A Novel Vision Transformer Approach with Adaptive Segmentation for Early Plant Disease Detection

II B.Tech II Semester
A641X
Minor Project

Submitted By:

21211A05X3	Vaishnavi Veruva
21211A05X6	Yadla Yogesh
23211A05W6	Vallepu Sai Soumya
24211A0532	Madas Vivek

Under the kind guidance of

Dr.D.Vivek

Domain Name: Computer Vision

Department of Computer Science and Engineering

B V Raju Institute of Technology

(UGC Autonomous, Accrediated by NBA and NAAC)

Vishnupur,Narsapur, Medak(Dist), Telangana State, India-502313.

CERTIFICATE OF APPROVAL

This project work (**23E9**) entitled " **A Novel Vision Transformer Approach with Adaptive Segmentation for Early Plant Disease Detection** " by Ms. Vaishnavi Veruva, Registration No. 23211A05x1 , Yadla Yogesh, Registration No. 23211A05X6 , Vallepu Sai Soumya, Registration No. 21211A05W6, Madas Vivek , Registration No. 2421A50532 under the supervision of **Dr.D.Vivek** in the Department of Computer Science and Engineering, B V Raju Institute of Technology, Narsapur, is here by submitted for the partial fulfillment of completing Minor Project during II B.Tech II Semester (2024 - 2025 EVEN). This report has been accepted by Research Domain Computational Intelligence and forwarded to the Controller of Examination, B V Raju Institute of Technology, also submitted to Department Special Lab " Artificial Intelligence Machine Learning" for the further procedures.

Dr.D.Vivek
Associate Professor and Supervisor
Department of CSE
B V Raju Institute of Technology
Narsapur.

Dr.CH.Madhu Babu
Professor and Dept.Head
Department of CSE
B V Raju Institute of Technology
Narsapur

External Examiner

Internal Examiner

Dr.T.Subba Reddy
Domain Incharge - "Computer Vision"
Department of Computer Science and Engineering
B V Raju Institute of Technology
Narsapur

DECLARATION

We, the members of Research Group domain **Computer Vision**, declare that this report titled: **A Novel Vision Transformer Approach with Adaptive Segmentation for Early Plant Disease Detection** is our original work and has been submitted in whole or in parts for International conference or journal **ICCDS2025**. All sources of information used in this report have been acknowledged and referenced respectively.

This project was undertaken as a requirement for the completion of our **II B.Tech II Sem Minor project** in Department of **Computer Science and Engineering** at
B V Raju Institute of Technology, Narsapur. The project was carried out between 23-Dec-2024 and 26-April-2025. During this time, we as a team were responsible for the process model selection, development of the micro document and designing of the project.

This paper introduces an improved Vision Transformer (ViT) model with adaptive segmentation for efficient and accurate plant disease detection. It uses deep learning to analyze leaf images and identify disease symptoms early. Adaptive segmentation helps the model to focus more on important regions in real-world environments.The model is trained and tested on real farm data to make it more reliable.It is robust on multiple environments and crops.This approach gives farmers an intelligent, simple, and reliable tool to manage plant health.

We would like to express our gratitude to our project supervisor **Dr.D.Vivek** for his guidance and support throughout this project. We would also like to thank our Department Head Dr.CH.Madhu babu and Domain Incharge **Dr.T.Subba Reddy** for his help and efforts.

We declare that this report represents Our own work, and any assistance received from others has been acknowledged and appropriately referenced.

Vaishnavi Veruva	(23211A05X3)	____________________
Yadla Yogesh	(23211A055X6)	____________________
Vallepu Sai Soumya	(23211A05W6)	____________________
Madas Vivek	(24211A0532)	____________________

Guide **Project Coordinator** **Domain Incharge** **HOD/CSE**

ACKNOWLEDGEMENT

This project is prepared in fulfilment of the requirement for Research Domain **Computer Vision**. We owe our deepest gratitude to the Department of Computer Science and Engineering
B V Raju Institute of Technology, Narsapur for providing us with an opportunity to work on this project.

We also extend our gratitude towards our project supervisor **Dr.D.Vivek** and Department Head **Dr.CH.Madhu babu** , whose guidance and expertise have been invaluable in steering us towards success. We also thank our Research Domain Incharge **Dr.T.Subba Reddy** and other faculty members of the Department for their valuable feedback and suggestions.

Finally, we would like to thank our family and friends for their continuous support and encouragement throughout the project. We acknowledge the contributions of everyone who supported us in the creation of this project report.

Thank you all for your assistance and support.

The experience of working on this project will surely enrich our technical knowledge and also give us hands on experience of working on a project and help develop our team's skill set to a great extent.

ABSTRACT

Early and accurate detection of plant diseases plays a vital role in ensuring agricultural productivity and food security. Traditional disease detection methods often lack speed and precision, while existing AI models may struggle to localize diseased regions effectively or adapt across different crop types. To address these challenges, this project proposes the Adaptive Vision Transformer (AViT), an enhanced model that integrates Adaptive Attention Segmentation (AAS) to highlight diseased areas and suppress background noise. Additionally, the model incorporates Multi-Crop Adaptation Transformer (MCAT), which enables it to dynamically adjust to various plant species, improving its versatility in real-world farming conditions. To further enhance transparency and usability, Explainable Transformer Integration (ExTi) is introduced, helping users understand the model's predictions through intuitive outputs. The combined framework significantly boosts detection accuracy and interpretability, offering a reliable and scalable solution for intelligent crop management.

Keywords: Adaptive Vision Transformer, Adaptive Attention Segmentation, convolutional neural network

List of Figures

List of Tables

LIST OF ACRONYMS AND ABBREVIATIONS

AAS Adaptive Attention Segmentation

AVIT Adaptive Vision Transformer

CNN convolutional neural network

DL deep learning

ML Machine learning

ResNets Residual Neural Networks

ViT Vision Transformer

MHSA Multi-Head Self-Attention

TABLE OF CONTENTS

CHAPTER - 1

1. INTRODUCTION

In this report, we present an Enhanced Vision Transformer (ViT)-based model for accurate plant disease detection using Adaptive Attention Segmentation (AAS), Multi-Crop Adaptation (MCAT), and Explainable Transformer Insights (ExTi). The goal of this study is to build a reliable system that can identify plant diseases across different crop types and offer clear explanations for its predictions. AAS helps the model focus on affected regions of the leaf by reducing background noise, improving detection accuracy. MCAT enables the system to adjust and perform well on multiple plant species. ExTi is used to provide understandable insights into the model's decisions, making the results more transparent for users.

1.1. Background

The field of plant disease detection has seen rapid growth with the integration of deep learning and artificial intelligence (AI). However, challenges remain in accurately identifying diseases from complex real-world images that often include noise and varied backgrounds. Vision Transformer (ViT)-based models have shown promise in image classification tasks, but they require enhancements for precise plant disease detection. This project builds on recent advancements by introducing Adaptive Attention Segmentation (AAS) to focus on disease-affected regions, Multi-Crop Adaptation (MCAT) to handle different crop types, and Explainable Transformer Insights (ExTi) to provide interpretability. Together, these techniques aim to deliver a robust and scalable solution for intelligent plant health monitoring.

1.2. Motivation

The motivation for developing this Vision Transformer-based model arises from the increasing need for accurate and automated plant disease detection in agriculture. Manual inspection of crops is time-consuming, error-prone, and often fails to detect diseases at an early stage, especially across large farmlands. Traditional models struggle with noisy backgrounds and variations across different crops.

This project seeks to overcome these limitations by using Adaptive Attention Segmentation (AAS) for precise focus on diseased regions, Multi-Crop Adaptation (MCAT) for broader crop applicability, and Explainable Transformer Insights (ExTi) to make the model's decisions more understandable.

These enhancements aim to support farmers with a reliable, efficient, and intelligent solution for early disease detection and crop management.

1.3. Objectives

1. To develop an enhanced Vision Transformer (ViT) model with Adaptive Attention Segmentation (AAS) for accurate plant disease localization.
2. To implement Multi-Crop Adaptation (MCAT) enabling the model to detect diseases across different crop species effectively.
3. To integrate Explainable Transformer Insights (ExTi) for better interpretability of disease detection results.
4. To reduce misclassification by minimizing background noise during feature extraction.
5. To provide a fast, scalable, and automated system for early plant disease detection to support modern agriculture.

1.4. Problem statement

- Identifying plant diseases early is crucial for reducing crop loss and improving agricultural productivity.
- Segmenting diseased areas accurately is challenging due to complex leaf textures and background noise.
- Different crops exhibit different disease patterns, requiring models to adapt across various plant types.
- There is a need for a system that not only detects diseases but also helps users understand the prediction in a clear way.
- Ensuring the model is scalable and usable in real-world field conditions is a key challenge.

1.5. Scope of Project

To gain an insight into the scope of our project, let us first understand what the term scope of a project means in software development. The scope refers to the boundaries and limitations of a project. It also defines the features, functions, and requirements of the software being developed. The scope of a software project describes what the software will and will not do, and what is included and excluded from the project.

The Enhanced Vision Transformer with Adaptive Segmentation Model will consist of three primary functionalities:

The project aims to deliver the following key features:

- Disease detection in plant leaves using a Vision Transformer (ViT) architecture enhanced with Adaptive Attention Segmentation (AAS).

- The segmentation module focuses on diseased regions by filtering out background noise, improving the accuracy of feature localization.
- A Multi-Crop Adaptation (MCAT) module that allows the model to generalize across various crop types and diseases.
- Explainable Transformer Integration (ExTi) to help interpret model predictions in an understandable way for users.

The boundaries and limitations of the Enhanced ViT Model can also be defined by its scope, which outlines the features, functions, and requirements of the system. Some of these boundaries and limitations are:

- **Data Quality and Quantity**: The model's performance is dependent on the availability of diverse and high-quality annotated plant disease datasets. Limited data could affect the generalization of the model.
- **Hardware and Computational Resources**: The use of Vision Transformers and segmentation techniques requires significant computational power, which might limit deployment in low-resource environments like rural farms.
- **Environmental Challenges**: Variations in lighting, occlusion, and leaf quality in real field images can affect model performance under uncontrolled conditions.
- **Integration in Agricultural Workflows**: Effective adoption will require that the model be integrated seamlessly into tools used by farmers, agronomists, or agricultural apps.

Addressing these limitations through continuous training, dataset expansion, and collaboration with agricultural experts will be crucial to maximizing the model's utility in real-world plant disease management.

Overall, the project report will offer a comprehensive insight into the Enhanced Vision Transformer with Adaptive Segmentation model. The report will serve as a valuable resource for researchers and agricultural stakeholders seeking intelligent crop disease detection solutions.

CHAPTER - 2

2. LITERATURE SURVEY

Plant health monitoring has changed a lot with the help of machine learning (ML) and deep learning (DL) technologies. In the past, checking plant health meant looking at leaves and stems manually, which took a lot of time and could lead to mistakes. Now, DL models, especially convolutional neural networks (CNNs), can look at pictures of plants and accurately identify problems like diseases, pest damage, or nutrient deficiencies [12]–[15]. This helps farmers find issues early and take action quickly, which reduces crop loss and increases productivity.

One big advantage of using DL models for plant health is that they can be used on a large scale. These models can be added to mobile phones, drones, and other smart systems to give real-time updates on plant health. For example, drones with cameras and DL models can scan huge fields and find areas that need attention [13]. This helps farmers use pesticides and fertilizers more efficiently, saving money and reducing harm to the environment. Farmers can also use mobile apps to take pictures of their crops and get instant feedback, making it easier for them to manage their fields.

To make these models more accurate, data augmentation is often used. This means making small changes to training images, like rotating, flipping, or adding noise. These changes help the model learn to recognize plant problems in different situations, such as under different lighting conditions or from different angles [11], [14]. This makes the models more reliable, even when used in different regions or in changing weather conditions. It also helps the models work well when plants are at different growth stages.

Advanced neural network designs, like Residual Neural Networks (ResNets), Inception Networks, and DenseNets, have improved plant health detection even more. ResNets, for example, can be made very deep to learn complex patterns in plant images, which helps them find multiple problems like diseases, pest damage, and nutrient issues [10], [13]. These advanced models are very accurate and work well in both controlled labs and real fields, making them useful for large-scale farming. They can also identify health issues in many types of plants, even under challenging field conditions.

Besides using normal RGB images, adding multispectral and hyperspectral images has made plant health monitoring even better. These images capture light that the human eye cannot see, which helps the model tell the difference between healthy and unhealthy plants [8], [9]. For example, a plant with a nutrient deficiency may reflect light differently than a healthy one. By using these extra details, DL models become more accurate and can find problems earlier, helping farmers act quickly to protect their crops.

Transfer Learning is another technique that helps DL models learn faster. It involves using models that have already been trained on large image datasets and then fine-tuning them for specific tasks, like identifying diseases in a certain crop [7], [14]. This saves time and effort because it reduces the need for collecting a lot of labeled images. It also makes the models more accurate and adaptable, which is useful for farmers in different regions with different crops.

Combining environmental data like temperature, humidity, and soil moisture with DL models has made plant health monitoring even smarter. Plant problems are often linked to environmental conditions. For example, some diseases spread faster in humid weather, while nutrient deficiencies can be affected by soil moisture [1], [6]. By including this data, DL models can not only find existing problems but also predict future issues. This helps farmers take preventive measures, reducing chemical use and saving crops. It also allows farmers to plan better, making agriculture more sustainable and efficient.

Using real-world field images for training DL models has made them more reliable and accurate. In the past, models were trained using lab images, which did not capture the complexity of real agricultural fields [2], [13]. Field images include different lighting, weather conditions, and plant growth stages, which helps the models learn to handle real-world challenges. This makes the models more useful for farmers in different regions and farming practices. By training on field data, DL models can also learn to track disease progression, helping farmers monitor plant health over time.
DL models are also changing the way farmers manage plant health. By connecting these models with mobile apps, drones, and automated farming tools, farmers get real-time updates and advice [4], [5]. This helps them make better decisions without needing to rely on experts all the time. It also saves time and effort, making it easier to take quick action when needed. This combination of DL, remote sensing, and mobile technology is making agriculture smarter and more efficient.

However, there are still some challenges in building DL models that work well everywhere. Different plants have different needs, and they grow in different environmental conditions. To create accurate models, large and varied datasets are needed, covering different plant species, health problems, and regions [10], [11]. Sometimes, models struggle when used in new environments, showing the need for regular updates and improvements. Researchers are working to expand datasets, develop better network designs, and include expert knowledge to make these models more accurate and reliable.

The main goal of these advancements is to build powerful and easy-to-use plant health monitoring systems that help farmers manage crops more effectively. By using DL models, farmers can find and solve plant health issues faster and more accurately, which reduces chemical use and increases crop yields [3], [9], [15]. These systems also support sustainable farming by reducing environmental impact and using resources more efficiently. As DL technology keeps improving, its role in plant health monitoring will grow, helping farmers worldwide adopt smart, data-driven practices.

The future of plant health monitoring looks bright with the ongoing development of advanced DL models, multispectral imaging, and mobile technology. As these systems become smarter and easier to use, they will help farmers monitor their crops more accurately, make better decisions, and boost productivity [6]–[8]. By making these tools more accessible, DL technology will play a key role in improving food security, supporting sustainable agriculture, and increasing crop yields globally.

CHAPTER - 4

3. DESIGN SPECIFICATION

- Use Case Diagram
- Data Flow Diagram
- Class Diagram
- Sequence Diagram
- Activity Diagram
- State Chart Diagram

3.1. Use Case Diagram

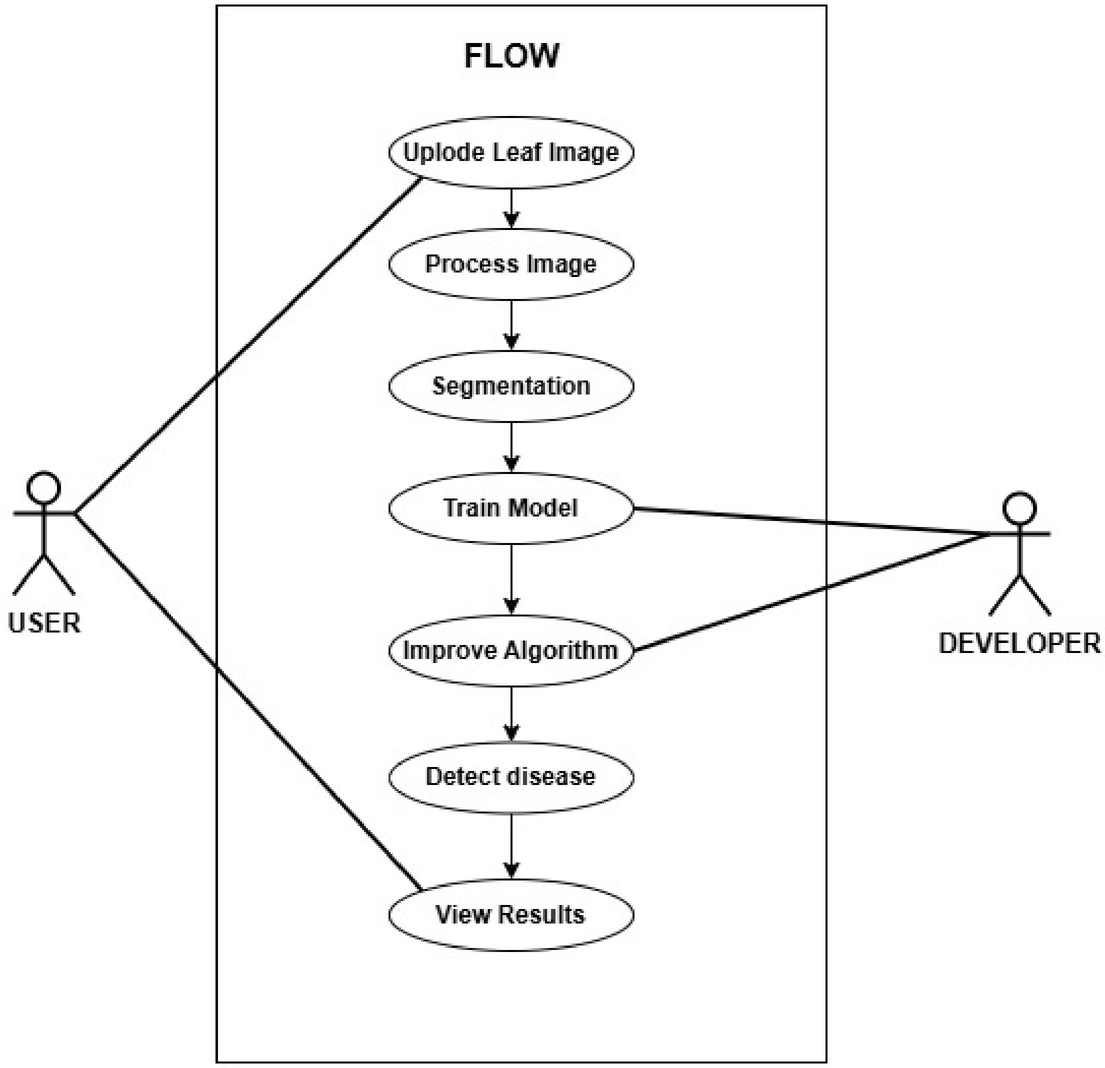

Figure 3.1: Use Case Diagram of Adaptive Attention Segmentation.

The Use Case Diagram refer to Figure 3.1 has two main actors: the User and the Developer. The User is supposed to upload an image of a leaf to the system and see the results of the disease detection. The interaction is made as simple and user-friendly as possible to enable users to get an easy response on the condition of the plant from the uploaded image.

The Developer, however, is tasked with training the AI model and enhancing the algorithm to enhance detection accuracy. Once the user uploads an image, the system processes and segments it. The trained model will then analyze the segmented image, identify any disease, and present the results to the user.

3.2. Data Flow Diagram

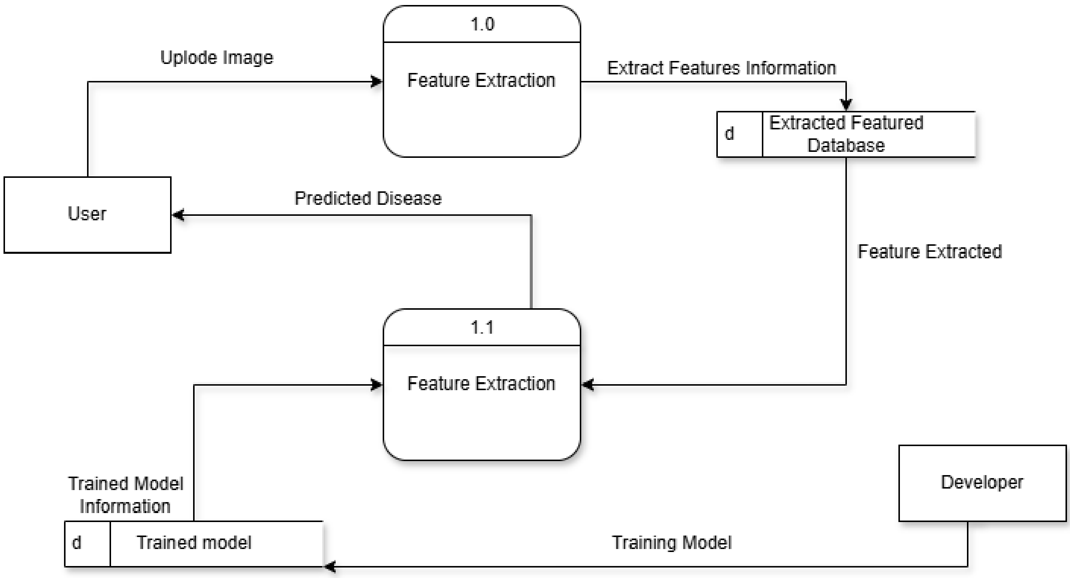

Figure 3.2: Data Flow Diagram of Adaptive Attention Segmentation.

As shown by Figure3.2, there exist two significant parties in the system: the Developer and the User.

The User enters the system by uploading a leaf image, which is processed in the Feature Extraction module (1.0). The module extracts relevant features of the image and stores them in the Extracted Feature Database. Once the features are extracted, they are passed into another Feature Extraction process (1.1) for further processing with the assistance of a trained model.

The Developer then trains the model on the gathered data, and the trained model is stored within the system. The trained model is then applied during the second phase of feature extraction to make a prediction of the plant disease, and the final output is presented back to the User.

3.3. Class Diagram

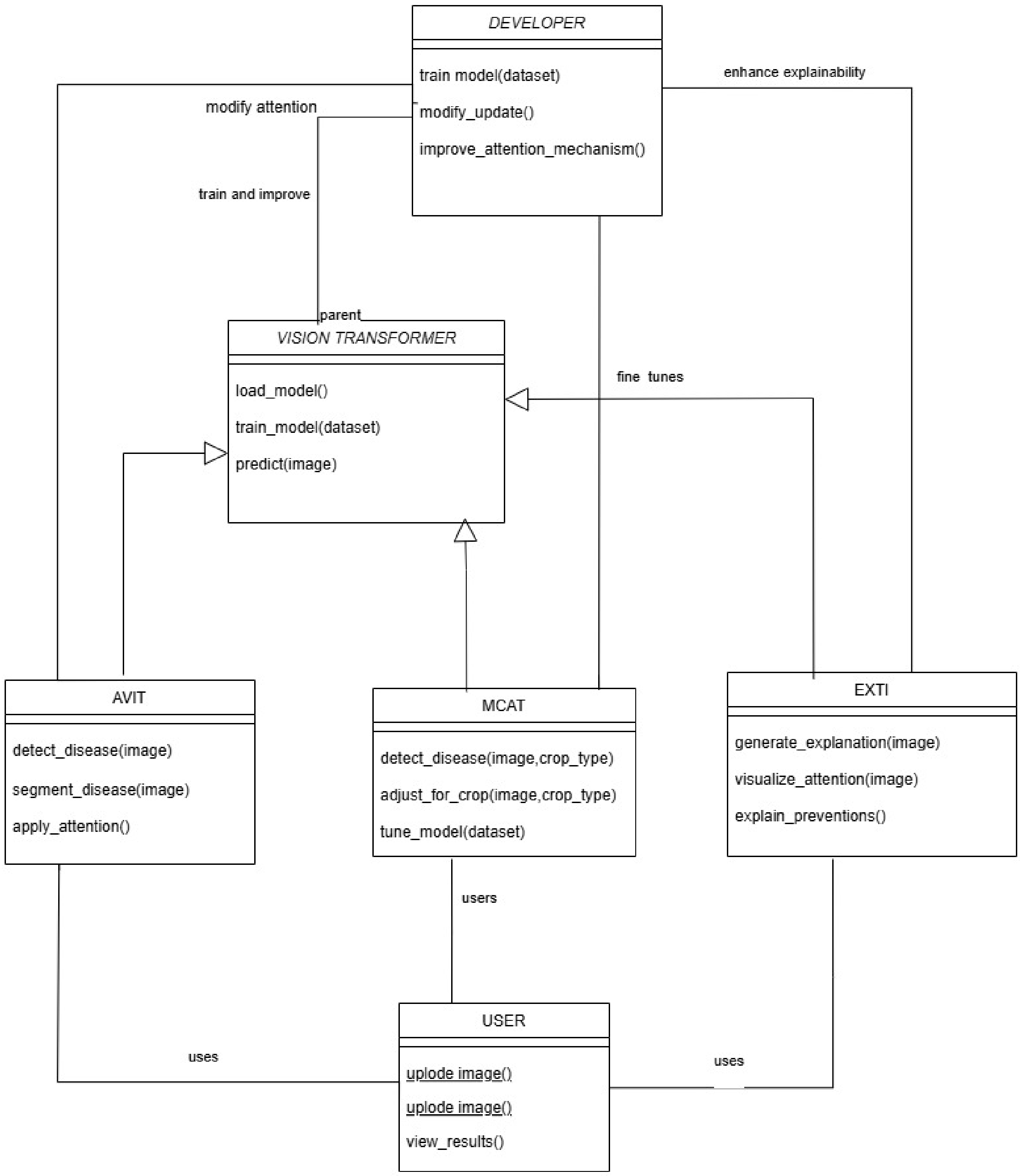

Figure 3.3: Class Diagram of Adaptive Attention Segmentation.

Figure 3.3 illustrates the class relationships within the proposed plant disease detection system. The primary classes include USER, DEVELOPER, and various model modules such as VISION TRANSFORMER, AVIT, MCAT, and EXTI.

The USER class interacts with the system by uploading leaf images and viewing results. These images are passed through different model classes. The AVIT and MCAT classes handle image-based disease detection, segmentation, and attention application—MCAT being crop-aware. The EXTI class is responsible for generating explainable outputs such as suggesting preventions. The DEVELOPER class is tasked with training, modifying, and improving the model's attention mechanism, interacting directly with the core VISION TRANSFORMER class that manages model loading, training, and prediction tasks.

3.4. Sequence Diagram

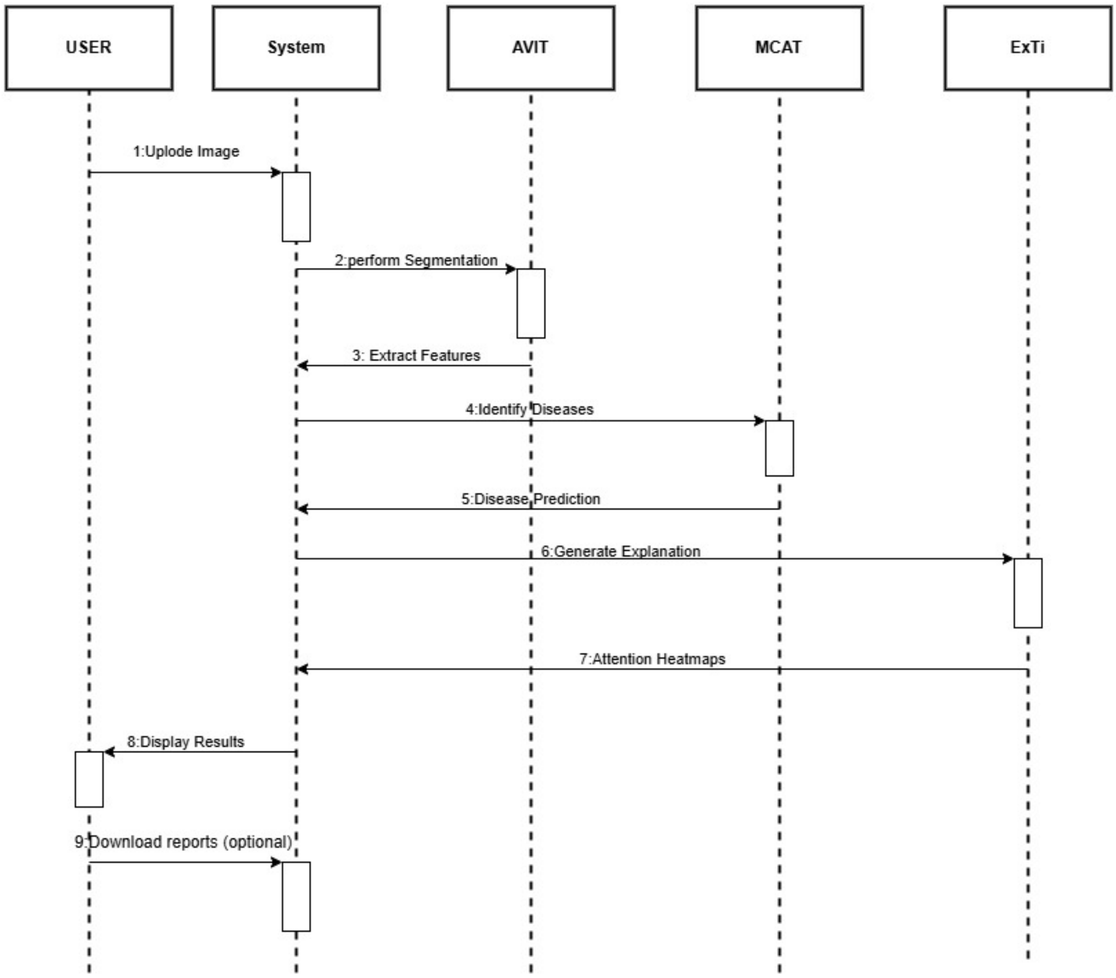

Figure 3.4: Sequence Diagram of Adaptive Attention Segmentation.

Figure 3.4 illustrates the sequence of interactions between different components of the plant disease detection system. The process begins with the user uploading an image. The system then performs segmentation (handled by AVIT), followed by feature extraction and disease identification using MCAT. Once the disease is predicted, an explanation is generated by the ExTI component. Finally, results are displayed to the user, and optionally, the user can download a report of the findings. The diagram captures the chronological order of communication and data flow among system entities.

3.5. Activity Diagram

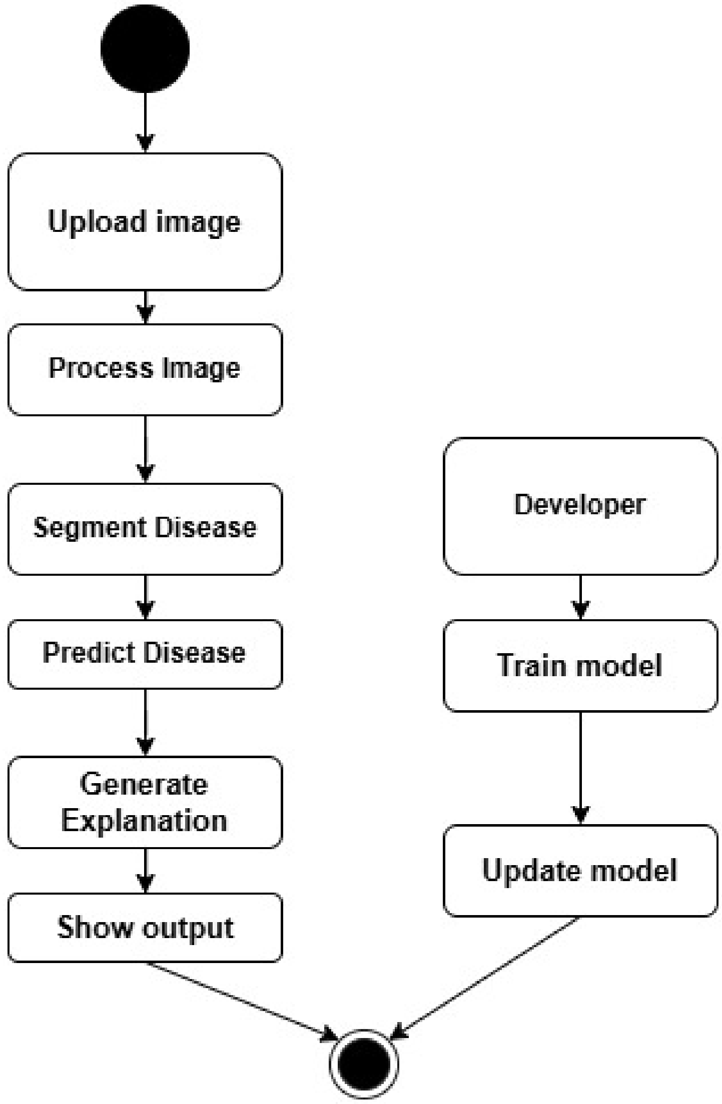

Figure 3.5: Activity Diagram of Adaptive Attention Segmentation.

Figure 3.5 shows the step-by-step workflow of the plant disease detection system. The process starts with the user uploading an image, followed by stages like processing the image, segmenting diseased areas, predicting the disease, generating explanations, and finally displaying the output. In parallel, the developer works on training and updating the model to ensure accurate predictions. This cyclical model enhancement helps maintain system performance and improve detection accuracy over time.

3.6. State Chart Diagram

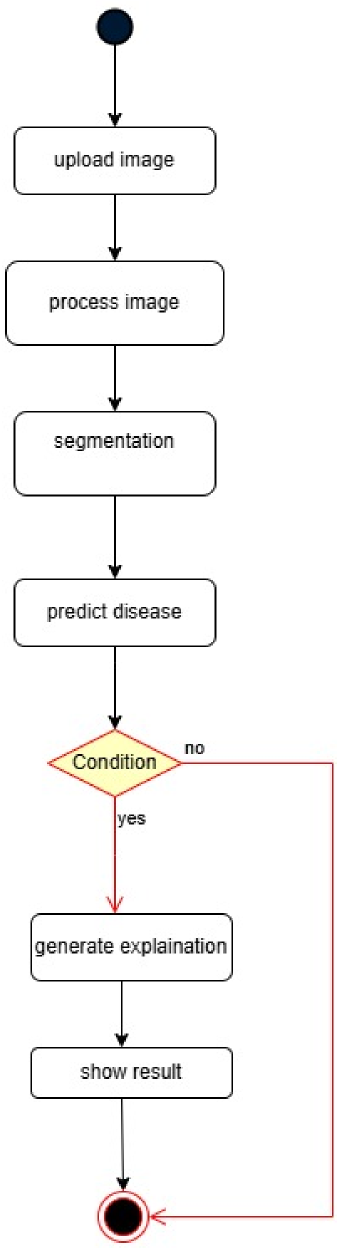

Figure 3.6: State Chart Diagram of Adaptive Attention Segmentation.

Figure 3.6 shows the state transitions in the plant disease detection system. The process starts when a user uploads an image. The system then processes the image, performs segmentation, and predicts the disease. A condition is evaluated to determine if an explanation needs to be generated. If yes, the system provides a textual explanation of the prediction before showing the result. If not, the system directly proceeds to display the result. This diagram clearly represents how the system responds to user input and internal decision-making at each stage.

CHAPTER - 5

4. METHODOLOGY

4.1. Dataset Collection:

This study uses a dataset with 76,000 images across 88 classes, compiled from 14 different sources. It includes both field and lab images of healthy and diseased plants. Images with poor quality, watermarks, or insufficient examples were removed to maintain consistency. The dataset covers a wide range of plant diseases, making it a valuable resource for identifying and understanding different plant health conditions.

4.2. Data Pre-Processing

Preparing the data is an important step to make sure the images are ready for training the Vision Transformer (ViT) model. First, all images are resized to a fixed size of 224×224 pixels to maintain consistency with the model's input requirements. This resizing ensures uniformity across all images, regardless of their original dimensions.

Next, pixel values are normalized to standardize the data distribution. Normalization is performed using either **Min-Max Normalization** or **Z-score Normalization**, depending on the requirements. Min-Max Normalization scales pixel values between 0 and 1, whereas Z-score Normalization adjusts them based on the dataset's mean (μ) and standard deviation (σ).

Additionally, image adjustments such as contrast and brightness modifications are applied to enhance visual clarity. This step ensures that key features, such as disease-affected regions, are more prominent while reducing unnecessary background noise. The preprocessing step results in a refined image, which is then fed into the model for further processing.

Algorithm: Image Preprocessing

Input: Image I

Step 1: Resize Image

$$I_r = \text{RESIZE}(I, (H, W)) \tag{4.1}$$

Step 2: Normalize Pixel Values If using Min-Max Normalization:

$$I_n = \frac{I_r - I_{\min}}{I_{\max} - I_{\min}} \tag{4.2}$$

where $I_{\min}$ is the minimum pixel value and $I_{\max}$ is the maximum pixel value of the image.
Else:

$$I_n = \frac{I_r - \mu}{\sigma} \tag{4.3}$$

where μ is the mean pixel value and σ is the standard deviation of pixel values.

Step 3: Apply Image Adjustments

$$I_a = \text{ADJUST}(I_n, \alpha, \beta) \tag{4.4}$$

Output: Preprocessed Image I_a

4.3. Adaptive Attention Segmentation(AAS):

A key challenge in detecting plant diseases is ensuring that the model focuses on diseased regions while ignoring unnecessary background elements. Most AI models treat all parts of an image equally, which can lead to lower accuracy and misclassification. To solve this, we propose Adaptive Attention Segmentation (AAS), a method that directs the model's attention toward affected areas while suppressing irrelevant details.

AAS starts by analyzing the input image to highlight the regions where disease symptoms are most likely to appear. Traditional deep learning models, such as CNNs, rely on generic feature extraction methods that do not always differentiate between healthy and diseased areas. Instead, AAS generates an attention mask that enhances disease-affected regions while downplaying distractions like soil, shadows, or healthy leaf parts. This helps the model focus only on important details for accurate classification.

Algorithm: Adaptive Attention Segmentation

Input: Preprocessed image I_a

Step 1: Generate Segmentation Mask

$$M_s = \text{SEGMENT}(I_a) \tag{4.5}$$

where M_s is the segmentation mask, which highlights regions of interest in the image.

Step 2: Suppress background and enhance disease regions

$$I_f = I_a \times A \tag{4.6}$$

where A is the layer that suppresses irrelevant regions (background) and enhances the regions with diseases.

Output: Focused image I_f

Once the attention mask is created, the Vision Transformer (ViT) processes the refined image. Unlike CNNs, which rely on fixed receptive fields, ViTs use self-attention mechanisms to analyze different image patches, making them more effective at identifying intricate patterns. With the help of AAS, ViT receives a cleaner input, free from background noise, allowing it to focus solely on disease-specific features.

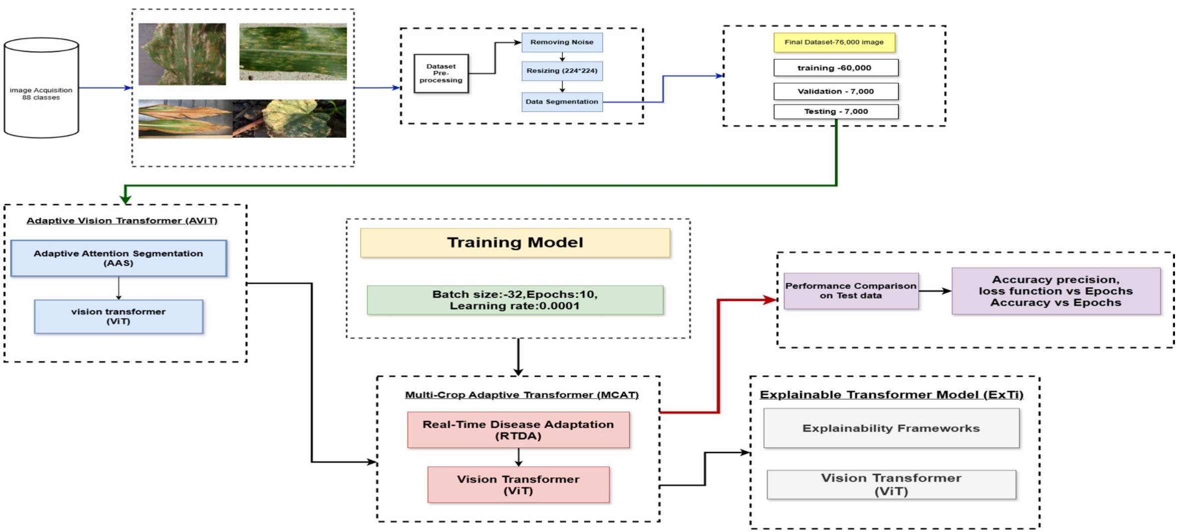

Figure 4.1: Architchure of Proposed Model

This significantly improves classification accuracy and reduces false positives and false negatives.By incorporating AAS, our model becomes more precise, better at isolating diseased regions, and more trustworthy. This technique strengthens deep learning applications in agriculture, enabling early disease detection and helping farmers take timely action to prevent crop losses. The synergy between AAS and ViT offers a reliable and high-performing approach to identifying plant diseases classification, making AI-powered agricultural tools more reliable and impactful.

4.4. Vision Transformer (ViT) for Feature Extraction:

Plant disease classification requires models to capture subtle differences between healthy and diseased plant regions. Traditional methods struggle with these subtle differences, but Vision Transformer (ViT) is a more suitable choice. Unlike traditional models, ViT processes images holistically and is hence more appropriate for detecting complex disease patterns.

ViT proceeds by treating an image as a sequence of small parts of it known as patches rather than processing the image as a single entity. This process allows the model to focus on fine-grained as well as large features without sacrificing important structural information. Through learning the relations among these patches, ViT can differentiate between diseases with similar visual features.

4.4.1. ViT Processing Steps:

Patch Embedding: The input image is split into patches of a set size, that are flattened and turned into vectors. These vectors are then passed through a linear projection layer to convert them into input tokens for processing.

Positional Encoding: Since images follow a fixed structure, positional encodings are added to patch embeddings. This step helps the model recognize spatial connections among various patches, preserving the overall layout of an image.

Self-Attention Mechanism: The key component of ViT is its Multi-Head Self-Attention (MHSA), which allows the model to analyze all patches of the image simultaneously. This helps it focus on the most important regions while minimizing the influence of less relevant areas.

Transformer Encoder: After processing, the tokens pass through multiple transformer layers, each consisting of self-attention and feed-forward networks. These layers enhance the extracted features, helping the model perform better in" distinguishing "different plant diseases.

Classification Head: The final output is passed through a classification layer after feature extraction,which classifies the image into a certain disease category. This ensures accurate identification of diseases in plants based on the extracted patterns.

By breaking down images into structured processing, ViT effectively captures local and global features, resulting in accurate classification of diseases. The capability to examine images as a whole makes it a strong tool for detecting subtle disease symptoms in plants.

4.5. AVIT Model:

The proposed model follows a structured approach to accurately detect and classify plant diseases by processing input images through a series of well-defined steps. The model consists of preprocessing, adaptive attention segmentation (AAS), and Vision Transformer (ViT)-based classification to ensure precise identification of disease-affected areas.

Initial Preprocessing: The input images go through preprocessing to standardize their size and pixel distribution. Each image is resized to 224×224 pixels for uniformity across the dataset. Normalization is performed to adjust all pixel values to a consistent range, either using Min-Max Scaling or Z-score Normalization, based on the dataset's characteristics.

After normalization, image adjustments such as contrast enhancement and brightness correction are applied. These adjustments improve visual clarity, making the disease-affected areas more distinguishable.

Adaptive Attention Segmentation (AAS): After preprocessing, the model utilizes Adaptive Attention Segmentation (AAS) "to highlight the most important areas of the image. This process creates an attention mask that enhances disease-specific regions while minimizing unnecessary background details.

The attention map is created by analyzing pixel intensities and structure, ensuring that only the necessary regions contribute to the classification process. The final output of this stage is an enhanced image that highlights disease symptoms clearly.

Vision Transformer (ViT) Processing: After segmentation, the processed image is passed into the Vision Transformer (ViT), which handles feature extraction and disease identification. The image is split into fixed-size patches, and each patch is converted into a feature representation.

The model follows these key steps:

Image Tokenization: The segmented image is split into patches, which are then converted into feature vectors. Position Encoding: To retain spatial relationships within the image, position embeddings are added to the tokenized patches. Self-Attention Mechanism: The model processes these tokens using self-attention layers, allowing it to extract features that represent disease patterns accurately.

Classification Head: The extracted features go through fully connected layers, which determine the disease class based on learned patterns.

Segmentation Refinement and Final Classification The final output of the model consists of both disease classification results and visual attention maps. These maps provide a clear indication of which parts of the image contributed to the classification, ensuring interpretability and transparency. The refined segmentation further enhances the model's accuracy by reducing noise and concentrating only on relevant infected areas.

This approach ensures a robust and reliable plant disease detection system by integrating adaptive segmentation and attention-based feature extraction. The structured pipeline enhances accuracy and efficiency, making it ideal for practical use in agriculture.

4.5.1. Multi-Crop Adaptation Transformer (MCAT)

To ensure robustness across a wide range of plant species, the model integrates a Multi-Crop Adaptation Transformer (MCAT). This component helps the system adjust dynamically to varying leaf structures, color profiles, and disease patterns across different crops.
The MCAT module operates by analyzing general visual patterns across multiple crop datasets. This cross-species learning allows the model to generalize its understanding, reducing the need for crop-specific retraining.
By adapting to inter-crop variations, MCAT enhances the model's flexibility and improves detection performance in real-world agricultural environments where disease symptoms may appear differently across plants. This enables scalable deployment across diverse farming systems.

4.5.2. Explainable Transformer Integration (ExTi)

Explainable Transformer Integration (ExTi) is included in the model pipeline to ensure that the plant disease detection process remains transparent and understandable to end users such as farmers and agronomists.

After classification, ExTi generates a visual explanation by indicating the regions of the image that influenced the model's prediction. These explanatory cues help users trust and interpret the results without needing in-depth knowledge of machine learning.
This interpretability is critical in high-stakes environments like agriculture, where understanding the "why" behind a diagnosis can inform treatment decisions and foster greater user confidence in AI-assisted systems.

4.6. System Block Diagram

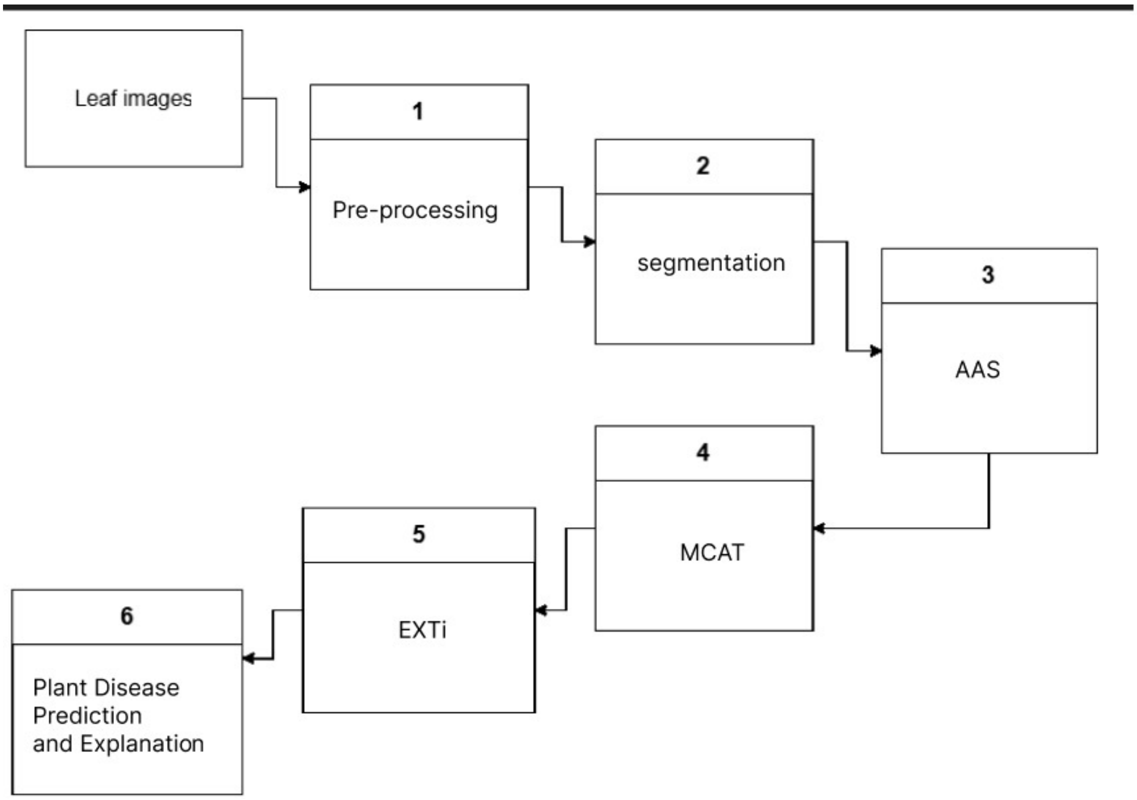

Figure 4.2: system block diagram

Figure 4.2 represents the block-level architecture of the plant disease prediction and explanation system. The process starts with input leaf images which undergo several stages:

1. **Pre-processing:** Enhances image quality for better analysis.
2. **Segmentation:** Separates the diseased region from the rest of the leaf.
3. **AAS Module:** Extracts critical features from segmented regions.
4. **MCAT Module:** Performs classification based on extracted features.
5. **EXTi Module:** Interprets the prediction results and generates explanations.
6. **Plant Disease Prediction and Explanation:** Final output is generated and displayed to the user.

CHAPTER - 6

5. Implementation Details

The Adaptive Vision Transformer (AViT)-based system is designed to improve the early detection and classification of plant diseases by identifying and highlighting diseased regions on plant leaves while filtering out irrelevant background noise. It uses Adaptive Attention Segmentation (AAS) to focus on infected areas, and the Multi-Crop Adaptation Transformer (MCAT) allows it to adapt effectively across different plant species and environments. Additionally, a basic rule-based system is integrated to assist users with queries regarding symptoms, treatment, and prevention of plant diseases. This section explains the implementation approach used to develop and combine AViT, AAS, MCAT, and the rule-based system for efficient disease diagnosis.

5.1. Technology Stack

The technology stack for the Plant Disease Detection and Rule-Based Assistant model includes Python as the core programming language, leveraging its rich ecosystem of machine learning and computer vision libraries. The model uses deep learning techniques for plant disease detection, employing frameworks like PyTorch for training and inference. OpenCV is used for image processing tasks, while PIL (Python Imaging Library) is integrated for further image enhancements such as blurring and highlighting. Gradio provides a simple and user-friendly interface for interaction and visualization. A rule-based system is incorporated to manage user queries related to symptoms, treatment, and prevention of plant diseases. The entire model and interface are deployed on Google Colab for easy access to computational resources, with Google Drive utilized for model storage and retrieval.

5.2. System Architecture

In designing the architecture of our Plant Disease Detection and Rule-Based Assistant model, we focused on ensuring high efficiency and scalability. The system is built to handle large volumes of data and manage increasing traffic, while maintaining optimal performance. To achieve this, we adopted a modular architecture that combines the detection and assistant components seamlessly. The core of the system is a multi-tier structure that includes a user interface, model inference layer, and backend services for data processing and rule-based responses. The inference layer, which hosts the deep learning models, interacts with the data processing components for image pre-processing and post-processing tasks. The backend services facilitate the integration of a rule-based assistant to handle user queries related to plant diseases. Gradio provides a lightweight web interface for interaction, while the backend is powered by Python and cloud-based services for scalability and ease of model deployment. This architecture ensures that the system remains responsive and adaptable, even under high demand.

5.3. User Interface

The user interface of the Plant Disease Detection and Rule-Based Assistant model is designed to be simple and intuitive. Users can easily upload plant images for disease detection, and the system highlights the diseased regions by blurring the non-diseased parts of the image. The integrated rule-based assistant provides responses to user queries regarding symptoms, treatment, and prevention of plant diseases. Built with Gradio, the interface ensures a smooth user experience without requiring technical expertise

5.4. Integration

The Plant Disease Detection and Rule-Based Assistant model requires seamless integration between the plant disease detection model and the rule-based assistant. The detected diseased regions from the plant images will be used to trigger relevant information from the assistant about symptoms, treatment, and prevention. The system will be integrated into a Gradio-based web interface, allowing users to upload images for detection and get responses from the assistant. Deployment will be handled on Google Colab for ease of access and scalability, and user feedback will be essential for refining the model and ensuring its compatibility across various plant species and disease types.

5.5. Testing and Deployment

The Plant Disease Detection and Rule-Based Assistant model will undergo extensive testing to ensure its accuracy, robustness, and reliability in real-world applications. Testing techniques such as unit testing, integration testing, and user acceptance testing will be used to verify the model's ability to accurately detect plant diseases and provide relevant information through the rule-based system. The model and its interface will be deployed on Google Colab for scalability and easy access to computational resources. Continuous monitoring will be implemented post-deployment to ensure consistent performance, and user feedback will be utilized to refine the system for improved adaptability across different plant species and disease scenarios.

CHAPTER - 7

6. OBSERVATIONS

6.1. Time Domain - Gann Chart

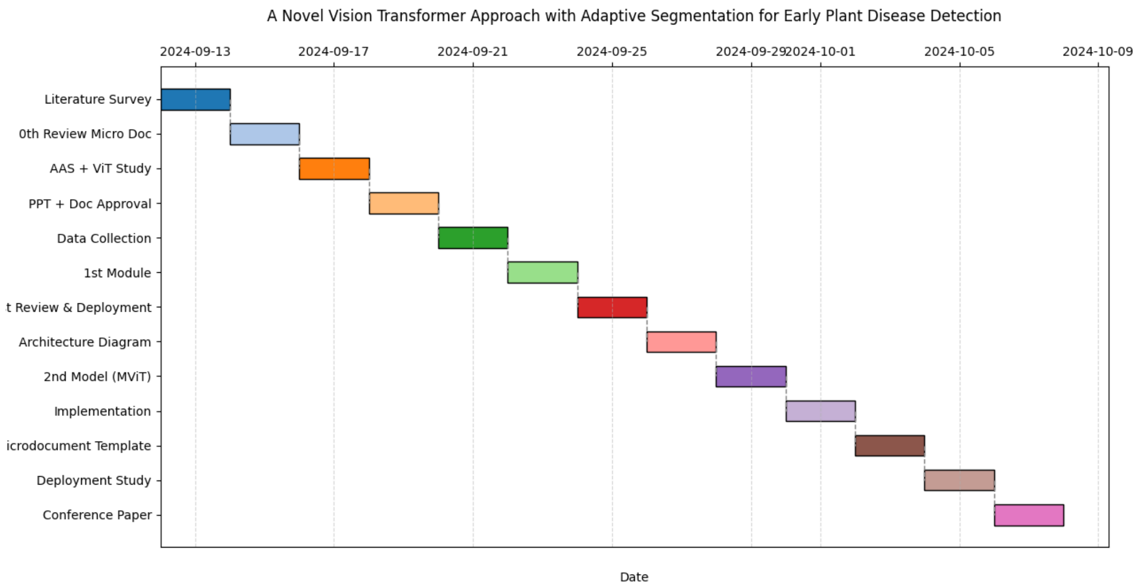

Figure 6.1: Gann Chart.

The Gantt chart presented illustrates the sequential workflow of the project titled "A Novel Vision Transformer Approach with Adaptive Segmentation for Early Plant Disease Detection." Each bar represents a specific task in the project, beginning with the literature survey and progressing through data collection, model implementation, deployment study, and conference paper preparation. The horizontal bars indicate the duration of each task, while their position along the timeline reflects their scheduled start and end dates.

All tasks are displayed in a clearly stacked format with distinct colors, making it easy to track the project's flow from one phase to the next. The timeline is shown along the top, offering a clear visual reference for the planning and progress of each project milestone.

6.2. Results and Comparitive Study

Table 6.1: Model Evaluation Metrics

Model	Accuracy	Precision	Recall	F1-Score
CoAtNets and Swin Transformer V2 (Wheat only)	98.00	98.00	98.00	98.00
MobileNet-V2	96.89	94.22	94.22	94.22
Deep CNN	96.46	96.47	99.89	98.15
Vision Transformers (ViTs)	94.00	90.00	90.00	90.00
Proposed Model	**98.54**	**98.54**	**98.54**	**98.54**

The table above presents a comparative evaluation of multiple deep learning models used for plant disease classification. Metrics such as accuracy, precision, recall, and F1-score were used to assess the performance of each model. The proposed model outperformed all others with an accuracy of 98.54%, demonstrating superior consistency across all metrics. Among the existing models, CoAtNets and Swin Transformer V2 also performed competitively, especially in the wheat-specific classification task. The results validate the robustness and effectiveness of the proposed architecture for early and accurate plant disease detection.

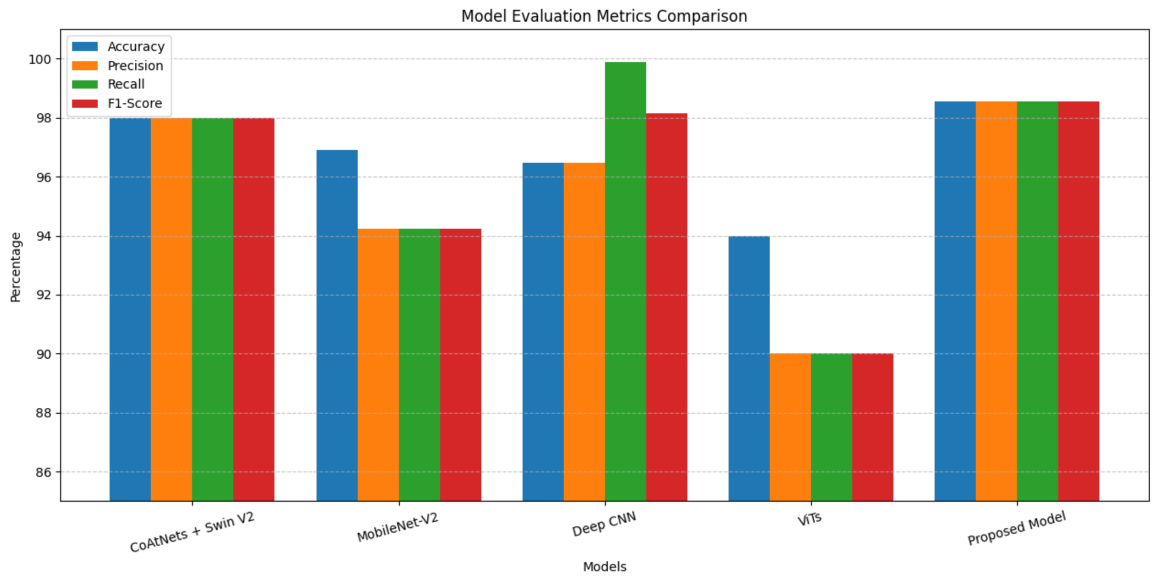

Figure 6.2: Metrics Graph

6.3. Gradio Interface

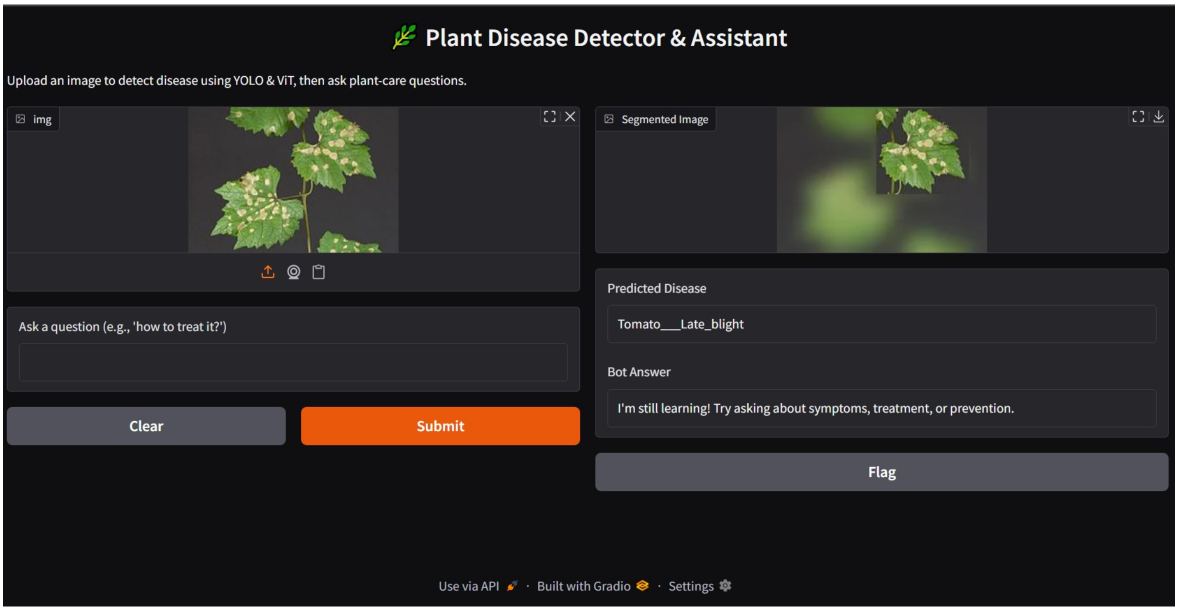

Figure 6.3: Prediction Results Displayed on the Gradio Interface

The above figure shows the Gradio-based web interface used for plant disease detection. Users can upload leaf images directly via the browser, and the system displays the predicted disease class along with the segmented image that highlights affected areas. This intuitive interface allows real-time interaction and visualization, enhancing accessibility for agricultural experts and farmers.

7. Discussion

In this paper, we proposed a model for plant disease detection using the Adaptive Vision Transformer (AViT), Multi-Crop Adaptation Transformer (MCAT), and a rule-based system for user interaction. The model efficiently localizes diseased areas and adapts to various plant species. In future work, we plan to integrate Grad-CAM for generating heat maps to enhance the model's interpretability, allowing users to visualize the detection process. Additionally, we aim to improve the rule-based chatbot by expanding its knowledge base and incorporating more advanced natural language processing techniques to provide more accurate and interactive support for plant disease diagnosis and management. These enhancements will make the system more effective and user-friendly in agricultural applications.

CHAPTER - 8

8. CONCLUSION

The Plant Disease Detection and Rule-Based Assistant model offers significant potential in agricultural applications by automating the process of plant disease detection and providing a rule-based assistant for user queries. By implementing these features:

- Disease detection and classification
- Rule-based system for symptom, treatment, and prevention queries

The model can assist farmers and agricultural experts in early disease detection, reducing crop loss and improving overall crop health management. During the development of this project, we successfully collaborated as a team and gained a deeper understanding of machine learning, computer vision, and rule-based systems. This experience highlighted the real-world implications of developing intelligent systems and the importance of thorough documentation in the model development process. Although the system has not yet been deployed on a large scale, we believe it will significantly enhance disease detection, increase accessibility for farmers, and streamline decision-making processes in agriculture. In conclusion, the Plant Disease Detection and Rule-Based Assistant model is a valuable tool that can help revolutionize plant health management by automating disease detection and providing useful insights for effective treatment and prevention.

CHAPTER - 9

9. LIMITATIONS AND FUTURE ENHANCEMENTS

9.1. Limitations

While the Plant Disease Detection and Rule-Based Assistant model has been carefully planned and executed, several limitations were encountered throughout the project. These limitations stem from various factors, including data characteristics, system complexity, and external requirements. Key limitations include:-

1. Complexity of Model Implementation: Developing deep learning models, especially for those with limited experience, can be challenging. Proper architecture, hyperparameter tuning, and training require significant expertise and can be time-consuming.

2. Data Preparation Challenges: Preparing the plant disease dataset, including cleaning, normalization, and augmentation, is crucial for model performance. Inadequate preprocessing can lead to suboptimal results.

3. Computational Requirements: Training deep learning models on large datasets demands substantial computational resources, often requiring GPUs or cloud-based platforms, which may lead to additional costs.

4. System Maintenance and Deployment: Managing model version control and ensuring smooth deployment can be complex, especially with continuous updates and bug fixes. Adequate testing and rollback strategies are necessary for stable production.

5. Hardware Constraints: The model's performance on target devices, such as edge devices for plant disease detection, may require optimization for efficient execution.

6. User Adoption: The effectiveness of the system depends on user acceptance and proper training. Ensuring user-friendly interfaces and comprehensive training materials for users is crucial for successful integration into workflows.

7. Ongoing Maintenance: Regular updates and support are needed to ensure the system remains effective and adapts to new challenges.

9.2. Future Enhancements

The Plant Disease Detection and Rule-Based Assistant model has the potential for further enhancement to address emerging needs and improve its capabilities. Some key areas for future development include:

1. Disease Detection and Classification: Integrating the ability to detect and classify various plant diseases will expand the system's functionality. This will help in providing more detailed insights into plant health and enable early intervention for better disease management.

2. To enable immediate feedback during field inspections, the model can be optimized for real-time inference on edge devices or plant monitoring systems. This would allow quicker analysis and decision-making in the field.

3. Expanding the model to support various plant imaging techniques such as thermal imaging or multispectral data would provide a more comprehensive analysis of plant health and environmental factors.

4. Developing a cloud-based platform for researchers, agricultural experts, and farmers to securely upload images and share analysis results would promote collaboration and assist in remote monitoring and diagnostics.

5. A mobile app would enhance accessibility by enabling farmers and agricultural experts to access real-time disease detection and model predictions on their smartphones or tablets.

6. Incorporating explainability into the model's predictions would help users understand how the system identifies diseases, fostering trust and ensuring that experts can interpret the results effectively.

7. Partnering with agricultural research organizations to test and validate the model on larger, more diverse datasets would enhance its accuracy and reliability in varied farming environments, ensuring its broader applicability.

A. APPENDIX

A.1. References

[1] J. Chaki and D. Ghosh, "Deep learning in leaf disease detection (2014–2024): A visualization-based bibliometric analysis," *IEEE Access*, 2024.

[2] M. Fahim-Ul-Islam, A. Chakrabarty, S. T. Ahmed, R. Rahman, H. H. Kwon, and M. J. Piran, "A comprehensive approach towards wheat leaf disease identification leveraging transformer models and federated learning," *IEEE Access*, 2024.

[3] S. Hemalatha and J. J. B. Jayachandran, "A multitask learning-based vision transformer for plant disease localization and classification," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 188, 2024.

[4] N. Nigar, H. M. Faisal, M. Umer, O. Oki, and J. Lukose, "Improving plant disease classification with deep learning based prediction model using explainable artificial intelligence," *IEEE Access*, 2024.

[5] H. Wang, T.-H. Nguyen, T. N. Nguyen, and M. Dang, "Pd-tr: End-to-end plant diseases detection using a transformer," *Computers and Electronics in Agriculture*, vol. 224, p. 109 123, 2024.

[6] A. Haridasan, J. Thomas, and E. D. Raj, "Deep learning system for paddy plant disease detection and classification," *Environmental monitoring and assessment*, vol. 195, no. 1, p. 120, 2023.

[7] M. Shoaib, B. Shah, S. Ei-Sappagh, *et al.*, "An advanced deep learning models-based plant disease detection: A review of recent research," *Frontiers in Plant Science*, vol. 14, p. 1 158 933, 2023.

[8] S. Yu, L. Xie, and Q. Huang, "Inception convolutional vision transformers for plant disease identification," *Internet of Things*, vol. 21, p. 100 650, 2023.

[9] Y. Borhani, J. Khoramdel, and E. Najafi, "A deep learning based approach for automated plant disease classification using vision transformer," *Scientific Reports*, vol. 12, no. 1, p. 11 554, 2022.

[10] C. Jackulin and S. Murugavalli, "A comprehensive review on detection of plant disease using machine learning and deep learning approaches," *Measurement: Sensors*, vol. 24, p. 100 441, 2022.

[11] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[12] G. Geetharamani and A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Computers & Electrical Engineering*, vol. 76, pp. 323–338, 2019.

[13] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, and A. Johannes, "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild," *Computers and Electronics in Agriculture*, vol. 161, pp. 280–290, 2019.

[14] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019.

[15] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and electronics in agriculture*, vol. 145, pp. 311–318, 2018.

A.2. Project Timeline Table

Tri-Level Segmented CNN project timeline: 12 September 2024 to 03 June 2025

date what discussed what actions taken

Table A.1: Project timeline.

Date	What discussed	What actions taken
13-09-2024	How to write literature survey	read the papers thoroughly and completed the survey
09-11-2024	micro document for 0th review	completed the document
13-11-2024	AAS algorithm and ViT transformer	studied the models and types of plant diseases
20-11-2024	ppt and document approval for review	made the necessary changes
06-01-2025	Dataset collection, preprocessing	Collected data
20-01-2025	make the 1st module	trained the module and got results
27-01-2025	announcement of 1st review	Made the ppt and started deployment
29-01-2025	Architechure diagram	drawn the diagram using online tool
06-02-2025	make the 2nd model MViT	started the code and got the resulst
20-02-2025	Implementation steps are discussed	Started implementing
01-03-2025	Complete microdocument template	used overleaf for document
10-03-2025	deployment	started learning gradio
19-03-2025	Conference paper must be ready	written a conference paper