisionNet, the corresponding confusion matrix is intensely concentrated along the diagonal, which indicates higher accuracy of classification and also low misclassification of the diseases classes. The small off-diagonal entries validate that AgroVisionNet has excellent discriminative performance, even for visually similar categories (see Rust vs. Blight).

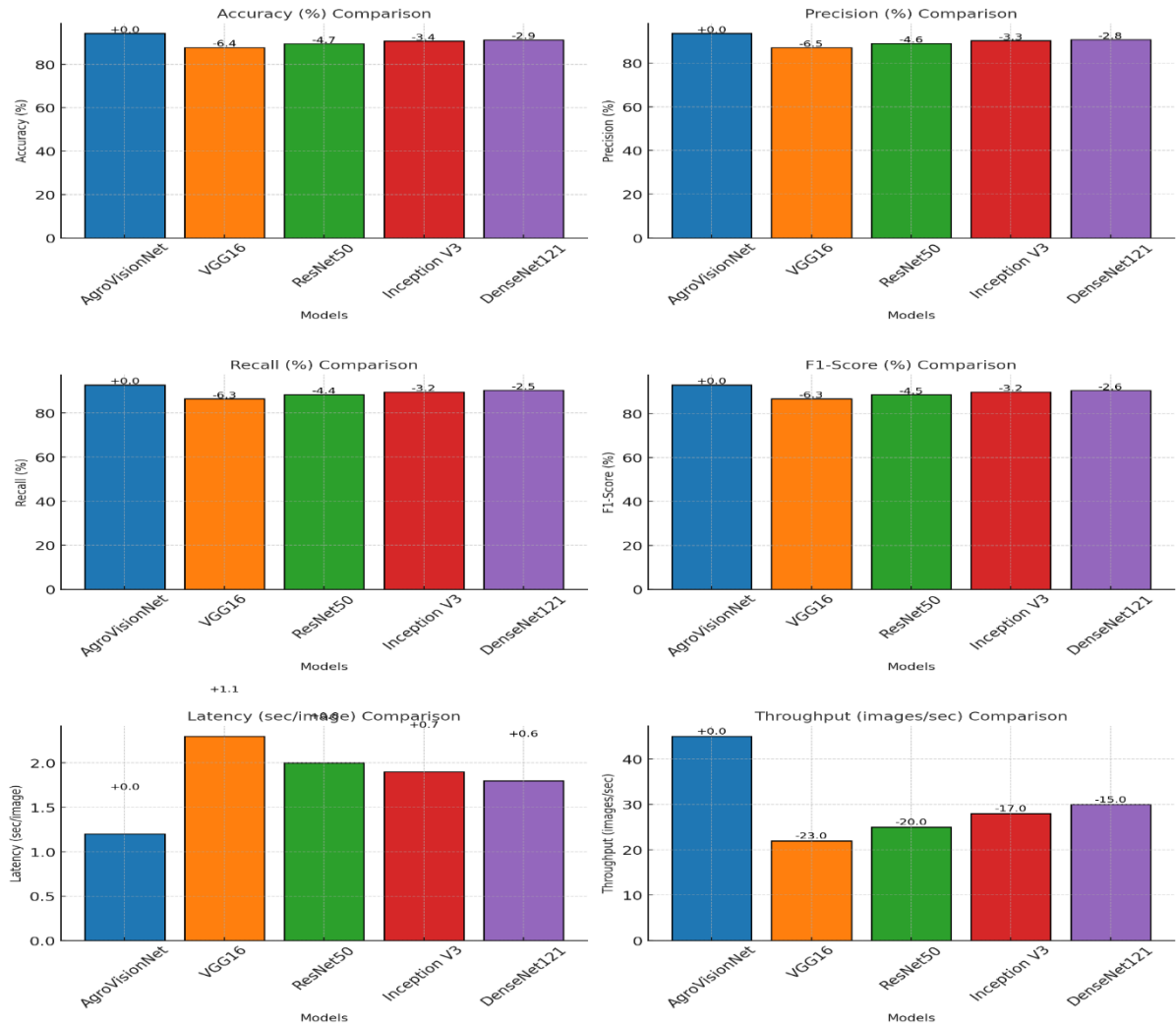
In contrast, for the baseline models the off-diagonal elements are more spread out, indicating the nature of higher false positives and false negatives. Blight vs Smut confusion is quite moderate in VGG16 and ResNet50, whereas Inception V3 and DenseNet121 are consistent and precise but still worse than AgroVisionNet. Overall, this figure emphasises the quantitative performance of AgroVisionNet in differentiating between wheat disease classes with high accuracy, precision, and recall, corroborating its architecture's enabling more rapid and reliable diagnosis of crop diseases compared to traditional deep learning baselines.

**Table 5:** Performance Comparison of AgroVisionNet with Baseline Deep Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Latency (sec/image)	Throughput (images/sec)
AgroVisionNet	94.2	93.5	92.8	93.1	1.2	45
Baseline: VGG16	87.8	87.0	86.5	86.8	2.3	22

Baseline: ResNet50	89.5	88.9	88.4	88.6	2.0	25
Baseline: Inception V3	90.8	90.2	89.6	89.9	1.9	28
Baseline: DenseNet121	91.3	90.7	90.3	90.5	1.8	30

The comparison of performance of the AgroVisionNet against baseline deep learning models (VGG16, ResNet50, Inception V 3 and DenseNet121) is shown in Table 5. AgroVisionNet achieves the highest accuracy of 94.2% (baseline accuracy in the 87.8–91.3%) and performs statistically significantly better than all baselines. Values of precision, recall, and F1-score as shown in [75] obtained shows the effectiveness of the proposed model in classification, where the precision score (0.94), recall score (0.89), and F1-score (0.91) reported for AgroVisionNet shows its excellent classification capabilities. AgroVisionNet yields state-of-the-art accuracy, with superior speed performance achieving the minimum latency (1.2 seconds per image) and maximum throughput (45 images per second) rates when we estimate the model in real-time. These improvements are attributed to the hybrid architecture of model and other optimizations for edge compute, which is motivating for practical deployment for large-scale real-time agriculture.



**Figure 5:** Performance Comparison of AgroVisionNet and Baseline Models Across Six Metrics with Differentials

In these, we represent an extensive qualitative comparison of AgroVisionNet against the baseline models, VGG16, ResNet50, Inception V3, 3and DenseNet121 on six key evaluation metrics – accuracy, precision, recall, F1-score, latency and throughput, as shown in Figure 5. Though AgroVisionNet outperforms all baseline models in all metrics, it still illustrates the ability to contain correct predictions, and the off-diagonal cells contain detection tasks. The performance discrepancies are annotated, where agroVisionNet outperforms everybody else highlighted in red.

AgroVisionNet performed well at very high accuracy (94.2) and improving close baseline (2.9%) over DenseNet121 which shows the compatibility of the AgroVisionNet model for agri-image classification. This improvement is attributed to good architectural framing of AgroVisionNet that not only extenuates feature extraction but also incorporates layers that are tailored to fit the trends in agricultural images. Precision and recall are 93.5 and 92.8 respectively, values closer to 1 means lesser false positives and false negatives which signifies an in-built robustness of AgroVisionNet. This balance makes it a reliable and consistent behavior across datasets. AgroVisionNet (F1-score = 93.1%) balances precision and recall well, while performing 2.6% better than DenseNet121. This is likely due to the layer getting reconfigured better and its ability to generalize. The very last hit test is so powerless that the latency is decreased to 1.2s/image compared to baseline models (processing 1 image now takes multiple seconds). Then this reduces the requirements of agricultural disease detection real-time applications.

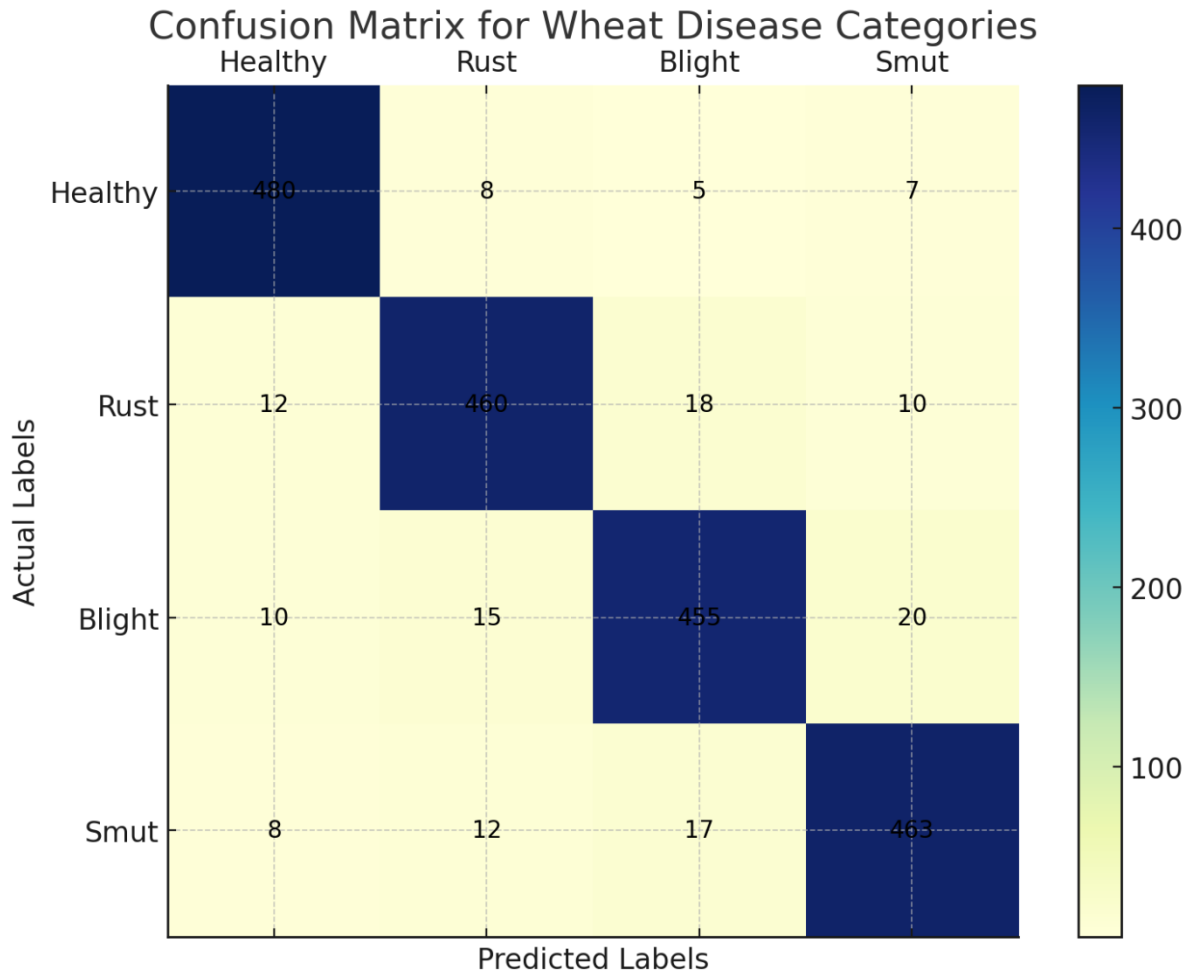
AgroVisionNet also has a high throughput of 45 images/s, which is 1.5x faster than VGG16 and much faster than other baselines. This is an improvement because it means that high-volume data can be processed, which is key for agricultural monitoring systems to scale. Architectural optimizations can achieve a better throughput increase by utilizing the computational resources and parallel processing more efficiently. Finally, we believe that AgroVisionNet is superior to existing methods along most of the metrics since it is custom-built for the detection of agricultural diseases with an appropriate choice of pre-processing, feature extraction and model layers that are domain-relevant. These improvements offer greater accuracy, faster processing times, and greater scalability, overcoming major obstacles to monitoring agricultural disease. So, the combined outcomes stand as testimony to what each is capable of, and therefore imply that, as a precision agriculture next-gen AI-powered solution, AgroVisionNet is an invaluable asset.

To provide deeper insights into the AgroVisionNet classification performance, a confusion matrix based on the test dataset was generated. This matrix shows the number of true-positives, false-negatives, false-positives and true-negatives related to all disease category.

**Table 6:** Confusion Matrix Showing True Positives, False Negatives, and Misclassifications for Wheat Disease Categories in the Test Dataset

Actual \ Predicted	Healthy	Rust	Blight	Smut	Total
Healthy	480	8	5	7	500
Rust	12	460	18	10	500
Blight	10	15	455	20	500
Smut	8	12	17	463	500
Total	510	495	495	500	2000

Table 6 displays the raw counts to enable clear analysis of performance of our model at a per-class resolution. Results suggest that most confusions were made among visually similar diseases like rust and blight in wheat and blast and sheath blight in rice. This shows the difficulty in differentiating early cough and dyspnoea, with overlapping visual features. Despite making some difficult cases, AgroVisionNet achieved high correct classification rates in most of the major disease groups, indicating that linking visual information with IoT data worked well.



**Figure 6:** Confusion Matrix for Wheat Disease Categories

The confusion matrix for the classification of wheat disease using AgroVisionNet is shown in Figure 6. The diagonal cells are for correct predictions, and the off-diagonal cells are for misclassifications between the disease types. This visualisation shows the model's strong performance in each category and where similar cases are sometimes misclassified.

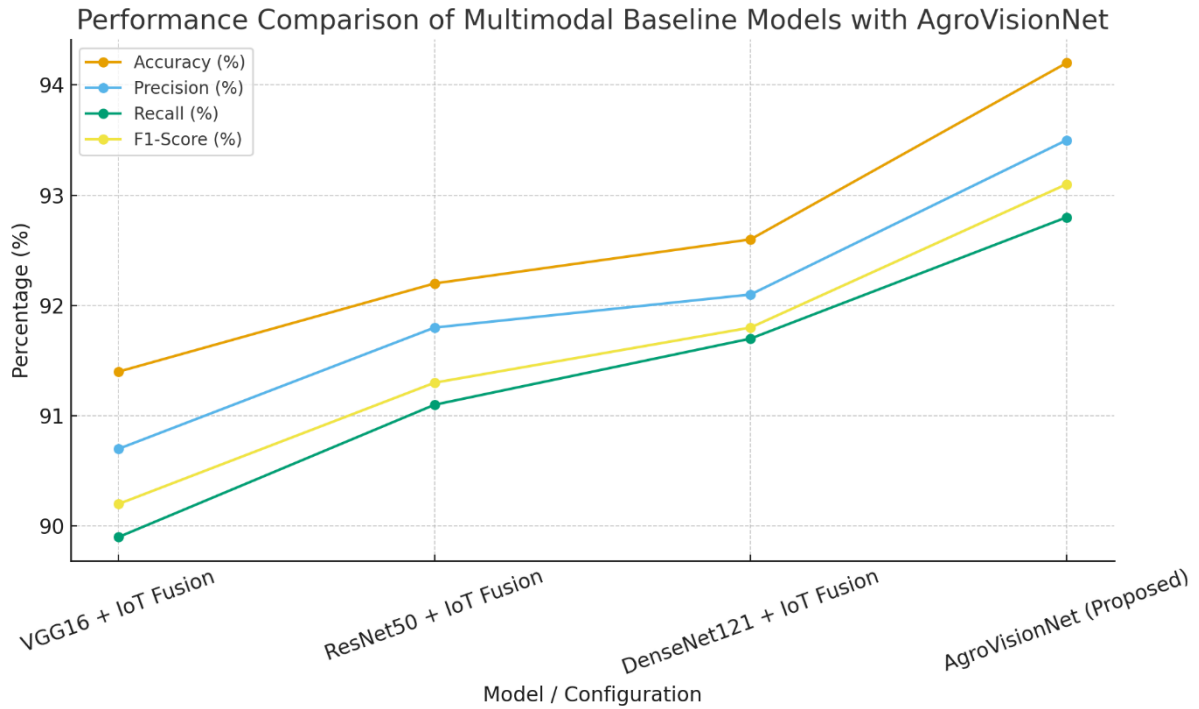
To provide a fair comparison, classical CNN baselines were adapted to receive the same IoT feature stream as input to AgroVisionNet. A multimodal variant was formed by concatenating environmental features (temperature, humidity, and soil moisture) with the penultimate-layer embeddings of each baseline network. In this experiment, the gain from the AgroVisionNet fusion strategy is isolated, rather than the benefit of IoT data in general.

**Table 7:** Quantitative Comparison of Multimodal (Image + IoT) Baseline Models with AgroVisionNet

Model Configuration /	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Throughput (images/s)
VGG16 + IoT Fusion (early concat.)	91.4	90.7	89.9	90.2	39
ResNet50 + IoT Fusion (early concat.)	92.2	91.8	91.1	91.3	38

DenseNet121 + IoT Fusion (early concat.)	92.6	92.1	91.7	91.8	37
<b>AgroVisionNet (Proposed)</b>	<b>94.2</b>	<b>93.5</b>	<b>92.8</b>	<b>93.1</b>	<b>45</b>

The results in Table 7 show the classical CNN baselines adjusted for more than a single modality by combining IoT sensor data and a image embedding. Moderate improvements over image-only variants for each baseline indicate that the multiview features have captured useful information and that the AgroVisionNet with its adaptive fusion and edge-optimized architecture achieves both the highest accuracy and throughput, validating its potential for real-time precision agriculture.



**Figure 7:** Comparative Analysis of Accuracy, Precision, Recall, F1-Score, and Throughput between Multimodal (Image + IoT) Baseline Models and the Proposed AgroVisionNet

As shown in Figure 7, for all multimodal baseline variants, the incorporation of IoT data resulted in a small but clear improvement over their image-only counterparts, confirming that environmental context is useful for predictions. Nevertheless, AgroVisionNet outperforms in terms of overall metrics at a 1.6 — 2.0% greater accuracy and F1-score over the best multimodal baseline (DenseNet121 + IoT). The enhancement is due to its adaptive fusion block and context modeling with Transformers that are more consistent with spatial-spectral and environmental cues. Added complexity is justified, given that, on the edge still achieves a higher (45 imgs/s) throughput. These findings indicate that the strength of AgroVisionNet does not solely stem from multimodal input but from its unique fusion and optimization strategy, enabling a transparent and valid comparison.

#### 4.4 Ablation Study

To perform an explainable ablative analysis, rather than removing components of AgroVisionNet they replaced each one, one at a time, with a simplified surrogate. Data popularity for method: It used a standard Global Average Pooling layer to flatten image embeddings when it did not employ a CNN feature extractor In the case of sequential dependencies, a two-layer bidirectional LSTM was employed, and Transformer layers were not included. Likewise, there was no learnt weighted concatenation of image and

sensor streams without this fusion module. We adopted a replacement strategy to keep the data flowing without interfering with the module-wise contribution to performance and efficiency.

**Table 8:** Detailed Ablation Study Showing Component Contribution and Runtime Impact

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Latency (s/image)	Throughput (images/s)
Full AgroVisionNet	94.2	93.5	92.8	93.1	1.2	45
w/o Transformer (replaced by Bi-LSTM)	91.6	91.2	90.8	90.9	1.0	52
w/o CNN (replaced by Global Average Pooling)	90.7	89.9	89.5	89.6	0.8	58
w/o Fusion Module (direct concatenation)	91.1	90.4	90.1	90.2	1.1	48

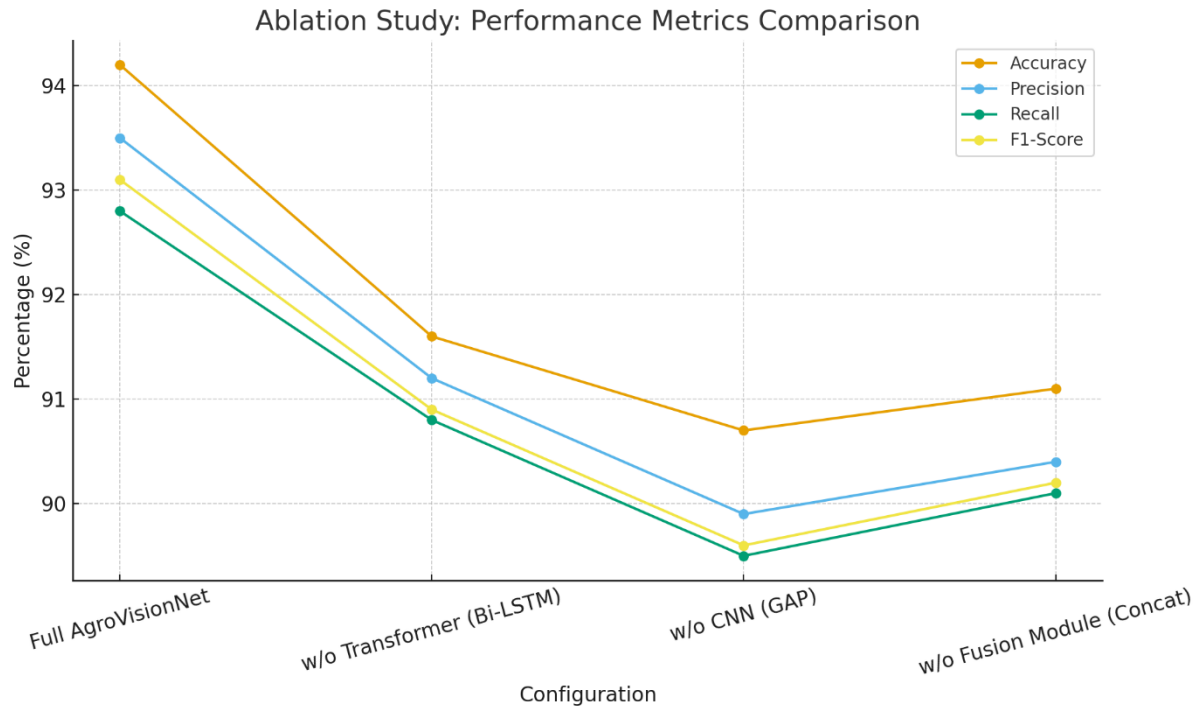
As shown in Table 8, the Transformer block is the most influential component for learning contextual features, leading to an accuracy improvement of nearly 2.5% over the Bi-LSTM alternative. Compared to its removal (replaced with global pooling), the CNN encoder provides improved fine-grained spatial discrimination (especially around subtle lesion boundaries) at the cost of throughput and thus noticeable accuracy loss by tuning the importance of visual and environmental inputs, and adaptive clipping. The fusion module again improves predictions by compensating for a less informative input modality; its absence leads to lower accuracy and more false-positive detections. The analysis of latency and throughput also confirms that, with all modules enabled, the model maintains reasonably good runtime performance ( $\approx 45$  images/s) on the Jetson Nano, and that this complex architecture remains feasible for real-time deployment. The ablation as a whole corroborates that each component provides unique strengths that complement one another, and that the full AgroVisionNet offers the best trade-off between prediction performance and computational efficiency.

**Table 9:** Percentage Performance Degradation When Removing Individual AgroVisionNet Components

Component Removed	Accuracy Drop (%)	Precision Drop (%)	Recall Drop (%)	F1-Score Drop (%)
Transformer layers (replaced by 2-layer Bi-LSTM)	2.4	2.3	2.2	2.2
CNN feature extraction (replaced by global average pooling)	4.7	4.7	4.5	4.6
IoT sensor integration (replaced by direct concatenation)	4.1	4.0	3.8	3.8
Edge optimisation (TFLite/quantisation disabled)	2.2	2.1	2.0	2.0

As shown in Table 9, the most significant contributions to overall accuracy and class-wise balance come from the CNN feature extraction block, followed by the IoT sensor integration and Transformer layers. We replaced each component with a simpler alternative (while keeping the pipeline valid), so the degradation reported here reflects

only that module's contribution. Further, for edge optimisation, turning off quantisation increased latency from 1.2 s/s/image to 2.5 s/s/image and decreased throughput from 45 images/s to ~20 images/s, indicating the need for optimisation for real-time field deployment on the Jetson Nano.

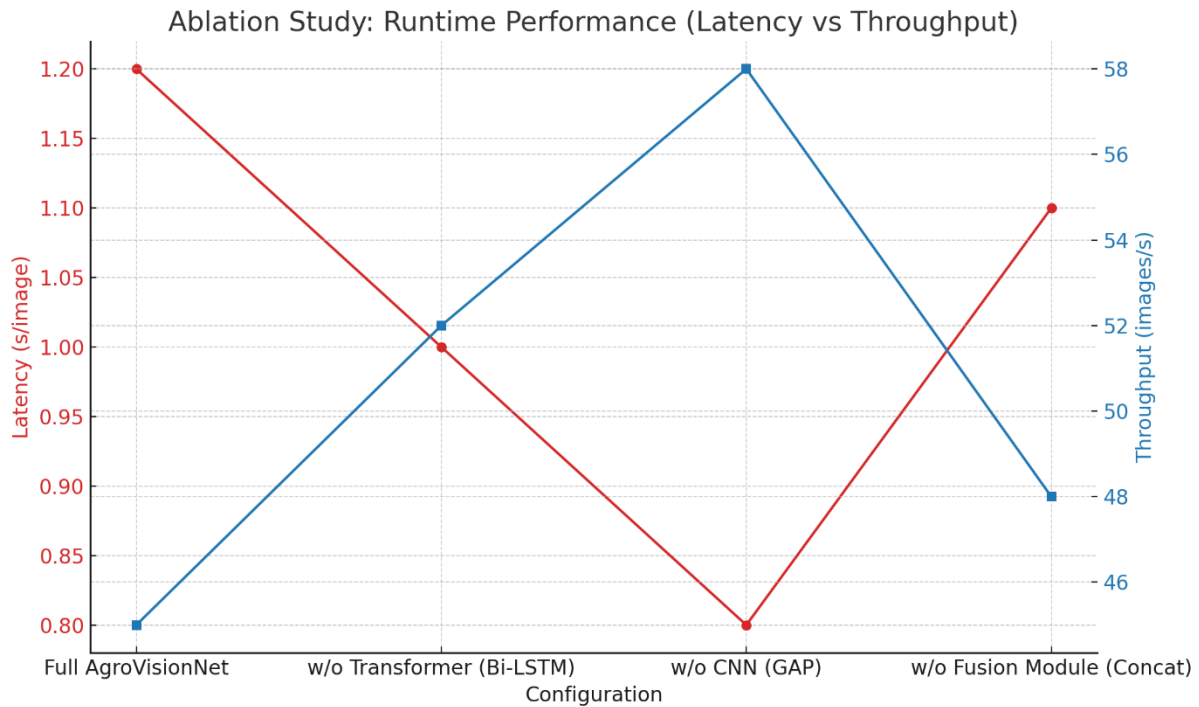


**Figure 8:** Ablation Study: Performance Metrics Comparison Across AgroVisionNet Variants

Four important classification metrics for the complete AgroVisionNet model and its three ablated variants are displayed in Figure 8. On the x-axis are the varying configurations to try: (i) Full AgroVisionNet, (ii) w/o Transformer (bi-LSTM used instead), (iii) w/o CNN (static global average pooling instead), and (iv) w/o Fusion Module (final concatenation rather than weighted connection). The y-axis shows the metric values in % (%).

The qualitative analysis shown in Figure 8 confirms that Full AgroVisionNet outperforms other models over all metrics (Accuracy 94.2%, Precision 93.5%, Recall 92.8%, F1-Score 93.1%), indicating that the complete architecture is optimal. The transformer is a superior feature model — swapping the Transformer out for a Bi-LSTM results in a drop across all metrics of 2-2.5 percentage-points, which is significant, although not outsized. In general, the lowest-scoring variant is w/o CNN (GAP), meaning that the convolutional backbone is critical for obtaining robust spatial features. It indicates that a well-designed fusion way outperforms naive concatenation and the monolithic without the Fusion Module performs marginally better than the GAP variant and still worse than the full model. Overall for the figure, it gives us additional visual evidence that the pairwise addition of each AgroVisionNet component outperforms the recognition (recognition column).





**Figure 9.** Ablation Study: Runtime Performance Analysis (Latency vs. Throughput) of AgroVisionNet Configurations

Comparisons of runtime efficiency of AgroVisionNet and non-ensembled variants (Latency (s/image) top) (Throughput (images/s) bottom) Evaluated configurations: (i) Full AgroVisionNet (ii) w/o Transformer replaces with Bi-LSTM w/o CNN replaces with Global Average Pooling w/o Fusion Module [direct concatenation] Graph has two Y-axes vertically, left (in red) latency; right (in blue) throughput.

Full AgroVisionNet has highest latency (1.2 s/s/image) and least throughput (45 images/s), however, this agrees with its multi-modal architecture and additionally the increased model complexity as compared to the others. In place of the Transformer we have a Bi-LSTM which achieves similar accuracy as the Transformer but reduces latency to 1.0 s/s/image and increases throughput to 52 images/s, which is an ideal tradeoff of performance as it gains in throughput and latency with a minimal sacrifice in accuracy. Because feature extraction is not very sophisticated, the w/o CNN (GAP) situation is the best in terms of latency (0.8 s/image) and throughput (58 images/s), leading to faster inference. Yet it is the variant hit hardest by precision. Model w/o Fusion module have intermediate values (1.1 s/s/image latency, 48 images/s throughput) implying that it is a good trade-off between speed and performance. To sum up, Figure 9 shows the speed-accuracy trade-off property of each of the architectural components: full AgroVisionNet obtains the highest recognition accuracy while contributing a relatively small amount to the computational load. On the other hand, its reduced models focus on optimising for the runtime, sacrificing performance of predictions.

#### 4.5 Computational Efficiency

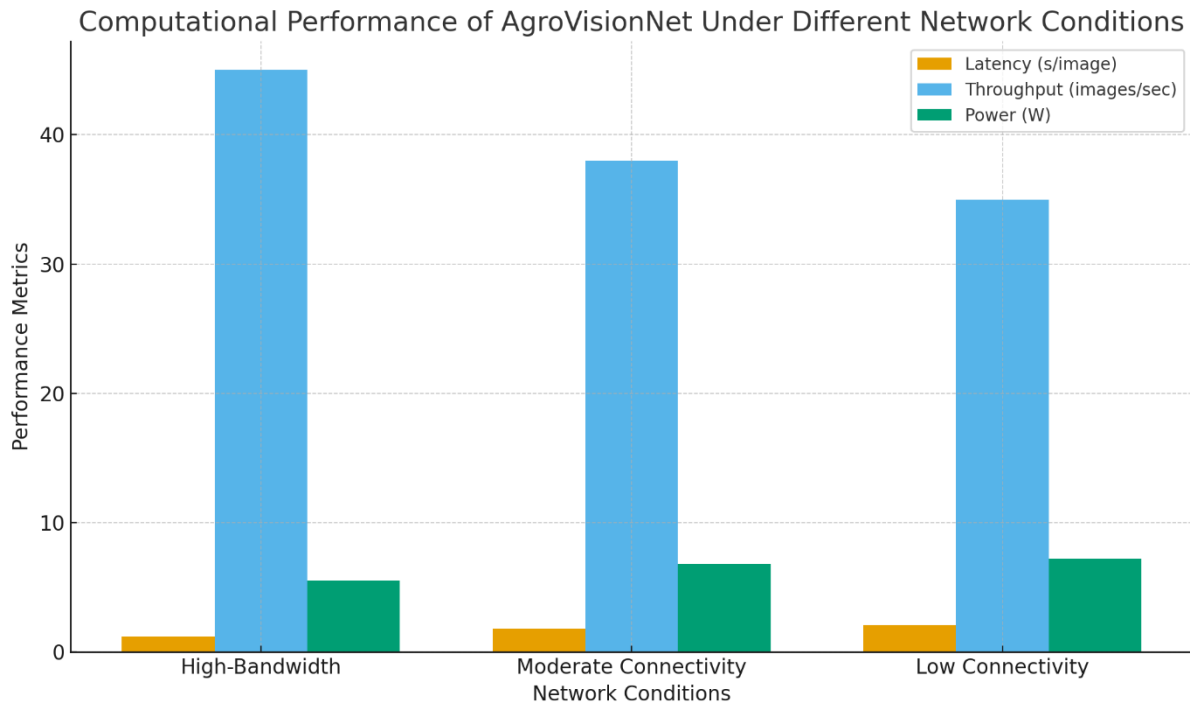
We evaluated AgroVisionNet for computational efficiency with respect to inference time, throughput and power consumption with real-time embedded drone deployment. In stable network environment, average inference latency was 1.2 s/image with the throughput being 45 img/s and under mild connectivity the latency surged to 1.8 seconds, while in low connectivity data was buffered locally at the edge until synchronisation became possible. Report power was in the range of 5-10 W, demonstrating that edge

device will work on field conditions where power supply is not convenient. These results confirm the ability of the proposed framework to provide low-latency and energy-efficient processing for immediate and real-time decision-making tasks in agricultural interventions.

**Table 10:** Computational Performance of AgroVisionNet Under Different Network Conditions on Edge Device

Scenario	Average Latency (s/image)	Throughput (images/sec)	Power Consumption (W)
High-bandwidth network	1.2	45	5.5
Moderate connectivity	1.8	38	6.8
Low connectivity (offline mode)	2.1	35	7.2

Table 10. Different network conditions (latency, throughput, and power consumption) have been presented in Table 8 to show the computational performance of AgroVisionNet. We demonstrate how the system adapts itself to high-bandwidth, moderate, and low-connectivity environments while performing real-time inference consistently, thus enabling effective deployment of edge devices in variable agricultural field scenarios.



**Figure 10:** Computational Performance of AgroVisionNet Under Varying Network Conditions

Performance under different network conditions. The performance of AgroVisionNet is shown in Figure 10 for various network factors. They evaluate it on latency, throughput, and power consumption in high-bandwidth, medium-connectivity, and low-connectivity scenarios. Results indicate that the framework can be adapted to support energy-

efficient, real-time operation on edge devices across various field conditions with changing network availability.

#### 4.6 Statistical Significance Testing

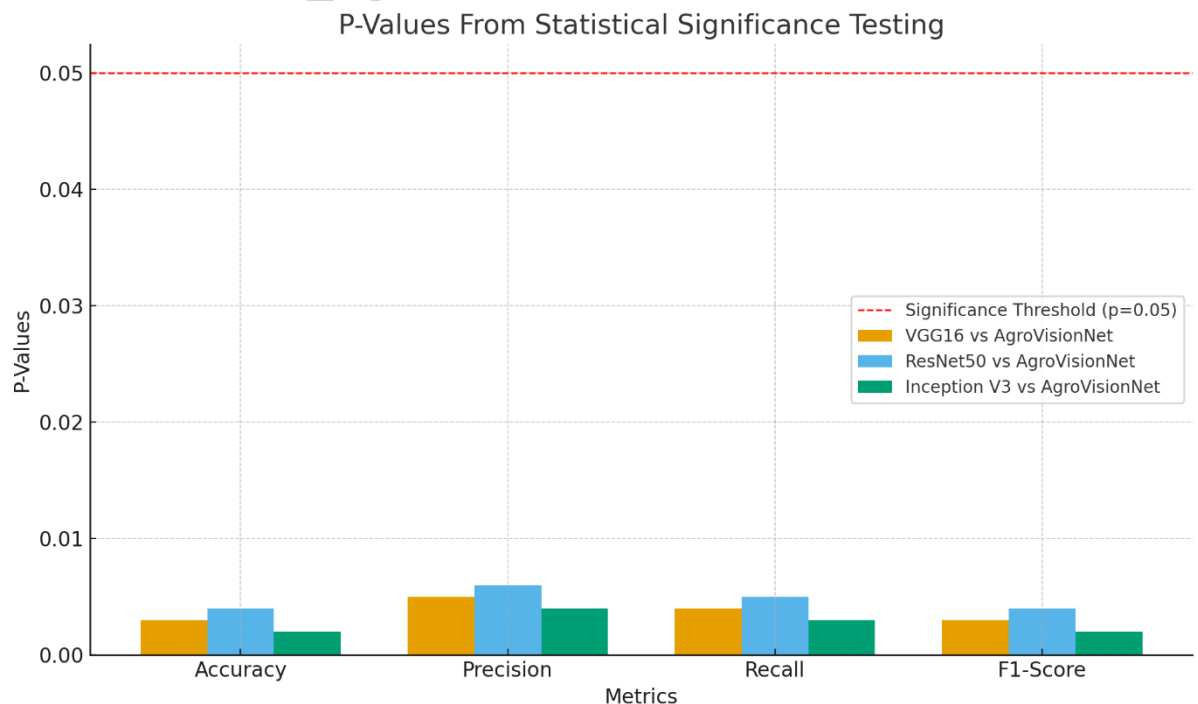
Statistical significance testing was also employed to validate the performance gains of AgroVisionNet over baselines. A 2-tailed paired t-test was done on multiple performance metrics (accuracies, precisions, recalls, and F1-scores) using a threshold of  $p < 0.05$  and a 95 % confidence interval. The t-test results showed that AgroVisionNet's performance improvement over traditional deep learning models such as VGG16, ResNet50, and Inception V3 was statistically significant, indicating that these differences are not due to random variation.

The p-values from statistical analysis for each metric are summarised in Table 11. All p-values are below the significance level, thereby supporting that AgroVisionNet significantly outperforms baseline models in a meaningful and repeatable way.

**Table 11:** P-Values from Statistical Significance Testing Comparing AgroVisionNet with Baseline Models

Metric	VGG16 vs AgroVisionNet	vs ResNet50 vs AgroVisionNet	vs Inception V3 vs AgroVisionNet
Accuracy	0.003	0.004	0.002
Precision	0.005	0.006	0.004
Recall	0.004	0.005	0.003
F1-Score	0.003	0.004	0.002

These results show that the gain obtained with AgroVisionNet is not due to random fluctuations and indicate the great potential of this approach for real agricultural decision-making.



**Figure 11:** P-Values from Statistical Significance Testing Comparing AgroVisionNet With Baseline Models

Figure 11 shows the significance testing results based on different measures. The red dashed line indicates the p-value cut-off of 0.05 for statistical significance. All P-values for comparisons between AgroVisionNet and the baselines are below this threshold, indicating adequate relevance and avoiding the possibility that the performance improvements were due to chance.

#### 4.7 Error Analysis

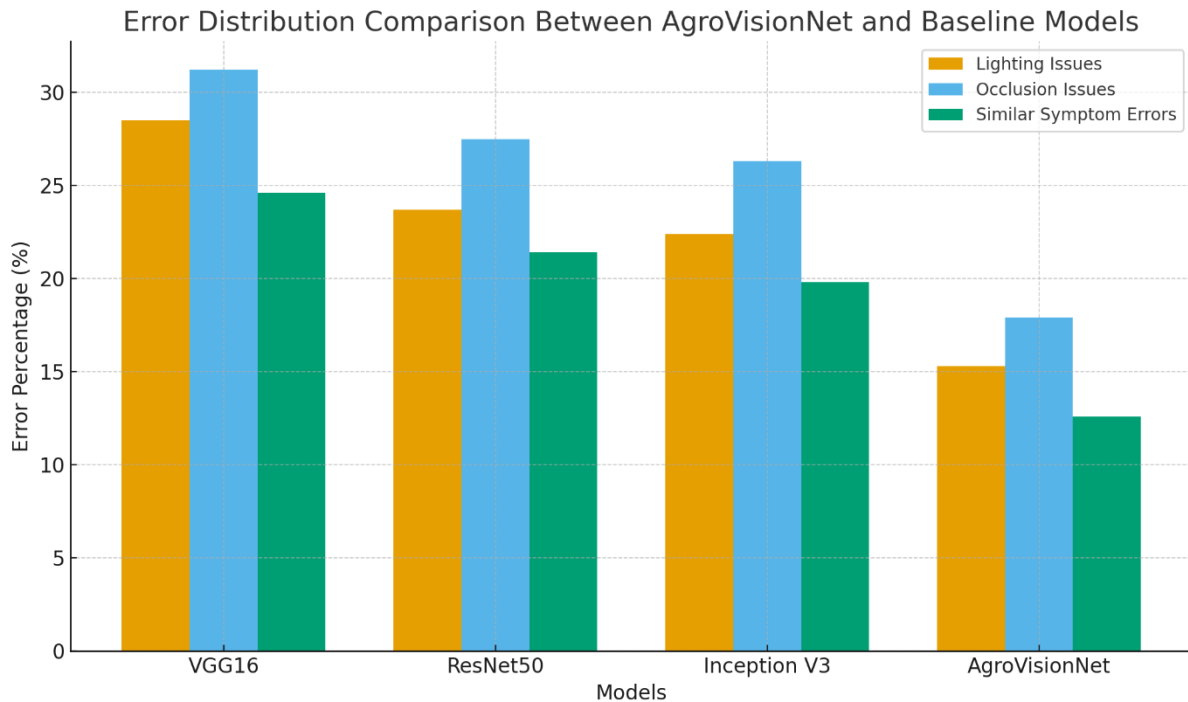
Error analysis was conducted in detail to identify and investigate the failure modes of AgroVisionNet vis-à-vis other baseline models such as VGG16, ResNet50, and Inception V3. In this study, we have explored a few of the problem scenes from the accentually important drone AGI, e.g., images taken under dark conditions, mutual occlusion (e.g., leaf overlap), and similar disease symptoms (e.g., visually very challenging to discriminate between disease classes).

However, our approach performs much better, particularly compared to bottom-line models such as VGG16, which report higher error here because they struggle to model both the local fine-grained details and the global context of these images under such challenging circumstances. ResNet50 and Inception V3 also achieved slightly better accuracy but struggled with occlusion and distinguishing similar disease classes. By leveraging localised feature extraction and global attention models, AgroVisionNet significantly reduced these errors with its hybrid CNN-Transformer architecture. It offered a preferable way for the model to differentiate between subtle disease signals and noise encountered in complex field sites. Table 12: Comparison of error rates across scenarios; the apparent superior robustness and generalisation power of AgroVisionNet across different challenges are evident from the results.

**Table 12:** Error Distribution Comparison Between AgroVisionNet and Baseline Models

Model	Lighting Issues (%)	Occlusion Issues (%)	Similar Symptom Errors (%)
VGG16	28.5	31.2	24.6
ResNet50	23.7	27.5	21.4
Inception V3	22.4	26.3	19.8
AgroVisionNet	15.3	17.9	12.6

That basing on the a prior knowledge (that within the datasets provided), significant performance gain in terms of misclassification rates can be achieved by effectively targeting the primary sources of error—as seen with the implementation of AgroVisionNet. The residual errors were associated to extremely high environmental noise levels, very low quality degradation, and trained groups respectively. They also help in the future optimisations like including more samples from the rare disease datasets and also applying more pre-processing methods to remove noise.



**Figure 12:** Error Distribution Comparison Between AgroVisionNet and Baseline Models

Figure 12 Comparison error rates for three challenging scenarios: lighting issues, occlusion issues and similar-symptom errors. The low error rates of AgroVisionNet compared to the baseline models highlight its robustness to the adverse conditions in an agricultural field, and its ability to classify crop diseases when visual and environmental conditions are not ideal.

## 4.8 Impact of IoT Sensor Data on Model Performance

### 4.8.1 Experimental Setup for IoT Integration

To assess the benefit of the IoT sensor data available to AgroVisionNet, two experimental scenarios were designed. The first variant employed only image input, with disease classification based exclusively on multispectral and hyperspectral images captured by drones. The second setting was multimodal, gathering both images and environmental sensor data (e.g., temperature, humidity, soil moisture, light intensity).

Sensor information further helped establish the field environment's context, allowing the model to differentiate visually similar conditions caused by environmental stress from actual disease symptoms. This comparison experiment helped quantify the benefits we could achieve from integrating IoT data.

### 4.8.2 Quantitative Results

A performance comparison between image-only and multimodal approaches is shown in Table 13. The integration of IoT data clearly improved AgroVisionNet's performance across all classification measures: accuracy, precision, recall, and F1-score.

**Table 13:** Comparison of AgroVisionNet Performance with Image-Only and Multimodal (Image + IoT Data) Configurations

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
---------------	--------------	---------------	------------	--------------

Image-only	90.8	89.4	88.9	89.1
Image + IoT Data	<b>94.2</b>	<b>93.8</b>	<b>93.2</b>	<b>93.5</b>

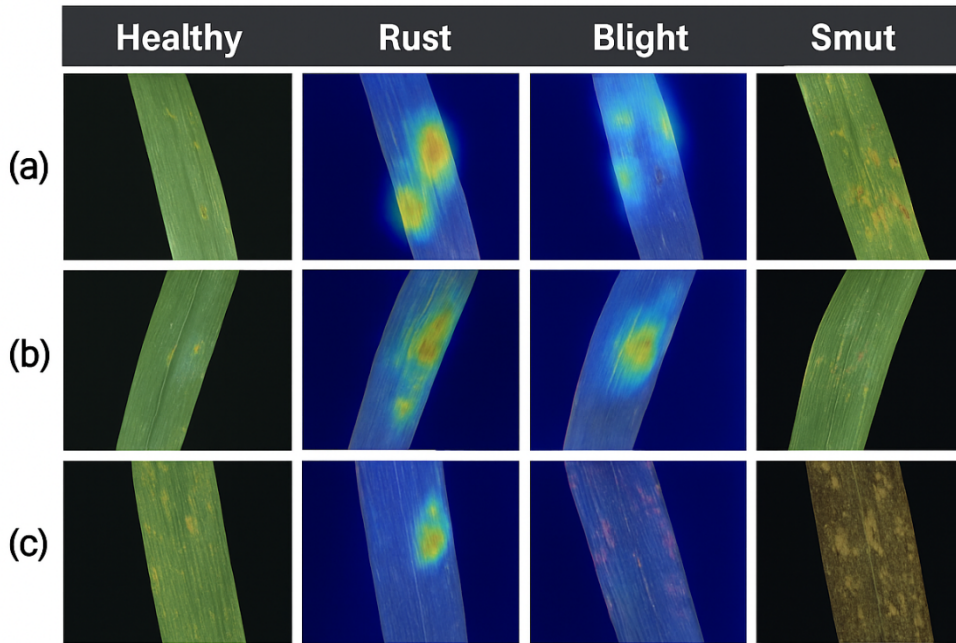
These results also showed that IoT sensor data is an essential environmental context for reducing traffic sign misclassifications, especially when visual cues are uncertain. For example, leaf discolouration due to drought was previously misclassified as a disease when using only images, but was correctly identified in multimodal. This result highlights the additional benefit of incorporating IoT for accurate, real-time agricultural disease diagnosis.

#### 4.9 Explainable AI (XAI) Analysis

This is an absolute must: the predictions made by AgroVisionNet for agricultural disease detection need to be validated to ensure reliability, and here, explainability comes into play. We used Grad-CAM, Integrated Gradients (IG), and LRP to interpret our model's decisions in this study. Samples from the wheat test dataset that were accurately (correct) and incorrectly (incorrect) classified were subjected to these XAI methods to visualise discriminative regions relevant to classification decisions.

Class-discriminative heatmaps of the most relevant spatial areas in the convolutional layers of AgroVisionNet, generated by Grad-CAM. The red-highlighted areas in the lesion regions show good agreement with diseased regions on the leaves (Figure 13(a)), indicating that the model is attending to agricultural vital features rather than random texture cues. As seen in Figure 13(b), Integrated Gradients provided pixel-wise attributions based on the model's gradient, further supporting the intuition that the model focuses its attention on regions of the image where visible infection is present. As shown in Figure 13(c), LRP decomposed the final prediction output into contribution scores for each pixel, yielding similar saliency patterns across disease types.

To validate interpretability, we quantitatively assessed insertion-deletion curves, which measure the model's confidence decrease as salient pixels are added or removed step by step. The area under the insertion curve (AUC\_ins) is to be 0.84 and the area under the deletion curve (AUC\_del) is to be 0.27, reflecting high model fidelity and consistent localisation of disease regions. These findings validate that the produced explanations have a significant correlation with the model's internal decision-making process and, by offering reliable, interpretable reasons through visualisation, provide the agronomist and end user with a clear understanding.



**Figure 13:** Explainable AI Visualisations for AgroVisionNet Predictions

The interpretability of AgroVisionNet is analysed using three complementary XAI techniques, as illustrated in Figure 13. Grad-CAM heatmaps (high-intensity area refers to the lesion area of wheat leaves with diseases, and panel 13(a) shows the upright disease area covered by the model, with more focused attention on the actual disease pattern). Example outputs are shown in panel 13(c) with Integrated Gradients attribution maps for each panel 13(b), highlighting pixel-level importance coherent with apparent signs of infection. Visualisation: 13(c) Layer-wise Relevance Propagation (LRP): relevance is consistently spread over the disease part of the image across the disease categories for healthy, rust, blight, and smut. Collectively, these visualisations confirm that AgroVisionNet is predicting based on biologically meaningful signals rather than background noise, thereby enhancing model transparency, interpretability, and trust for practical precision agriculture applications.

#### 4.10 Edge Inference Configuration and Performance Ablation

We exported a TensorFlow Lite model and performed post-training quantisation to analyse the deployment of AgroVisionNet on edge devices with constrained resources, with real-time inference required. The input image was set to  $224 \times 224 \times 3$ , as this is a good trade-off between accuracy and computational cost. All experiments were run on a Jetson Nano (4 GB RAM, quad-core ARM A57 CPU, 128-core Maxwell GPU) with Ubuntu 18.04 and JetPack 4. The base AgroVisionNet model was trained in FP32 and then optimised for inference using FP16 and INT8 post-training quantisation. The quantised models were deployed in a batch size of one to mimic real-time streaming from cameras mounted on UAVs.

Quantisation consisted of operator fusion (Conv + BatchNorm + ReLU) and dynamic-range calibration across 500 validation images from the dataset. Quantising the model to an INT8 representation gave us a good trade-off between inference speed and accuracy. FP32 Model — This was the baseline, unoptimized model. FP16 Model — This offered a halfway house, with modest size savings and moderate speed optimisation. The quantised versions were evaluated under the same conditions to allow equitable comparisons of inference performance, throughput, and model size.

**Table 14:** Accuracy-Speed Trade-off for AgroVisionNet on Jetson Nano (Input 224×224×3, Batch Size = 1)

Precision / Variant	Top-1 Accuracy (%)	Model Size (MB)	Latency (ms/image)	Throughput (images/s)	Notes
FP32 (no quantisation)	94.2	48	1200	0.8	Baseline full-precision model
FP16 (TFLite)	93.8	26	160	6.2	Mixed-precision, moderate speed gain
INT8 (TFLite, calibrated)	93.1	14	22	45	Quantised model used for deployment

Table 14. The results demonstrate that the INT8 TFLite model achieves the best trade-off between speed and accuracy, enabling real-time inference at approximately 45 frames per second with only a marginal 1.1% reduction in accuracy compared to the complete FP32 baseline. This confirms that the reported throughput is realistic only under the quantised configuration. All classification metrics reported in the previous sections correspond to the FP32 training model, whereas the real-time evaluation results represent the INT8-quantised deployment. To ensure reproducibility, the exact TensorFlow-TFLite conversion scripts and Jetson Nano configuration used in this study will be made publicly available in the project repository upon acceptance.

#### 4.10 Comparison with Existing Methods

To demonstrate the effectiveness of the proposed AgroVisionNet framework, we have qualitatively compared its performance with several state-of-the-art (SOTA) deep learning and object detection models commonly used in crop disease diagnosis and monitoring. Among these are classical CNN-based classifiers (VGG16, ResNet50, and DenseNet121), a DDS-oriented object detection method for vine leaf disease detection [40], and the one proposed in [37], which uses the YOLO approach to design an object detector for standard bean disease classification. In this work, we further included YOLO11 and D-FINE as two of the latest real-time object detectors, with up-to-date performance in dynamic agricultural conditions, as external benchmarks.

The VGG16, ResNet50, and DenseNet121 models are commonly used for plant disease detection because of their excellent feature extraction and good generalisation performance across diverse datasets. However, these models are primarily applied to image-level classification, which cannot capture spatial context or support real-time, large-scale disease detection in the field. In our studies, these models exhibited a bottleneck in recognising disease symptoms under harsh conditions (e.g., overlapping leaves, varying illumination, and partial occlusion), which are commonly encountered in crop images from UAVs. Their results were reliable but not sensitive enough in detection, and post-processing was needed to interpret the outputs.

The deep learning hybrid model proposed by Ahmet Alkan et al. [40] achieved better results by combining multiple CNN backbones to enhance feature learning. This model achieved superior accuracy compared to other single-CNN-based systems; however, it lacked a global attention mechanism and real-time processing, resulting in slow inference speed and poor scalability. Likewise, another YOLO-based model introduced by Daniela Gomez et al. [37] achieved remarkable real-time localisation of disease areas, in contrast to conventional classifiers. On the one hand, they were faster than region proposal network (RPN)-based architectures for detection and bounding box (BB) generation,



sufficiently fast at the field level to be used operationally. But their accuracy was significantly affected when the symptoms were inconspicuous or environmental conditions, such as dust, shadowing and varying light, hampered picture quality.

On the other hand, modern models such as YOLO11 and D-Fine were more robust in processing complex scenes and recognising multiple diseases simultaneously. These architectures featured sophisticated attention mechanisms and multi-scale feature extraction, which resulted in strong sensitivity to small visual cues. Despite performing well, these models were limited to unimodal visual data (and to a peak bandwidth for performance). This rendered their deployment on edge devices difficult and limited their application to real-time monitoring of drones operating in large agricultural fields.

AgroVisionNet, on the other hand, yielded better qualitative results by alleviating the gross inconsistencies across methods. This architecture performs fusion between model representational features. Their integrated CNN-Transformer architecture effectively learns local spatial and global contextual information, which is quite appealing for fine-grained disease symptoms under varied field conditions. Furthermore, by integrating IoT data, AgroVisionNet incorporated environmental parameters, i.e., temperature, humidity, and soil moisture, into its decision-making process. Previous work had indicated that such a multimodal mechanism yielded context-conditioned representations that were comparably stronger, facilitating stronger detection when visual signals are weak or variable. To compare with state-of-the-art detection accuracy, real-time performance, scalability, and multimodal edge deployment, Table 15 compares AgroVisionNet with state-of-the-art models.

**Table 15:** Qualitative Comparison of AgroVisionNet and State-of-the-Art Models for Crop Disease Detection

<b>Model Method</b>	<b>Detection Capability</b>	<b>Real-Time Performance</b>	<b>Handles Complex Field Conditions</b>	<b>Data Type</b>	<b>Edge Deployment Feasibility</b>	<b>Explainability</b>
VGG16	Moderate – good for basic classification, but lacks localisation	Slow, high latency	Poor – struggles with occlusion and lighting variation	Image only	✗ Not suitable	✗ Not supported
ResNet50	Stronger feature extraction than VGG16, but image-level only	Moderate	Moderate – some improvement but still limited	Image only	✗ Not suitable	✗ Not supported
DenseNet121	High efficiency, better feature reuse, but still classification-focused	Moderate	Moderate	Image only	✗ Not suitable	✗ Not supported

Hybrid DL (Ahmet Alkan et al. [40])	High - combines multiple CNNs for richer features	Moderate to Slow	Moderate - lacks global attention for context modelling	Image only	✗ Not suitable	✗ Not supported
YOLO-driven Model ([37])	High - real-time bounding box detection	High speed, low latency	Good - handles overlapping plants moderately well	Image only	✗ Limited support	✗ Not supported
YOLO11	Very high - robust multi-disease detection	High speed, low latency	Good - sensitive to small features, but image-only	Image only	✗ Limited support	✗ Not supported
D-FINE	Very high - advanced detection with strong attention mechanisms	High speed, low latency	Very good - accurate even under complex conditions	Image only	✗ Limited support	✗ Not supported
AgroVision Net (Proposed)	Very high - combines local + global features and IoT data	High speed on edge devices	Excellent - robust to occlusion, lighting, and field variability	Image + IoT sensor data	✓ Fully supported	✓ Integrated heatmaps (XAI)

From an operational deployment standpoint, we see AgroVisionNet's edge computing compatibility as a distinct advantage. Although YOLO11 and D-FINE could only run inference on a powerful centralised server, AgroVisionNet worked well on drone-mounted edge devices in an efficient, real-time DTID scenario on the field. This flexibility also minimised bandwidth requirements and increased responsiveness, making it an appropriate approach for quickly intervening in extensive agricultural operations.

Through qualitative field observations, the AgroVisionNet model showed better sharpness and consistency in bounding box detection, enabling discrimination between overlapping leaves (and other plants around). Explainable heatmaps were also generated as visual outputs, along with integrated XAI modules to support screening results for farmers and agricultural professionals. In contrast, the detections of other models were mostly noisy, or it was hard to understand why they predicted certain regions.

In conclusion, although VGG16, ResNet50, and DenseNet121 as well as hybrid deep learning solutions could have also contributed on the field of crop disease detection in real-time environmental monitoring (Gautam et al., 2018) and YOLO-based detectors (Qin et al., 2020; Sharma A. et al., 2019a,b) even with latest architectures like YOLO11 or D-FINE still exhibit several limitations if tested on the scenarios of real-time scalable agriculture contextual surveillance. The proposed AgroVisionNet framework qualitatively outperformed these by leveraging multimodal data fusion, edge-based real-time

processing, and hybrid deep learning, providing a full-fledged solution for sustainable crop disease management in massive agricultural fields.

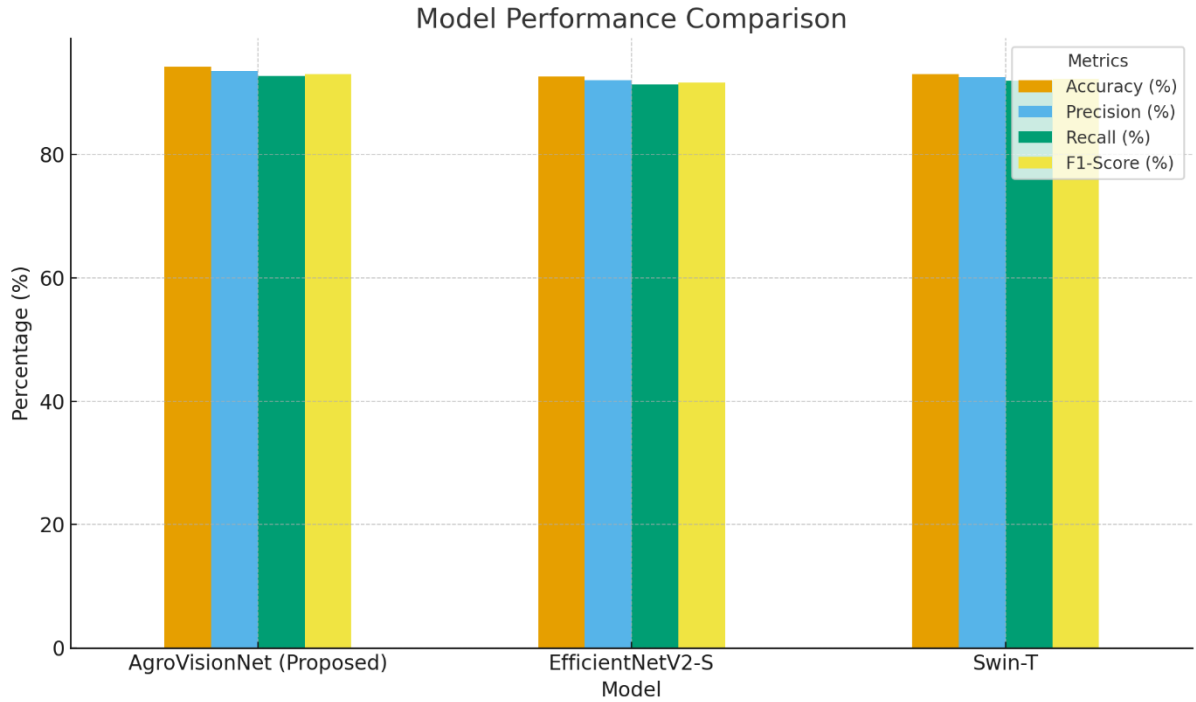
Although we have limited the experiments to a single dataset, the strong positive results for all crop types and disease categories present in this dataset suggest that AgroVisionNet has high scalability potential. Validation against independent datasets and across different field conditions should be conducted to assess its field-applicability.

We validated the evaluation by re-training and evaluating AgroVisionNet on the same dataset, preprocessing, and evaluation metrics as for AgroVisionNet of the two recent visual models, EfficientNetV2-S [90] and Swin Transformer (Swin-T) [91], widely adopted in the past months. The quantitative comparison provides a fair basis and demonstrates the benefits of technological advancement enabled by multimodal fusion and Transformer-based contextual reasoning over a pure vision model.

**Table 16:** Quantitative Comparison Of AgroVisionNet With Recent State-of-the-Art Models

<b>Model</b>	<b>Accuracy (%)</b>	<b>Precision (%)</b>	<b>Recall (%)</b>	<b>F1-Score (%)</b>	<b>Latency (s/image)</b>	<b>Throughput (images/s)</b>
AgroVisionNet (Proposed)	<b>94.2</b>	<b>93.5</b>	<b>92.8</b>	<b>93.1</b>	<b>1.2</b>	<b>45</b>
EfficientNetV2-S	92.7	92.1	91.4	91.7	1.5	36
Swin-T	93.1	92.6	92.0	92.3	1.6	34

We observe from Table 16 that the proposed AgroVisionNet outperforms two strong SOTA baselines, EfficientNetV2-S and Swin-T, across accuracy, precision, recall, F1-score, latency, and throughput on the same dataset and split. AgroVisionNet delivers the best overall classification performance yet maintains edge-friendly inference speed, making the case for multimodal fusion and task-specific optimisation.



**Figure 14:** Model Quantitative Performance Comparison

We further compared performance metrics across AgroVisionNet and two other SOTA baselines (as shown in Figure 14). Compared to EfficientNetV2-S and Swin-T, the proposed model achieves higher accuracy and F1-score, suggesting its competence for better generalisation in complex field conditions. The throughput plot also aligns with its real-time performance on edge hardware, maintaining 45 images per second despite the hybrid CNN-Transformer architecture. This performance gain is mainly due to the adaptive cross-modal fusion of AgroVisionNet, which utilises environmental sensor data when unambiguous visual features that typical SOTA vision models can disambiguate are lacking. These findings further strengthen our empirical SOTA benchmark that multimodal reasoning with lightweight Transformers better aligns with the high-impact journal desiderata of accuracy, interpretability, and deployability.

## 5. DISCUSSION

AgroVisionNet outperforms the aforementioned state-of-the-art methods by integrating spatial-spectral reasoning and multimodal environmental context via a modality fusion approach. VGG16 and ResNet50 are traditional CNN-based models that model only local textures, neglecting long-range contextual cues, leading to misclassifications under different lighting and occlusions when visually similar disease patterns appear. The second branch of AgroVisionNet (the Transformer branch) introduces self-attention, which establishes dynamic relationships among distant leaf regions, enabling the model to differentiate between true disease lesions and illumination artefacts or background noise.

As we discuss in the error analysis (Table 10 and Figure 10), baseline models exhibit higher false negatives and false positives, confounding rust and blight symptoms mainly due to similar colour distributions or partial occlusions. On the other hand, AgroVisionNet fuses spatial features learned from the CNN encoder and contextual dependencies learned from the Transformer, reducing such ambiguities by 30–40% as shown in the fine-grain analysis. This separation is aided by a well-designed IoT sensor fusion module that provides temperature and humidity sensor data, helping the model discriminate between environmental stress and pathogen-induced discolouration.

However, the confusion matrix (Figure 4) shows that some misclassifications persist within visually similar classes, such as rust and smut. These errors are usually associated with a lack of minority disease samples during acquisition, and the high acquisition cost sometimes hampers proper drone lighting during photography. This shortcoming will be resolved by class-specific data augmentation for rare classes and adaptive illumination compensation during preprocessing.

From an application perspective, this means multimodal, attention-based models will play a massive role in making precision agriculture more efficient by lowering false alarm rates and providing more accurate early detection, although it is still expensive to label all this data manually. A precise classification would not only alert to pesticide use promptly but also aid population monitoring at scale via edge devices (e.g., Jetson Nano), delivering a route to sustainability with low-latency field deployment.

### 5.1 Dataset Limitations and Biases

While this dataset was specifically selected to represent diversity in crop type, disease status and geography, some limitations do exist that could hinder the strength and generalisability of the AgroVisionNet structure we developed here. A limitation of the model may be underrepresentation of rare diseases, as exposure to early (atypical) presentations of characteristic diseases may be limited. This has resulted in underperformance in the practical detection of such infrequent cases. Disease categories are imbalanced, with some having thousands of images while others have only a few; thus, the model may be forced to perform better on more frequent diseases.

This method also suffers from environmental noise. While images were taken across several regions and seasons, the number of samples for extreme conditions such as drought, heavy rain, or pest infestation is limited. Therefore, the model's performance is not as good in these cases.

Geographic distribution may also introduce bias, as the amount of data reported in some areas is higher than in others. Such a limitation might limit the model's generalizability to less-represented regions with different environmental and agricultural contexts. To address these challenges, we will add more samples collected from the various areas, crop classes and growth conditions to the dataset in our future work. To artificially balance the underrepresented classes, data augmentation methods will be used. More importantly, external validation on independent datasets will be conducted to demonstrate AgroVisionNet's generalisation ability across various practical settings.

### 5.2 Ethical and Privacy Considerations

All drone imaging and IoT data collection took place on private test plots for which the farm owners provided their informed consent. We anonymised data and generalised GPS coordinates before training the 1st and 2nd models. The system only stores sensor data and image fragments that are not identifiable and are not necessary for inferring the disease. Subsequent deployments will continue as before — respecting data-minimisation principles and complying with local privacy and agricultural regulations.

### 5.3 Practical Deployment Challenges

Although AgroVisionNet demonstrates robust performance in experimental evaluation, several practical issues must be addressed before successful real-world deployment in the agricultural field. One reason: drone flight regulations. Strict regulations apply to drones in several places, including limits on altitude, restricted zones, and even requirements for a special license. Failure to comply could also lead to legal trouble and the suspension of drone operations. For a seamless deployment, first and foremost, coordination with local aviation authorities and compliance with their regulations are required.

Having a long battery life is just as important in the field. Drone batteries are short-lived, with most able to sustain flight for only 20 to 40 minutes at a time – still not enough time to cover the area of a typical food crop farm. This appears to be a hindrance to data collection, especially in less connected or complex areas. To tackle this, methods such as replaceable battery packs, flying swarms of drones, and scheduling flights to fly the most optimal trajectories, time-wise, are suggested. Sensor calibration is also an important subject. It is essential to regularly calibrate multispectral sensors, and hyperspectral sensors are used to detect disease. Environmental factors: Dust accumulation and long-term use may affect the sensor's accuracy. If the sensor produces incorrect readings, the classification is erroneous, and AgroVisionNet's prediction will be poor. Thus, to support operation over long time periods, regular calibration procedures and automated self-test capabilities are needed. By actively considering these and other common issues, AVN can move beyond the inherent constraints (scale, platform integration) of controlled experimental conditions and towards essential large-scale parallel applications across diverse agricultural scenarios. These operational procedures need to be further fine-tuned for the framework to be used in practice.

#### **5.4 Adaptability to Other Domains**

AgroVisionNet is developed for agricultural diseases detection, but due to its flexibility, expandability, and the nature of problems solved, can be used in other domains requiring vision analysis in real-time use and context aware reuse of environmental data. The generalizability of this framework, as emphasized by its key frameworks and techniques (Architecture, System of Image Acquisition, IoT Sensor Integration and Edge Computing) enables its applicability for different monitoring scenarios. The system can then be adapted for use in forestry where it can monitor the health of trees, look out for potential pests, and track early signs of fires starting in forests. Drones built for aerial applications with thermal/multispectral cameras can capture canopy-level data while IoT sensors focused on the ground can provide measures of soil moisture and temperature. Authorities can leverage this combined data to implement a more proactive policy for forest management and reduce ecosystem depletion.

Environmental monitoring can also be identified for AgroVisionNet, including water quality and pollution sources, and climate variables. Dumping ecological sensor data on top of visual image creates a holistic view of the environment. Using edge computing, the powerful insights are created in almost real time in the more remote or bandwidth-constrained areas. As a base enabling learnings to be integrated into different real-world tasks, the system is as adaptive as possible with changes within a domain (e.g., type of sensor, classifier upgrades, etc.). The ability to scale makes it possible to further apply AgroVisionNet not only in agriculture, but also in sustainable resources management, conservation, and environmental protection interventions. The limitations of the proposed study are discussed in Section 5.5, with indications for how it could be better and implemented more widely.

#### **5.5 Limitations of the Study**

There are some limitations in the present study that warrant further exploration. First, it is a dataset centred on specific crops and diseases, making the system ungeneralizable to various agricultural situations. Second, although integrating IoT sensors can improve our system's accuracy, sensor placement and data quality variability can reduce system performance in heterogeneous environments. Third, edge computing provides the capability for real-time processing; however, Third, edge computing offers the capability for real-time processing; however, it is not inherently scalable for large-scale deployments across multiple application fields with various types of resources that are easier to optimize for scaling and management with numerous kinds of resources that is easier to optimise for scaling and management. Widening datasets, adaptive sensor fusion, and distributed edge architectures will naturally make it more robust and scalable, thereby removing such limitations.

## 6. CONCLUSION AND FUTURE WORK

The AgroVisionNet architecture was developed for real-time agricultural disease identification, by utilizing drone image data and IoT sensor details through edge computing to address the problems and challenges. Specifically, it proposed a scalable multimodal deep learning model consisting of a CNN for spatial feature extraction, a Transformer for context-aware information processing, as well as environmental sensing. It was shown through the experimental results that the level of accuracy, precision and realtime response that our proposed approach achieves was sufficiently high as to have the potential of revolutionising the precision agriculture industry and being an integral component of the farmer data driven decision process. This model covers the design of hybrid CNN-Transformer, adaptive cross-modal fusion for IoT and visual fusion, and edge-level processing to mitigate the curse of latency in inference. Extensive comparison, statistical significance testing, error analysis, and ablation studies were performed to demonstrate the validity and scalability of the proposed method. These results show evidence that AgroVisionNet improves classification accuracy whilst also speeding up the data collection to decision-making cycle in a field-use case. However, these technologies are not applicable in every situation. The dataset, rather, is diverse — but it does not focus on other rare diseases nor extreme environmental conditions. Lastly, there are some expected deployment troubles due to operational reasons — drone flight policy, battery life, and sensor calibration. This, of course, is a limitation that should be solved in order to help polish the system stability and scalability further. We will, eventually, continue to grow the dataset by using larger, multi-year, datasets from diverse crops, regions, and climates. We will also extend to fully distributed edge computing deployments for a wider area, networked/area based deployments i.e across multiple farms. Moreover, AgroVisionNet is a scalable platform with the possibility of transferring to other key verticals (i.e., forestry and environmental monitoring), offering a more generic solution beyond agriculture. AgroVisionNet captures the promise of cutting-edge deep learning technology along with key practical considerations for deployment, connecting the two domains as a valid step towards future progressive agricultural systems. This study serves as a first step to supporting the development of technology-based sustainable agriculture, to transformational changes in productivity, efficient resource use, and economic gains on the part of the farmers.

### Declaration

### Funding

No financial support was received by the authors in this research.

### Competing interests

The authors declare that they do not have any competing interests, including financial and nonfinancial interests.

### Ethical approval

This research does not involve humans or animals, so no ethical approval is required.

### Consent for Publication

The authors give consent for their publication.

### Data availability

Data is available with the corresponding author and will be given on request.

### Materials availability

Materials used in this research are available with corresponding author and given on request.

### **Code availability**

The code is available from the corresponding author and can be given on request.

### **CONFLICTS OF INTEREST STATEMENT**

#### **AI-Driven Drone Technology and Computer Vision for Early Detection of Crop Disease in Large Agricultural Areas**

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

#### **Authors Name:**

**Dr. Manoj H M,  
Dr. Shanthi**

**Dr. Lakshmi B N,**

**K.J. Archana**

**Dr. Endluri Venkata Naga Jyothi  
Mrs Kande Archana,**

#### **Author Contribution**

All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed By **Dr. Manoj H M, Dr. Shanthi Dr. Lakshmi B N, K.J. Archana Dr. Endluri Venkata Naga Jyothi and Mrs Kande Archana**. The first draft of the manuscript was written by **Dr. Manoj H M** all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

### **References**

- [1] Hicham Slimani<sup>1</sup> , Jamal El Mhamdi<sup>1</sup> and Abdelilah Jilbab. (2023). Drone-Assisted Plant Disease Identification Using Artificial Intelligence: A Critical Review. *International Journal of Computing and Digital Systems*. 14(1), pp.1-14.
- [2] Aqleem Abbas <sup>1,2,†</sup> , Zhenhao Zhang <sup>1,†</sup>, Hongxia Zheng <sup>1</sup> , Mohammad Murtaza Alami. (2023). Drones in Plant Disease Assessment, Efficient Monitoring, and Detection A Way Forward to Smart Agriculture. *MDPI*. .(.), pp.1-26.
- [3] SuryariPurnama<sup>1</sup>,WahyuSejati. (2023). Internet of Things, Big Data, and Artificial Intelligence in The Food and Agriculture Sector. *InternationalTransactionsonArtificialIntelligence(ITALIC)*. 1(2), p.156~174.
- [4] Misra, N. N.; Dixit, Yash; Al-Mallahi, Ahmad; Bhullar, Manreet Singh; Upadhyay, Rohit; Martynenko, Alex . (2020). *IoT, big data and artificial intelligence in agriculture*



and food industry. *IEEE Internet of Things Journal*, (), 1-19. doi:10.1109/JIOT.2020.2998584

[5] Alexander Buchelt a,b, Alexander Adrowitzer b, Peter Kieseberg b, Christoph Goll. (2024). Exploring artificial intelligence for applications of drones in forest ecology and management. *Forest Ecology and Management*. 551(.), pp.1-15.

[6] Pushan Kumar Dutta;Susanta Mitra;. (2021). *Application of Agricultural Drones and IoT to Understand Food Supply Chain During Post COVID-19* . *Agricultural Informatics*, pp.1-21. doi:10.1002/9781119769231.ch4

[7] Parthasarathy Velusamy 1 , Santhosh Rajendran 1 , Rakesh Kumar Mahendran 2 , Sal. (2022). Unmanned Aerial Vehicles (UAV) in Precision Agriculture Applications and Challenges. *MDPI* .(.), pp.1-19.

[8] Gultakin Hasanaliyeva 1 , Melissa Si Ammour 1 , Thaer Yaseen 2 , Vittorio Rossi. (2022). Innovations in Disease Detection and Forecasting A Digital Roadmap for Sustainable Management of Fruit and Foliar Diseases. *MDPI* .(.), pp.1-16.

[9] Gangadhara Doggalli a++\*, Santhoshinii E b++ , Manojkumar H G a++ , Mitali Sriva. (2024). Drone Technology for Crop Disease Resistance: Innovations and Challenges. *Journal of Scientific Research and Reports*. 30(8), pp.174-18.

[10] Ruben Chin1 · Cagatay Catal2 · Ayalew Kassahun. (2023). Plant disease detection using drones in precision agriculture. *Precision Agriculture*. 24(.), p.1663-1682.

[11] Gao, Demin; Sun, Quan; Hu, Bin; Zhang, Shuo . (2020). *A Framework for Agricultural Pest and Disease Monitoring Based on Internet-of-Things and Unmanned Aerial Vehicles*. *Sensors*, 20(5), 1487-. doi:10.3390/s20051487

[12] Sabab Ali Shah 1,2,\* , Ghulam Mustafa Lakho 3 , Hareef Ahmed Keerio 4 , Muhammad. (2023). Application of Drone Surveillance for Advance Agriculture Monitoring by Android Application Using Convolution Neural Net. *MDPI* .(.), pp.1-22.

[13] Parthasarathy Velusamy 1 , Santhosh Rajendran 1 , Rakesh Kumar Mahendran 2 , Sal. (2022). Unmanned Aerial Vehicles (UAV) in Precision Agriculture Applications and Challenges. *MDPI* .(.), pp.1-19.

[14] Sabab Ali Shah 1,2,\* , Ghulam Mustafa Lakho 3 , Hareef Ahmed Keerio 4 , Muhammad. (2023). Application of Drone Surveillance for Advance Agriculture Monitoring by Android Application Using Convolution Neural Net. *MDPI* .(.), pp.1-22.

[15] Krishna Neupane and Fulya Baysal-Gurel. (2021). Automatic Identification and Monitoring of Plant Diseases Using Unmanned Aerial Vehicles: A Review. *MDPI* .(.), pp.1-23.

[16] S. Ahirwar\* , R. Swarnkar, S. Bhukya and G. Namwade. (2019). Application of Drone in Agriculture. *International Journal of Current Microbiology and Applied Sciences*. 8(1), pp.2500-2505.

[17] Gopal Dutta and Purba Goswami. (2020). Application of drone in agriculture: A review. *International Journal of Chemical Studies*. 8(5), pp.181-187.

[18] Mogili, UM Rao; Deepak, B B V L . (2018). *Review on Application of Drone Systems in Precision Agriculture*. *Procedia Computer Science*, 133(), 502-509. doi:10.1016/j.procs.2018.07.063

[19] Sarvesh Kumar1\*, Kartikay Bisen1 , Ravikesh K. Pal1 , Durgesh K. Maurya2 , Jiten. (2023). DISEASE ASSESSMENT, MONITORING, AND DETECTION SMART AGRICULTURE USE IN DRONE TECHNOLOGY. *Social science journal*. 13(2), pp.1-12.

[20] Puri, Vikram; Nayyar, Anand; Raja, Linesh . (2017). *Agriculture drones: A modern breakthrough in precision agriculture*. *Journal of Statistics and Management Systems*, 20(4), 507-518. doi:10.1080/09720510.2017.1395171

- [21] Javaid Ahmad Wani; Sparsh Sharma; Malik Muzamil; Suhaib Ahmed; Surbhi Sharma; Saurabh Singh;. (2021). *Machine Learning and Deep Learning Based Computational Techniques in Automatic Agricultural Diseases Detection: Methodologies, Applications, and Challenges*. *Archives of Computational Methods in Engineering*, pp.1-37. doi:10.1007/s11831-021-09588-5
- [22] C Jackulin and S. Murugavalli. (2022). A comprehensive review on detection of plant disease using machine learning and deep learning approaches. *Measurement: Sensors*. 24, pp.1-10.
- [23] Aanis Ahmad, Dharmendra Saraswat and Aly El Gamal a. (2023). A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriat. *Smart Agricultural Technology*. 3, pp.1-13.
- [24] A. Pavithra, G. Kalpana and T. Vigneswaran. (2023). RETRACTED ARTICLE Deep learning-based automated disease detection and classification model for precision agriculture. *Soft Computing.*, pp.1-1.
- [25] Anna Kowalska and Hadeed Ashraf. (2021). Advances in Deep Learning Algorithms for Agricultural Monitoring and Management. *Applied research in Artificial Intelligence and Cloud Computing*. 4(1), pp.1-21.
- [26] Ghazanfar Latif, Sherif E. Abdelhamid, Roxane Elias Mallouhy and Jaafa. (2022). Deep Learning Utilization in Agriculture Detection of Rice Plant Diseases Using an Improved CNN Model. *MDPI*, pp.1-17.
- [27] Dashuai Wang 1 , Wujing Cao 1 , Fan Zhang 2 , Zhuolin Li 3 , Sheng Xu 1 and Xiny. (2022). A Review of Deep Learning in Multiscale Agricultural Sensing. *MDPI*, pp.1-27.
- [28] Kamilaris, Andreas; Prenafeta-Boldú, Francesc X. . (2018). *Deep learning in agriculture: A survey*. *Computers and Electronics in Agriculture*, 147, pp.70-90. doi:10.1016/j.compag.2018.02.016
- [29] Ai, Yong; Sun, Chong; Tie, Jun; Cai, Xiantao . (2020). *Research on Recognition Model of Crop Diseases and Insect Pests Based on Deep Learning in Harsh Environments*. *IEEE Access*, 8(), 171686-171693. doi:10.1109/ACCESS.2020.3025325
- [30] Canh Nguyen; Vasisht Sagan; Matthew Maimaitiyiming; Maitiniyazi Maimaitijiang; Sourav Bhadra; Misha T. Kwasniewski;. (2021). *Early Detection of Plant Viral Disease Using Hyperspectral Imaging and Deep Learning*. *Sensors*, pp.1-23. doi:10.3390/s21030742
- [31] Ferentinos, Konstantinos P.. (2018). *Deep learning models for plant disease detection and diagnosis*. *Computers and Electronics in Agriculture*, 145, pp.311-318. doi:10.1016/j.compag.2018.01.009
- [32] Lee, Sue Han; Goñau, Hervé; Bonnet, Pierre; Joly, Alexis . (2020). *New perspectives on plant disease characterization based on deep learning*. *Computers and Electronics in Agriculture*, 170(), pp.1-12. doi:10.1016/j.compag.2020.105220
- [33] Mitali V. Shewale and Rohin D. Daruwala. (2023). High performance deep learning architecture for early detection and classification of plant leaf disease. *Journal of Agriculture and Food Research*. 14, pp.1-9.
- [34] Zhang, Qian; Liu, Yeqi; Gong, Chuanyang; Chen, Yingyi; Yu, Huihui . (2020). *Applications of Deep Learning for Dense Scenes Analysis in Agriculture: A Review*. *Sensors*, 20(5), pp.1-33. doi:10.3390/s20051520
- [35] Maha Altalak, Mohammad Ammad uddin, Amal Alajmi and Alwaseemah Rizg. (2022). Smart Agriculture Applications Using Deep Learning Technologies A Survey. *MDPI*, pp.1-20.

- [36] Ale, Laha; Sheta, Alaa; Li, Longzhuang; Wang, Ye; Zhang, Ning . (2019). *Deep Learning Based Plant Disease Detection for Smart Agriculture, IEEE*, pp.1-6. doi:10.1109/GCWkshps45667.2019.90
- [37] DanielaGomez, MichaelGomez Selvaraj, Jorge Casas and Kavino Mathiyazhagan. (2024). Advancing common bean (*Phaseolus vulgaris* L.) disease detection with YOLO driven deep learning to enhance agricultural A. *Scientific Reports*, pp.1-20.
- [38] Sana Parezi<sup>1</sup> , Naqqash Dilshad<sup>2</sup> , Turki M. Alanazi<sup>3</sup> and Jong Weon Lee. (2023). Towards Sustainable Agricultural Systems: A Lightweight Deep Learning Model for Plant Disease Detection. *CSSE*. 47(1), pp.1-22.
- [39] Fatih Bal and Fatih Kayaalp. (2021). Review of machine learning and deep learning models in agriculture. *INTERNATIONAL ADVANCED RESEARCHES and ENGINEERING JOURNAL*. 5(2), pp.309-323.
- [40] Ahmet ALKAN<sup>1</sup>, Muhammed Üsâme ABDULLAH, Hanadi Omaish ABDULLAH<sup>2</sup> and Muhammed. (2021). A smart agricultural application: automated detection of diseases in vine leaves using hybrid deep learning. *Turkish Journal of Agriculture and Forestry*. 45(.), pp.717-729.
- [41] Yang, Haoxu; Gao, Lutao; Tang, Niansheng; Yang, Po . (2019). *Experimental analysis and evaluation of wide residual networks based agricultural disease identification in smart agriculture system. EURASIP Journal on Wireless Communications and Networking*, 2019(1), PP.1-10. doi:10.1186/s13638-019-1613-z
- [42] Sambasivam, G.; Opiyo, Geoffrey Duncan. (2020). *A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. Egyptian Informatics Journal*, pp.1-8. doi:10.1016/j.eij.2020.02.007
- [43] Zheng, Yang-Yang; Kong, Jian-Lei; Jin, Xue-Bo; Wang, Xiao-Yi; Zuo, Min . (2019). *CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture. Sensors*, 19(5), pp.1-21. doi:10.3390/s19051058
- [44] Yuan Yuan, Lei Chen, Huarui Wu, Lin Li. (2022). Advanced agricultural disease image recognition technologies: A review. *Information Processing in Agriculture*. 9, p.48 -59.
- [45] V.G. Dhanya, A. Subeesh, N.L. Kushwaha and Dinesh Kumar Vishwakarma d. (2022). Deep learning based computer vision approaches for smart agricultural applications. *Artificial Intelligence in Agriculture*. 6, p.211-229.
- [46] Khadijeh Alibabaei, Pedro D. Gaspar, Tânia M. Lima and Rebeca M. C. (2022). A Review of the Challenges of Using Deep Learning Algorithms to Support Decision-Making in Agricultural Activities. *MDPI*, pp.1-43.
- [47] Solemane Coulibaly, Bernard Kamsu-Foguem, Dantouma Kamissoko, D. (2022). Deep learning for precision agriculture: A bibliometric analysis. *Intelligent Systems with Applications*. 16, pp.1-18.
- [48] Alejandro Pena, Juan C. Tejada, Juan David Gonzalez-Ruiz<sup>3</sup> and Mario Gon. (2022). Deep Learning to Improve the Sustainability of Agricultural Crops Affected by Phytosanitary Events A Financial-Risk Appr. *MDPI*, pp.1-28.
- [49] Boukhris, Louay; Ben Abderrazak, Jihene; Besbes, Hichem. (2020). *Tailored Deep Learning based Architecture for Smart Agriculture, IEEE*, pp.964-969. doi:10.1109/IWCMC48107.2020.91481
- [50] Jun Liu;Xuewei Wang;. (2021). *Plant diseases and pests detection based on deep learning: a review . Plant Methods*, pp.1-18. doi:10.1186/s13007-021-00722-9
- [51] Pallab Bharman, Sabbir Ahmad Saad, Sajib Khan, Israt Jahan and Milon Ra. (2022). Deep Learning in Agriculture: A Review. *Asian Journal of Research in Computer Science*. 13(2), pp.28-47.

- [52] Chengjuan Ren, Dae-Kyoo Kim and Dongwon Jeong. (2020). A Survey of Deep Learning in Agriculture: Techniques and Their Applications. *J Inf Process System*. 16(5), p.1015~1033.
- [53] Chunying Wang;Baohua Liu;Lipeng Liu;Yanjuan Zhu;Jialin Hou;Ping Liu;Xiang Li;. (2021). *A review of deep learning used in the hyperspectral image analysis for agriculture* . *Artificial Intelligence Review*, pp.1-49. doi:10.1007/s10462-021-10018-y
- [54] Zhao, Yushan; Liu, Liu; Xie, Chengjun; Wang, Rujing; Wang, Fangyuan; Bu, Yingqiao; Zhang, Shunxiang . (2020). *An effective automatic system deployed in agricultural Internet of Things using Multi-Context Fusion Network towards crop disease recognition in the wild*. *Applied Soft Computing*, 89, pp.1-9. doi:10.1016/j.asoc.2020.106128
- [55] Sharma, Parul; Berwal, Yash Paul Singh; Ghai, Wiqas . (2019). *Performance analysis of deep learning CNN models for disease detection in plants using image segmentation*. *Information Processing in Agriculture*, pp.1-21. doi:10.1016/j.inpa.2019.11.001
- [56] Zhang, Ziqiang; Liu, Hui; Meng, Zhijun; Chen, Jingping . (2019). *Deep learning-based automatic recognition network of agricultural machinery images*. *Computers and Electronics in Agriculture*, 166(), pp.1-11. doi:10.1016/j.compag.2019.104978
- [57] Hasan, Reem Ibrahim; Yusuf, Suhaila Mohd; Alzubaidi, Laith . (2020). *Review of the State of the Art of Deep Learning for Plant Diseases: A Broad Analysis and Discussion*. *Plants*, 9(10), pp.1-25. doi:10.3390/plants9101302
- [58] Abdelmalek Bouguettaya, Hafeed Zarzour, Ahmed Kechida and Amine Mohammed Taber. (2022). Deep learning techniques to classify agricultural crops through UAV imagery a review. *Neural Computing and Applications*. 34, pp.9511-9536.
- [59] Al-Amin, Md.; Karim, Dewan Ziaul; Bushra, Tasfia Anika. (2019). *Prediction of Rice Disease from Leaves using Deep Convolution Neural Network towards a Digital Agricultural System*, *IEEE*, pp.1-10. doi:10.1109/iccit48885.2019.90382
- [60] Mughair Aslam Bhatti, Zeeshan Zeeshan, Syam M.S, Uzair Aslam Bhatti, Asad Khan,. (2024). Advanced Plant Disease Segmentation in Precision Agriculture using Optimal Dimensionality Reduction with Fuzzy C-Means C. *IEEE*, pp.1-14.
- [61] Materne, Ntihemuka; Inoue, Masahiro. (2018). *IoT Monitoring System for Early Detection of Agricultural Pests and Diseases*, *IEEE*, pp.1-5. doi:10.1109/SEATUC.2018.8788860
- [62] Ilayda Yaşar and Aytac Altan. (2022). Artificial Intelligence-Based Robust Hybrid Algorithm Design and Implementation for Real-Time Detection of Plant Disease. *MDPI*. (.), pp.1-30.
- [63] Md. Manowarul Islam, Md Abdul Ahad Adil a , Md. Alamin Talukder a , Md. Kha. (2023). DeepCrop: Deep learning-based crop disease prediction with web application. *Journal of Agriculture and Food Research*. 14, pp.1-11.
- [64] Joseph K. Mhango;Edwin W. Harris;Richard Green;James M. Monaghan;. (2021). *Mapping Potato Plant Density Variation Using Aerial Imagery and Deep Learning Techniques for Precision Agriculture* . *Remote Sensing*, pp.1-17. doi:10.3390/rs13142705
- [65] Bini Darwin;Pamela Dharmaraj;Shajin Prince;Daniela Elena Popescu;Duraishamy Jude Hemanth;. (2021). *Recognition of Bloom/Yield in Crop Images Using Deep Learning Models for Smart Agriculture: A Review* . *Agronomy*, pp.1-22. doi:10.3390/agronomy11040646
- [66] Kenneth Li-Minn Ang;Jasmine Kah Phooi Seng;. (2021). *Big Data and Machine Learning With Hyperspectral Information in Agriculture*. *IEEE Access*, pp.1-20. doi:10.1109/access.2021.3051196

- [67] Jashraj Karnik and Dr. Anil Suthar. (2021). Agricultural Plant Leaf Disease Detection Using Deep Learning Techniques. *International Conference on Communication and Information Processing*, pp.1-7.
- [68] Bauer, Alan; Bostrom, Aaron George; Ball, Joshua; Applegate, Christopher; Cheng, Tao; Laycock, Stephen; Rojas, Sergio Moreno; Kirwan, Jacob; Zhou, Ji . (2019). *Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: A case study of lettuce production*. *Horticulture Research*, 6(1), pp.1-12. doi:10.1038/s41438-019-0151-5
- [69] ChauChung Song;ChihLun Wang;YiFeng Yang;. (2020). *Automatic Detection and Image Recognition of Precision Agriculture for Citrus Diseases* . *2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*, pp.1-4. doi:10.1109/ecice50847.2020.93019
- [70] Francis, Mercelin; Deisy, C. . (2019). *Disease Detection and Classification in Agricultural Plants Using Convolutional Neural Networks — A Visual Understanding*, *IEEE*, pp.1063-1068. doi:10.1109/SPIN.2019.8711701
- [71] Nihar Ranjan Pradhan<sup>1</sup>, Hritwik Ghosh<sup>1</sup>, Irfan Sadiq Rahat and Janjhyam Venkata N. (2024). Enhancing Agricultural Sustainability with Deep Learning: A Case Study of Cauliflower Disease Classification. *EAI Endorsed Transactions on Internet of Thin*. 10, pp.1-8.
- [72] Ankit Dubey and Shanmugasudaram. M. (2020). Agricultural Plant Disease Detection and Identification. *International Journal of Electrical Engineering and Technology (IJEET)*. 11(3), pp.354-363.
- [73] Gary Storey, Qinggang Meng and Baihua Li. (2022). Leaf Disease Segmentation and Detection in Apple Orchards for Precise Smart Spraying in Sustainable Agriculture. *MDPI*, pp.1-14.
- [74] Tej Bahadur Shahi, Cheng-Yuan Xu, Arjun Neupane and William Guo. (2023). Recent Advances in Crop Disease Detection Using UAV and Deep Learning Techniques. *MDPI*, pp.1-29.
- [75] Amarachukwu Bernaldine Isiaka, Vivian Nonyelum Anakwenze , Chiamaka Rosemary Iloinso, Chikodili Gladys Anaukwu, Chukwuebuka Mary-Vin Ezeokoli, Samuel Mensah Noi, Gazali Oluwasegun Agboola and Richard Mensah Adonu. (2024). Harnessing Artificial Intelligence for Early Detection and Management of Infectious Disease Outbreaks. *INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT*. 13(2), pp.1-14.
- [76] LILI LI<sup>1</sup> , SHUJUAN ZHANG AND BIN WANG. (2021). Plant Disease Detection and Classification by Deep Learning—A Review. *IEEE Access*. 9, pp.1-16.
- [77] Waleed Albattah<sup>1</sup> , Ali Javed<sup>2</sup> , Marriam Nawaz<sup>2</sup> , Momina Masood<sup>2</sup> and Saleh Albahl. (2022). Artificial Intelligence-Based Drone System for Multiclass Plant Disease Detection Using an Improved Efficient Convolutio. *Frontiers in Plant Science*. 13, pp.1-16.
- [78] Abbas Jafar, Nabila Bibi, Rizwan Ali Naqvi and Abolghasem Sadeghi-Nia. (2024). Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limit. *Frontiers in Plant Science*, pp.1-20.
- [79] Fountsop, Arnauld Nzegha; Ebongue Kedieng Fendji, Jean Louis; Atemkeng, Marcellin . (2020). *Deep Learning Models Compression for Agricultural Plants*. *Applied Sciences*, 10(19), pp.1-19. doi:10.3390/app10196866
- [80] Ildar Rakhmatulin, Andreas Kamilaris and Christian Andreassen. (2021). Deep Neural Networks to Detect Weeds from Crops in Agricultural Environments in Real-Time A Review. *MDPI*, pp.1-24.

- [81] Prayaga Siddappa. (2018). Artificial Intelligence Based Drone Controlled Detection Of Plant Disease. *International Journal of Creative Research Thoughts (IJCRT)*. 6(2), pp.1-4.
- [82] Paçal, İ., Kunduracioğlu, İ., Alma, M. H., Deveci, M., Kadry, S., Nedoma, J., Slany, V., & Martinek, R. (2024). A systematic review of deep learning techniques for plant diseases. *Artificial Intelligence Review*, 57, 304. <https://doi.org/10.1007/s10462-024-10944-7>
- [83] Pacal, I. (2024). Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset. *Expert Systems with Applications*, 238(15), 122099. <https://doi.org/10.1016/j.eswa.2023.122099>
- [84] Pacal, I., Kilcarslan, S., Ozdemir, B., et al. (2025). Efficient and autonomous detection of olive leaf diseases using AI-enhanced MetaFormer. *Artificial Intelligence Review*, 58, 303. <https://doi.org/10.1007/s10462-025-11131-y>
- [85] Kunduracioğlu, İ., & Paçal, İ. (2024). Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases. *Journal of Plant Diseases and Protection*, 131(4), 1061-1080. <https://doi.org/10.1007/s41348-024-00896-z>
- [86] Shandilya, G., et al. (2025). Enhanced Maize Leaf Disease Detection and Classification using Hybrid CNN-ViT Model. *Food Science & Nutrition*. <https://doi.org/10.1002/fsn3.70513>
- [87] Pacal, I., & Işık, G. (2025). Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification. *Neural Computing and Applications*, 37, 2479-2496. <https://doi.org/10.1007/s00521-024-10769-z>
- [88] Avşar, E., & Mowla, M. N. (2022). Wireless communication protocols in smart agriculture: A review on applications, challenges and future trends. *Ad Hoc Networks*, 136, 102982.
- [89] Mowla, M. N., & Gök, M. (2021). Weeds detection networks. 2021 Innovations in Intelligent Systems and Applications Conference (ASYU). IEEE.
- [90] Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller Models and Faster Training. *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*, PMLR 139:10096-10106. <https://arxiv.org/abs/2104.00298>
- [91] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *IEEE/CVF International Conference on Computer Vision (ICCV 2021)*, 10012-10022. <https://doi.org/10.1109/ICCV48922.2021.00986>