**Development of a mobile application for the detection and classification of tomato plant diseases based on Swin Transformer**

Oscar David Valencia - 20192182074

Maria José Capera Firigua – 20201188262

Facultad de Ingeniería, Universidad Surcolombiana

Electiva Ciencias de la computación II - Visión Artificial

Ferley Medina Rojas

Agosto de 2023

**TABLE OF CONTENTS**

[**INTRODUCTION 3**](#_tycrd2ytbpn5)

[**PROBLEM 4**](#_76696e9e9q2l)

[**OBJECTIVES 5**](#_5wmsknvme73h)

[General objective 5](#_f6dlx77kymqv)

[Specific objectives 5](#_d6r42owmwevb)

[**LITERATURE REVIEW 6**](#_r2lsqc1cu1it)

[Machine Learning Approach towards Tomato Leaf Disease Classification. International Journal of Advanced Trends in Computer Science and Engineering, 9(1). 6](#_xa2sjoqh2ody)

[Tomato plant disease detection using transfer learning with C-GAN synthetic images 7](#_nhdykxx5j871)

[Tomato leaf disease classification using supervised learning techniques contrasting. 9](#_93hzfrq8tjbd)

[Artificial intelligence in tomato leaf disease detection 11](#_1pm0ksx7pk1k)

[Computer-aided fusion-based neural network in application to categorize tomato plants. Springer-Verlag London Ltd., part of Springer Nature 2023 14](#_obxgds67bnog)

[**THEORETICAL FRAMEWORK 17**](#_osxtvooujh3s)

[**REFERENCES 22**](#_3irfze7s0g8d)

# INTRODUCTION

Early and accurate detection of tomato diseases has become an imperative to mitigate their devastating impact. Although advances in computer vision and machine learning have provided potential solutions, these methods have been insufficient in terms of speed, cost, and accuracy. Promising Deep Learning techniques, particularly convolutional neural networks (CNNs), have not re-proposed new architectures or even a paradigm shift, calling for innovative rethinking.

This study aims to explore and validate an innovative approach by applying the "Swin Transformer", a novel architecture in the computer vision domain. Through a systematic analysis and rigorous methodology, it will seek to develop a model that overcomes existing limitations in tomato disease detection and classification.

It will focus on answering critical questions related to the development and validation of this new model, as well as its comparison with CNN-based methods. The research will address essential aspects such as data augmentation, completeness of results, optimal configuration and training of the Swin Transformer architecture, and visual interpretation of model decisions.

The proposed research has significant potential to transform current practice in tomato disease detection and classification by providing a more accurate and efficient model.

The paper is structured in introduction, problem, objectives, state of the art, theoretical framework, methodology, materials and methods, results, conclusions and future work.

# PROBLEM

Tomato cultivation is essential for food and the global agricultural economy because it is one of the most widely produced vegetables worldwide. According to FAO (2021), world production exceeded 189 million tons, covering more than 5 million hectares under cultivation. Nutritionally, it is a rich source of phytochemicals, nutrients, antioxidants and phenolic compounds (Collins et al., 2022). Its consumption has been found to provide anti-cancer benefits and protection against cardiovascular and neurodegenerative diseases (Collins et al., 2022).

However, tomato is vulnerable to a wide variety of diseases caused by fungi, bacteria, phytoplasmas, viruses and viroids (Panno et al., 2021) Table 1. This vulnerability is increased by factors such as its limited genetic diversity resulting from intensive selection, the tendency to monoculture, international exchange of infected material, and climate change (Panno et al., 2021). These diseases represent economic losses exceeding US$30 billion per year (Caruso et al., 2022). Such losses impact agricultural economics, public health, and environmental sustainability by affecting crop yield and quality, destabilizing ecosystems, and raising production costs due to the intensive use of pesticides and other control measures (Panno et al., 2021).

Early and accurate detection of diseases in tomato is essential to mitigate their impact, but despite significant advances in computer vision over the last decade, there have been limitations. In machine learning, methods often rely on manual feature extraction and heuristic algorithms, which can be slow, laborious, costly and error prone (Zahangir Alom et al., 2018). These approaches have been overtaken by Deep Learning techniques such as convolutional neural networks (CNNs) (Rawat et al. 2022; Thangaraj et al. 2022). However, the latter seem to have reached a stagnation point following the development of architectures such as ResNet, DenseNet and EfficientNet (Li, 2020).

Therefore, the present research suggests an innovative approach based on the "Swin Transformer" architecture, which has proven to be highly effective in computer vision tasks but has never been applied to tomato disease classification. The objective is to develop and validate a more accurate model than previous solutions using the PlantVillage dataset, which contains nine tomato disease classes and one healthy plant class. Subsequently, its performance will be compared with previous CNN-based studies.

In this context, the following questions are raised:

* How to develop and validate a tomato disease detection and classification model using Swin Transformer architecture that surpasses the accuracy of convolutional neural networks (CNN)?
* How to build and prepare a balanced dataset across the different tomato disease classes to ensure the integrity of the model predictions?
* How to optimally configure and train the Swin Transformer architecture?
* How to evaluate and interpret the Swin Transformer model predictions for each class?
* How does the accuracy of the Swin Transformer model compare to models based on convolutional neural networks?

# OBJECTIVES

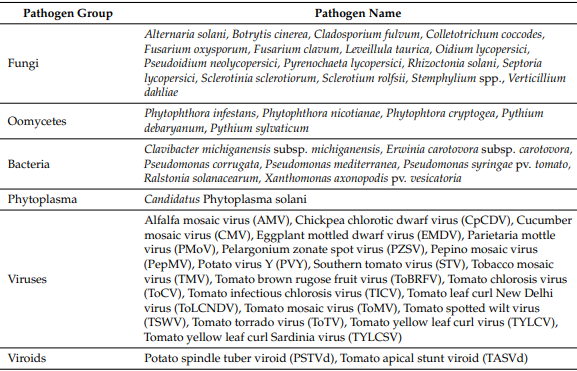
## General Objective

To develop and validate a tomato plant disease detection and classification model using the Swin Transformer architecture, with the purpose of achieving superior accuracy compared to convolutional neural networks.

## Specific Objectives

1. Build and prepare a balanced dataset on the different tomato disease classes to ensure the integrity of the model predictions.
2. Optimally configure and train the Swin Transformer architecture.
3. Evaluate and interpret the predictions of the Swin Transformer model for each class.
4. Compare and analyze the accuracy of the Swin Transformer model with models based on convolutional neural networks.

**Table 1.** List of tomato plant pathogens (Panno et al., 2021).



**LITERATURE REVIEW**

**2020:**

## Title: Machine Learning Approach towards Tomato Leaf Disease Classification. International Journal of Advanced Trends in Computer Science and Engineering, 9(1).

**Authors:** Gadade, H. D., & Kirange, D. K.

**Methodology:**

The methodology consists of several steps, each aimed at enhancing the detection and classification of tomato leaf diseases.

1. **Data Collection:** Utilized 9,000 images of tomato leaves from the PlantVillage Dataset, focusing on seven types of diseased images.
2. **Preprocessing:** Tomato leaf disease images were processed to remove noise using a Median Filter.
3. **Feature Extraction:** Used GLCM (Gray-Level Co-Occurrence Matrix), Gabor, and SURF (Speeded Up Robust Features) for feature extraction.
4. **Classification:** Deployed various classifiers, including Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Naïve Bayes (NB), and Decision Trees, to classify the tomato leaf images into normal or diseased.
5. **Performance Evaluation:** Used accuracy, precision, recall, and F-measure to evaluate the performance of different classifiers.

**Results:**

The results are presented in terms of accuracy, recall, F-measure and run time for different extraction techniques and classifiers. The main findings are:

The following tables summarize the performance measures with different classifiers:

* **Decision Trees:**
  + GLCM: Accuracy 0.6497, Precision 0.1715, Recall 0.722, F-Measure 0.2772
  + Gabor: Accuracy 0.6759, Precision 0.2047, Recall 0.861, F-Measure 0.3307
* **SVM:**
  + Gabor: Accuracy 0.7339, Precision 0.2525, Recall 0.9492, F-Measure 0.3989
* **KNN:**
  + Gabor: Accuracy 0.732, Precision 0.2555, Recall 0.9831, F-Measure 0.4056
* **Naïve Bayes:**
  + Gabor: Accuracy 0.675, Precision 0.2187, Recall 0.9695, F-Measure 0.3568

1. **Gabor features** were found to be effective with all classifiers for tomato leaf disease classification.
2. **SVM** showed better accuracy compared to other classifiers but required more execution time.
3. **KNN** with Gabor features offers better accuracy with less execution time compared to SVM. [9]

**Conclusion:**

The paper concludes that a KNN classification framework with Gabor features is effective for tomato leaf disease classification. The method is deemed to be efficient and accurate enough to be employed in real-time applications. The authors also suggest that more sophisticated techniques like Adaptive neuro fuzzy, Neural Networks, and Genetic algorithms could be used for plant image recognition and classification.

**2021:**

## Title: Tomato plant disease detection using transfer learning with C-GAN synthetic images

**Authors:** Amreen Abbas, Sweta Jain, Mahesh Gour

**Methodology**

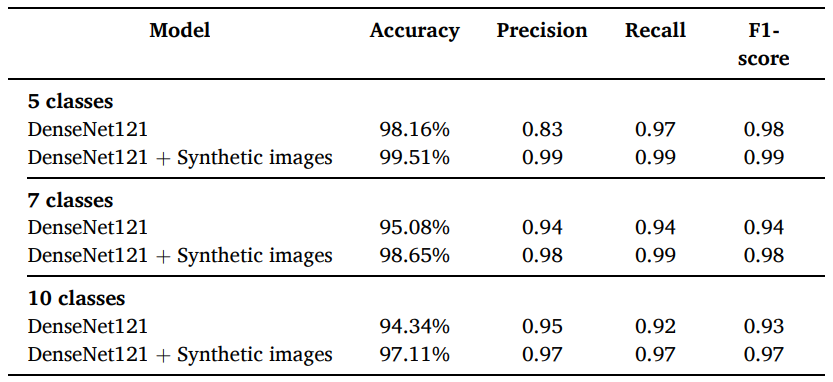
A method based on deep learning has been presented to detect diseases in tomato plants. This technique makes use of a Conditional Adversarial Generative Network (C-GAN) to generate synthetic visual representations of tomato leaves. Subsequently, training of a model called DenseNet121 is carried out, which is fed with both real and synthetic images. This training process is performed through knowledge transfer, allowing the model to accurately classify tomato leaf images into up to ten different disease categories.

The proposed model has undergone an exhaustive training and evaluation phase using the PlantVillage public dataset, which focuses on plant-related issues. The results obtained are highly promising: an accuracy of 99.51% in classifying images in five disease categories, 98.65% in seven categories and 97.11% in ten categories has been achieved. These results highlight the effectiveness of the approach in accurately detecting tomato leaf diseases, which could potentially have a significant impact on improving the health and yield of agricultural crops.

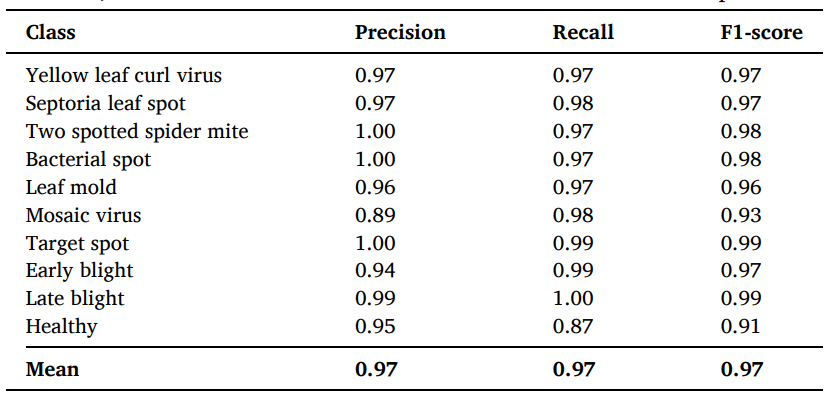
**Results**

Table represents the classification performance of the proposed method on the PlantVillage dataset and augmented dataset (PlantVillage + Synthetic images dataset). The proposed method achieved a classification accuracy of 99.51%, 98.65%, and 97.11% for 5-class

**Table 1. Accuracy, Precision, Recall, and F1-score for the model with and without augmentation. [10]**

****

**Table 2. Precision, Recall and F1-score for different disease classes of tomato plant. [10]**

****

**Conclusion**

The proposed data augmentation technique effectively enhances the neural network's generalization capacity, thereby mitigating the risk of overfitting. The achieved results with the proposed model are highly encouraging, achieving precision rates of 98.16%, 95.08%, and 94.34% on the original PlantVillage dataset for 5, 7, and 10-class classification tasks, respectively. Moreover, when utilizing the expanded dataset that includes synthetic PlantVillage+ images, precision rates increase to 99.51%, 98.65%, and 97.11% for the same classification tasks.

The experimental outcomes unequivocally underscore the superiority of the proposed method compared to existing approaches. In future work, the intention is to extend this methodology to encompass disease identification and classification across various parts of the plant, including fruits, stems, and branches. Additionally, there is a plan to undertake the identification of distinct stages of disease development in plants.

**2022**

## Title: Tomato leaf disease classification using supervised learning techniques: contrasting analysis. International Conference on Advances in Computing, Communication and Materials (ICACCM).

**Authors:** Vandana Rawat, Neelam Singh, Bhavleen Kaur, Saksham Bora

**Methodology**

The methodology of the study consisted of the following components:

1. **Dataset**: They utilized all the images of tomatoes from the "Plant Village" collection, which contained 10 categories, among them pictures of healthy tomatoes.
2. **Preprocessing**: Obtaining and preprocessing images of tomato leaf disease, including noise removal.
3. **Machine Learning Techniques**: Applied various algorithms including SVM, K-Nearest Neighbor, Naïve Bayes, Decision Tree, Feed Forward Neural Network, Back Propagation Neural Network, Deep Neural Network, Conventional Neural Network, and Multi Kernel SVM.
4. **Evaluation Metrics**: Utilized Accuracy, Precision, F1-Score, and Recall for evaluating the performance of the learning algorithms.

**Results**

**Table 2. Comparison of ML Algorithms [11]**

|  |  |  |  |
| --- | --- | --- | --- |
| **Author and Year** | **Dataset** | **Algorithm** | **Accuracy (%)** |
| H. D. Gadade [2020] | PlantVillage dataset | Decision Tree | 67,00 |
| Jayanthi M.G [2020] | PlantVillage dataset | SVM | 90,00 |
| Nishant Vijay [2021] | PlantVillage dataset | KNN | 83.6 |
| Dr. D.K. Kirange [2020] | PlantVillage dataset | Naive Bayes | 67,00 |
| Dr. Sreelatha P [2021] | PlantVillage dataset | FFNN | 56.89 |
| Sridhar Udaya kumar [2021] | PlantVillage dataset | BPNN | 57.19 |
| K. Ch. Sri Kavya [2021] | PlantVillage dataset | Linear Kernel SVM | 59.28 |
| S. Karthick [2021] | PlantVillage dataset | RNN | 59.54 |
| Nishant Vijay [2021] | PlantVillage dataset | CNN | 98.5 |
| Shashikumar D.R. [2020] | PlantVillage dataset | MKSVM | 97.34 |

The results were presented as a comparative analysis of various machine learning techniques applied to the hydroponics dataset. The Convolution Neural Network (CNN) achieved the highest accuracy of 98.5%. Other algorithms' accuracies were also presented, highlighting the effectiveness of different methods in the tomato leaf disease classification. [11]

**Conclusión**

The study concluded that machine learning technologies are frequently utilized for tomato leaf disease prediction and prognosis. Among the various machine learning techniques compared, the Convolution Neural Network (CNN) method was identified as the best for detecting tomato leaf disease early on, with an accuracy of 98.5%. The study also emphasized the potential of hydroponics and the application of data-driven techniques in modern agriculture.

## Title: Artificial intelligence in tomato leaf disease detection: a comprehensive review and discussion

**Authors:**  Rajasekaran Thangaraj, S. Anandamurugan

**Methodology**

In the study, the researchers primarily searched for articles on tomato leaf disease from major electronic databases like IEEE Xplore, ScienceDirect, Google Scholar, and ACM library. These sources were selected because they contain a significant number of studies on tomato leaf disease detection using image processing, including machine learning (ML) and deep learning (DL) techniques.

The focus has been on publications from 2015 onward due to the growing interest in tomato plant disease detection using AI techniques in recent years. The search was initiated using specific keywords related to tomato leaf disease and AI techniques.

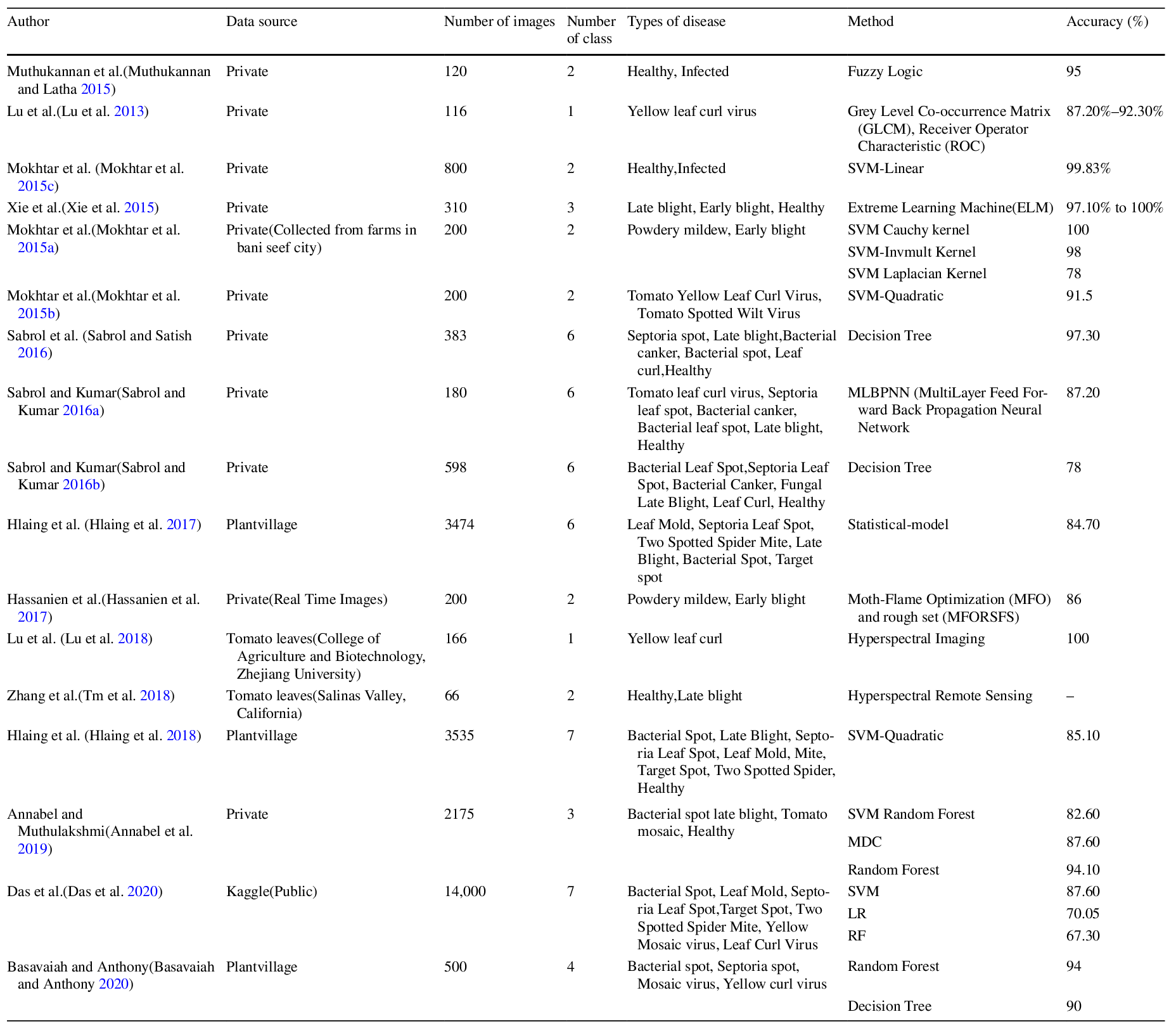
The research process is depicted in Fig. 1. The initial step involved downloading papers relevant to tomato leaf disease detection using AI. After reading, they were categorized based on the techniques used: traditional image processing, ML, or DL. From the search, 79 papers were identified, which was later refined down to 44 after thorough optimization and analysis.

Each article was reviewed in detail, considering:

* The specific plant diseases discussed.
* The AI model used.
* The dataset employed.
* The performance of the ML and DL techniques in the research.

**Results**

**Table 3. Comparison table showing the performance of machine learning algorithm applied to the detection of tomato plant disease through leaf images [12]**

****

**Table 4. Comparison table showing performance of deep learning algorithm applied to the detection of tomato plant disease through leaf images [12]**

****

**Conclusion:**

The agricultural domain has grappled with various challenges in recent times. This study offers a comprehensive review of the latest research on tomato leaf disease identification using artificial intelligence. We examined 44 related studies, focusing on datasets, pre-processing methods, models, and prediction accuracy. Our analysis prioritized data sources, top accuracy rates, and methodologies. The findings indicate that deep learning models surpass traditional methods like image processing and neural networks in identifying tomato diseases from leaf images.

Early detection of these diseases can minimize costs by avoiding unnecessary pesticide use. Emerging techniques like combining deep learning with hyperspectral imaging are particularly promising for early disease detection. As tomato diseases intensify over time, specialized deep learning models can be utilized to track and categorize these diseases throughout their lifecycle. To enhance prediction speed and accuracy, the integration of features from CNN models is beneficial. Looking ahead, the integration of agricultural robots and drones to automatically capture and classify diseased plant images is a promising avenue.

**2023:**

## Title: Computer-aided fusion-based neural network in application to categorize tomato plants. Springer-Verlag London Ltd., part of Springer Nature 2023

**Authors:** Rajyalakshmi Uppada · D. V. A. N. Ravi Kumar

**Methodology**

The methodology employed can be broken down into the following key steps:

1. **Image Pre-processing**: Utilizing Non-Subsampled Contourlet to acquire energy-detail components from the image dataset (Kaggle PlantVillage and Mendeley dataset).
2. **Cluster Extraction**: Modified K-means processing is used to extract colored clusters for tomato-leaf detection.
3. **Classification using SVM**: 38 extracted features are utilized by the Multiclass SVM (M-SVM) classifier for accurate categorization.
4. **Proposed M-CNN model**: A custom deep learning architecture is established for optimized unsupervised categorization. The model is based on the DenseNet-201 architecture and involves pre-processing stages followed by CNN categorization.

**Results**

**Table 5** Distinguished performance of NSC + K-Means + M-CNN classifier on specific tomato-leaves [13]

|  |  |  |  |
| --- | --- | --- | --- |
| **Category of tomato-leaf** | **Number of images** | **Average accuracy** | **M-CNN categorization** |
| Healthy Tomato-Leaf (D0) | 1528 | 99.32 | Healthy |
| Early-Blight (D1) | 1600 | 98.14 | Infected |
| Septoria-Leaf-Spot (D2) | 1701 | 99.17 | Infected |
| Leaf-Mold (D3) | 1523 | 99.39 | Infected |
| Bacterial-Spot (D4) | 1702 | 99.18 | Infected |
| Spider-Mites (D5) | 1610 | 99.32 | Infected |
| Late-Blight (D6) | 1527 | 99.27 | Infected |
| Target-Spot (D7) | 2247 | 98.53 | Infected |
| Mosaic-Virus (D8) | 1555 | 99.7 | Infected |
| Yellow-Leaf-Curl-Virus (D9) | 4286 | 99.55 | Infected |
| Powdery\_Mildew (D10) | 1004 | 99.09 | Infected |

**Table 6** Normalized confusion matrix for M-CNN classifier [13]

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Actual class | Predicted class | | | | | | | | | | |
|  | **D0** | **D1** | **D2** | **D3** | **D4** | **D5** | **D6** | **D7** | **D8** | **D9** | **D10** |
| **D0** | **98.98** | 0 | 0 | 0 | 0.13 | 0 | 0.13 | 0.76 | 0 | 0 | 0.13 |
| **D1** | 0.13 | **95.80** | 0.64 | 0.51 | 0 | 0.25 | 0.25 | 1.65 | 0.76 | 0 | 0.13 |
| **D2** | 0.64 | 0.91 | **94.48** | 0.82 | 0.78 | 0.67 | 0.64 | 0.80 | 0.51 | 0.76 | 0 |
| **D3** | 0.38 | 0.13 | 1.27 | **96.18** | 0.64 | 0.13 | 0.25 | 0.76 | 0.13 | 0.13 | 0.89 |
| **D4** | 0.25 | 0.25 | 1.27 | 0.25 | **96.82** | 0.13 | 0.64 | 0.25 | 0 | 0.13 | 0 |
| **D5** | 0.51 | 0.78 | 0.27 | 0.64 | 0.91 | **98.17** | 0.89 | 0.56 | 0.38 | 0.89 | 0.27 |
| **D6** | 0.16 | 0.13 | 0 | 0.13 | 0 | 0 | **99.95** | 0 | 0.13 | 0 | 0.27 |
| **D7** | 3.56 | 0.25 | 0.89 | 0.25 | 0.51 | 1.27 | 2.54 | **90.33** | 0.13 | 0.25 | 1.27 |
| **D8** | 0.13 | 1.40 | 0 | 0 | 0 | 0 | 0.51 | 0 | **97.96** | 0 | 0 |
| **D9** | 0 | 0 | 0.13 | 0 | 0 | 0 | 0.13 | 0.13 | 0.13 | **99.49** | 0.64 |
| **D10** | 0.64 | 0.91 | 0.82 | 0.78 | 0.67 | 0.64 | 0.80 | 0.51 | 0.76 | 0 | **94.48** |

These are the results of the proposed method:

* Achieved a model accuracy of 99.15% and an average precision of 95.6%.
* The combination of NSC, K-means, and M-CNN classifiers produced average sensitivity and specificity of 98.78% and 97.34% respectively.
* A detailed comparison with state-of-the-art approaches showed the model's superiority in terms of accuracy, precision, and handling more disease classes.

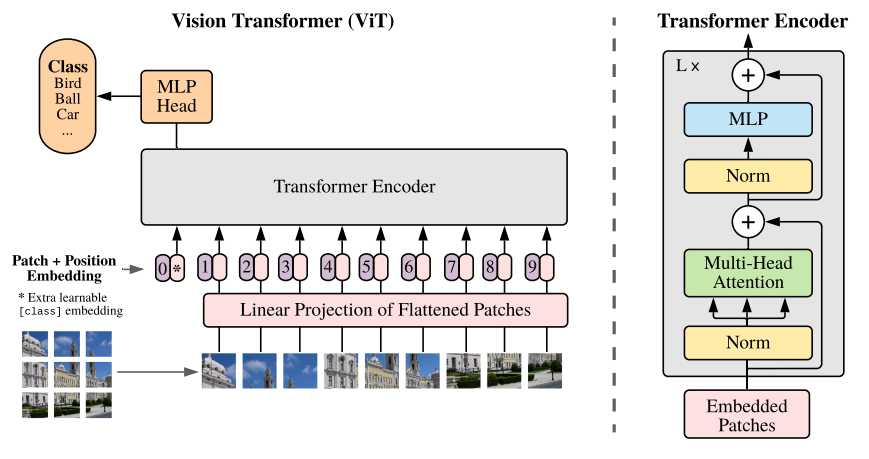
**Conclusion**

The study successfully introduces an efficient and accurate approach to detect and categorize diseased and healthy tomato plants. The proposed method, employing image pre-processing, clustering, SVM, and a customized CNN model, significantly outperformed traditional methods and other state-of-the-art approaches.

# THEORETICAL FRAMEWORK

**Introduction to the Vision Transformers (ViTs)**

Before getting into the Swin Transformer, it is essential to understand the concept of Vision Transformers (ViTs). These models adapt the transformer architecture, originally designed for natural language processing, for computer vision tasks. The key idea is to divide an image into patches, flatten them and feed them to a series of transformer layers, allowing the model to capture long-range relationships between different parts of the image [14].



**Figure 1:** Model overview of the first visual transformer [14]

**Limitations of Traditional ViTs**

Although ViTs have proven to be effective, they have certain limitations. Primarily, they require enormous amounts of data and computational power to train from scratch. In addition, the global attention used in traditional ViTs may not be optimal in terms of computational efficiency, especially for high-resolution images [14][15].

**General Architecture of the Swin Transformer: A Paradigm Shift**

The Swin Transformer, which stands for "Shifted Window Transformer," addresses the limitations of ViTs by introducing two key concepts: sliding windows and pyramid structure.

The Swin Transformer, like other Vision Transformers (ViTs), starts by dividing an image into non-overlapping patches using a splitting module. Each patch is considered a "token" and its feature is set by concatenating the RGB values of the raw pixels. In the Swin Transformer implementation, a patch size of 4x4 is used, resulting in a feature dimension of 48 (4 × 4 × 3). These features are subsequently projected to an arbitrary dimension (denoted as *C*) by means of a linear embedding layer.

The term "Stage" refers to the different stages of image processing through the architecture. In the Swin Transformer, the image resolution is successively reduced, from "Stage 1" to "Stage 4". This hierarchical structure is similar to traditional convolutional neural networks such as VGG and ResNet, where features are processed at different resolutions.

**Some detailed explanations:**

**MSA (Multi-Head Self-Attention):** a type of attention mechanism that allows the model to pay attention to different parts of the input simultaneously. It is essential to the transformer's ability to capture long-range relationships.

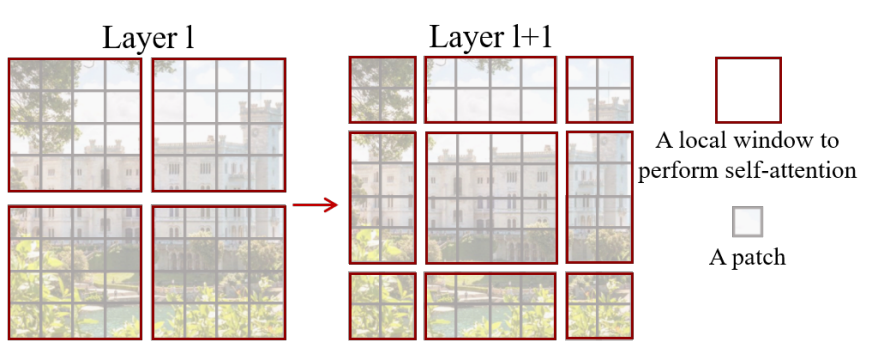
**MLP (Multi-Layer Perceptron):** Refers to a fully connected neural network. In the context of transformers, MLPs are used to transform features after layers of attention.

**GELU (Gaussian Error Linear Unit):** An activation function used in neural networks. It has been shown to have beneficial properties in transformer models.

**LN (LayerNorm or Layer Normalization):** It is a normalization technique used to stabilize and accelerate the training of neural networks.

**Ω(MSA) and Ω(W-MSA):** These are notations to represent the computational complexity of global attention and window-based attention, respectively.

**- Sliding Windows:** Instead of performing global attention over the entire image (which would have a quadratic complexity with respect to the number of tokens), the Swin Transformer introduces the concept of attention within non-overlapping local windows. This significantly reduces the computational complexity [16].



**Figura 2.** The shifted window approach to computing self-attention in the proposed Swin Transformer architecture.

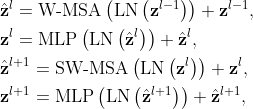
Equations (1) and (2) show the difference in complexity between global care and window-based care. Window-based attention is scalable and computationally more efficient [16].

(1) 

(2)

However, in order not to lose the ability to model relationships between patches in different windows, the Swin Transformer introduces the idea of windows shifted in successive blocks. This strategy alternates between two partitioning configurations in consecutive blocks of transformers, allowing each patch to relate to patches in neighboring windows.

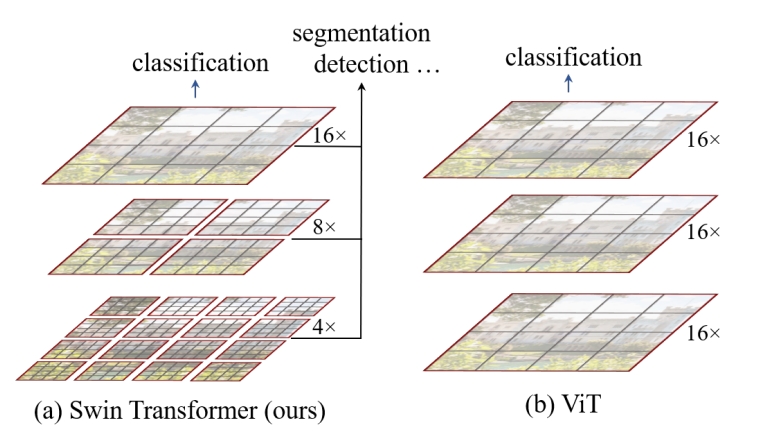
Equations (3) show how features are computed in successive blocks using different window-based attention configurations [16].

(3) 

In this context,  and  represent the resulting characteristics from the (S)W-MSA module and the MLP module corresponding to block . W-MSA and SW-MSA refer to window-based multi-head self-attention utilizing standard and shifted window partitioning setups, respectively.

Relative position bias is introduced to account for spatial relationships between patches within a window. This bias significantly improves model performance and is preferable to absolute position embeddings.

**- Pyramid Structure:** The Swin Transformer processes images in a series of resolutions, similar to a pyramid. It starts with small windows at a higher resolution and gradually clusters patches while reducing the resolution, allowing the model to capture features at different scales [16].



**Figure 3.** Comparison of the pyramidal structure method used by Swin tranformer and the method used by Vit.

**Architectural Details**

**- Tokenization and Embeddings:** Like other ViTs, the Swin Transformer starts by dividing the image into patches, which are then flattened and converted into embeddings through a linear layer [16].

**- Attention within Windows:** The attention layer in Swin Transformer operates only within sliding windows, using standard attention mechanisms. After each attention layer, the windows are scrolled to ensure coverage of the entire image [16].

**-Swin Transformer Block:** A Swin Transformer block is a fundamental unit in the architecture. It is where the shifted window-based attention occurs, which is essential for the efficiency and effectiveness of the model.

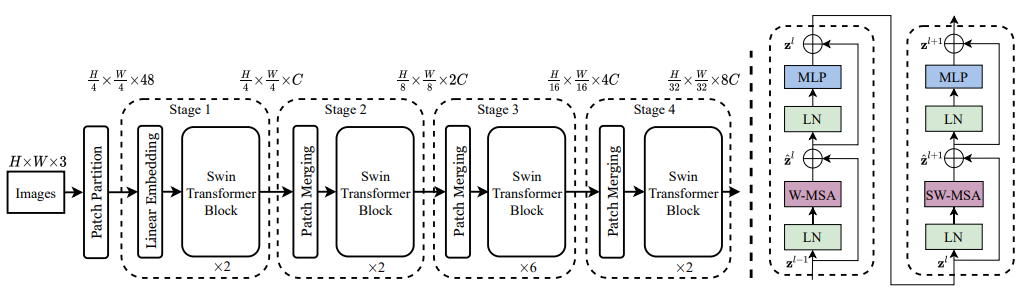
**-GELU:** GELU, or Gaussian Error Linear Unit, is a nonlinear activation function. It is used in neural networks to introduce nonlinearities into the model. In the Swin Transformer block, the GELU is used as an activation function in the 2-layer multilayer perceptron (MLP).

**-LN (LayerNorm):** LN, or Layer Normalization, is a normalization technique that is applied to features in a specific layer, rather than to a specific mini-batch. In Swin Transformer, LayerNorm is applied before each auto-tuned multi-head attention module (MSA) and each MLP. Normalization helps stabilize and accelerate training.

**-Residual Connection:** In deep networks, residual connections help avoid the gradient fading problem by allowing activations to skip one or more layers. In the Swin Transformer, a residual connection is applied after each attention module and MLP.

**- Resolution Reduction and Upscaling:** After processing the image at one resolution, the Swin Transformer reduces the resolution by grouping tokens and processing them in larger windows. This process is repeated several times, forming the pyramidal structure of the model [16].

**- Patch Merging:** "Patch Merging" is a technique used in the Swin Transformer to reduce the spatial resolution of the feature map and simultaneously increase the feature dimension, allowing the model to capture more abstract representations as it goes deeper. It works by aggregating information from neighboring patches and creating a new patch with a higher feature dimension.



**Figure 4.** On the left is the architecture of a Swin transformer (Swin-T) and on the right are two successive blocks of the Swin transformer W-MSA and SW-MSA which are multihead self-attention modules with regular and offset window configurations, respectively.

**Advantages of the Swin Transformer**

**- Computational Efficiency:** By avoiding global focus on the entire image, the Swin Transformer is significantly more efficient than traditional ViTs, especially for high-resolution images.

**- Modeling Capability:** The pyramidal structure allows the Swin Transformer to capture features at different scales, which can be crucial for tasks such as plant disease detection where patterns can vary in size.

**- Benchmarking:** Tests have shown that Swin Transformer performs very well and better on various data sets, outperforming other ViTs and CNNs in several benchmarks.

# REFERENCES

* Collins, E.J.; Bowyer, C.; Tsouza, A.; Chopra, M. Tomatoes: An Extensive Review of the Associated Health Impacts of Tomatoes and Factors That Can Affect Their Cultivation. Biology 2022, 11, 239.
* Caruso, A. G., Bertacca, S., Parrella, G., Rizzo, R., Davino, S., & Panno, S. (2022). Tomato brown rugose fruit virus: A pathogen that is changing the tomato production worldwide. Annals of Applied Biology, 181(3), 258–274.
* Panno, S., Davino, S., Caruso, A. G., Bertacca, S., Crnogorac, A., Mandić, A., Noris, E., & Matić, S. (2021). A Review of the Most Common and Economically Important Diseases That Undermine the Cultivation of Tomato Crop in the Mediterranean Basin. Agronomy, 11(11), 2188. https://doi.org/10.3390/agronomy11112188
* Singh, V. K., Singh, A. K., & Kumar, A. (2017). Disease management of tomato through PGPB: current trends and future perspective. 3 Biotech, 7(4). https://doi.org/10.1007/s13205-017-0896-1
* Zahangir Alom, Taha, T. M., Yakopcic, C. G., Westberg, S., Sidike, P., Nasrin, Mst Shamima, Esesn, V., Abdul, & Asari, V. K. (2018). The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. ArXiv E-Prints, 2. https://doi.org/10.48550/arxiv.1803.01164
* Boulent, J., Foucher, S., Théau, J., & St-Charles, P.-L. (2019). Convolutional Neural Networks for the Automatic Identification of Plant Diseases. Frontiers in Plant Science, 10. https://doi.org/10.3389/fpls.2019.00941
* Li, E. Y. (2020, August 1). 10 Papers You Should Read to Understand Image Classification in the Deep Learning Era. Medium. https://towardsdatascience.com/10-papers-you-should-read-to-understand-image-classification-in-the-deep-learning-era-4b9d792f45a7
* Rawat, V., Singh, N., Kaur, B., & Bora, S. (2022). Tomato Leaf Disease Classification Using Supervised Learning Techniques: Contrasting Analysis. 2022 International Conference on Advances in Computing, Communication and Materials (ICACCM), 1(2642-7354). https://doi.org/10.1109/icaccm56405.2022.10009617
* Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2. https://doi.org/10.1109/iccv48922.2021.00986
* Rawat, Vandana & Singh, Neelam & Kaur, Bhavleen & Bora, Saksham. (2022). Tomato Leaf Disease Classification Using Supervised Learning Techniques: Contrasting Analysis. 1-7. 10.1109/ICACCM56405.2022.10009617.
* Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 9992-10002, doi: 10.1109/ICCV48922.2021.00986.
* Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... & Asari, V. K. (2018). The history began from alexnet: A comprehensive survey on deep learning approaches.
* Boulent, J., Foucher, S., Théau, J., & St-Charles, P. L. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Frontiers in plant science*, *10*, 941.
* Gadade, H. D., & Kirange, D. D. (2020). Machine learning approach towards tomato leaf disease classification. International Journal of Advanced Trends in Computer Science and Engineering, 9(1), 490-495.
* Abbas, A., Jain, S., Gour, M., & Vankudothu, S. (2021). Tomato plant disease detection using transfer learning with C-GAN synthetic images. Computers and Electronics in Agriculture, 187, 106279.
* Rawat, V., Singh, N., Kaur, B., & Bora, S. (2022, November). Tomato Leaf Disease Classification Using Supervised Learning Techniques: Contrasting Analysis. In 2022 International Conference on Advances in Computing, Communication and Materials (ICACCM) (pp. 1-7). IEEE.
* Thangaraj, R., Anandamurugan, S., Pandiyan, P., & Kaliappan, V. K. (2022). Artificial intelligence in tomato leaf disease detection: a comprehensive review and discussion. Journal of Plant Diseases and Protection, 129(3), 469-488.
* Uppada, R., & Kumar, D. R. (2023). Computer-aided fusion-based neural network in application to categorize tomato plants.Springer-Verlag London Ltd., part of Springer Nature 2023, 1-9.
* Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale.
* Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. (2021, July). Training data-efficient image transformers & distillation through attention. In International conference on machine learning (pp. 10347-10357). PMLR.