**Desarrollo de un modelo para la detección y clasificación de enfermedades en la planta del Tomate utilizándo 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

**TABLA DE CONTENIDOS**

[**INTRODUCCIÓN 3**](#_tycrd2ytbpn5)

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

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

[Objetivo general 5](#_f6dlx77kymqv)

[Objetivos especificos 5](#_d6r42owmwevb)

[**REVISIÓN DE LITERATURA 6**](#_r2lsqc1cu1it)

[**MARCO TEÓRICO 17**](#_osxtvooujh3s)

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

# INTRODUCCIÓN

La detección temprana y precisa de enfermedades en el tomate se ha convertido en una necesidad imperante para mitigar su impacto devastador. Aunque los avances en la visión artificial y el aprendizaje automático han proporcionado soluciones potenciales, estos métodos han sido insuficientes en términos de velocidad, costo, y precisión. Las prometedoras técnicas de Deep Learning, particularmente las redes neuronales convolucionales (CNN), no han vuelto a plantear nuevas arquitecturas ni tampoco un cambio de paradigma, lo que exige un replanteamiento innovador.

Este estudio se propone explorar y validar un enfoque innovador mediante la aplicación del "Swin Transformer", una arquitectura novedosa en el dominio de la visión por computadora. A través de un análisis sistemático y una metodología rigurosa, se buscará desarrollar un modelo que supere las limitaciones existentes en la detección y clasificación de enfermedades del tomate.

Se enfocará en responder preguntas críticas relacionadas con el desarrollo y validación de este nuevo modelo, así como en su comparación con los métodos basados en CNN. La investigación abordará aspectos esenciales como el aumento de datos, la integridad de los resultados, la configuración y entrenamiento óptimo de la arquitectura Swin Transformer, y la interpretación visual de las decisiones del modelo.

La investigación propuesta tiene un potencial significativo para transformar la práctica actual en la detección y clasificación de enfermedades del tomate, ofreciendo un modelo más preciso y eficiente.

El trabajo se estructura en introducción, problema, objetivos, estado del arte, marco teórico, metodología, materiales y métodos, resultados, conclusiones y trabajo futuro.

# PROBLEMA

El cultivo del tomate es esencial para la alimentación y la economía agrícola global debido a que es una de las hortalizas más producidas mundialmente. De acuerdo con la FAO (2021), la producción mundial superó los 189 millones de toneladas, abarcando más de 5 millones de hectáreas cultivadas. Nutricionalmente es una fuente rica de fitoquímicos, nutrientes, antioxidantes y compuestos fenólicos (Collins et al., 2022). Se ha determinado que su consumo aporta beneficios anticancerígenos y protección contra enfermedades cardiovasculares y neurodegenerativas (Collins et al., 2022).

No obstante, el tomate es vulnerable a una amplia variedad de enfermedades causadas por hongos, bacterias, fitoplasmas, virus y viroides (Panno et al., 2021) Tabla 1. Esta vulnerabilidad se ve incrementada por factores como su limitada diversidad genética resultado de la selección intensiva, la tendencia al monocultivo, el intercambio de material infectado a nivel internacional y el cambio climático (Panno et al., 2021). Estas enfermedades representan pérdidas económicas que superan los 30.000 millones de dólares al año (Caruso et al., 2022). Dichas pérdidas repercuten en la economía agrícola, la salud pública y la sostenibilidad ambiental, al afectar el rendimiento y la calidad del cultivo, desestabilizar ecosistemas y elevar los costos de producción por el uso intensivo de pesticidas y otras medidas de control (Panno et al., 2021).

La detección temprana y precisa de enfermedades en el tomate es primordial para mitigar su impacto, pero a pesar de los avances significativos en visión artificial en la última década, estos han presentado limitaciones. En el ámbito del machine learning, los métodos suelen recurrir a la extracción manual de características y algoritmos heurísticos, lo que puede ser lento, laborioso, costoso y susceptible a errores (Zahangir Alom et al., 2018). Estos enfoques han sido superados por técnicas de Deep Learning como las redes neuronales convolucionales (CNN) (Rawat et al. 2022). Sin embargo, estas últimas parecen haber llegado a un punto de estancamiento tras el desarrollo de arquitecturas como ResNet, DenseNet y EfficientNet (Li, 2020).

Es por ello, que la presente investigación sugiere un enfoque innovador basado la arquitectura "Swin Transformer", el cual ha demostrado ser altamente efectivo en tareas de visión por computador pero que nunca ha sido aplicado en la clasificación de enfermedades del tomate.

En este contexto, se plantea las siguientes preguntas:

* ¿Cómo desarrollar y validar un modelo de detección y clasificación de enfermedades del tomate utilizando la arquitectura Swin Transformer que supere la precisión de las redes neuronales convolucionales (CNN)?
* ¿Cómo se puede construir un dataset balanceado en las diferentes clases de las enfermedades del tomate?
* ¿Cómo configurar y entrenar de manera óptima la arquitectura Swin Transformer?
* ¿Cómo evaluar e interpretar las predicciones del modelo Swin Transformer para cada clase?
* ¿Cómo se compara la exactitud (accuracy) del modelo Swin Transformer con los modelos basados en redes neuronales convolucionales?

# OBJETIVOS

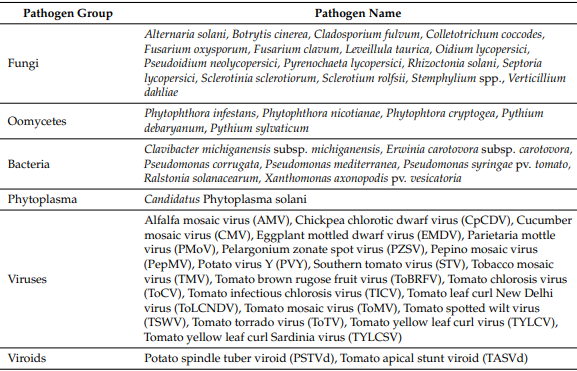
## Objetivo general

Desarrollar y validar un modelo de detección y clasificación de enfermedades de plantas de tomate utilizando la arquitectura Swin Transformer, con el propósito de conseguir una precisión superior en comparación con las redes neuronales convolucionales.

## Objetivos específicos

1. Construir un dataset balanceado en las diferentes clases de las enfermedades del tomate.
2. Configurar y entrenar de manera óptima la arquitectura Swin Transformer.
3. Evaluar e interpretar las predicciones del modelo Swin Transformer para cada clase.
4. Comparar y analizar la exactitud (accuracy) del modelo Swin Transformer con modelos basados en redes neuronales convolucionales.

**Tabla 1.** Lista de patógenos de plantas de tomate (Panno et al., 2021).



# ESTADO DEL ARTE

En el artículo ***"Machine Learning Approach towards Tomato Leaf Disease Classification"***, Gadade y Kirange, (2020) presentan un enfoque de machine learning para la identificación y clasificación de las enfermedades de las hojas del tomate. El proceso de investigación inició con la recolección de 9,000 imágenes de hojas de tomate del dataset de PlantVillage, enfocándose en siete clases de plantas enfermas y una clase de planta sana.

En la fase de preprocesamiento se minimizó el ruido de las imágenes utilizando un filtro de mediana, después se realizó la extracción de características mediante las técnicas GLCM (Gray-Level Co-Occurrence Matrix), Gabor y SURF (Speeded Up Robust Features). Para clasificar las imágenes en categorías normales o enfermas, se emplearon varios algoritmos de clasificación como la máquina de soporte vectorial (SVM), k-Nearest Neighbors (KNN), Naïve Bayes (NB) y árboles de decisión.

Los resultados mostraron que la combinación de SVM y Gabor obtuvo métricas superiores a todas las demás con un 'accuracy' de 0,7339, 'precision' de 0,2525, 'recall' de 0,9492 y 'f1-score' de 0,3989. A pesar de estos resultados, los autores indican que técnicas más avanzadas, como Adaptive Neuro Fuzzy, Neural Networks y algoritmos genéticos, podrían mejorar la clasificación.

No obstante, este trabajo evidencia que los métodos tradicionales de machine learning muestran métricas inferiores en comparación con las técnicas modernas de Deep learning basadas en redes neuronales convolucionales. Esta observación señala una posible obsolescencia de los métodos tradicionales en el ámbito de la clasificación de enfermedades en hojas de tomate a través de técnicas de visión por computadora.

En el estudio titulado ***"ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network"*** de Mohit Agarwala et al. (2020) se propone un enfoque basado en una Red Neuronal Convolucional (CNN) para la detección y clasificación de enfermedades en las hojas del tomate. El estudio utiliza el dataset de Plant Village que contiene imágenes de hojas de tomate afectadas por nueve tipos de enfermedades, además de una clase de hojas sanas. Para equilibrar el dataset, se aplicaron técnicas de aumento de datos resultando en 10,000 imágenes para el entrenamiento, 7,000 para la validación y 500 para las pruebas.

La arquitectura CNN desarrollada en el estudio consta de 3 capas convolucionales, 3 capas de max-pooling y 2 capas completamente conectadas. Se evaluaron también modelos CNN preentrenados a través de transferencia de aprendizaje, empleando el mismo dataset, con el objetivo de contrastar la eficacia del modelo propuesto. Las métricas resultantes mostraron una precisión en un rango de 76% a 100% para diversas clases, con una precisión promedio de 91.2%. Este rendimiento supera al de modelos preentrenados como VGG16, con una precisión del 77.2%, InceptionV3, con 63.4%, y MobileNet, con 63.7%

El modelo propuesto tiene un número menor de parámetros entrenables comparado al de los modelos preentrenados, sugiriendo potenciales ventajas en almacenamiento y computación. No obstante, la variabilidad en la exactitud entre clases y la comparación con trabajos más actuales indican que el modelo, aunque eficaz, está superado por enfoques más recientes.

El artículo ***"Tomato plant disease detection using transfer learning with C-GAN synthetic images"*** de Abbas et al. (2021) aborda un enfoque de Deep learning para la detección y clasificación de enfermedades de la planta del tomate. La metodología implementa dos fases principales, la generación de imágenes sintéticas mediante Redes Antagónicas Generativas Condicionales (C-GAN) para extender el dataset y la clasificación de enfermedades utilizando un modelo DenseNet121 previamente entrenado.

El desempeño del método se evaluó con el dataset de PlantVillage, que incluye nueve clases de hojas de tomate con distintas enfermedades además de una clase de hojas sanas. Las imágenes sintéticas generadas con C-GAN se añadieron al dataset existente. Luego, se entrenó el modelo DenseNet121 con el conjunto de imágenes extendido.

Se realizaron dos series de experimentos, el primero centrado en el entrenamiento de C-GAN y la generación de imágenes sintéticas; el segundo, en el entrenamiento del modelo DenseNet121 utilizando el dataset ampliado. Un análisis comparativo con modelos preentrenados como VGG19, ResNet50, Inception-V3, Xception y MobileNet mostró que el modelo propuesto logró el más alto nivel de exactitud, con un 97.11% en la clasificación de las 10 clases del estudio.

No obstante, aunque el modelo propuesto demostró una alta eficacia en la clasificación de enfermedades de la planta de tomate y la aplicación de C-GAN ayudó a evitar el sobreajuste, es importante evaluar la fidelidad de las imágenes sintéticas generadas, ya que podrían no tener el nivel de detalle que presentan las imágenes reales y, por lo tanto, afectar negativamente el rendimiento del modelo en tareas de generalización.

El estudio ***"Early Detection and Classification of Tomato Leaf Disease Using High-Performance Deep Neural Network"*** de Trivedi et al. (2021) aborda la identificación y clasificación de enfermedades en hojas de tomate mediante Convolutional Neural Networks (CNN). Se empleó un conjunto de datos compuesto por 3000 imágenes, distribuidas en nueve clases de hojas afectadas y una clase de hojas sanas. En la fase de preprocesamiento, las imágenes se normalizaron y redimensionaron.

El proceso comprende etapas de preprocesamiento, segmentación y clasificación a través de CNN. La arquitectura CNN especificada consta de ocho capas convolucionales, ocho capas de max pooling, tres capas fully connected, con una tasa de dropout de 0.5 y función de activación Relu. Se realizaron experimentos variando parámetros como la tasa de aprendizaje y el número de épocas.

Los resultados de la métrica de exactitud fueron del 98.49%. Los experimentos adicionales mostraron que la precisión varía levemente con diferentes tasas de aprendizaje y épocas, oscilando en general alrededor del 98.4% al 98.58%. La comparación con modelos estándar como MobileNet, VGG16 e InceptionV3 indicó que el modelo propuesto supera a estos modelos CNN en exactitud.

El estudio resalta la eficacia del modelo propuesto, especialmente en comparación con otros modelos. Sin embargo, omite discusiones sobre limitaciones, aplicabilidad futura o la justificación para la elección de la arquitectura específica empleada.

El artículo titulado ***"Tomato leaf disease classification using supervised learning techniques: contrasting analysis"***, presentado por Rawat et al., (2022), ofrece una contribución al campo de la clasificación de enfermedades en hojas de tomate a través de técnicas de aprendizaje supervisado.

El estudio se articula en torno a cuatro componentes principales: selección del dataset de imágenes de tomates del archivo "Plant Village", el preprocesamiento de estas imágenes para la eliminación de ruido mediante un filtro de mediana, la aplicación de diversos algoritmos de aprendizaje automático como SVM, K-Nearest Neighbour, Naïve Bayes, Decision Tree, Feed Forward Neural Network, Back Propagation Neural Network, Deep Neural Network, Convolutional Neural Network y Multi Kernel SVM, y por último la evaluación de la efectividad de estos algoritmos mediante métricas como Precisión, F1-Score y Recall.

El estudio realiza un análisis comparativo de la eficacia de los diferentes algoritmos. En este análisis, la Red Neuronal Convolucional (CNN) logra una exactitud de 98,5%, superando a otros algoritmos como SVM, que alcanza un 90% de exactitud, y KNN con un 83,6%. La investigación concluye que las CNN ofrecen un rendimiento superior en la clasificación de enfermedades de hojas de tomate en comparación con métodos tradicionales de machine learning.

El estudio titulado ***"AlexNet Convolutional Neural Network for Disease Detection and Classification of Tomato Leaf"*** (Chen et al., 2022) aborda la detección y clasificación de enfermedades en hojas de tomate mediante una red neuronal convolucional (CNN) basada en la arquitectura de AlexNet. Los autores utilizaron un conjunto de datos de 18,345 imágenes de entrenamiento y 4,585 imágenes de prueba, distribuidas en nueve clases correspondientes a diversas enfermedades de la hoja del tomate y una clase de hoja sana.

La modificación realizada a la arquitectura de la red AlexNet involucra tres capas de convolución, tres capas completamente conectadas y una capa de salida. Se utilizó el optimizador Adam con una tasa de aprendizaje de 0.0005, 75 épocas y un tamaño de lote de 128. La función de pérdida fue la entropía cruzada. El modelo mostró una exactitud del 96%, una tasa de precisión del 98%, un valor de recall del 95% y un F1-score del 97%.

Las conclusiones destacan la eficacia del modelo para la clasificación de enfermedades en hojas de tomate. Se señala que el modelo es lo suficientemente ligero como para implementarse en dispositivos móviles, lo que es relevante dadas las limitaciones de capacidad de memoria en estos dispositivos.

El estudio ***"BotanicX-AI: Identification of Tomato Leaf Diseases Using an Explanation-Driven Deep-Learning Model"*** de Bhandari et al., (2023), aborda la clasificación de nueve clases de enfermedades en hojas de tomate y una clase para las hojas sanas utilizando un modelo de aprendizaje profundo basado en EfficientNetB5. El dataset se construye preprocesando las imagenes mediante redimensionamiento y aumentos de datos resultando en 11,000 imágenes de hojas de tomate distribuidas en 10 clases.

La metodología consiste en aplicar transferencia de aprendizaje utilizando un modelo de EfficientNetB5 al que se le agregaron capas adicionales como la normalización por lotes, capas densas y técnicas de regularización como dropout. Para incrementar la interpretabilidad del modelo se integran técnicas como GradCAM y LIME, que ayudan a identificar las regiones de la imagen que más contribuyen a la clasificación.

En cuanto a métricas el modelo propuesto alcanzó una exactitud del 99.07%, superando a otros modelos preentrenados, como MobileNet (94,00%), Xception (95,32%), VGG16 (93,35%), ResNet50 (96,03%) y DenseNet121 (96,30%). Las técnicas de GradCAM y LIME pudieron identificar las regiones más importantes en las imágenes de las hojas para la clasificación. Aunque se observó que GradCAM tenía limitaciones en la identificación de ciertas regiones en algunas clases de enfermedades.

Finalmente, los autores recomiendan la consideración de otras técnicas de interpretabilidad, como HiResCAM y SHAP, para trabajos futuros, argumentando que estas pueden aumentar la confianza en la aplicabilidad del modelo.

Los estudios revisados abordan el problema de la detección y clasificación de enfermedades en hojas de tomate utilizando diversas técnicas que van desde machine learning tradicional hasta deep learning. Los algoritmos de machine learning, como SVM y KNN, muestran métricas de desempeño inferiores a los modelos basados en deep learning. Entre los modelos de deep learning, las Redes Neuronales Convolucionales (CNN) y sus variaciones, como DenseNet121 y EfficientNetB5, muestran métricas superiores en términos de precisión, exactitud y F1-score.

Los métodos que incorporan imágenes sintéticas generadas mediante Redes Antagónicas Generativas Condicionales (C-GAN) y técnicas de transferencia de aprendizaje parecen ofrecer ventajas en términos de desempeño y prevención de sobreajuste. Sin embargo, la calidad de las imágenes sintéticas generadas es una variable a considerar para tareas de generalización.

La aplicabilidad en dispositivos móviles y la interpretabilidad del modelo son aspectos que se resaltan en algunos trabajos, evidenciando la importancia de estos factores en aplicaciones prácticas. Las técnicas de explicabilidad como GradCAM y LIME ofrecen posibilidades para entender las decisiones del modelo, aunque se sugiere que métodos adicionales pueden mejorar la confianza en el modelo.

Los métodos basados en deep learning muestran un desempeño superior y se destacan como el método más efectivo en la clasificación de enfermedades de la planta de tomate. Sin embargo, no han presentado innovaciones en su arquitectura recientemente. Además, no se ha investigado la aplicación de arquitecturas de visión transformer para esta problemática específica, lo cual representa un área no explorada que podría ofrecer soluciones a limitaciones como la interpretabilidad o una mayor exactitud y que por tanto justifica una investigación más exhaustiva.

# MARCO TEÓRICO

## Enfermedades de la planta de tomate

***Definición***

Las enfermedades en plantas de tomate son alteraciones fisiológicas o estructurales de la planta originados por entidades biológicas conocidas como patógenos y que afectan negativamente la salud o el rendimiento de la planta. La manifestación de estas enfermedades requiere la concurrencia de tres elementos fundamentales; presencia de agentes patogénicos capaces de inducir enfermedad (por ejemplo, hongos, bacterias o virus); un huésped (la planta de tomate) que se encuentra en un estado susceptible, comúnmente debido a condiciones de estrés como sobrefertilización, salinidad del suelo, competencia por espacio o excesiva irrigación; y por ultimo las condiciones ambientales que facilitan la acción del agente patogénico, tales como temperatura y humedad (Flores, 2012; Bernal, 2010).

### Enfermedades más comunes en la planta de tomate

Las enfermedades en tomates pueden ser causadas por una variedad de patógenos, incluidos hongos, bacterias, fitoplasmas, virus y viroides (Panno et al., 2021). Algunas de las enfermedades más comunes que afectan a los tomates en la región del Mediterráneo incluyen:

**Hongos**:

* Early blight of tomato causada por Alternaria solani.
* Septoria leaf spot causada por Septoria lycopersici.
* Grey mould causada por Botrytis cinerea.
* Fusarium wilt of tomato y Crown and root rot causadas por Fusarium oxysporum.
* Verticillium wilt of tomato causada por Verticillium dahliae.

**Bacterias**:

* Bacterial canker causada por Clavibacter michiganensis subsp. michiganensis.
* Bacterial speck causada por Pseudomonas syringae pv. tomato.

**Phytoplasma**:

* Stolbur causada por Candidatus Phytoplasma solani.

**Virus**:

* Spotted wilt disease of tomato causada por Tomato spotted wilt virus.
* Tomato fern leaf causada por Cucumber mosaic virus.
* Tomato yellow leaf curl disease causada por Tomato yellow leaf curl virus y Tomato yellow leaf curl Sardinia virus.
* Tomato brown rugose fruit disease causada por Tomato brown rugose fruit virus.
* Tomato mosaic disease causada por Tomato mosaic virus.

**Viroides**:

* Bunchy top of tomato causado por Potato spindle tuber viroid.

Según Corrales Paredes (2021), a nivel mundial existen más de 200 patógenos que afectan a las plantas de tomate. Las enfermedades de mayor importancia incluyen las causadas por Alternaria solani (tizón temprano), Phytophthora infestans (tizón tardío), Fusarium oxysporum y Ralstonia solanacearum (marchitamientos), así como diversas enfermedades virales.

Blancard (2011) resalta que, en regiones tropicales húmedas, el marchitamiento bacteriano causado por Ralstonia solanacearum es prevalente. Otras enfermedades bacterianas incluyen la sarna bacteriana debido a varias especies de Xanthomonas y el moteado por Pseudomonas tomato. Virus transmitidos por Bemisia tabaci, como el Tomato yellow curl virus (TYLCV) y otros "Geminivirus", son problemáticos en regiones tropicales. En regiones templadas y tropicales, el Tomato spotted wilt virus (TSWV) y el Pepino mosaic virus (PepMV) afectan a los tomates. En el ámbito de los hongos, se destacan enfermedades causadas por Phytophthora infestans, Alternaria tomatophila, Verticillium dahliae y Fusarium oxysporum. Otros agentes patogénicos incluyen nematodos y diversos insectos y ácaros que pueden causar daño a la planta.

### Síntomas

Los síntomas de las enfermedades en plantas de tomate varían considerablemente y están estrechamente ligados al tipo de patógeno involucrado y a la enfermedad específica que se manifiesta. Los síntomas pueden incluir, pero no están limitados a, decoloración de hojas, marchitamiento, necrosis, manchas foliares, deformación de frutos y presencia de hongos o esporas en diversas partes de la planta. Además, la severidad y la aparición de síntomas pueden verse influenciadas por factores externos como condiciones ambientales y prácticas de manejo del cultivo. Es crucial para el manejo efectivo de la enfermedad identificar correctamente los síntomas asociados con cada patógeno o enfermedad específica (Blancard, 2011; Panno et al., 2021).

### Síntomas de enfermedades comunes del tomate:

* **Alternaria solani (early blight)** - Manchas oscuras concéntricas en las hojas, defoliación, lesiones necróticas en flores y tallos, lesiones necróticas en bayas.
* **Septoria lycopersici (Septoria leaf spot)** - Lesiones necróticas circulares a ovales con centros grises en las hojas, defoliación, picnidios en las lesiones.
* **Botrytis cinerea (grey mould)** - Podredumbres blandas, encharcamiento de tejidos, moho gris en tallos, flores y frutos.
* **Fusarium oxysporum f. sp. lycopersici (Fusarium wilt)** - Amarilleamiento de las hojas inferiores, pardeamiento vascular, marchitamiento, colapso de la planta.
* **Fusarium oxysporum f. sp. radicis-lycopersici (crown and root rot)** - Decoloración vascular y cortical de raíces y tallos, necrosis en las raíces, marchitamiento de la planta.
* **Verticillium dahliae (Verticillium wilt)** - Amarilleamiento y necrosis de las hojas inferiores progresando hacia arriba, pardeamiento vascular, marchitamiento de la planta.
* **Clavibacter michiganensis (bacterial canker)** - Enrollamiento marginal de las hojas, cancros en el tallo, lesiones en forma de ojo de pájaro en los frutos.
* **Pseudomonas syringae pv. tomato (bacterial speck)** - Manchas angulares en las hojas, manchas en el tallo, manchas en los frutos.
* **Candidatus Phytoplasma solani (stolbur phytoplasma)** - Amarilleamiento, poca hoja, crecimiento arbustivo, virescencia, frutos pequeños malformados.
* **Tomato spotted wilt virus (TSWV)** - Bronceado de las hojas, manchas necróticas en hojas, tallos y frutos, muerte de la planta.
* **Cucumber mosaic virus (CMV)** - Mosaico, moteado, deformación foliar, lesiones necróticas y deformación del fruto.
* **Tomato yellow leaf curl virus (TYLCV)** - Curvatura de la hoja, amarilleamiento, poca hoja, atrofia de la planta.
* **Tomato brown rugose fruit virus (ToBRFV)** - Deformación de las hojas, amarilleamiento, jaspeado y deformación de los frutos.
* **Tomato mosaic virus (ToMV)** - Moteado y manchado de las hojas, manchado y deformación de los frutos.
* **Parietaria mottle virus (PMoV)** - Necrosis foliar, necrosis del ápice del tallo, manchas anulares en los frutos.
* **Pepino mosaic virus (PepMV)** - Moteado de hojas, jaspeado de frutos.
* **Potato spindle tuber viroid (PSTVd)** - Retraso del crecimiento, enrollamiento de las hojas, clorosis, necrosis venosa, deformidad de los frutos.

### Métodos tradicionales de detección de enfermedades

**La detección manual** de enfermedades en plantas de tomate se lleva a cabo por expertos que deben tener un conocimiento amplio sobre las causas de las enfermedades en los cultivos. Este enfoque es común en áreas rurales y requiere una considerable cantidad de tiempo y recursos humanos para su implementación en grandes fincas. Además, la evaluación manual es susceptible a errores humanos como sesgos o impresiones psicológicas (Thangaraj et al., 2022; Garcia, 2016; Blancard, 2011).

**Los métodos de laboratorio** incluyen técnicas inmunológicas, cultivo de patógenos y microscopía. Aunque estos métodos son precisos, son laboriosos, costosos y requieren tiempo, lo que puede no ser viable para un diagnóstico rápido y eficaz (Verma et al., 2018; Valdivieso, 2021).

### Desafíos en la detección de enfermedades

La detección manual y los análisis de laboratorio presentan desafíos como el tiempo requerido para obtener resultados, el costo y la necesidad de expertos altamente capacitados. Además, la subjetividad en la evaluación manual puede llevar a errores (Garcia, 2016; Valdivieso, 2021).

**Procesamiento de imágenes en la agricultura**

El procesamiento de imágenes contribuye a la optimización sostenible de la producción agrícola mediante la incorporación de diversas tecnologías, dispositivos y enfoques computacionales, y hace parte de lo que se conoce como agricultura 4.0 (J. Gómez-Camperos et al., 2021). Las aplicaciones de esta tecnología en el sector agrícola abarcan: detección de malezas y aplicación focalizada de herbicidas (López, Andrés F Jiménez et al., 2020); gestión de riego de precisión (Capraro & Tosetti, 2020); análisis de imágenes de drones para cubrir áreas extensas, generar orto-mosaicos y visualizar variables agronómicas clave (Castello et al., 2018); utilización de imágenes de sensores remotos y satélites para modelar espacialmente variables fenológicas en zonas cultivables (Jiménez et al., 2013); desarrollo de algoritmos para evaluar el crecimiento vegetal (Alberto et al., 2017); identificación de plagas y enfermedades (J. Gómez-Camperos et al., 2021); y diagnóstico de estados de cultivos a través del procesamiento digital de imágenes (Jiménez et al., 2015).

Estas aplicaciones no solo permiten la recopilación de datos en tiempo real, sino también el procesamiento inmediato de dichos datos. Esto posibilita la rápida identificación de eventos y un conocimiento detallado de la variabilidad en áreas cultivadas. El objetivo final es maximizar la productividad mientras se minimiza el uso de agroquímicos y se mitigan las implicaciones sociales y económicas. Es en este contexto donde se resalta que la detección temprana de enfermedades de las plantas es particularmente relevante para la seguridad alimentaria ya que facilita intervenciones más efectivas, contribuyendo a una producción de alimentos más segura y abundante.

**Seguridad alimentaria (food safety)**

La seguridad alimentaria se centra en prevenir enfermedades transmitidas por alimentos a través de la gestión de toda la cadena de producción, desde el cultivo hasta el consumo. Las fuentes comunes de contaminación incluyen bacterias, virus, parásitos y sustancias químicas (Guennouni et al., 2022; FAO, 2022). Esta contaminación amenaza tanto la salud individual como pública y puede comprometer la calidad nutricional de los alimentos. Por lo tanto, es necesario implementar medidas de control en cada etapa de la cadena alimentaria (Uçar et al., 2016).

Es por eso que en el ámbito de la producción de plantas de cultivo como el tomate la detección temprana de enfermedades cumple un rol fundamental para asegurar tanto la salud del cultivo como la de los consumidores. Ante esta necesidad, la tecnología de visión artificial se presenta como un método efectivo para identificar enfermedades de forma automática, no destructiva y económica. Esta tecnología contribuye a la biovigilancia al permitir una identificación más rápida y precisa de microorganismos patógenos (Khan et al., 2021).

Las tecnologías de visión artificial permiten analizar múltiples variables que influyen en la calidad de los alimentos. Esto incluye el examen de atributos visuales como tamaño, color, forma y textura. Antes de la adopción de estos sistemas tecnológicos, estos análisis eran realizados manualmente, lo que resultaba en un proceso más costoso y subjetivo. Sin embargo, los sistemas de visión artificial ofrecen ventajas tales como mayor velocidad de operación, consistencia, fiabilidad y adaptabilidad en entornos industriales (Khan et al., 2021).

**Visión artificial**

La visión artificial o visión por computador es un campo de estudio que se centra en capacitar a los ordenadores para interpretar y comprender los datos visuales del mundo que les rodea. Implica el uso de algoritmos y modelos matemáticos para analizar y extraer información de imágenes y vídeos. La visión artificial tiene una amplia gama de aplicaciones, como el reconocimiento de objetos, la segmentación de imágenes, la detección de rostros y los vehículos autónomos. Se trata de un campo en rápido crecimiento que evoluciona constantemente con nuevos avances tecnológicos y de investigación (Mahadevkar et al., 2022). En la presente investigación, el enfoque se dirigirá precisamente hacia la aplicación de visión artificial en la detección y clasificación de enfermedades del tomate.

En el campo de la agronomía, el uso de visión artificial se focaliza principalmente en el diagnóstico temprano de enfermedades en plantas. La detección se realiza generalmente a través del análisis de síntomas visibles en las hojas, que a menudo son los primeros indicadores de la presencia de enfermedades. Con los avances en el reconocimiento de patrones y aprendizaje automático, se han desarrollado nuevos métodos para el diagnóstico de enfermedades en plantas. Estos avances han despertado un mayor interés entre los investigadores, por lo que nuevos modelos y técnicas de visión artificial se siguen explorando e implementando (Nanehkaran et al., 2020).

**Modelos de visión artificial**

Existen múltiples tipos de modelos en el campo de la visión por computador. Estos incluyen modelos de aprendizaje profundo como las redes neuronales convolucionales (CNNs), que se han vuelto dominantes en tareas como el reconocimiento de imágenes, la detección de objetos y la segmentación (Yamashita et al., 2018). Las CNNs han demostrado ser una aproximación efectiva y eficiente (Hanhirova et al., 2018). Además de las CNNs, los transformadores de visión (ViTs) son una variante especializada de transformadores que también se emplean en aplicaciones de visión por computador como el reconocimiento de imágenes (Jamil et al., 2023). Otros modelos de visión artificial que no son tan conocidos son los modelos Bayesianos de aprendizaje profundo (Kendall & Gal, 2017), y las redes de codificador-decodificador (Joo & Kärkkäinen, 2020). Estos modelos, incluidos los de aprendizaje profundo, han mostrado un progreso significativo en tareas discriminativas como la clasificación de imágenes, la detección de objetos y la segmentación de imágenes (Shorten & Khoshgoftaar, 2019).

### Técnicas de visión artificial para la Detección de Enfermedades

Los métodos Basados en Inteligencia Artificial se han desarrollado para abordar los desafíos asociados con métodos tradicionales. Estos utilizan imágenes de hojas de plantas de tomate para identificar enfermedades de manera automática. Los métodos más utilizados en este ámbito van desde técnicas de machine learning como algoritmos de extracción de características y de clasificación hasta algoritmos de Deep learning como redes neuronales, redes neuronales convolucionales y redes adversarias generativas (Thangaraj et al., 2022; Valdivieso, 2021; Verma et al., 2018; Abbas et al., 2021).

**Visión artificial mediante machine learning**

Los algoritmos de machine learning o aprendizaje automático han encontrado aplicaciones significativas en el campo de la visión artificial incluyendo la clasificación de enfermedades en hojas de plantas, como el tomate. En la literatura científica, se observa la prevalencia de ciertos algoritmos en este dominio específico (Thangaraj et al., 2022).

Según Sood y Singh (2021), las técnicas populares para el análisis de imágenes en machine learning incluyen Support Vector Machine (SVM), Decision Tree, Random Forest, Redes Neuronales Artificiales (ANN), etc. Aunque subrayan que el deep learning mejora el estado del arte en reconocimiento de objetos y texto, y es particularmente útil en la seguridad alimentaria.

1. **Máquinas de Vectores de Soporte (SVM):** SVM es un algoritmo utilizado con frecuencia que se destaca por su capacidad para manejar espacios de alta dimensión. En el contexto de enfermedades de hojas de tomate, se ha empleado para clasificar diversas enfermedades con alta precisión, utilizando diversas funciones de núcleo como lineal, polinómico y radial.
2. **Árboles de Decisión y Bosques Aleatorios (RF):** Estos algoritmos son comunes debido a su interpretabilidad y capacidad para manejar características categóricas y numéricas. Los bosques aleatorios, una extensión de árboles de decisión, se han utilizado para mejorar la robustez y precisión del modelo.
3. **Redes Neuronales Artificiales y Perceptrón Multicapa (ANN):** Las redes neuronales proporcionan un enfoque flexible y potente, capaz de aprender características complejas. En el ámbito de la detección de enfermedades en hojas de tomate, se ha utilizado para extraer características de textura y color para la clasificación.
4. **Algoritmos Genéticos y Optimización de Enjambre de Partículas (PSO):** Estos algoritmos de búsqueda y optimización han sido aplicados para seleccionar características y optimizar parámetros de modelos, con el fin de mejorar la precisión de la clasificación.
5. **Modelos Estadísticos y Regresión Logística:** Se utilizan para situaciones en las que se busca una solución más simple y fácilmente interpretable, aunque con la posible concesión de una menor precisión en comparación con otros algoritmos más complejos.
6. **Técnicas de Segmentación como Otsu:** Aunque no son algoritmos de clasificación per se, las técnicas de segmentación se utilizan a menudo en una etapa previa para aislar las regiones de interés en las imágenes de las hojas.

**Redes Neuronales Convolucionales (CNN)**

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) son una variante de las redes neuronales artificiales tradicionales (ANN, por sus siglas en inglés). Ambas tipos de redes contienen neuronas que se optimizan a través del aprendizaje. En una CNN, cada neurona recibe una entrada y realiza una operación, tal como un producto escalar seguido de una función no lineal, similar a una red neuronal tradicional. La red completa expresa una única función de puntuación perceptiva, relacionada con los pesos de la red. La última capa de la CNN contiene funciones de pérdida asociadas a las clases del problema que se está resolviendo. Las técnicas desarrolladas para las ANN tradicionales son en general aplicables a las CNN.

La principal diferencia entre CNN y redes neuronales tradicionales es la especialización de las CNN en el reconocimiento de patrones dentro de imágenes. Esto se logra codificando características específicas de imagen en la arquitectura de la red, lo que permite reducir el número de parámetros necesarios para configurar el modelo. Esta especialización permite a las CNN manejar de forma más eficiente la complejidad computacional asociada con el procesamiento de datos de imagen, que puede ser problemática para las ANN tradicionales. Por ejemplo, un dataset como MNIST con dimensiones de imagen de 28x28 es manejable para la mayoría de los ANNs. Pero si se consideran imágenes a color de mayor tamaño, como 64x64, la cantidad de pesos en una sola neurona de la primera capa oculta aumenta significativamente, lo que eleva los requisitos computacionales. Los CNNs ofrecen una solución más eficiente para tratar con este tipo de datos de alta dimensionalidad (O’Shea & Nash, 2015).

Khan et al. (2021) señalan que la diversidad en la forma, cantidad y contenido de los alimentos hace que la identificación de productos alimenticios sea un proceso complejo. La red neuronal convolucional (CNN) es el modelo más comúnmente utilizado para la identificación y categorización de alimentos debido a su eficacia en el análisis de imágenes.

Nanehkaran et al. (2020) resaltan que las CNN se han convertido en una técnica dominante para abordar dificultades en la identificación de enfermedades en plantas. Estas redes han superado a otros métodos de clasificación comunes como k-NN, SVM, y ANN en este ámbito. Las CNN han mostrado avances significativos en problemas relacionados con métodos de clasificación en la detección de enfermedades en plantas.

Sood y Singh (2021) y Xiao et al. (2022) mencionan que, además de las CNN, existen otras técnicas como SVM, ANNs, y NDVI para el análisis de imágenes en la seguridad alimentaria y la detección de enfermedades en plantas. Sin embargo, las CNN tienen la ventaja de la extracción automática de características, lo que reduce la necesidad de ingeniería manual de características. Este aspecto es particularmente útil para manejar tareas complejas en visión por computadora, como la clasificación de imágenes, detección de objetos y segmentación semántica (Xiao et al., 2022).

Thangaraj et al. (2022) resaltan que las CNN han demostrado ser eficaces en la detección y clasificación de enfermedades de plantas, como las hojas de tomate. Enumeran diversas arquitecturas de CNN, incluidas AlexNet, LeNet, GoogLeNet, VGGNet, ResNet, DenseNet, Inception V3 y Xception, que han sido empleadas para identificar enfermedades en las hojas de tomate con gran precisión (Thangaraj et al., 2022).

Claude 2

Redes Neuronales Convolucionales (CNN)

Las CNN son populares para el reconocimiento y clasificación de imágenes en aplicaciones de food safety y detección de enfermedades en plantas debido a su habilidad para manejar eficientemente la alta dimensionalidad de los datos de imágenes (Khan et al., 2021; Sood & Singh, 2021). Comparado con métodos tradicionales, las CNN pueden extraer automáticamente características de las imágenes a través del aprendizaje, lo que permite una identificación más inteligente (Xiao et al., 2022).

Se han utilizado varias arquitecturas CNN populares para la detección y clasificación de defectos y enfermedades en alimentos y plantas, incluyendo VGG, Inception, ResNet y DenseNet (Nanehkaran et al., 2020; Xiao et al., 2022). Los modelos CNN entrenados en grandes conjuntos de datos de imágenes han demostrado un alto rendimiento en tareas como identificación de variedades de frutas, detección de defectos en vegetales, y diagnóstico de enfermedades en plantas (Thangaraj et al., 2022).

Las CNN son una herramienta poderosa para el análisis automatizado de imágenes en aplicaciones de seguridad alimentaria y agricultura, permitiendo la implementación de sistemas de visión artificial más inteligentes y efectivos. Su habilidad para extraer características de forma automática a través del aprendizaje profundo las hace ideales para el procesamiento de grandes volúmenes de datos de imágenes.

**Vision Transformer**

**Redes Neuronales Convolucionales VS Vision Transformer**

Este aspecto ofrece una imagen contradictoria sobre si los transformadores de visión son mejores que las redes neuronales convolucionales. Mientras que algunos estudios han demostrado que los transformadores de visión pueden superar a las redes neuronales convolucionales en determinadas tareas [1][2], otros han descubierto que las redes neuronales convolucionales siguen superando a los transformadores de visión en otras tareas [7][8]. Además, algunos estudios han explorado la combinación de los puntos fuertes de los transformadores de visión y de las redes neuronales convolucionales para crear modelos híbridos que funcionan mejor que cualquiera de las dos arquitecturas por sí solas [2][4]. Por lo tanto, es difícil hacer una afirmación definitiva sobre qué arquitectura es mejor en general, ya que depende de la tarea específica y del conjunto de datos que se utilice.

**Swin Transformer**

**Dataset**

**Dataset de plantVillage y plantDoc**

**Aumento de datos**

**Métricas de evaluación**

**Interpretabilidad de Modelos**

**Arquitectura Swin Transformer**

Swin Transformer, que significa "Shifted Window Transformer", introduce una arquitectura innovadora que aborda las limitaciones de los ViTs y las CNN tradicionales. Comienza dividiendo la imagen en parches y utiliza la atención de ventanas deslizantes en lugar de atención global, lo que reduce significativamente la complejidad computacional. Además, implementa una estructura piramidal para procesar imágenes a diferentes resoluciones, permitiendo capturar características a diferentes escalas (Liu et al., 2021).

**Dataset PlantVillage**

Para entrenar modelos de detección de enfermedades del tomate, es esencial contar con un conjunto de datos de alta calidad. El dataset PlantVillage proporciona una variedad de clases de enfermedades del tomate y refleja la complejidad de la detección en el campo. La calidad de los datos etiquetados es fundamental para el éxito de la capacitación de modelos.

**Clases del dataset de plantVillage:**

Tomato Yellow Leaf Curl Virus (YLCV)

Tomato Bacterial Spot (Bctsp)

Tomato Late Blight (TLB)

Tomato Septoria leaf spot (SptL)

Tomato Two Spotted Spider Mite (SpdM)

Tomato Target Spot (TISS)

Tomato Early Blight (TEB)

Tomato Leaf Mold (LMld)

Tomato Mosaic Virus (MscV)

Tomato healthy (Hlth)

**Interpretabilidad en Modelos de Visión Artificial**

La interpretabilidad en modelos de visión artificial es crucial para evaluar la calidad de las decisiones del modelo en la clasificación de enfermedades de los tomates. Esta interpretabilidad permite comprender cómo se llega a cada decisión y puede guiar acciones correctivas.

**Clasificación de Imágenes**

Esta tarea pertenece al ámbito de la visión por computadora. Consiste en asignar una etiqueta a una imagen de entrada seleccionada de un conjunto de categorías predefinidas. El proceso generalmente implica el uso de modelos de aprendizaje profundo como las redes neuronales convolucionales.

**Matriz de Confusión**

Es una tabla que presenta el rendimiento de un algoritmo de clasificación. Las filas y columnas representan las clases reales y las clases predichas, respectivamente. Los elementos de la matriz son:

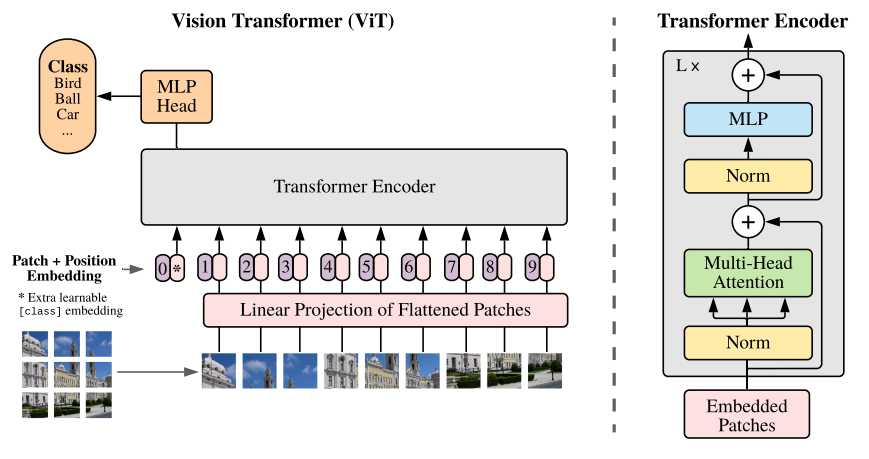
* Verdaderos Positivos (TP): Casos positivos correctamente identificados.
* Verdaderos Negativos (TN): Casos negativos correctamente identificados.
* Falsos Positivos (FP): Casos negativos incorrectamente identificados como positivos.
* Falsos Negativos (FN): Casos positivos incorrectamente identificados como negativos.

**Métricas de Evaluación:** Parámetros que cuantifican la calidad de un modelo en tareas específicas. Se definen matemáticamente como:

* 
* 
* 
* 

**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 (Alexey Dosovitskiy et al., 2020).



**Figure 1:** Model overview of the first visual transformer (Alexey Dosovitskiy et al., 2020)

**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 (Alexey Dosovitskiy et al., 2020; Touvron et al., 2021).

**General Architecture of the Swin Transformer**

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.

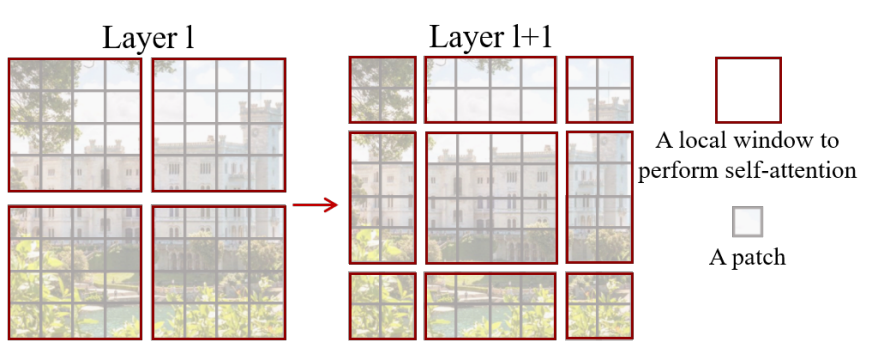
**GELU (Gaussian Error Linear Unit):** An activation function used in neural networks. It has been shown to have beneficial properties in transformer models.

**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 (Liu et al., 2021).



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

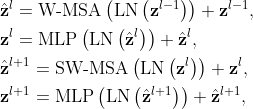
Equations (1) and (2) show the difference in complexity between global care and window-based care. Window-based attention is scalable and computationally more efficient (Liu et al., 2021).

(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 (Liu et al., 2021).

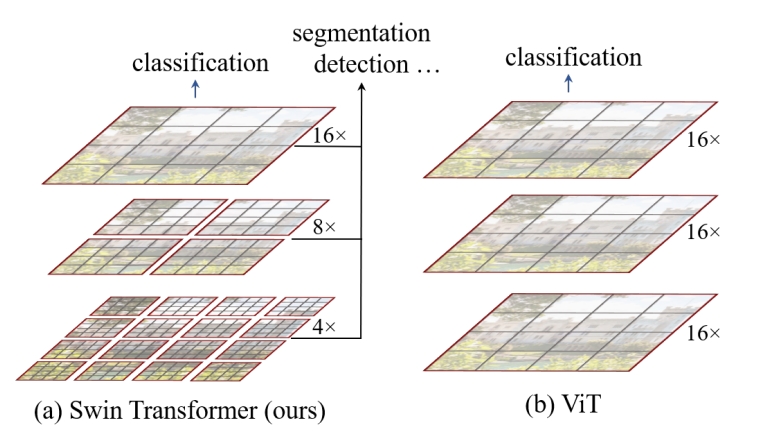
(3) 

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 (Liu et al., 2021).



**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 (Liu et al., 2021).

**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 (Liu et al., 2021).

**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.

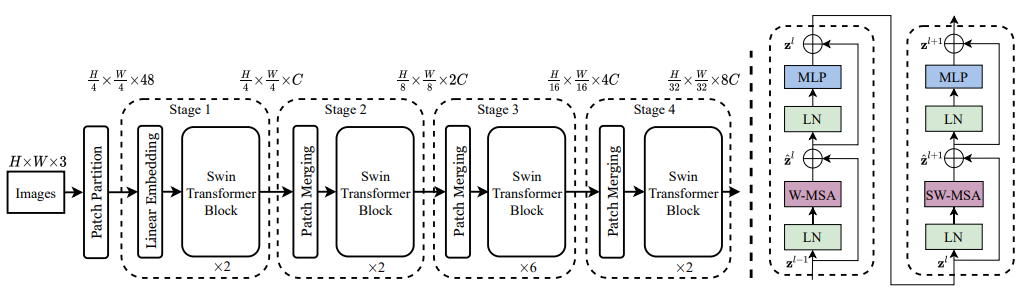
**GELU:** GELU, or Gaussian Error Linear Unit, is a nonlinear activation function. It is used in neural networks to introduce nonlinearities into the model. In the Swin Transformer block, the GELU is used as an activation function in the 2-layer multilayer perceptron (MLP).

**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 (Liu et al., 2021).

**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.

# REFERENCIAS

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.

Agarwal, Mohit, et al. “ToLeD: Tomato Leaf Disease Detection Using Convolution Neural Network.” *Procedia Computer Science*, vol. 167, no. 293–301, 2020, pp. 293–301, https://doi.org/10.1016/j.procs.2020.03.225.

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). https://doi.org/10.1016/j.compag.2021.106279

Chen, H.-C., Widodo, A. M., Wisnujati, A., Rahaman, M., Lin, J. C.-W., Chen, L., & Weng, C.-E. (2022). AlexNet Convolutional Neural Network for Disease Detection and Classification of Tomato Leaf. Electronics, 11(6), 951. https://doi.org/10.3390/electronics11060951

Bhandari, M., Shahi, T. B., Neupane, A., & Walsh, K. B. (2023). BotanicX-AI: Identification of Tomato Leaf Diseases Using an Explanation-Driven Deep-Learning Model. Journal of Imaging, 9(2), 53. https://doi.org/10.3390/jimaging9020053

FAO. (2021). Crops and livestock products. Www.fao.org. https://www.fao.org/faostat/en/#data/QCL

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

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.

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.

Flores, C. (2012). Enfermedades de tomate: Guía de consulta (S. Buono & S. Giorgioni, Eds.; 1st ed.). Ediciones Instituto Nacional De Tecnología Agropecuaria De Argentina INTA.

Bernal, R. (2010). Enfermedades de tomate (lycopersicum esculentum mill.) en invernadero en las zonas de salto y bella unión. Serie Técnica. INIA. Montevideo, Editorial Hemisferio Sur SRL, 181, 1–71.

Thangaraj, R., Anandamurugan, S., Pandiyan, P., & Kaliappan, Vishnu Kumar. (2022). Artificial intelligence in tomato leaf disease detection: a comprehensive review and discussion. *Journal of Plant Diseases and Protection*, *129*, 3.

Garcia, J. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. Biosystems Engineering, 144, 52–60.

Valdivieso. (2021). Estudio comparativo de las técnicas de inteligencia artificial para el diagnóstico de enfermedades en la agricultura.

Blancard, D. (2011). Enfermedades del tomate. Mundi-Prensa Libros.

Verma, S., Chug, A., & Singh, A. P. (2018). Prediction models for identification and diagnosis of tomato plant diseases. IEEE, 1557–1563.

Corrales Paredes, J. (2021). Uso de depredadores, parasitoides y entomopatógenos para el control biológico de plagas y enfermedades de hortalizas con énfasis en tomate (Solanum lycopersicum l.).

Gómez-Camperos, J., Jaramillo, H., & Guerrero-Gómez, G. (2022). Técnicas de procesamiento digital de imágenes para detección de plagas y enfermedades en cultivos: una revisión. Ingeniería Y Competitividad, 24, 1.

Castello, F. A., Albornoz, E., & Martínez, C. E. (2018). Desarrollo de un sistema integral de adquisición de imágenes por dron, procesamiento y análisis para agricultura de precisión. X Congreso de Agro Informática (CAI)-JAIIO 47 (CABA, 2018 López, Andrés F Jiménez, Andrea, D., & Ramírez, Dayra Yisel García. (2020). Sistema inteligente para el manejo de malezas en el cultivo de piña con conceptos de agricultura de precisión. Ciencia Y Agricultura, 17, 3.).

Capraro, F., & Tosetti, S. (2020). Herramientas modernas de gestión en riego de precisión basadas en dispositivos electrónicos, programas informáticos y técnicas de control automático. Electronic Journal of SADIO (EJS), 19, 1.

Jiménez, F., Jisell, M., Eduardo, O., & Salamanca, J. M. (2015). Diagnóstico de cultivos utilizando procesamiento digital de imágenes y tecnologías de agricultura de precisión. INGE CUC, 11, 1.

Alberto, C., Mendoza, N. F., Díaz, N. G., & Rebolledo, A. G. (2017). Procesamiento de imágenes de plantas ornamentales multi-escala para calcular su crecimiento. 3 c TIC: Cuadernos de Desarrollo Aplicados a Las TIC, 6, 3.

Jiménez, F., Jiménez, R., & Pérez, Elkyn Fagua. (2013). Procesamiento digital de imágenes de sensores remotos para aplicaciones de agricultura de precisión. REVISTA COLOMBIANA de TECNOLOGIAS de AVANZADA (RCTA), 1, 21.

Guennouni, M., Admou, B., Bourrhouat, A., El Khoudri, Noureddine, Zkhiri, W., Talha, I., Hazime, R., & Hilali, A. (2022). Knowledge and practices of food safety among health care professionals and handlers working in the kitchen of a moroccan university hospital. Journal of Food Protection, 85, 4.

Uçar, A., Yilmaz, M. V., & Çakiroglu, Funda Pinar. (2016). Food safety–problems and solutions. Significance, Prevention and Control of Food Related Diseases, 3.

Nanehkaran, Y., Zhang, D., Chen, J., Tian, Y., & Al-Nabhan, N. (2020). Recognition of plant leaf diseases based on computer vision. In Journal of Ambient Intelligence and Humanized Computing (pp. 1–18). Springer.

Khan, R., Kumar, S., Dhingra, N., & Bhati, N. (2021). The use of different image recognition techniques in food safety: a study. Journal of Food Quality, 2021, 1–10.

Mahadevkar, S. V., Khemani, B., Patil, S., Kotecha, K., Vora, D. R., Abraham, A., & Gabralla, L. A. (2022). A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions. IEEE Access, 10, 107293–107329. https://doi.org/10.1109/access.2022.3209825

Sood, S., & Singh, H. (2021). Computer vision and machine learning based approaches for food security: A review. Multimedia Tools and Applications, 80, 18.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights into Imaging, 9(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9

Hanhirova, J., Kämäräinen, T., Seppälä, S., Siekkinen, M., Hirvisalo, V., & Ylä-Jääski, A. (2018). Latency and throughput characterization of convolutional neural networks for mobile computer vision. 204–215.

Jamil, S., Md. Jalil Piran, & Kwon, O.-J. (2023). A Comprehensive Survey of Transformers for Computer Vision. Drones, 7(5), 287–287. https://doi.org/10.3390/drones7050287

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, *30*.

Joo, J., & Kärkkäinen, K. (2020). *Gender slopes: Counterfactual fairness for computer vision models by attribute manipulation*. 1–5.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0197-0

Xiao, Z., Wang, J., Han, L., Guo, S., & Cui, Q. (2022). Application of machine vision system in food detection. *Frontiers in Nutrition*, *9*, 888245.