Olive Leaf Disease Detection Using Transformer-Based Deep Learning Approach

Shahzada Muhammad Junaid[1], Rabia Tehseen[1*], Uzma Omer[3], Muhammad Inam Ul Haq[3], Ayesha Zaheer[4]
[1]Department of Data Science, University of Central Punjab, Lahore, Pakistan.
[2]Department of Computer Science, University of Central Punjab, Lahore, Pakistan.
[3]Department of Computer Science, University of Education, Lahore, Pakistan.
[4]Department of Computer Science and Bioinformatics, Khushal Khan Khattak University, Karak, Pakistan
*Correspondence: rabia.tehseen@ucp.edu.pk

Citation | Junaid. S. M, Tehseen. R, Omer. U, Haq. M. U. I, "Olive Leaf Disease Detection Using Transformer-Based Deep Learning Approach", IJIST, Vol. 7, Issue. 4 pp 3048-3062, December 2025
Received | November 04, 2025 **Revised |** November 20, 2025 **Accepted |** November 25, 2025 **Published |** December 04, 2025.

The use of AI and DL in automated crop health monitoring and disease diagnosis, especially relevant to Pakistan's burgeoning olive growing industry, has gained momentum. This paper proposes a transformer-based deep learning approach for the detection of olive leaf diseases due to significant shortcomings in the robustness and generalization of traditional convolutional neural networks. The proposed system makes use of a Vision Transformer (ViT) architecture to extract both local and global contextual features from the images of leaves using multi-head self-attention mechanisms. The developed Optimized ViT-Small model identifies olive leaves into three classes: Healthy, Aculus olearius, and Olive Peacock Spot. It is trained and tested on a pre-processed dataset of 3,400 high-resolution olive leaf images collected from olive-growing regions of Pakistan. Experimental results show strong performance with a test accuracy of 97% while demonstrating high precision, recall, and F1-scores throughout the classes. Moreover, performance assessment through confusion matrix analysis, ROC AUC, and precision-recall curves supports the developed model's effectiveness. Although the dataset's geographical coverage is limited, the results indicate that transformer-based architectures are an attractive alternative for the applications of precision agriculture in Pakistan.

Keywords: Artificial Intelligence (AI), Deep Learning, Vision Transformer (ViT), Olive Leaf Disease Detection, Explainable AI, Smart Agriculture.

Introduction:

Artificial Intelligence (AI) has been considered as one of the most robust transformation technologies used in agriculture to address challenges related to plant health. In this regard, deep learning (DL) methods that rely on AI can automate the detection of diseases and monitoring of plant health, and reduce dependence on human inspection and expert evaluation [1][2][3][4]. Unique visual features of the dataset were automatically extracted from plant images using deep convolutional neural networks (CNNs), which learn to identify patterns such as leaf texture, shape, and color variations relevant to disease symptoms. This automated feature extraction enables a highly effective and reliable process for olive leaf disease identification. By integrating both feature extraction and classification within the CNN framework, the proposed approach optimizes detection accuracy and efficiency [1][2][5][6][7].

New developments have propelled computer vision and transformer-based DL into the zone of precision agriculture through increased image-based interpretation and classification. Computer vision enables an AI-based system to comprehend visual clues of leaf images highly important for early diagnosis and preventive treatment against crop diseases [3][8]. The transformer was first developed for natural language processing (NLP), which changed image classification due to its global and local dependencies in the visual representation [9][10][11][12][13]. Previously proposed frameworks like ViT and hybrid CNN-Transformer can actually perform efficiently and can be scaled to complex agricultural scenarios. Multiclass olive leaf disease recognition has been performed by multiple researchers in the literature [10][12][13].

Globally, in the Mediterranean and subtropical regions, it has been reported that productivity and quality of olive plants are being greatly affected by bacteria and fungi [14][15][8] due to excessive rain. Traditionally, disease diagnosis through direct inspection is a lengthy process, invariably subject to inconsistency and often dependent on the availability of experts, thus rendering any effective intervention impossible [3]. DL provides scalable and automated methods for classifying diseases with a high degree of accuracy [14][16]. With very few instances of disease samples in their training data and with minimal human supervision, the CNN-based architectures have produced state-of-the-art results in the identification of olive leaf diseases [5][14][15][17].

Further, depending on the parameters of illumination and other conditions in which the images are captured, traditional CNNs cannot relate the contextual entities, including leaf textures, colors, and lesion patterns [17][18][5]. The hybrid DL architectures have been proposed in different related works to further improve the extraction of the features and performance [16][18][19]. While these hybrid models adopt both spatial and sequential learning, their robustness on different datasets still has timely detection limitations [20]. The motivation of this research is to develop an advanced architecture that learns both the local and global contextual dependencies existing in the olive leaf images and improves its diagnostic accuracy and scalability.

The proposed work intends to present a DL based framework using the Transformer for olive leaf disease. Although the CNNs and hybrid models have attained noteworthy performances [16][19][21], all reported their limitations of degradation in performance when datasets are complex or even when environmental conditions are changed. Most hybrid CNN-LSTM and CNN-Transformer models have maintained superior accuracy performance in these classification tasks, but poor generalizability across domains and a lack of interpretability have plagued the applicability of their methods [21][19].

The proposed research incorporates a multi-head self-attention mechanism with a Transformer backbone that will facilitate enhanced spatial feature extraction, followed by better interpretability. Although recent research work confirmed the efficacy of Vision Transformers and MetaFormer variants [22], it lacks generalizability for the quality classification of olive diseases. Generatability and state-of-the-art data augmentation with feature fusion methods have been integrated into the proposed framework to improve its robustness and

diagnostic reliability.

Smart and sustainable agriculture is impacted by the proposed Transformer-based model that exploits the potential of transfer learning with attention mechanisms [12]. Its complete evaluation has been conducted in multiple studies using several benchmark datasets, thereby adding to the cross-domain generalization credentials [17][18][19]. Important determinants in the development of reliable agricultural AI systems are variations in datasets, optimized training pipelines, and models' explainability [18][20][23][24][25][26]. Ultimately, this method provides early assessment of diseases in olive leaves but still have limitation of interpretability while maintaining their accuracy, thus impacting negatively on yield and environmental sustainability.

Olive leaf disease detection within Pakistan is significantly progressing. Olive farming already features in national initiatives within the provinces of Punjab and Khyber Pakhtunkhwa; the crops face severe yield and quality reductions due to diseases. The current state-of-the-art systems based on CNN and a hybrid CNN-Transformer approach usually lack robustness and interpretability, enhanced in this work concerning model accuracy and applicability in various environments through optimized attention mechanisms, effective training strategies, and feature fusion methods. The proposed research provides the way forward by producing the most optimized Transformer- based deep learning models for in-time crop disease detection, ultimately increasing per-yield production. This work will lead towards sustainable cultivation, providing a reliable and automated framework for disease detection that would assist farmers, researchers, and policymakers in the field of agriculture in Pakistan.

Large-scale plantation of olives has been recently encouraged in Khyber Pakhtunkhwa and Punjab, and the cultivation of olives is now an upcoming industry in Pakistan. However, olive productivity is being threatened by bacterial and fungal leaf diseases, often misdiagnosed or caught too late in the process for effective intervention due to reliance on manual visual inspection by the field personnel. Traditional methods are time-consuming, subjective, inconsistent between different inspectors, and incapable of early-stage disease identification at a time when symptoms may be minimal.

While CNN-based models are the dominant deep learning approaches that have advanced automated disease recognition, they still suffer from serious technical limitations. The general limitations of CNNs include poor generalization across environmental conditions, sensitivities to changes in illumination and background, and inability to model long-range dependencies and variations in global texture that are indicative of the respective phenotypes of olive leaves. Each factor affects robustness and generally tends to result in lower-than-expected performance in challenging agricultural environments such as those found in Pakistani farms. Plant disease detection has significantly improved with recent transformer-based methods. Vision Transformer models have outperformed traditional CNN-based techniques when applied to datasets of cotton leaf disease [27] and general plant leaf disease [28]. In particular, it has been reported that olive leaf disease can be detected by combining CNN and ViT models [9], demonstrating the increasing importance of transformer-based architectures in precision agriculture.

In particular, compared with CNNs, the Vision Transformers have demonstrated improved modeling of global spatial relationships, enhanced interpretability through attention mechanisms, and greater robustness to challenging variations in vision. Their capabilities, however, have not yet been investigated for olive leaf disease detection in Pakistan under conditions of heterogeneous leaf morphologies, seasonal color variations, and high inter-class similarity. Thus, the fundamental technical challenge to be addressed by this research involves the design of a Transformer-based deep learning framework that overcomes CNNs' limitations by effectively modeling long-range dependencies between textures, improving generalization across diverse field conditions, and enabling accurate, robust, and interpretable olive leaf disease classification for precision agriculture in Pakistan.

Literature Review:

With the advancement in AI, DL has been taking over the future of plant pathology [29]. In olive leaf disease detection, deep-learning-based architectures have greatly improved automation [21], early disease detection [12], and classification accuracy [4]. Much improved accuracy and robustness came from dynamic clustering techniques without overfitting [13]. Deep CNN architecture has also performed exceptionally well under field conditions, classifying multiple diseases with high reliability but challenges in terms of interpretability and generalization across datasets [10]. Some early works using CNNs laid the groundwork for image classification in agriculture, hence establishing the benchmark for deep learning in that area.

CNN-Based Approaches:

Recent systems of olive leaf disease recognition has built around CNN-based architectures. It has been found that CNNs, as opposed to conventional algorithms, perform better in extracting visual features in that they have deep hierarchical feature representation capability. Baseline models using convolutional layers with a different number of pre-processing stages provided a good classification accuracy for small datasets. Furthermore, the depth of the network, along with filter and pooling size, should be varied for better model performance while reducing the model's computational complexity. Generally, these methods have the capacity to generalize well for application even in uncontrolled scenarios and be majorly dependent on large balanced datasets for stable performance.

Hybrid Deep Learning Models:

Hybrid DL models brought together the different advantages of the diverse architectural styles for their added merit in performance. For instance, the development of MobivRes-Net, a hybrid combining Mobile Net and ResNet, has not only improved the detection accuracy but also increased the computational efficiency in classifying olive leaves. Other models merged CNN and LSTM layers for simultaneous consideration of both spatial and temporal dependencies. This greatly enhanced the possibility of earlier detection, as well as preventive diagnosis. Self-attention mechanisms and convolutive feature extraction were part of the incorporated systems in the latest hybrid CNN-Transformer frameworks, which reported better classification and generalization abilities, but these models typically require more complex tuning and, hence, more processing power.

Transformer and Vision Transformer (ViT) Models:

The research on disease detection of olives has evolved with transformer-based architectures [30]. The performance achievement on high accuracy for multiple datasets made transformer- based models like MetaFormer achieve the beating performance and flexibility compared to CNNs. Vision Transformers have recognition capabilities, which can model long-range dependencies in image data. State-of-the-art, accurate models in hybrid CNN-Transformer frameworks have used the integration of spatial and attention-based learning; however, they are practically implementable with limited model complexity. Fully, this shows that methodologies based on transformers constitute the most recent means of development in agricultural image analysis, with a leaning towards robustness and generalization. Recent research has shown how successful Vision Transformers are at identifying plant diseases. ViTs have been used to classify cotton leaf diseases with a high degree of accuracy [27]. Additionally, they have been applied to the diagnosis of general plant leaf diseases, and hybrid ViT-CNN models have enhanced the classification of olive diseases [28][9]. These studies show that transformer-based architectures are strong frameworks for agricultural image analysis because they improve accuracy and generalization across datasets.

Transfer Learning Techniques:

Multiple pre-trained CNN architectures have already been fine-tuned on olive leaf datasets and delivered high accuracy, achieved quickly with fewer training hours for models

[15][10][30]. One of those state-of-the-art methods is the EfficientNet models, which optimizes all three scales of depth, width, and resolution [25]. These methods improved model efficiency and accuracy; however, many of them inherited from source domains [30].

Optimization and Clustering Enhancements:

Techniques aimed at infusion of efficiency into CNN are optimization of parameters and adaptive clustering. The optimization-based hyperparameter tuning increased the model's accuracy and stability during training in all the deep learning models. Dynamic clustering reduced the intra-class variance in improving the robustness of feature learning. The methods in the frequency domain, like spectral analysis and Fourier transformation, provided much more discriminative power. Hence, enhances robustness to noise and environmental variability [31]. These studies' orientation towards optimizations proves the necessity of fine-tuning the architecture and learning strategies for their better performance [32].

Traditional Machine Learning Approaches:

Traditional machine learning methods provide cheap and interpretable alternatives, even though deep learning occupies the greater share of current research. Among them are classical algorithms such as decision trees, Naïve Bayes, and support vector machines. They record applications for predicting olive disease, making their baseline comparisons on accuracy against comparative costs.

Explainable and Interpretable AI:

One of the key concerns regarding AI for agriculture has been the question of interpretability. Some explainable DL models have thus been proposed to foster transparency of the automated disease diagnosis systems. These models offer either a visual or a semantic explanation for the predictions using CNN, thus aiding in decision-making by agronomists and farmers. Although performance may often take a hit in exchange for predictability, there are other cases where the arrival of explainable versions has vastly improved the functional usability of DL [32].

Multimodal and Frequency-Domain Approaches:

The robustness of the detection process was raised by the merging of visual and spectral data for disease detection with multimodal data fusion. The one disadvantage of the models was that they were expensive and complicated in data collection from various environmental settings. Generalizations and noise resistance were enhanced by employing Fourier-based features from spectral and frequency-domain analysis techniques. AI-supported imaging systems have widened the diagnostic scope for detecting disease and nutrient content in agricultural setups.

Smart Agriculture and Sustainability Integration:

Deep learning has been embedded in smart agriculture for precision farming. AI techniques are thus at the forefront of promoting olive production through disease prevention, yield prediction, and early detection of the actual symptoms observed. Thus, the DL-based sustainable agricultural frameworks have led to resource-efficient agricultural methods. This has also used CNNs and remote sensing to map olive diseases spatially through multispectral imagery, although resolution is still a constraint. The lightweight CNN for mobile and embedded devices has further facilitated real-time monitoring in agriculture. Altogether, this shows how AI could be harnessed to create precision and sustainable farming.

Summary and Research Gaps:

The reviewed studies provide preferential proof about the transition of CNN for the olive leaf disease detection toward transformer-based and hybrid deep learning approaches. After the introduction of hybrid and transformer models, following the success of early CNN-based models, very high accuracies were obtained. Nevertheless, issues such as scarcity of data, environmental variations, interpretability, and computational costs remain to be addressed. Therefore, in ensuring scalable and sustainable agricultural intelligence, future endeavours will have to look into three issues: explainable AI, standard open datasets, and lightweight

transformer architecture.

Methodology:

Research Design:

The proposed research has been conceptualized as a quantitative and experimental research design to build a model of deep learning that would automatically identify and classify diseases in olive leaves using a Transformer. The proposed framework introduces the Vision Transformer (ViT) architecture that extends the Transformer concept from natural language processing to computer vision. A step-by-step methodology developed in the proposed study includes data collection, image data preprocessing and augmentation, model training, and performance evaluation. The technical focus of the proposed model is the classification of diseases of olive leaves using intricate visual pattern analysis, including color discoloration, lesion textures, and shape deformations. Unlike many other CNN-based methods, which tend to focus on local features, the Transformer model exploits self-attention to describe the global dependencies between all image patches so that the resulting representations can understand a broader view of the symptoms of the disease. In establishing its reliability and generalizability, the performance of the model is verified empirically in terms of accuracy, precision, recall, and F1-score measures.

Dataset Description:

The olive leaf image dataset utilized in this research comprises 3,400 high-resolution images of olive leaves gathered in the spring and summer seasons. The dataset is relevant to Pakistan's growing olive cultivation sector, even though the data was collected outside of Pakistan. It can be used for image classification tasks that aim to automatically find olive leaf diseases using transformer-based deep learning models. The dataset contains three classes: Healthy leaves, Olive Peacock Spot-infected leaves caused by Spilocaea oleagina, and Aculus Olearius-affected leaves resulting from olive gall mite infestations. The images are organized into **train** and **test** directories, with class-specific folders to facilitate preprocessing and transformer-based deep learning model training. The dataset has a moderate class imbalance problem, since the Healthy class has more samples compared to the other disease classes. In order to reduce this imbalance, data augmentation was done while training with rotation, flipping, scaling, and color jittering. These augmentations increase sample diversity and contribute to the better generalization of deep learning models across all classes. Table 2 summarizes the distribution of images across the classes and dataset splits.

Table 1. Olive Leaf Dataset Distribution

Class Name	Train Set	Test Set	Total
Healthy	830	220	1,050
Olive Peacock Spot	1,200	260	1,460
Aculus Olearius	690	200	890
Total	2,720	680	**3,400**

Data Preprocessing and Augmentation:

Before training the Transformer model, all images were normalized as per standard 224×224 pixel specifications before their min-max normalization to the range (0,1). To achieve higher robustness in the model and less tendency towards overfitting, various augmentation techniques were utilized, which include: random rotation, random horizontal and vertical flipping, changing brightness and contrast, introducing Gaussian noise, and random cropping. This provided some real-world variation in the orientation of a leaf, its lighting conditions, and texture, under which these will ensure learning invariant representations for the model. Further, each image is divided into smaller patches, and all these were flattened and embedded into input tokens to the Transformer model. Each token was put through positional encoding to preserve the spatial relationships between patches owing to the lack of sequential awareness in the

attention mechanism. The above pipeline on preprocessing and augmentation ultimately prepares the present input data to be balanced and diverse, within the standards of the architectural specifications of Vision Transformer. Figure 2 presents the pre-processing stages conducted in the proposed work.

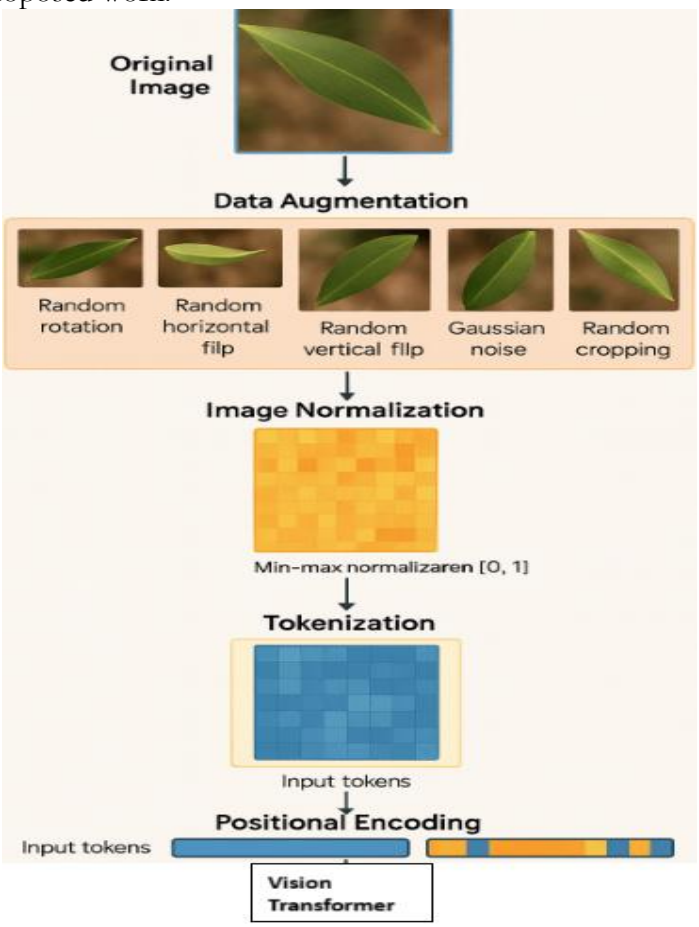

Figure 1. Preprocessing stage

Detailed Description of Transformer Model Layers:

The full architecture of the proposed framework has been illustrated in Figure 3. It consists of two main parts, namely, the encoder and decoder, which are connected with an attention mechanism. First of all, the model flattens each input image into smaller fixed-size patches using linear projection layers, after which these patches are projected into dense vector embedding spaces. To further augment these embedded representations with respect to the lack of positional-order recognition inherent to the transformer architecture, positional encodings are added in order to preserve spatial information. The multiple identical layers in the encoder section then allow attention on different locations of the image simultaneously, and the capturing of both local and global relations between patches. Pursuant to augmenting nonlinear representational capability, a Feed-Forward Network (FFN) and Add-and-Normalize functions follow after each attention layer. In turn, the output of the encoder is sent to several layers of the decoder. In this framework, the decoder is used to improve classification accuracy by refining and aggregating encoder-derived visual features into class-discriminative representations rather than generating sequences.

After the completion of masked multi-head attention in every layer of the decoder, a cross-attention mechanism enables the decoder to attend to contextual outputs from the encoder. Such attention regulates the information flow from encoder to decoder, thus blocking access to future tokens. These layers and the respective Add-and-Normalize operations convert

the encoded representations into useful class-level features. The output of the decoder is finally passed through softmax activation and a linear layer in order to generate probability distributions for each disease class. The encoder learns detailed visual patterns from the input images, and the decoder converts these learned representations to precise disease predictions, creating an effective end-to-end system for classifying diseases of olive leaves.

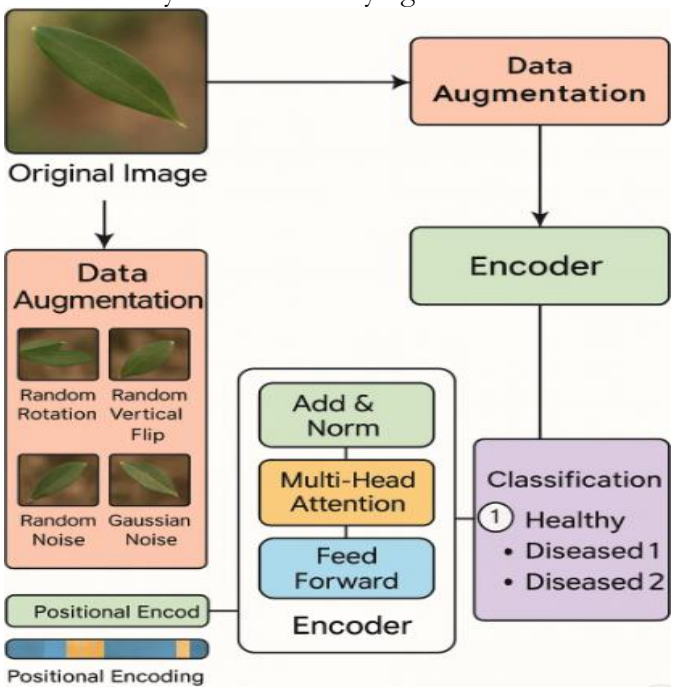

Figure 2. Overall Workflow of the Proposed Transformer-Based Olive Leaf Disease Detection Framework

The suggested model mainly uses an encoder-based Vision Transformer architecture for image classification, even though the transformer framework is described using encoder-decoder terminology for conceptual clarity. The decoder-style attention blocks are not used for autoregressive prediction or sequence generation, but only for improved contextual aggregation and feature refinement. An elaborate description of the internal flow and processing in the stacks of Transformers' encoders and decoders has been provided in Figure 4. It identifies layer-by-layer hierarchical advancement of the data through the architecture. The positional encodings are combined at the bottom of the model with the embedded image patches prior to passing through the encoder so as to maintain spatial coherence. Multi-head self-attention in each encoder layer is used to compute attention weights across the patches, enabling the model to relate features from different regions of the leaf image, such as edges, veins, or disease spots.

To avoid any non-linearity in feature representation, the output of the attention layer is passed through a feed-forward network, which increases and then decreases the dimension of the features. These processes, performed in several encoder layers, lead to the gradual extraction of more abstract and discriminative feature representations. A cross-attention bridge passes the encoder's final output to the decoder so that the decoder can utilize the global image context, which was recorded by the encoder. Inside the decoder, each layer masked self-attention and, in doing so, made any information leakage impossible, followed by a cross-attention layer that selectively attends to relevant encoder outputs corresponding to significant disease regions. The feed-forward sublayer of the decoder further refines these features, and the final output is projected linearly and activated via softmax to provide classification probabilities. To summarize, Figure 4 illustrates the ongoing interactive information flow within and between encoder and decoder processes. This allows the model to analyze the broad global structures and fine-grained local textures of olive leaves, providing superior recognition performance.

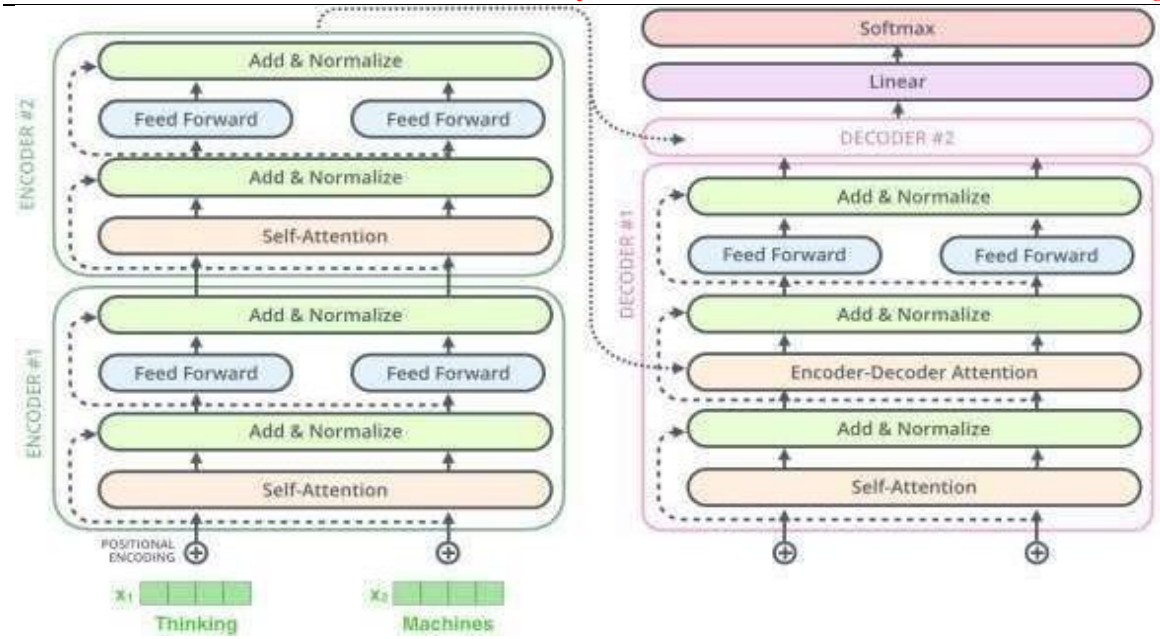

Figure 3. Detailed Transformer Encoder–Decoder Architecture Used in the Proposed Model.

Experimental:

Model Training and Evaluation:

Using PyTorch, the ViT model was trained on the preprocessed olive leaf dataset. The dataset was evenly divided to guarantee equivalency during performance evaluation into training, validation, and test subsets (70% will be used for training; 15% each for validation & test subsets). This will speed up the training process because the computations were done on a workstation with an NVIDIA RTX GPU (12 GB RAM). The categorical cross-entropy loss function will be utilized to minimize class prediction errors during optimization, and the Adam optimizer will be chosen with an initial learning rate of 3×10^{-5}. The model will have trained for 10 epochs with a batch size of 32 to reduce overfitting and increase generalization capabilities. Early stopping and learning-rate scheduling were used to support this training. Validation metrics were recorded at the end of each epoch to monitor convergence and model stability. During testing, the best model weights are saved based on validation accuracy. The performance of all the disease categories was evaluated on the test set by employing standard performance metrics, including accuracy, precision, recall, F1-score, and confusion matrix.

Experimental Setup and Tools

Python-based tools and libraries were used to implement the transformer model in a simulated environment, as presented in Table 3.

Table 2. Experimental Environment: Hardware and Software Specifications

Component	Details
Hardware Details	Device Name: DESKTOP-QI6H2EA Processor: Intel Core i5-6300U, 2.40–2.50 GHz RAM: 8 GB (7.88 GB usable) Storage: 466 GB HDD, 119 GB SSD Graphics Card: Intel HD Graphics 520, 128 MB System Type: 64-bit OS, x64-based processor
Software Details	Programming Language: Python Frameworks & Libraries: Tensor Flow Federated (TFF), Hugging Face Transformers, Pandas, Scikit-learn Cloud Platform: Google Cloud Platform (GCP) Compute Engine, Cloud Storage Operating System: Linux (Ubuntu)

The desktop system that will be used for the experiments includes DESKTOP-QI6H2EA, which has an Intel Core i5-6300U processor, 8 GB of RAM, 466 GB of HDD, 119 GB of SSD, and Intel HD Graphics 520. The software environment for Linux (Ubuntu) comprised Python with TensorFlow Federated, Hugging Face Transformers, Pandas, and Scikit-learn. Furthermore, Cloud and Compute Engine services helped to complete computationally intensive tasks on the Google Cloud Platform, making it much easier to train and assess Transformer-based models on the olive leaf dataset.

Model Interpretability

Interpretability techniques were applied to penetrate deeper into the internal mechanisms of the transformer**.** Attention visualization maps were generated in order to highlight certain areas of the image that were important for classification by the model. In these attention heat maps, it could be inferred that instead of focusing on irrelevant parts of the background, the vision transformer teaches important disease features on the leaves, such as dark lesions, color discolorations, and texture differences. The use of the t-distributed stochastic neighbor embedding technique, the feature embedding extracted from encoder layers of the transformer, was analyzed on the strictly unseen test dataset collected from olive growing regions of Pakistan, and was shown to present completely separated clusters between healthy and diseased classes. This interpretability analysis takes a step further in understanding how the inference is made between the predictions by providing biological as well as visual relevance to the model's prediction**.** Based on this analysis, the model becomes much more trustworthy for assimilation into real systems of agricultural disease monitoring.

Model Evaluation:

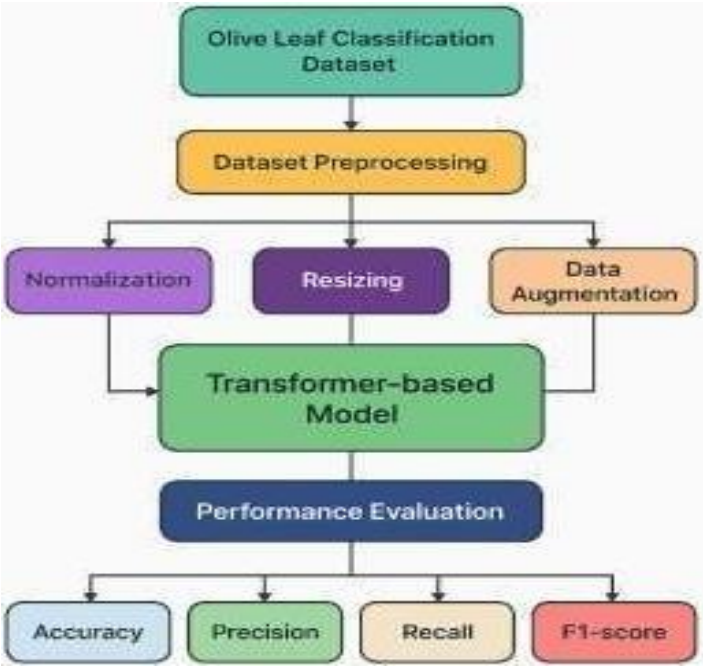

Figure 4. Performance Evaluation

The proposed study follows a systematic and replicable workflow presented in Figure 5, incorporating factors of experimentation for consistency and reliability. To eliminate duplicate and low-quality samples, the dataset was first cleaned to remove duplicates and low-quality images. To add diversity of data among samples and lessen overfitting, various preprocessing measures, including normalization, resizing, and data augmentation techniques such as rotation, flipping, and brightness adjustment, were then performed. Subsequently, this preprocessed dataset, collected from olive growing regions of Pakistan and split into training, validation, and test sets, with the test set strictly held out and never used during training or validation, was

utilized to train a Transformer-based model (Vision Transformer) concerning olive leaf disease classification. The learning performance of the model was closely monitored through training and validation sessions. Finally, the trained model was evaluated under standard accepted performance metrics, including F1-score, Accuracy, Precision, and Recall. This structured and methodologically sound workflow ensures transparency and reproducibility in future investigations into the automated detection of olive leaf disease.

Results and Discussion:

The proposed ViT-Small model, trained on the augmented olive leaf dataset using automatic mixed precision (AMP), demonstrated rapid convergence, with the training loss decreasing from 0.3684 to 0.0002 and reaching a training accuracy of 100% by the tenth epoch. On the independent test set comprising 593 images, the model achieved an overall accuracy of 97%. Class-wise performance was consistently high, with precision values of 0.98 for Healthy leaves, 0.96 for Aculus Olearius, and 0.97 for Olive Peacock Spot leaves. Corresponding recall values ranged from 0.96 to 0.98, while F1-scores varied between 0.96 and 0.97, indicating balanced performance across all classes (see Table 4).

Analysis of the confusion matrix (Figure 6) revealed very few misclassifications, predominantly between Healthy and Olive Peacock Spot leaves. Specifically, four Healthy samples were misclassified as Olive Peacock Spot, while one Olive Peacock Spot sample was predicted as Healthy. This limited overlap is likely due to early-stage disease symptoms exhibiting visual characteristics similar to healthy leaves, demonstrating the model's ability to identify subtle disease patterns despite minor appearance variations.

The model's strong discriminative ability was further validated by the ROC curves (Figure 7) and Precision Recall (PR) curves (Figure 8). The average precision values were ≥0.993, and all classes had an AUC of roughly 0.997. Classes with visually similar textures showed a slight decrease in precision at high recall, which is consistent with the small misclassifications. Confidence intervals for the ROC AUC and PR curves were not calculated, despite the fact that these metrics show excellent performance; reporting them in subsequent evaluations would provide statistical reliability and bolster the robustness of the model's performance claims.

The ViT-Small architecture successfully captures spatial and textural features important for olive leaf disease detection, according to the multi-metric evaluation and high accuracy. Although the perfect training accuracy suggests possible overfitting, which could be addressed with regularization strategies like dropout or weight decay, pre-applied data augmentation helped to improve generalization. ViT-Small uses patch-based attention mechanisms to enhance feature representation in comparison to traditional CNNs, especially for high-resolution leaf images. Overall, the suggested model's suitability for actual olive leaf disease detection and precision agriculture systems is supported by the thorough assessment of accuracy, precision, recall, F1-score, confusion matrix, ROC AUC, and PR curves.

Table 3. Proposed ViT-Small Model Classification Performance Metrics on the Test Dataset

Class	Precision	Recall	F1-Score
Healthy	0.98	0.96	0.97
Aculus Olearius	0.96	0.96	0.96
Olive Peacock Spot	0.97	0.98	0.97
Overall Accuracy	**0.97**	–	–

Comparison with CNN-Based Baseline Models:

Table 4. Comparison of CNN-Based Baseline Models with the Proposed Method

Ref.	Method	Model Type	Dataset	Classes	Accuracy (%)
[1]	Dynamic Clustering CNN	CNN	Olive leaf images	3	93.4
[2]	Deep CNN	CNN	Olive leaf dataset	3	94.8

[14]	Conventional CNN	CNN	Small olive dataset	3	91.2
[25]	EfficientNet (TL)	CNN	Olive leaf images	3	94.5
[30]	EfficientNet (TL)	CNN	Augmented olive dataset	3	95.0
[17]	Optimized CNN	CNN	Olive leaf dataset	3	94.2
Proposed	**ViT-Small**	**Transformer**	**3,400 olive images**	**3**	**97.0**

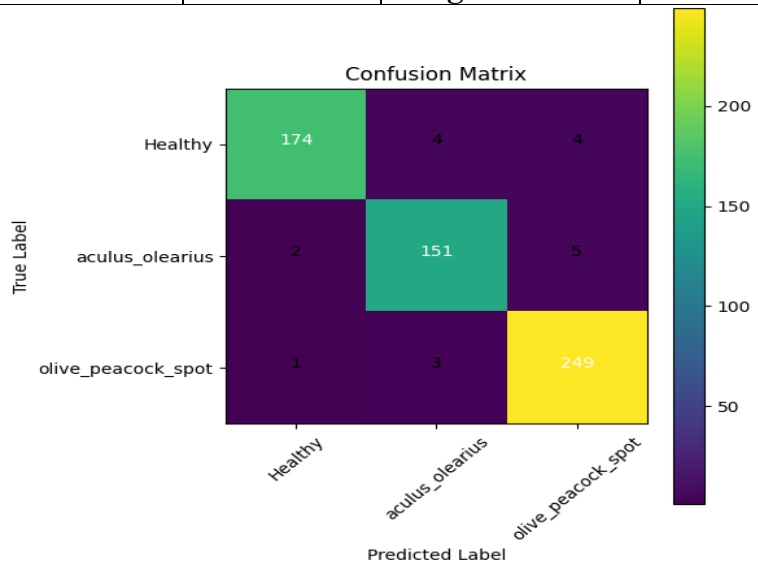

Figure 5. Confusion matrix of the ViT-Small model on the test set.

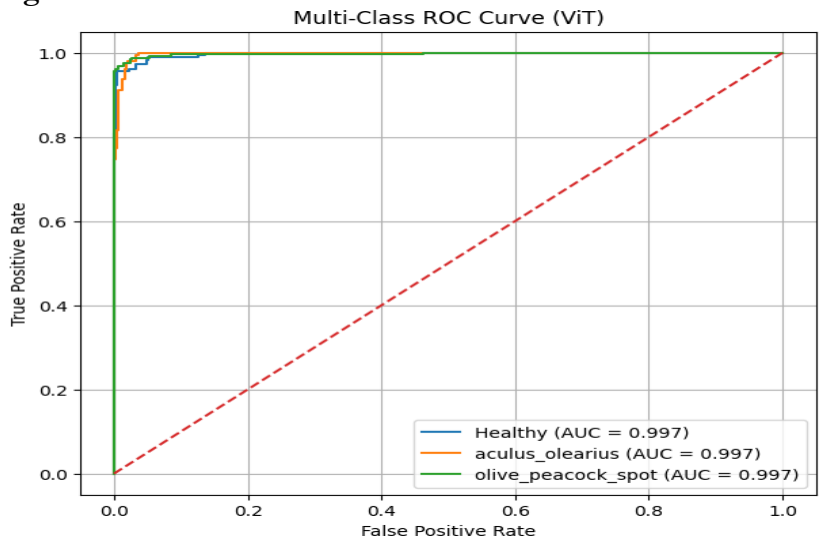

Figure 6. Multi-class ROC curves for the ViT-Small model.

CNN-based models achieve competitive performance through efficient local feature extraction [33], as Table 5 illustrates. However, inter-class similarity, background complexity, and illumination variation frequently have an impact on their performance. With a test accuracy of 97.0%, the suggested ViT-Small model outperforms these CNN baselines, demonstrating an enhanced capacity to capture subtle texture variations and global contextual dependencies crucial for the classification of olive leaf disease.

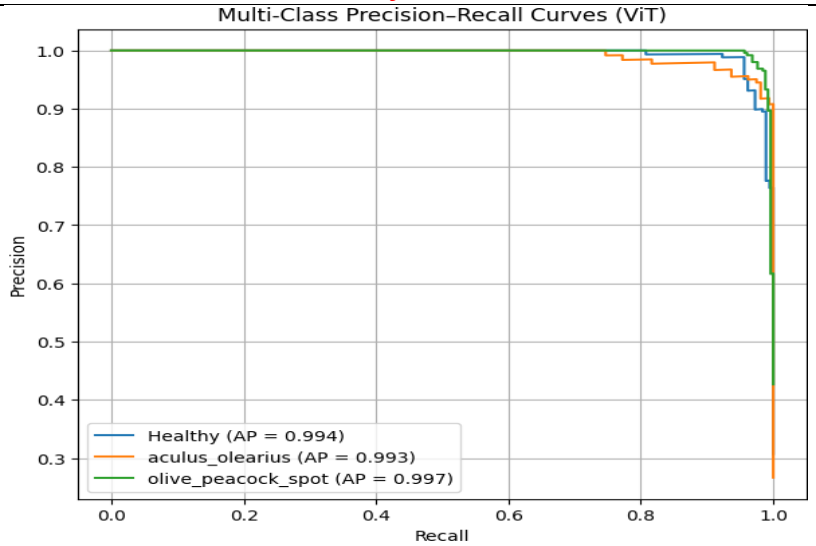

Figure 7. Multi-class Precision Recall curves for the ViT-Small model.

Conclusion:

An improved Vision Transformer (ViT-Small) model for automated olive leaf disease detection was presented in this study. Healthy, Aculus Olearius, and Olive Peacock Spot leaves were reliably classified by the model, which reached 97% test accuracy with high precision, recall, and F1-scores across all classes. The model's resilience and capacity for discrimination were further confirmed by confusion matrix analysis, ROC AUC, and precision-recall curves. These findings demonstrate how well ViT structures detect plant diseases and how they might be used in precision agriculture applications. The scope of the dataset, however, may not adequately represent the diversity of olive leaf diseases in Pakistan's various regions. As a result, additional validation on more varied datasets that reflect different environmental conditions and disease stages may be necessary for real-world field applicability. Future work will include cross-validation and confidence interval estimation to further enhance statistical reliability. Future research may concentrate on using explainable AI techniques, like attention map visualization, to increase generalization across various datasets and improve interpretability.

References:

[1] A. M. ; A.-J. A H ; Alsaeedi, "Dynamic Clustering Strategies Boosting Deep Learning in Olive Leaf Disease Diagnosis," *Sustainability*, vol. 15, no. 18, p. 13723, 20231, doi: https://doi.org/10.3390/su151813723.

[2] I. N. C.-C. Erbert F. Osco-Mamani, "Highly Accurate Deep Learning Model for Olive Leaf Disease Classification: A Study in Tacna-Per´u," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 4, 2023, [Online]. Available: https://thesai.org/Publications/ViewPaper?Volume=14&Issue=4&Code=IJACSA&SerialNo=94

[3] J. V. D. Singh, A. Kaur, "Machine vision for agricultural disease identification: a review," *Comput. Electron. Agric.*, vol. 210, pp. 107–120, 2023.

[4] S. Wang *et al.*, "Advances in Deep Learning Applications for Plant Disease and Pest Detection: A Review," *Remote Sens.*, vol. 17, no. 4, p. 698, Feb. 2025, doi: 10.3390/RS17040698/S1.

[5] J. L. João Mendes, "Impact of hyper-parameter tuning on CNN accuracy in agricultural image classification," *Smart Agric. Technol.*, vol. 11, p. 101016, 2025, doi: https://doi.org/10.1016/j.atech.2025.101016.

[6] A. A. Ibrahim Alrashdi, "An efficient deep-learning model for olive tree diseases diagnosis in Al-Jouf region," *Alexandria Eng. J.*, vol. 130, pp. 709–723, 2025, doi: https://doi.org/10.1016/j.aej.2025.09.059.

[7] M. D. & S. K. Ishak Pacal, Serhat Kilicarslan, Burhanettin Ozdemir, "Efficient and autonomous detection of olive leaf diseases using AI-enhanced MetaFormer," *Artif. Intell. Rev.*, vol. 58, no. 303, 2025, doi: https://doi.org/10.1007/s10462-025-11131-y.

[8] F. A. Attilio Di Nisio, "Fast Detection of Olive Trees Affected by Xylella Fastidiosa from UAVs Using Multispectral Imaging," *Sensors*, vol. 20, no. 17, p. 4915, 2020, doi: 10.3390/s20174915.

[9] H. Alshammari, K. Gasmi, I. Ben Ltaifa, M. Krichen, L. Ben Ammar, and M. A. Mahmood, "Olive Disease Classification Based on Vision Transformer and CNN Models," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3998193.

[10] Alaoui Fatima-Zahra, Zouheir Banou, Sanaa El Filali, Rachida Ait Abdelouahid, "Deep Learning for Plant Disease Detection: A Novel Convolutional Neural Network Approach Using the new dataset HealthySick Plants," *E3S Web Conf.*, 2025, doi: 10.1051/e3sconf/202568000005.

[11] F. A.-H. M. Al-Dhaheri, A. Al-Marzooqi, "Multimodal deep learning for olive disease classification using spectral and visual data," *Agriculture*, vol. 13, no. 7, pp. 1197–1212, 2023.

[12] A. A. M. Bashir, U. Mehmood, "Vision Transformer for olive leaf disease detection," *Comput. Mater. Contin.*, vol. 79, no. 2, pp. 2819–2835, 2024.

[13] S. K. H. Hussain, F. Raza, "CNN–Transformer hybrid model for high-accuracy olive leaf disease recognition," *Sensors*, vol. 24, no. 8, pp. 3285–3297, 2024.

[14] Z. A. T. Abbasi, M. Khan, "Olive leaf disease detection using convolutional neural networks," *Procedia Comput. Sci.*, vol. 193, pp. 252–260, 2021.

[15] F. U. A. Uysal, "Convolutional neural network-based feature extraction for plant disease detection," *Expert Syst. Appl.*, vol. 178, pp. 114–125, 2021.

[16] M. O. A. Mohamed, S. Ali, "MobiRes-Net: A hybrid deep learning architecture for olive leaf disease classification," *Neural Process. Lett.*, vol. 55, pp. 1543–1561, 20225.

[17] L. L. P. Paredes, R. Alves, "Olive tree disease detection using optimized convolutional models," *Sensors*, vol. 24, no. 5, pp. 2310–2321, 2024.

[18] S. T. N. Rana, M. Latif, "Sustainable deep learning approaches for olive disease prediction," *Sustainability*, vol. 16, no. 5, pp. 2752–2768, 2024.

[19] U. S. M. Qureshi, H. Ahmad, "Hybrid CNN–LSTM network for automatic olive leaf disease diagnosis," *Neural Comput. Appl.*, vol. 36, pp. 501–513, 2024.
[20] A. Y. R. Ameen, K. Shakir, "Performance analysis of machine learning algorithms for olive leaf disease classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 3, pp. 54–61, 2023.
[21] F. J. T. Hamid, S. Bukhari, "CNN–LSTM hybrid architecture for agricultural disease detection under temporal variability," *Comput. Electron. Agric.*, vol. 209, no. 107939, 2023.
[22] F. C. A. Rahman, K. Uddin, "Comparative evaluation of machine learning techniques for olive leaf disease prediction," *Comput. Intell. Neurosci.*, vol. 2024, no. 991245, 2024.
[23] F. G. J. Torres, R. García, "AI-based early detection of olive diseases for precision agriculture," *Comput. Electron. Agric.*, vol. 210, pp. 107–115, 2024.
[24] N. W. A. Amin, M. Khalid, "Explainable AI framework for plant disease diagnosis using CNN and Grad-CAM," *IEEE Access*, vol. 11, pp. 45902–45915, 2023.
[25] P. C. K. Patel, R. Singh, "Efficient Net based olive disease classification with transfer learning," *IEEE Access*, vol. 12, pp. 90201–90214, 2024.
[26] M. J. A. Dali, R. Mourad, "AI-based imaging for nutrient and disease diagnosis in olive leaves," *Comput. Electron. Agric.*, vol. 209, pp. 107933–107948, 2023.
[27] I. B. M. Ahmad, F. Ullah, A. R. A. Hamza, M. Usman, M. Imran, "Cotton Leaf Disease Detection Using Vision Transformers: A Deep Learning Approach," *Crops*, vol. 1, no. 3, 2024.
[28] A. Sundaraj, D. P. Isravel, and J. P. M. Dhas, "Diagnosis of Plant Leaf Disease using Vision Transformer," *Proc. 2024 10th Int. Conf. Commun. Signal Process. ICCSP 2024*, pp. 82–87, 2024, doi: 10.1109/ICCSP60870.2024.10543969.
[29] H. A. S. Najeeb, U. Rafiq, "Comprehensive analysis of CNN architectures for olive leaf disease detection," *Appl. Sci.*, vol. 15, no. 2, pp. 542–557, 2025.
[30] S. A. M. Iqbal, H. Rehman, "Transfer learning-based olive leaf disease classification using EfficientNet," *IEEE Access*, vol. 12, pp. 78140–78153, 20241.
[31] K. L. Y. Li, J. Zhang, "Frequency-domain CNN for robust plant disease classification," *Comput. Electron. Agric.*, vol. 224, p. 109298, 2024.
[32] Z. N. A. Javed, U. Khan, "Lightweight CNNs for mobile-based olive leaf disease detection," *Neural Process. Lett.*, vol. 57, pp. 1189–1202, 2024.
[33] E. A. T. Eleyan, M. Ozturk, "Machine learning approaches for olive disease classification: Comparative analysis," *Procedia Comput. Sci.*, vol. 180, pp. 124–132, 2021.