



UNIVERSIDAD ESAN
FACULTAD DE INGENIERÍA
INGENIERÍA DE TECNOLOGÍAS DE INFORMACIÓN Y SISTEMAS

**Implementación de técnicas de Visión Computacional para la detección temprana de
Rancha”(Phytophthora infestans) en hojas de papa peruana**

Trabajo de investigación para el curso de Trabajo de Tesis I

Sebastian Puruguay
Asesor: Marks Calderón

Lima, 22 de junio de 2024

Índice general

1. PLANTEAMIENTO DEL PROBLEMA	4
1.1. Descripción de la Realidad Problemática	4
1.2. Formulación del Problema	7
1.2.1. Problema General	7
1.2.2. Problemas Específicos	7
1.3. Objetivos de la Investigación	8
1.3.1. Objetivo General	8
1.3.2. Objetivos Específicos	8
1.4. Hipótesis	8
1.4.1. Hipótesis General	8
1.4.2. Hipótesis Específicas	8
1.5. Justificación de la Investigación	9
1.5.1. Teórica	9
1.5.2. Práctica	10
1.5.3. Metodológica	10
1.6. Delimitación del Estudio	10
1.6.1. Espacial	10
1.6.2. Temporal	11
1.6.3. Conceptual	11

1.6.4. Matriz de Consistencia	11
2. MARCO TEÓRICO	12
2.1. Antecedentes de la investigación	12
2.1.1. Automated recognition of optical image based potato leaf blight diseases using deep learning (CHAKRABORTY2022101781)	12
2.1.2. Research and Validation of Potato Late Blight Detection Method Based on Deep Learning (antecedente2)	15
2.1.3. Supervised Learning-Based Image Classification for the Detection of Late Blight in Potato Crop (antecedente3)	19
2.1.4. Potato Blight Classification Android Application using Deep Learning (antecedente5)	22
2.1.5. Deep Convolutional Neural Networks for image based tomato leaf disease detection (antecedente6)	24
2.1.6. Potato Blight Classification Android Application using Deep Learning (antecedente7)	27
2.1.7. Investigation of <i>Phytophthora Infestans</i> Causing Potato Late Blight Disease: A Review (antecedente4)	30
2.2. Bases Teóricas	32
2.2.1. Inteligencia Artificial	32
2.2.2. Deep Learning	32
2.2.3. Support Vector Machine: A comprehensive survey on support vector machine classification: Applications, challenges and trends (tecnica3) .	33
2.2.4. Redes neuronales convolucionales: Review of Image Classification Algorithms Based on Convolutional Neural Networks (tecnica2)	46
2.2.5. Vision Computer	63
2.2.6. Vision Transformer: Explainability of Vision Transformers: A Comprehensive Review and New Perspectives (tecnica1)	63
2.2.7. You Only Look One (YOLO): You Only Look Once: Unified, Real-Time Object Detection (tecnica4)	74

2.3. Marco Conceptual	80
2.3.1. Vision Computacional (vc1)	80
2.3.2. Rancha "Phytophthora Infestans" Tizon Tardio (prom2)	84
3. METODOLOGÍA DE LA INVESTIGACIÓN	87
3.1. Diseño de la investigación	87
3.1.1. Diseño no experimental	87
3.1.2. Tipo explicativo	87
3.1.3. Enfoque cuantitativo	88
3.2. Población y Muestra	88
3.3. Operacionalización de Variables	90
3.4. Técnicas de recolección de Datos	90
3.4.1. Captura de Imágenes	90
3.4.2. Anotación de Imágenes	91
3.5. Técnicas para el Procesamiento y Análisis de Información	91
3.5.1. Metodología de la implementación de la solución	91
3.5.2. Metodología para la medición de resultados	92
A. Anexo I: Matriz de Consistencia	94
B. Anexo II: Resumen de Papers investigados	96
C. Anexo III: Árbol del problema	98
D. Anexo III: Árbol de objetivo	99

Capítulo 1

PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la Realidad Problemática

Hoy en día, la producción de papa es una de las actividades más importantes en nuestro país tanto a nivel económico como alimenticio y el Perú es el mayor productor de papa en América Latina contando con más de 4mil variedades.([cr'elcomercioproduccionpapa](#)).

De acuerdo al reporte del Ministerio de Desarrollo Agrario y Riego (Midagri) del 2023, el consumo de papa por persona es de, aproximadamente, 92 kilos por año, cifra que ha venido incrementándose desde hace dos décadas debido a la importancia que tiene este tubérculo en la dieta de los peruanos. La papa es considerada uno de los alimentos más importantes y nutritivos en el Perú, rico en carbohidratos, proteínas, vitaminas y minerales. Mientras que otros productos relevantes como el arroz no se cultivan en más de 7 regiones, lo que evidencia la amplia distribución geográfica del cultivo de papa en el territorio peruano ([cr'agroinforma1](#)). Asimismo, el tubérculo se siembra en 19 distintas regiones del país, resaltando en departamentos como Puno y Huánuco, que lideran la producción nacional con 20.6% y 12.6% respectivamente ([minagri'estadisticas'2022](#)). También, según el Censo Nacional Agropecuario 2022, el cultivo de papa es el sustento de al menos 710,000 familias en el Perú, representando aproximadamente el 25% de los hogares dedicados a la agricultura y aportando el 6.5% al PBI agropecuario del país ([inei'cenagro'2022](#)).

Sin embargo, en los últimos años, la producción se ha visto afectada por una de las patologías más mortales que es Phytophthora infestans, conocido por los agricultores como Rancha o tizón tardío, la cual es capaz de acabar con el 100% de cultivos enteros en solo unos días sino se antepone una medida de prevención. Según el técnico del Servicio para el Desarrollo Integral Rural, la rancha es peligrosa y puede arruinar toda la siembra. Es relevante

para el agricultor diagnosticarla y encontrar la mejor manera de controlarla. Las condiciones climáticas propicias para la aparición de la rancha o tizón tardío se dan cuando las temperaturas fluctúan entre los 15°C y 18°C, o bien, cuando la temperatura se mantiene por debajo de los 25°C durante un periodo de 7 días consecutivos. Asimismo, esta enfermedad se ve favorecida por la presencia de lluvias constantes o cuando los niveles de humedad relativa oscilan entre el 90 % y el 100% **(cr·rancha1)**. Los principales síntomas de la enfermedad en el tubérculo son: hojas, lesiones en los bordes mostrando manchas necróticas con halo amarillento y micelio blanquecino en el relieve de la hojas; tallos, lesiones que recorren el tallo de color marrón osucro; tubérculos, lesiones irregulares en de color marrón rojizo que no solo están presentes en los alrededores de la papa sino del al interior también **(cr·rancha2)**.

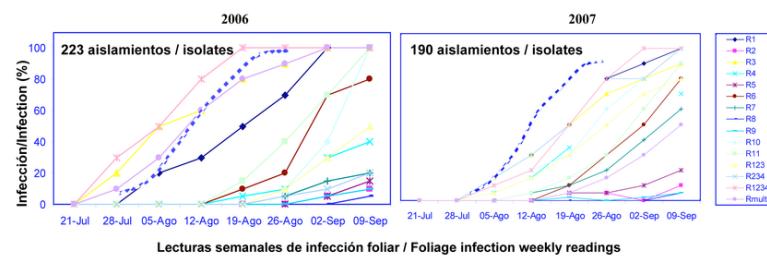


Figura 1.1: Lecturas semanales de infección foliar

Fuente: **cr·gestion2018emprend.** *El tizón tardío es capaz de llegar a un alto porcentaje de severidad en solo dos semanas.*

El caso más popular sobre rancha en Perú ocurrió en 2010 afectando a la región de Junín, donde al menos 26 mil de toneladas de papa fueron afectadas por la rancha en dos mil 660 hectáreas , lo cual representó una pérdida de 10 millones de nuevos soles. Esto se pudo evitar con fungicidas; sin embargo, estos son costosos entonces no todos los agricultores tiene acceso a este **(cr·rancha6)**.

En abril de 2024, la rancha afectó a sembríos de papa nativa en la sierra de Lima, según reportes el tizón tardío arrasó con tres custodios de papa autóctonos tres en Huarachorí, dos en Yauyos y uno en Cajatambo **(cr·rancha3)**. Asimismo, en marzo de 2024, la rancha afectó fuertemente en el departamento de Junín, haciendo que 2000 agricultores pierdan, aproximadamente, 550 hectareas de papa por el tizón tardío, valorizando esta pérdida en S/300 mil. Siendo este el principal cultivo agrícola en su sierra **(cr·rancha4)**. Finalmente, en abril de 2024, la rancha atacó el departamento de Apurímac, la cual perdió una hectárea de papa y una y media hectáreas de cultivos afectados **(cr·rancha5)**.

Ante esto, el gobierno se ha visto obligado a concientizar sobre la rancha, en marzo de 2024, el Servicio para el Desarrollo Integral Rural (Sedir) y Servicio Nacional de Sanidad

Agraria (Senasa) en Ancash, ya que varios agricultores locales temen perder sus sembríos debido a la rancha, principalmente porque se desarrolla en temperaturas entre 12°C y 16°C y con fuerte presencia de humedad. Según la charla, una forma de prevenir la rancha es rotando el cultivo, es decir cambiando a otros productos como maíz y alverja (**cr'agroinforma2**).

Por eso mismo, con el gran avance de la Inteligencia Artificial (IA) se ha perfeccionado y popularizado su uso en el sector agrícola de distintas formas, ya que es una herramienta útil y apta para procesar enormes cantidades de información rápidamente. La IA es una alternativa tecnología que puede brindar soluciones prácticas porque es eficiente creando algoritmos capaces de detectar patologías a tiempo, así como mejorar tanto la precisión del diagnóstico como la eficiencia del flujo de trabajo en el campo, además que reduce costes porque agiliza. La IA también se puede emplear para identificar a plantas que están en peligro de pescar una enfermedad específica, lo que facilita la intervención y prevención temprana. Por ejemplo, un estudio reciente mostró que un sistema de IA logró detectar la zona enferma de una hoja con una precision deñ 96.44 % (**cr'iaplanta**).

Añadiendo, la adopción de tecnología de IA en el ámbito de la agricultura del Perú puede ofrecer soluciones para una serie de problemas que enfrenta el sistema de siembra y cosecha, como el bajo acceso a asistencia técnica en este sector en el país, la escasez de personal, la falta de competencias y la insuficiencia de equipos tecnológicos para una mejor desempeño para el proceso agrícola. En muchas ocasiones, particularmente en las zonas rurales y de recursos limitados, la falta de tecnología agrícola adecuada puede restringir la productividad y la calidad de los cultivos en la agricultura peruana. Sin embargo, al emplear herramientas de inteligencia artificial, como sistemas de monitoreo automatizado de cultivos y análisis de datos agrícolas, se puede incrementar la eficiencia en la producción, reducir los costos y garantizar una mayor calidad de los productos agrícolas, beneficiando así a más agricultores y contribuyendo al desarrollo sostenible del sector. En muchos casos, especialmente en las áreas rurales y de bajos ingresos del Perú, la falta de equipamiento agrícola básico y adecuado limita el rendimiento y la productividad de los cultivos. Al utilizar herramientas de inteligencia artificial, como sistemas de monitoreo de cultivos, análisis de patologías en cultivos y predicción del clima, se puede mejorar la toma de decisiones en cuanto a la siembra, el riego y la aplicación de insumos, optimizando los recursos y aumentando los rendimientos, lo que permite una mayor producción agrícola para alimentar a más personas. Un porcentaje significativo de los trabajos se enfocan en la detección y diagnóstico de enfermedades (29.4%) y en la recomendación de fertilizantes (29.4%), sin embargo, este último aspecto no abarca específicamente el cultivo de maíz. Los sistemas expertos están diseñados principalmente para que el agricultor pueda tomar decisiones más acertadas durante el ciclo de cultivo, lo cual representa el 80.6% de los casos. Estos sistemas se consideran usuarios potenciales, ya que muchos agricultores no tienen los

recursos para contratar a un experto en el área. Los sistemas expertos ofrecen una alternativa accesible y de bajo costo (**criaplanta2**).

En este sentido, la presente investigación busca desarrollar e implementar un sistema robusto y escalable de visión computacional, aprovechando los avances tecnológicos más recientes, con el fin de proporcionar una solución innovadora, accesible y de alto impacto para los agricultores peruanos, facilitando la lucha contra esta devastadora enfermedad y promoviendo la resiliencia y productividad de los cultivos de papa a nivel nacional.

1.2. Formulación del Problema

Para formular los problemas de esta investigación, se creó un "árbol de problemas". (véase Anexo C.1).

1.2.1. Problema General

¿Es posible desarrollar un sistema de visión computacional que permita la clasificación temprana de la rancha (*Phytophthora infestans*) en hojas de papa peruana con alta precisión?

1.2.2. Problemas Específicos

- ¿Cómo podemos desarrollar un sistema de detección automática de "Rancha" *Phytophthora infestans* en las hojas de papa peruana que sea preciso, eficiente y robusto a la variabilidad de las lesiones y las condiciones ambientales?
- ¿Cómo obtener un conjunto de datos representativo y diverso de imágenes de hojas de papa con Rancha y sin ella?
- ¿Cuáles son las técnicas de visión computacional más apropiadas para la detección temprana de Rancha.^{en} hojas de papa peruano?
- ¿Cuáles son las técnicas de preprocesamiento de imágenes más apropiadas para la detección temprana de Rancha.^{en} hojas de papa peruano??

1.3. Objetivos de la Investigación

1.3.1. Objetivo General

Desarrollar un sistema de visión computacional basado en técnicas de aprendizaje profundo para la clasificación temprana de la rancha (*Phytophthora infestans*) en hojas de papa peruana.

1.3.2. Objetivos Específicos

- Desarrollar un sistema de detección automática de "Rancha" *Phytophthora infestans* que alcance una precisión alta en la detección de lesiones en las hojas de papa.
- Recolectar un conjunto de datos amplio y diverso de imágenes de hojas de papa que cubra una variedad de condiciones y escenarios relevantes para la detección de la Rancha.
- Identificar las técnicas de visión computacional más adecuadas para la detección temprana de "Rancha" en hojas de papa peruana.
- Identificar las técnicas de procesamiento de imágenes más adecuadas para la detección temprana de "Rancha" en hojas de papa peruana.

1.4. Hipótesis

1.4.1. Hipótesis General

Se sostiene que mediante el desarrollo de un sistema de clasificación automática de las lesiones de "Rancha" *Phytophthora infestans* en las hojas de papa peruana, se puede lograr un método de detección de la Rancha en las hojas de papa peruana.

1.4.2. Hipótesis Específicas

- HE1: La implementación de técnicas de aprendizaje automático y detección de imágenes permitirá desarrollar un sistema de detección automática de "Rancha" *Phytophthora infestans* con alta precisión.

- HE2: La recopilación de un conjunto de datos representativo y diverso proporcionará una base sólida para el entrenamiento y la validación de los algoritmos de visión computacional, lo que mejorará su capacidad para detectar con precisión la Rancha en las hojas de papa.
- HE3: Se sostiene que al investigar específicamente técnicas de visión computacional, se logrará una mejora significativa en la detección temprana de Rancha.^{en} hojas de papa peruano.
- HE4: Se sostiene que al investigar específicamente técnicas de preprocessamiento de imágenes, se logrará una mejora significativa en la detección temprana de Rancha.^{en} hojas de papa peruano.

Se ha verificado que los problemas, objetivos e hipótesis descritos anteriormente guardan una estrecha relación entre sí, como se puede apreciar en la Matriz de Consistencia del [A.11](#). Los objetivos específicos, por su parte, fueron el resultado de una lluvia de ideas realizada tras analizar los objetivos planteados en los antecedentes. El detalle de estos objetivos, junto con su correspondiente referencia, se encuentra en el [D.1](#).

1.5. Justificación de la Investigación

1.5.1. Teórica

El propósito de esta investigación es contribuir al avance en la clasificación temprana de la 'Rancha' *Phytophthora infestans* en las hojas de papa peruana mediante la aplicación de técnicas de Visión Computacional. Este problema reviste importancia debido a su impacto en la agricultura peruana, donde la detección temprana de la enfermedad es crucial para su manejo efectivo.

Cabe recalcar que cada vez este tipo de herramientas tecnológicas son más útiles para la clasificación de patologías visuales en las hojas de plantas. Asimismo, es importante resaltar que no hay este tipo de investigaciones en Perú, siendo uno de los países con más historia con la papa y que cuenta con miles de variedades.

Al implementar este enfoque multimodal, se espera proporcionar una herramienta útil para los agricultores y expertos en la detección temprana de la 'Rancha' *Phytophthora infestans*, lo que eventualmente podría contribuir a una mejor gestión de esta enfermedad y a la protección de los cultivos de papa en el Perú.

1.5.2. Práctica

Muchos de los trabajos previos, superaron su efectividad y precision esperada; sin embargo, en gran mayoría de la literatura mencionada (5 de 5) utilizaron técnicas comunes y antiguas, ninguno optó por utilizar técnicas innovadoras como Visual Transformer que en este caso sí se dará.

Al concluir esta investigación, las personas podrán hacer uso del sistema de clasificación temprana de la 'Rancha' *Phytophthora infestans* en las hojas de papa peruana con el fin de tomar las medidas apropiadas para su control y erradicación. De encontrarse, con una plantación infectada deberán utilizar pesticidas especializados para esos casos de manera profesional. Los beneficiados serán aquellas personas que viven de la cultivación de este tubérculo, tales como agricultores, campesinos, negociantes y el consumidor final. Esta investigación va a servir para evitar un tardía reacción a la Rancha en los cultivos de papa para que así no haya pérdida de cosecha. El trabajo presente demostrará que un sistema de vision computacional puede clasificar, mediante imágenes RGB de hojas de planta de papa, si el tubérculo presenta la enfermedad de tizón tardío, lo cual puede cambiar el proceso de cuidado de cultivo y no solo del agricultor sino de toda la caneda que se beneficia de este producto.

1.5.3. Metodológica

La aplicación del modelo expuesto ayudará a agricultores a clasificar tempranamente el tizón tardío en la papa haciendo su trabajo más rápido y eficiente, ya que una detección temprana promete ser una rápida toma de decisiones que ayudará solucionar y controlar este problema.

En este trabajo, se emplearon técnicas de Visión Computacional que fueron preparadas con un conjunto de datos conformado por diversas bases de datos reales, las cuales fueron recopiladas previamente.

1.6. Delimitación del Estudio

1.6.1. Espacial

Nuestro estudio abarcó proyectos tecnológicos de diversas localidades y naciones, con un fuerte énfasis en Latinoamérica. No obstante, al entrenar el modelo, solo se consideraron descripciones y comentarios en inglés.

1.6.2. Temporal

El periodo de tiempo tomara en cuenta desde el año 2024, fecha en la que se tiene mapeado los conjuntos de datos de plantas de papa con Rancha hasta el mes de diciembre de 2025, el cual se tendra las ultimas imagenes de plantas de papa con esta patología.

1.6.3. Conceptual

Esta investigación se orientará en la implementación de un modelo que logre clasificar si una o varias hojas de papa están infectadas con rancha. Para ello, se valió del uso de herramientas de Vision computacional para desarrollar los modelos de acuerdo a sus modalidades respectivas.

1.6.4. Matriz de Consistencia

A continuación se presenta la matriz de consistencia elaborada para la presente investigación (véase Anexo A.1).

Capítulo 2

MARCO TEÓRICO

2.1. Antecedentes de la investigación

En esta sección se mostrarán diferentes artículos de investigación y tesis que tratan sobre diversas técnicas y enfoques utilizados para enfrentar problemas similares a los abordados en esta tesis. Además, se incluye un cuadro resumen (véase Anexo B.1) con la información presentada en esta sección.

2.1.1. Automated recognition of optical image based potato leaf blight diseases using deep learning (CHAKRABORTY2022101781)

CHAKRABORTY2022101781 realizó un artículo de investigación el cual fue publicado en la revista «Physiological and Molecular Plant Pathology» en el año 2022. Este fue titulado **CHAKRABORTY2022101781** la cual traducida al español significa «Reconocimiento automatizado de enfermedades del tizón de la hoja de la papa basado en imágenes ópticas mediante aprendizaje profundo». La investigación sostiene que el Tizón tardío se presenta como manchas en las hojas de la papa y que para su detección los agricultores sospechan de la patología lo que arriesga los cultivos debido a esta subjetividad y enorme consumo de tiempo. Asimismo, el trabajo explora diferentes recientes modelos de Deep Learning, los cuales los compara con técnicas existentes.

2.1.1.1. Planteamiento del Problema y objetivo

El artículo examina la importancia de la papa como uno de los alimentos más consumidos globalmente, esencial en la dieta diaria de 1.5 mil millones de personas, pero que sufre de diversas enfermedades. Entre las más temidas se encuentra el Tizón tardío, que impacta negativamente en la producción del tubérculo. Por ello, se destaca la necesidad de una detección temprana para aplicar estrategias de manejo eficaces. Sin embargo, los métodos convencionales de diagnóstico, como la inspección visual, son subjetivos y requieren mucho tiempo. Así, el artículo sugiere la necesidad de modelos computacionales avanzados para la clasificación temprana de estas enfermedades, mejorando su gestión y control en la papa. Los objetivos principales son explorar y entrenar modelos de aprendizaje profundo, identificar el modelo de mejor desempeño, optimizar el modelo seleccionado, comparar el modelo propuesto con técnicas existentes y proporcionar una solución práctica y eficiente.

2.1.1.2. Fundamento Teórico usado por el Autor

El autor planteó emplear recientes modelos de Deep Learning, los principales modelos que utilizó fueron VGG16, VGG19, ResNet50 y MobileNet utilizando imágenes ópticas de hojas de papa del conjunto de datos PlantVillage.

2.1.1.3. Metodología empleada por los autores

La metodología empleada por el autor, para la creación de su chatbot consiste en 6 pasos:

1. Se adquirió la base de datos de imágenes del conjunto de datos PlantVillage, del cual solo extrajo imágenes de hoja de papa.
2. Realizaron un preprocesamiento de imágenes donde usaron las técnicas de Cambio de ancho, Cambio de Largo, Zoom y recorte aleatorios, Zoom y recortes rotativos y Giro horizontal
3. Los CNN candidatos fueron VGG16, VGG19, ResNet50 y MobileNet
4. Entrenamiento y testeо de los modelos.
5. Se eligió VGG16 como modelo final
6. Las clases que el modelo clasificaba son, Hoja de papa sana, Hoja de papa con Tizón temprano, Hoja de papa con Tizón tardío

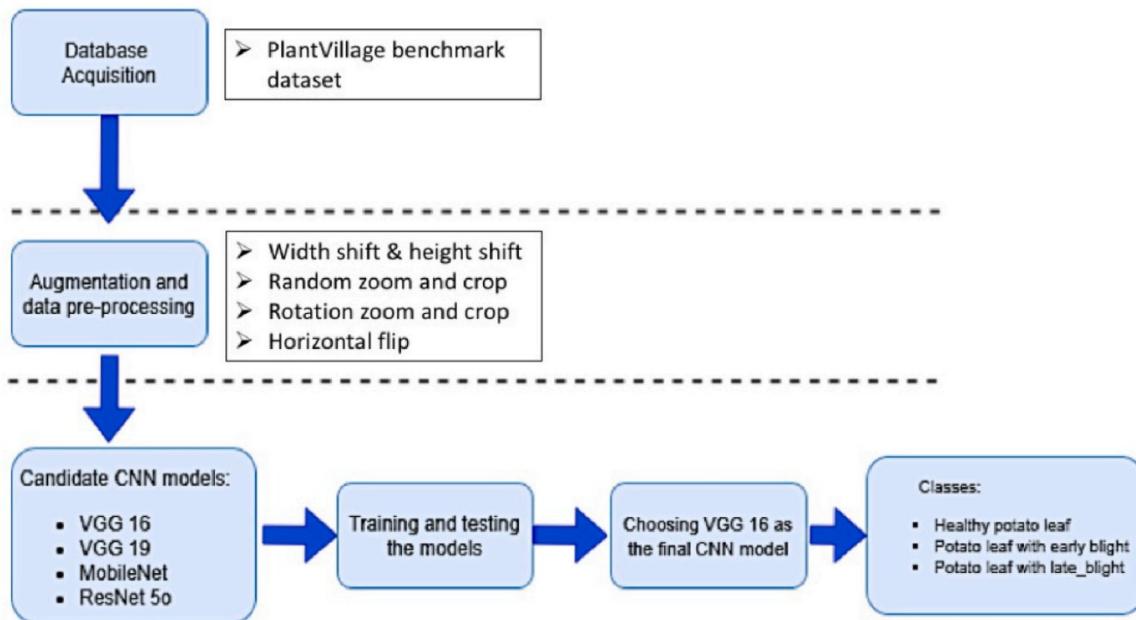


Figura 2.1: Metodología propuesta para el modelo (**CHAKRABORTY2022101781**)

2.1.1.4. Resultados obtenidos

En el paper, cuando se evalúan las distintas técnicas de CNN a considerar, la que sobresale y tiene un mayor desempeño logrando un accuracy promedio del 92.69 %. Es por eso, que se tunea y se escoge como la técnica principal para el desarrollo del modelo final, obteniendo un accuracy promedio del 97.89 %.

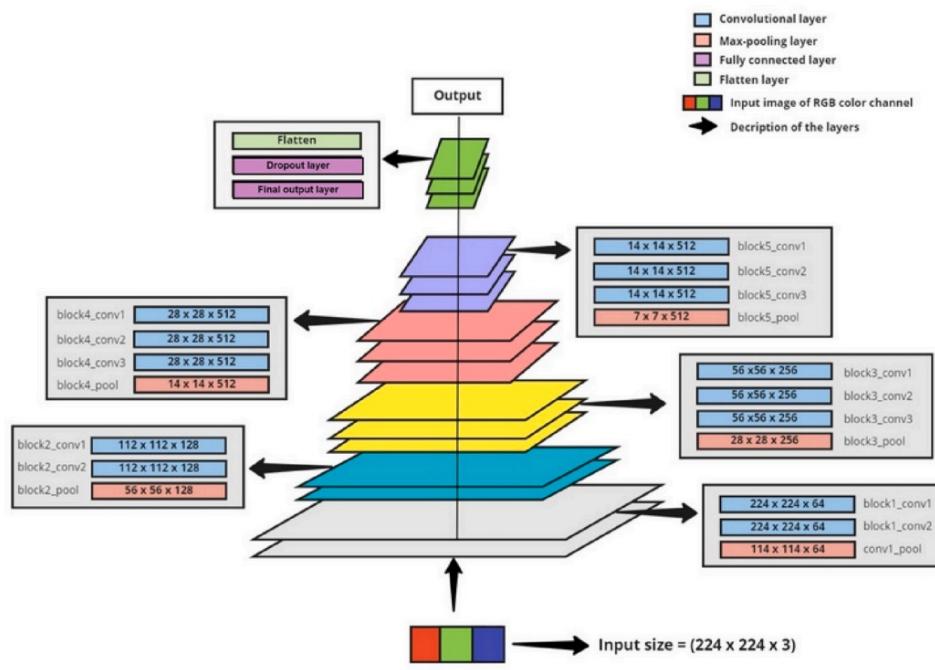


Figura 2.2: Arquitectura del modelo ajustado propuesto de VGG16, muestra las disntitas capas (convolucion). Tambien, se mencionana los respectivos tamaños del filtro convolucional (CHAKRABORTY2022101781)

2.1.2. Research and Validation of Potato Late Blight Detection Method Based on Deep Learning (antecedente2)

antecedente2 realizó este trabajo publicado en la revista Agronomy, para la sección de Precision and Digital Agriculture. Este fue titulado antecedente2 la cual traducida al español significa «Investigación y validación del método de detección del tizón tardío de la papa basado en aprendizaje profundo». La investigación nos dice que el Tizón tardío puede llevar al fracaso total del cultivo papa. Asimismo, construyeron un total de siete categorías de conjuntos de datos de enfermedades de las hojas de papa en fondos simples y complejos. Finalmente, la investigación se introduce en diferentes modelos de Deep Learning en distintas versiones.

2.1.2.1. Planteamiento del Problema y objetivo

El trabajo discute sobre el Tizón tardío como enfermedad muy grave para los cultivos de papa. Esta pone en riesgo el total del cultivo, además que los métodos tradicionales como la detección basada en la inspección visual suele ser subjetivo y demora tiempo. Asimismo, los sistemas de detección actuales suelen ser inefficientes debido a la variabilidad de luminosidad

y el sombreado de las hojas. Es crucial desarrollar un modelo de detección automatizada que pueda superar estas limitaciones, permitiendo una monitorización y prevención temprana del tizón tardío de la papa. El objetivo de este estudio es desarrollar y optimizar un modelo de aprendizaje profundo para la detección del tizón tardío en hojas de papa, que sea altamente preciso y rápido en su inferencia, y que pueda ser implementado en dispositivos móviles para la monitorización automática y la alerta temprana de enfermedades en cultivos. Para alcanzar este objetivo, se plantean las siguientes metas específicas: Lograr una clasificación detallada de enfermedades, Optimizar el modelo base elegido, Evaluar la viabilidad y efectividad del modelo en hardware.

2.1.2.2. Fundamento Teórico usado por el Autor

El autor planteó utilizar modelos de Deep Learning, se enfocó en modelos de ligeros y eficientes estos con el fin de implantar el modelo en un dispositivo móvil. Las técnicas que utilizó fueron MobileNet, ShuffleNet, GhostNet, and SqueezeNet pre-trained models usando imágenes de hojas de papa de los conjuntos de dato Plant Village y AI Challenger 2018.

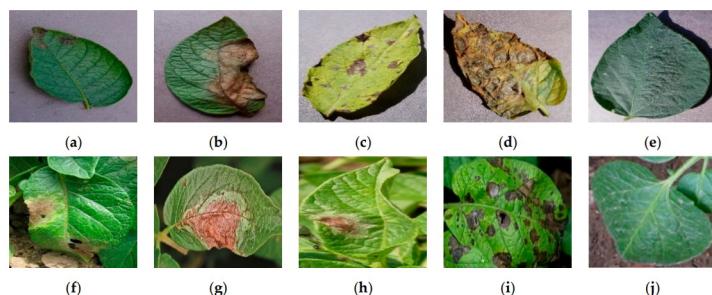


Figura 2.3: (a) Etapas tempranas de la hoja con tizón tardío en un contexto único. (b) Etapas finales de la hoja con tizón tardío en un contexto único. (c) Etapas tempranas de la hoja con tizón temprano en un contexto único. (d) Etapas finales de la hoja con tizón temprano en un contexto único. (e) Hoja sana en un contexto único. (f) Etapas tempranas de la hoja con tizón tardío en un contexto natural. (g) Etapas finales de la hoja con tizón tardío en un contexto natural. (h) Etapas tempranas de la hoja con tizón temprano en un contexto natural. (i) Etapas finales de la hoja con tizón temprano en un contexto natural. (j) Hoja sana en un contexto natural. (antecedente2)

2.1.2.3. Metodología empleada por los autores

La metodología empleada por el autor, para la creación de su modelo final de clasificación es el siguiente:

1. Se adquirió la base de datos de imágenes del conjunto de datos PlantVillage y IA Challenger 2018 dataset, del cual solo extrajo imágenes de hoja de papa.
2. Realizaron un preprocesamiento de imágenes donde usaron técnicas para la Anotación de hoja, Zona de enfermedad, Volteo de imagen, Mejora HSV, Ajuste de brillo, Agregar sombras
3. Los CNN candidatos fueron MobileNet, ShuffleNet, GhostNet y SqueezeNet
4. Entrenamiento y testeо de los modelos.
5. Se eligió ShuffleNetV2 como modelo final
6. Las clases que el modelo clasificaba son, Hoja de papa sana, Tizón tardío nivel 1, Tizón tardío nivel 2 y Tizón tardío nivel 3

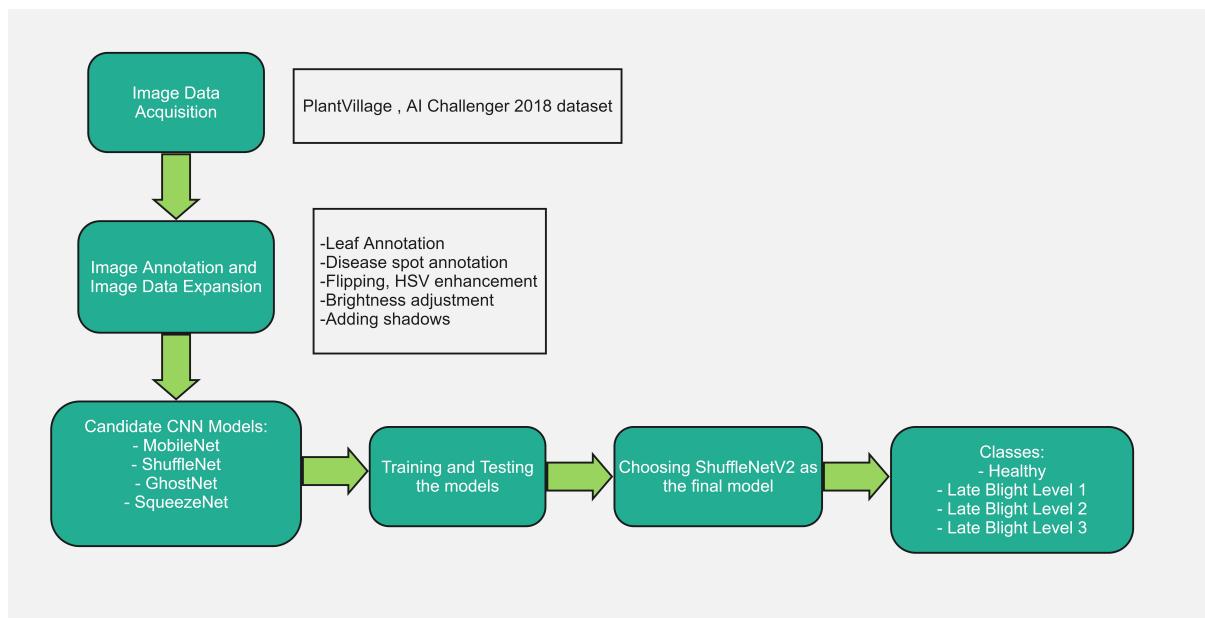


Figura 2.4: Arquitectura del bot de Azure (**antecedente2**)

2.1.2.4. Resultados obtenidos

En esta investigación, se analizaron distintos resultados con diferentes técnicas. Con la técnica final, ShuffleNetV2, se obtuvieron resultados un accuracy de 95.41 %. Finalmente, se consideró una mejoría para esta técnica, ya que se buscaba el mejor modelo final posible para un dispositivo móvil disminuyendo el tiempo de procesamiento y el costo computacional, tomando esos objetivos en cuenta, se logró un 95.04 %

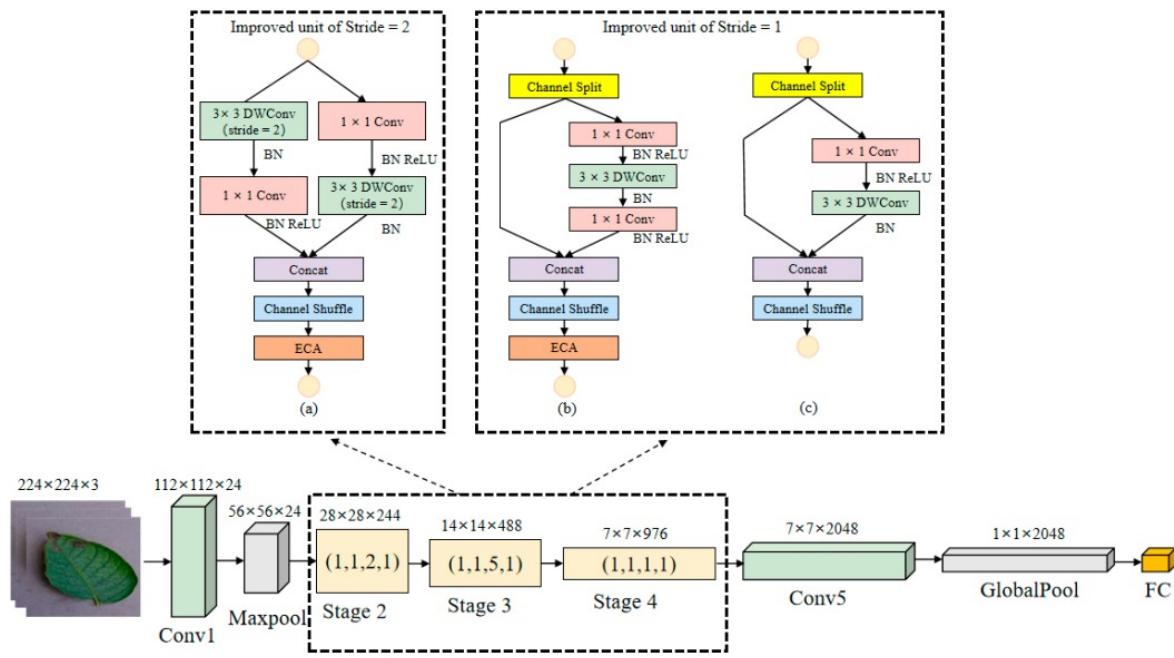


Figura 2.5: Arquitectura del modelo final. La estructura del modelo mejorado ShuffleNetV2 2x. Nota: La Etapa 2, la Etapa 3 y la Etapa 4 están compuestas por la unidad base original, con Stride = 1, y la unidad modificada está compuesta por una unidad con Stride = 1 y una unidad con Stride = 2. (1, 1, 2, 1) en la Etapa 2 indica una pila de la unidad (a) con Stride = 2 mejorado, una pila de la unidad (b) con Stride = 1 mejorado, dos pilas de la unidad básica con Stride = 1 original, y una pila de la unidad (c) con Stride = 1 mejorado; (1, 1, 5, 1) en la Etapa 3 indica una pila de la unidad (a) con Stride = 2 mejorado, una pila de la unidad (b) con Stride = 1 mejorado, cinco pilas de la unidad básica con Stride = 1 original, y una pila de la unidad (c) con Stride = 1 mejorado; (1, 1, 1, 1) en la Etapa 4 indica una pila de la unidad (a) con Stride = 2 mejorado, una pila de la unidad básica con Stride = 1 original, y una pila de la unidad (c) con Stride = 1 mejorado. ([antecedente2](#))

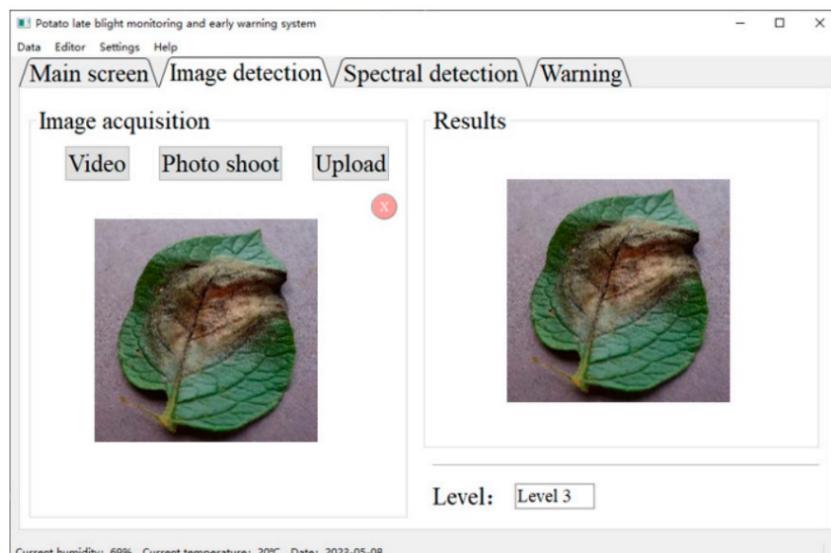


Figura 2.6: Interfaz de operación de detección de imágenes. ([antecedente2](#))

2.1.3. Supervised Learning-Based Image Classification for the Detection of Late Blight in Potato Crop (antecedente3)

antecedente3 realizó un trabajo con el fin de ser publicado en la revista Computer Vision and Pattern Recognition Based on Deep Learning. Este fue titulado **antecedente3** la cual traducida al español significa «Clasificación de imágenes basada en aprendizaje supervisado para la detección del tizón tardío en cultivos de papa». La investigación plantea el desarrollo de un modelo de detección temprana del Tizón tardío en la papa utilizando Redes Neuronales Convolucionales y Máquinas de Vectores de Soporte. Asimismo, el autor desarrolló, por sí mismo, su base de datos de imágenes de cultivos de papa. Finalmente, se aplicaron varias métricas de rendimiento, eficiencia y calidad en las tareas de aprendizaje y clasificación para determinar los mejores algoritmos de aprendizaje automático.

2.1.3.1. Planteamiento del Problema y objetivo

La investigación aborda la necesidad de encontrar métodos más eficaces en cuanto a la detección temprana del Tizón Tardío causada por Oomiceto *Phytophthora*, que afecta rápidamente las hojas, tallos y tubérculos de la papa que es un alimento crucial en la economía de Bógora. Asimismo, la detección tradicional, como pruebas de laboratorio, aunque es eficaz, toma mucho tiempo y es altamente costoso. En conclusión, el objetivo principal del estudio es aplicar técnicas de aprendizaje supervisado y clasificación de imágenes, específicamente mediante redes neuronales convolucionales (CNN) y máquinas de vectores de soporte (SVM), para la detección temprana del tizón tardío en cultivos de papa.

2.1.3.2. Fundamento Teórico usado por el Autor

El autor plantea hacer uso de CNN y SVM como sus posibles modelos finales. Utiliza una CNN, según el preprocesamiento de la data, no modificada y una aumentada, por otro lado, divide SVM en cuatro posibles modelos según característica del preprocesamiento: Color, Textura, PCA, Combined.

2.1.3.3. Metodología empleada por los autores

La metodología empleada por el autor, para la creación de su modelo consiste en los siguientes pasos:

1. Captura los datos imputados por parte del cliente.

2. Envía el query a Dialogflow para su procesamiento.
3. Se procesa la información mediante una API externa y el código implementado, así como la base de datos.
4. Retorna los resultados procesados y se agrega información extra requerida.
5. Envía al cliente o usuario la respuesta.

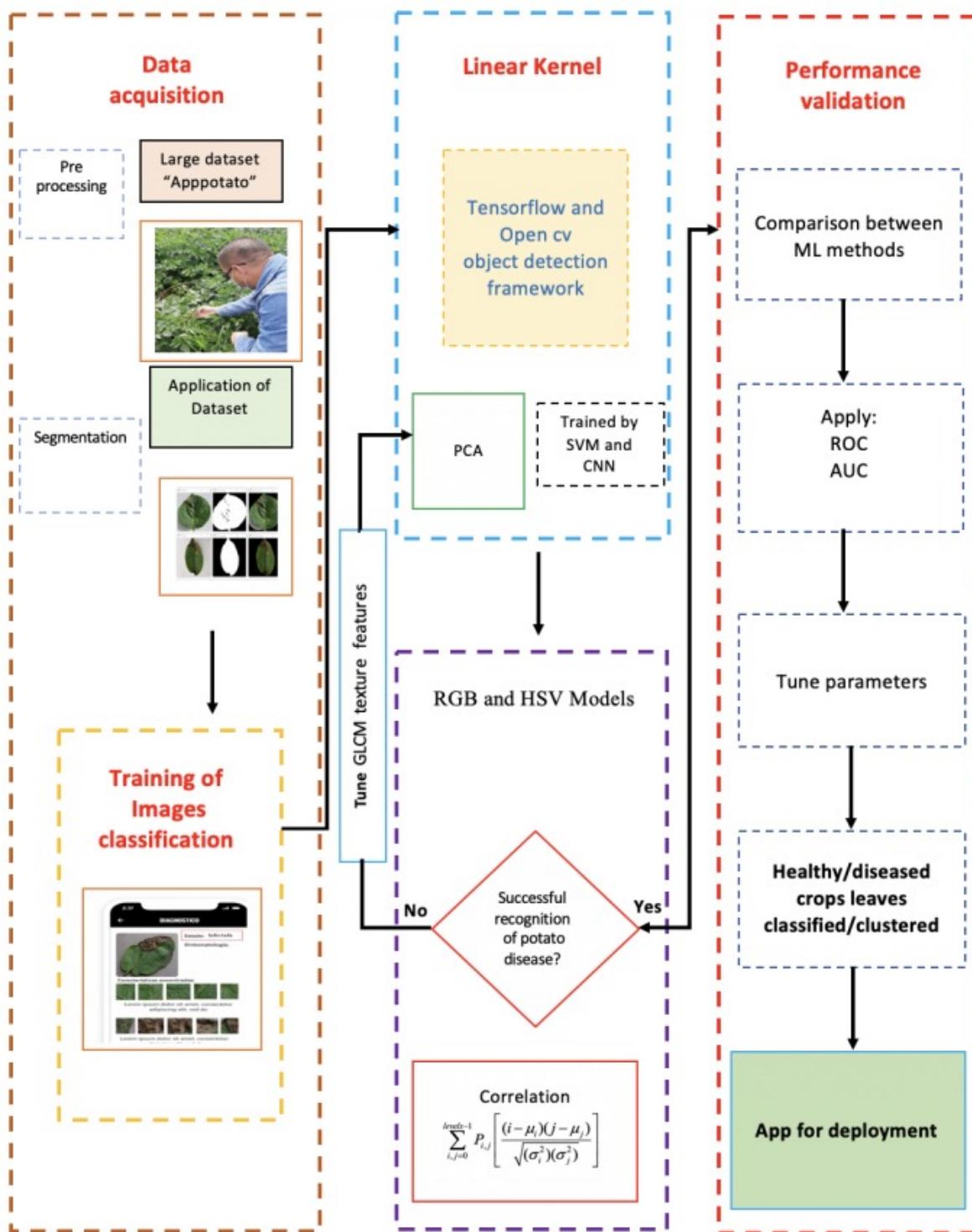


Figura 2.7: Diagrama de la metodología Flowchart(**antecedente3**)

2.1.3.4. Resultados obtenidos

Las CNN entrenadas con el conjunto de datos aumentado mostraron el mejor desempeño con una precisión del 93 % y una AUC de 0.97. Además, las SVM entrenadas con características de color obtuvieron mejores resultados en comparación con las SVM entrenadas con otras características. Finalmente, se propone desarrollar una aplicación móvil con características avanzadas para la agricultura de precisión que ayude a los agricultores a identificar la enfermedad del tizón tardío de manera no invasiva y en tiempo real..

2.1.4. Potato Blight Classification Android Application using Deep Learning (antecedente5)

antecedente5 realizó un artículo publicado en base de datos de ELSEVIER, siendo también parte de la revista Sustainable Chemistry and Pharmacy en el año 2022. Este fue titulado **antecedente5** la cual traducida al español significa «Redes neuronales convolucionales profundas para la detección de enfermedades de la hoja del tomate basada en imágenes». El trabajo nos dice que el reconocimiento de enfermedades foliares en las plantas representa un riesgo significativo para la seguridad alimentaria, ya que puede reducir la producción agrícola y, por ende, la economía nacional. Es crucial identificar estas enfermedades en etapas tempranas para mejorar la calidad y cantidad de los productos agrícolas. Por lo tanto, se requiere un sistema automático de reconocimiento de enfermedades foliares que pueda identificar y clasificar estas enfermedades en etapas tempranas. En este contexto, se han utilizado modelos de redes neuronales convolucionales profundas (DCNN) para el análisis de imágenes de hojas, con el objetivo de mejorar la precisión y reducir el tiempo de respuesta en la identificación de enfermedades foliares en tomates. Se propone un sistema automático de identificación de enfermedades foliares en tomates utilizando DCNN, con un conjunto de datos de 18160 imágenes de hojas de tomate. Este conjunto de datos se dividió en un 60 % para entrenamiento y un 40 % para pruebas, logrando una precisión del 98.40 % en el conjunto de pruebas con el modelo DCNN propuesto.

2.1.4.1. Planteamiento del Problema y objetivo

El artículo aborda principalmente que las enfermedades en las hojas de tomate causan pérdidas significativas en la producción, afectando tanto la calidad como la cantidad de los productos. Identificar y diagnosticar estas enfermedades de manera temprana es crucial, ya que pueden reducir drásticamente el crecimiento de los cultivos y, por lo tanto, la producción. Sin

embargo, el diagnóstico manual de las enfermedades foliares puede llevar a una disminución en la producción debido a la gravedad de las enfermedades y a la variabilidad en los síntomas causada por factores ambientales como la temperatura, el viento y la humedad. Por lo tanto, existe una necesidad de desarrollar un sistema automático que pueda identificar y diagnosticar estas enfermedades en etapas tempranas, permitiendo a los agricultores tomar medidas preventivas adecuadas para proteger sus cultivos. El objetivo es desarrollar una herramienta automática que diagnostique las enfermedades de las hojas de tomate tempranamente para mejorar la producción agrícola. Se utilizará un enfoque basado en redes neuronales convolucionales profundas (DCNN) para clasificar 10 tipos de enfermedades en los cultivos de tomate, con el fin de identificar los síntomas de las hojas en etapas tempranas y mejorar la eficiencia y precisión del modelo mediante técnicas de ajuste de parámetros.

2.1.4.2. Fundamento Teórico usado por el Autor

Los autores plantean utilizar como técnica principal una DCNN (Deep Convolutional Neural Networks) y compararla con técnicas como MLP (Multiplayer Layer Perceptron) y SVM (Support Vector Machine).

2.1.4.3. Metodología empleada por los autores

La metodología empleada por los autores, para la creación de su chatbot consiste en los siguientes pasos:

1. Se adquirió la base de datos de imágenes del conjunto de datos PlantVillage
2. Realizaron un preprocessamiento de imágenes donde se ajustó el tamaño de imágenes.
3. El modelo utilizó una DCNN
4. Entrenamiento y testeо de los modelos.
5. Se reajustó el modelo
6. Finalmente se compararon los resultados del modelo final con MLP y SVM

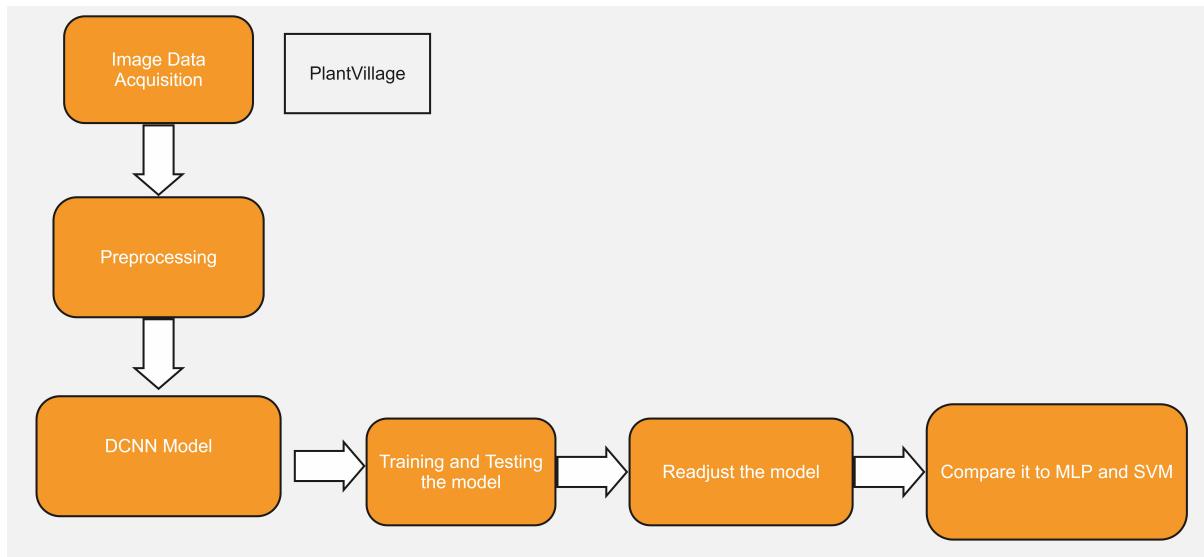


Figura 2.8: Metodología (antecedente5)

2.1.4.4. Resultados obtenidos

Como resultado del desarrollo del modelo final, este alcanzó un 98.40 % de accuracy promedio, mientras que SVM obtuvo un accuracy de 90.01 % y MLP un accuracy de 88.30 %.

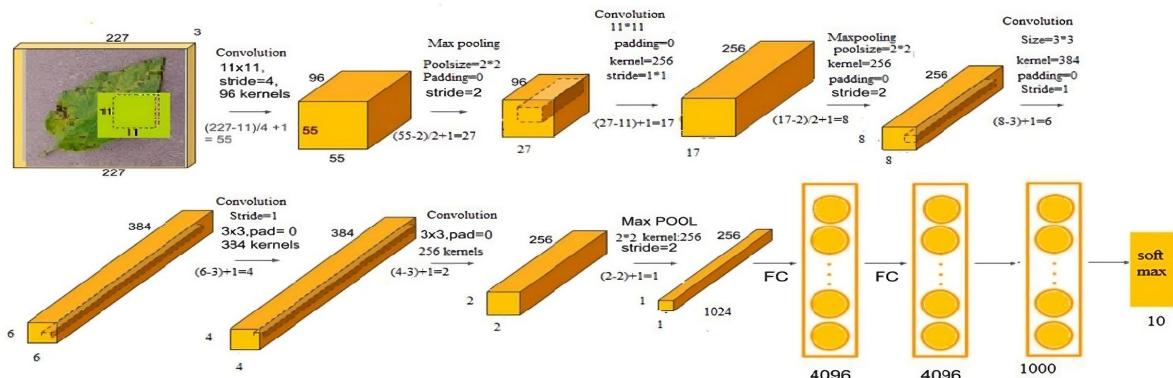


Figura 2.9: Arquitectura del modelo final (antecedente5)

2.1.5. Deep Convolutional Neural Networks for image based tomato leaf disease detection (antecedente6)

antecedente6 realizó un artículo publicado en base de datos de International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) en el año 2023. Este fue titulado **antecedente6** la cual traducida al español significa «Aplicación para Android

de clasificación del tizón de la papa usando el aprendizaje profundo». La investigación sostiene que los agricultores de papas sufren pérdidas económicas debido a enfermedades como el tizón temprano y tardío. La detección temprana y tratamiento adecuado pueden prevenir estas pérdidas, pero los métodos tradicionales son lentos y propensos a errores. Proponemos usar una Red Neuronal Convolutacional (CNN) personalizada para diagnosticar enfermedades de plantas de manera rápida y precisa, reduciendo el tiempo de computación y minimizando errores, lo que ayuda a los agricultores a tratar las enfermedades a tiempo y reducir pérdidas.

2.1.5.1. Planteamiento del Problema y objetivo

El trabajo sustenta que la industria de la papa enfrenta un desafío significativo en la prevención de pérdidas de cultivos debido a enfermedades como el tizón temprano y el tizón tardío. Estas enfermedades pueden causar pérdidas económicas sustanciales para los agricultores si no se detectan y tratan a tiempo. Los métodos tradicionales de inspección visual son lentos, propensos a errores y no son viables para aplicaciones en tiempo real debido a los largos tiempos de procesamiento de imágenes. Además, existe la necesidad de mejorar la precisión y reducir el tiempo de computación en los métodos actuales de diagnóstico de enfermedades de las plantas. Por eso, su objetivo es desarrollar una red neuronal convolucional (CNN) eficiente y precisa para detectar enfermedades en las plantas de papa, reduciendo el tiempo de computación y mejorando la precisión, con el fin de ayudar a los agricultores a tratar las enfermedades a tiempo y reducir pérdidas económicas.

2.1.5.2. Fundamento Teórico usado por el Autor

Los autores formulan utilizar como técnica principal un Red Neuronal Convolutacional, mejorarla para que sea una Red neuronal convolucional personalizada de enfermedades profundas (PDDCNN), ya que buscan lograr un modelo que se adapte a distintas regiones de donde se pueda obtener el dataset.

2.1.5.3. Metodología empleada por los autores

La metodología empleada por los autores, para la creación de su chatbot consiste en los siguientes pasos:

1. Se adquirió la base de datos de imágenes del conjunto de datos PlantVillage. Asimismo, añadieron una base de datos de imágenes propias tomadas con cámaras digitales

2. Realizaron un preprocessamiento de imágenes donde se ajusto el tamaño de imágenes segun el dataset..
3. El modelo utilizó una CNN
4. Entrenamiento y testeo de los modelos.
5. Se reajusto el modelo a una PDDCNN
6. Finalmente, el modelo final detecto las clases Tizon Temprao, Tizon Tardio, Hoja Saludable

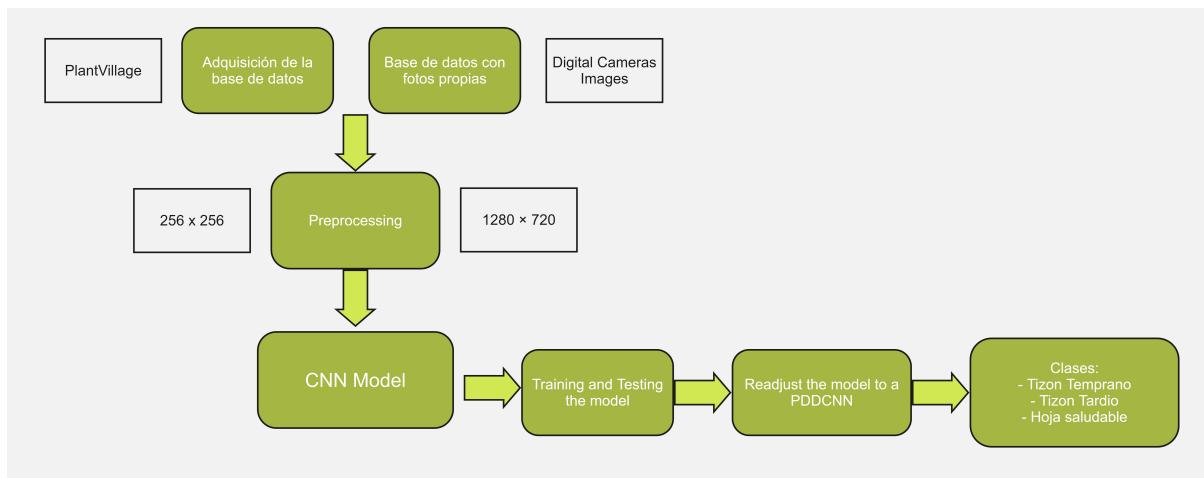


Figura 2.10: Metodología (antecedente6)

2.1.5.4. Resultados obtenidos

Los resultados finales fueron tomados sobre la base de datos que ellos crearon después de entrenar su modelo con el conjunto de datos PlantVillage. Obtuvo un accuracy de 99 % en el tizón tardío y tizón temprano, mientras que en las hojas sanas tuvieron un accuracy de 100 %

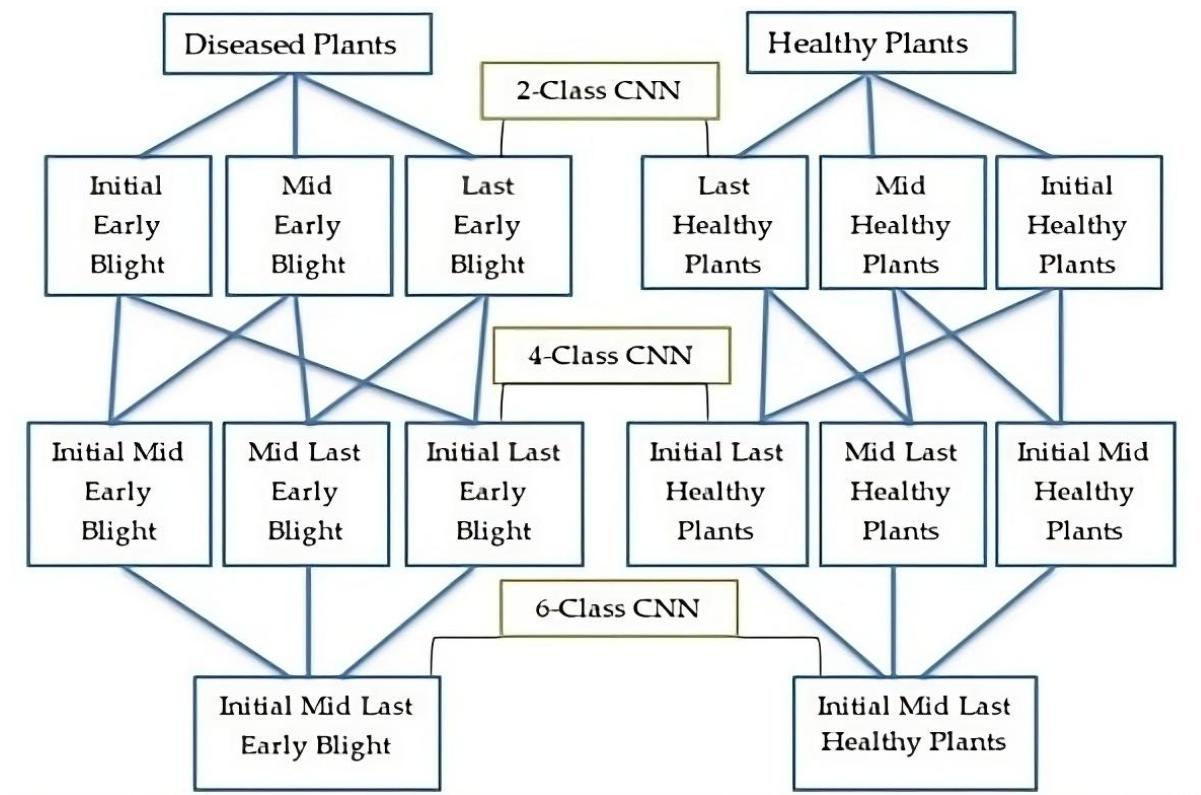


Figura 2.11: Arquitectura del modelo final (anteceidente6)

2.1.6. Potato Blight Classification Android Application using Deep Learning (anteceidente7)

anteceidente7 realizó un artículo publicado en base de datos de ELSEVIER, siendo también parte de la revista Sustainable Chemistry and Pharmacy en el año 2022. Este fue titulado **anteceidente7** la cual traducida al español significa «Redes neuronales convolucionales profundas para la detección de enfermedades de la hoja del tomate basada en imágenes». El trabajo nos dice que el reconocimiento de enfermedades foliares en las plantas representa un riesgo significativo para la seguridad alimentaria, ya que puede reducir la producción agrícola y, por ende, la economía nacional. Es crucial identificar estas enfermedades en etapas tempranas para mejorar la calidad y cantidad de los productos agrícolas. Por lo tanto, se requiere un sistema automático de reconocimiento de enfermedades foliares que pueda identificar y clasificar estas enfermedades en etapas tempranas. En este contexto, se han utilizado modelos de redes neuronales convolucionales profundas (DCNN) para el análisis de imágenes de hojas, con el objetivo de aumentar la exactitud y disminuir el tiempo de respuesta. Identificación de enfermedades foliares en tomates. Se propone un sistema automático de identificación de enfermedades foliares en tomates utilizando DCNN, con un conjunto de datos de 18160 imágenes de hojas

de tomate. Este conjunto de datos se dividió en un 60% para entrenamiento y un 40% para pruebas, logrando una precisión del 98.40% en el conjunto de pruebas con el modelo DCNN propuesto.

2.1.6.1. Planteamiento del Problema y objetivo

El artículo aborda principalmente que las enfermedades en las hojas de tomate causan pérdidas significativas en la producción, afectando tanto la calidad como la cantidad de los productos. Identificar y diagnosticar estas enfermedades de manera temprana es crucial, ya que pueden reducir drásticamente el crecimiento de los cultivos y, por lo tanto, la producción. Sin embargo, el diagnóstico manual de las enfermedades foliares puede llevar a una disminución en la producción debido a la gravedad de las enfermedades y a la variabilidad en los síntomas causada por factores ambientales como la temperatura, el viento y la humedad. Por lo tanto, existe una necesidad de desarrollar un sistema automático que pueda identificar y diagnosticar estas enfermedades en etapas tempranas, permitiendo a los agricultores tomar medidas preventivas adecuadas para proteger sus cultivos. El objetivo es desarrollar una herramienta automática que diagnostique las enfermedades de las hojas de tomate tempranamente para mejorar la producción agrícola. Se utilizará un enfoque basado en redes neuronales convolucionales profundas (DCNN) para clasificar 10 tipos de enfermedades en los cultivos de tomate, con el fin de identificar los síntomas de las hojas en etapas tempranas y mejorar la eficiencia y precisión del modelo mediante técnicas de ajuste de parámetros.

2.1.6.2. Fundamento Teórico usado por el Autor

Los autores plantean utilizar como técnica principal una DCNN (Deep Convolutional Neural Networks) y compararla con técnicas como MLP (Multiplayer Layer Perceptron) y SVM (Support Vector Machine).

2.1.6.3. Metodología empleada por los autores

La metodología empleada por los autores, para la creación de su modelo consiste en los siguientes pasos:

1. Se adquirió la base de datos de imágenes del conjunto de datos PlantVillage.
2. Realizaron un preprocessamiento de imágenes donde se hizo el labeling manualmente.
3. Se extrajo características de las imágenes usando VGG19

4. Los candidatos de CNN para esta investigación fueron VGG16, VGG19 e Inception v3
5. Se hizo el testeo y entrenamiento del modelo
6. Finalmente, se evaluó el modelo con las siguientes métricas: AUC, CA, F1, Precision y Recall

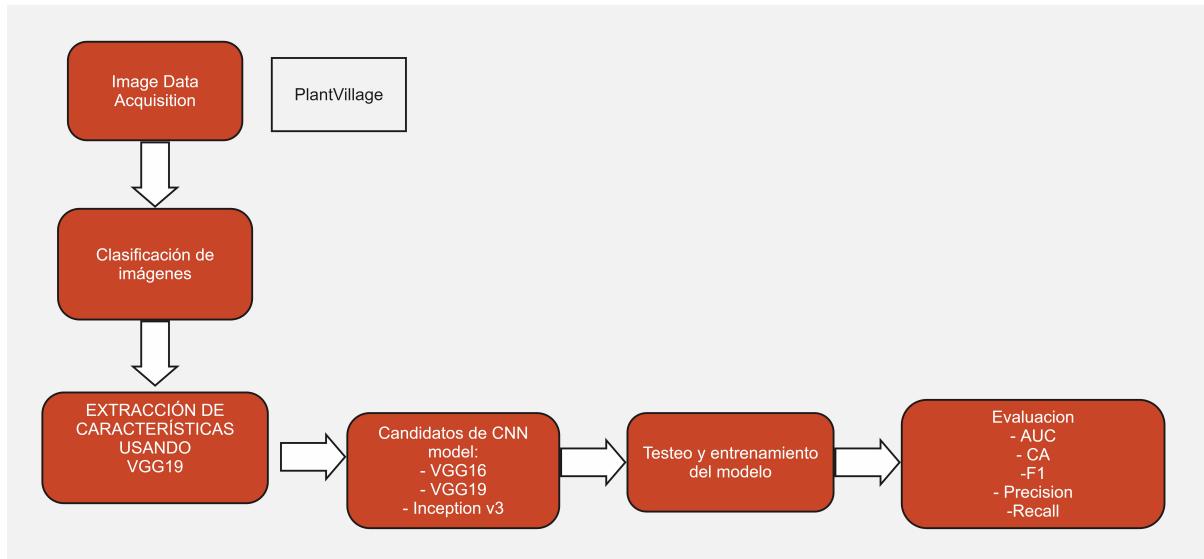


Figura 2.12: Metodología (antecedente7)

2.1.6.4. Resultados obtenidos

Para escoger su modelo consideraron VGG19 con la técnica de clasificación Logistic Regresion. Este alcanzó 97.8 % de accuracy sobre el testeo del dataset.

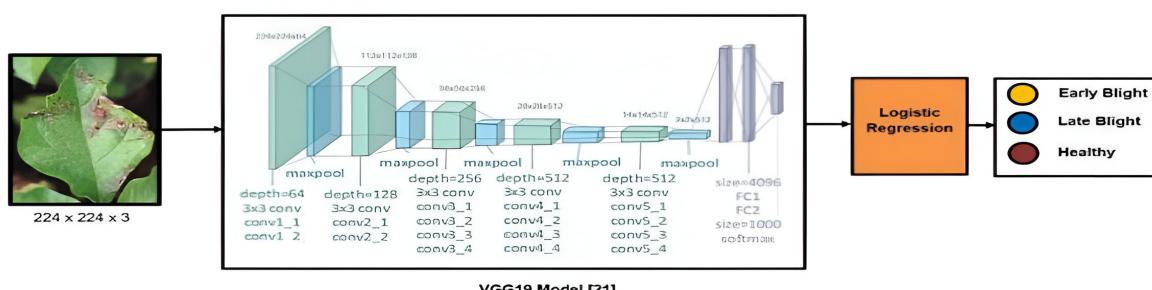


Figura 2.13: Arquitectura del modelo final (antecedente7)

2.1.7. Investigation of Phytophthora Infestans Causing Potato Late Blight Disease: A Review (antecedente4)

antecedente4 realizó un trabajo con el fin de ser publicado en la revista Biomedicine and Chemical Sciences. Este fue titulado **antecedente4** la cual traducida al español significa «Investigación de Phytophthora Infestans que causa el tizón tardío de la papa». La investigación sostiene Phytophthora infestans causa el tizón tardío de la papa, infectando raíces, tubérculos y brotes. La propagación se debe al cultivo de tubérculos infectados y restos de plantas en el campo. Las estructuras como micelio, zoosporas, oosporas y esporangios pueden causar infección, y las oosporas pueden sobrevivir de 3 a 4 años en bajas temperaturas. *P. infestans* puede causar pérdidas de hasta el 100% en condiciones óptimas. Hay dos patrones de apareamiento, A1 y A2, y varios patrones genéticos que complican el control de la enfermedad. Las estrategias de control incluyen químicos, rotación de cultivos, agentes biológicos y plantas resistentes, siendo más efectivo combinar plantas resistentes y fungicidas. El artículo analiza los factores de propagación y desafíos del tizón tardío.

2.1.7.1. Planteamiento del Problema y objetivo

La investigación aborda que el hongo *Phytophthora* sp, con más de 60-80 especies, infecta diversas plantas, incluidas las papas y los tomates. *Phytophthora infestans*, en particular, causa estragos en estos cultivos, provocando pérdidas significativas de rendimiento y desencadenando eventos históricos como la hambruna irlandesa en la década de 1840. A pesar de los esfuerzos de control, la enfermedad del tizón tardío sigue siendo una amenaza global para la producción de papas, con pérdidas que pueden variar del 50 al 100% dependiendo de diversos factores ambientales y de gestión. El objetivo principal es investigar *Phytophthora infestans* y su impacto en la enfermedad del tizón tardío en las papas. Se busca comprender mejor los modos de infección, las estrategias de reproducción del patógeno y los factores que contribuyen a su agresividad. A través de esta investigación, se espera identificar mejores métodos de control y gestión de la enfermedad, incluyendo el uso de fungicidas, prácticas culturales y la resistencia de las plantas hospederas, para mitigar las pérdidas económicas y asegurar la seguridad alimentaria en las regiones afectadas.

2.1.7.2. Fundamento Teórico usado por el Autor

El autor desarrolla su investigación en base a múltiples investigaciones pasadas. Desde los orígenes del Tizón tardío, hasta su condición en el mundo actual.

2.1.7.3. Metodología empleada por los autores

La metodología empleada por el autor, para la creación de su chatbot consiste en los siguientes pasos:

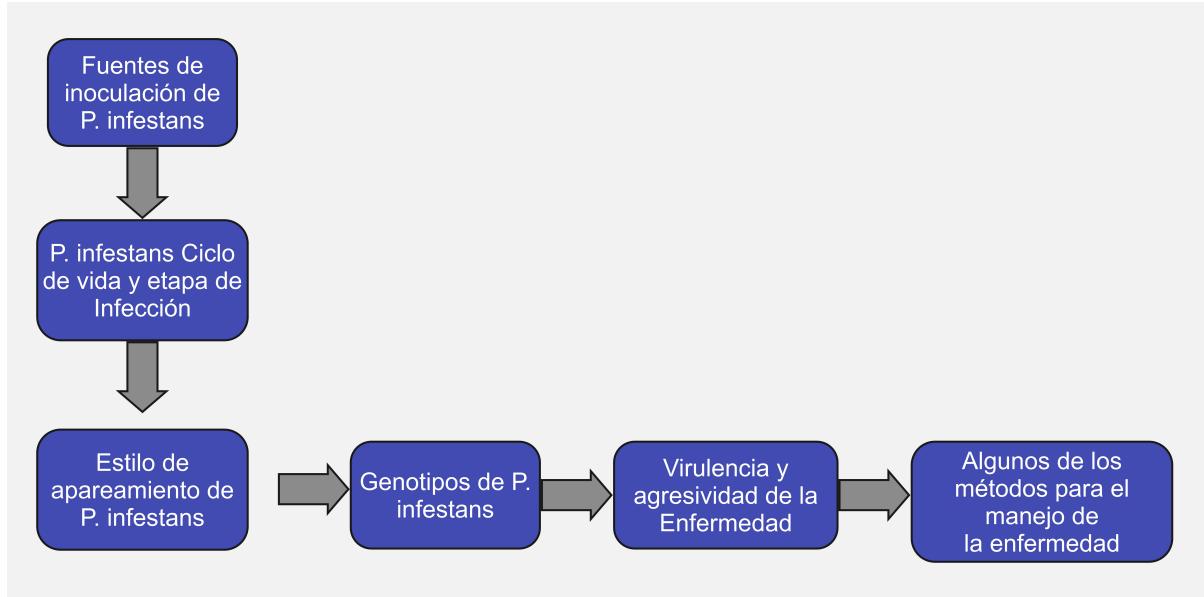


Figura 2.14: Diagrama de la metodología(**antecedente4**)

2.1.7.4. Resultados obtenidos

Los aislamientos de *P. infestans* tienen varias estructuras de infección que pueden infectar diferentes partes de la planta de papa y pueden sobrevivir durante un largo período. El hongo es capaz de reproducirse tanto sexual como asexualmente, y también presenta patrones de apareamiento A1 y A2 y diferentes genotipos. Todas estas características han llevado a la aparición de varios desafíos en el estudio de la virulencia y la agresividad. El segundo y más importante desafío incluye cómo elegir el mejor método para controlar la enfermedad (tizón tardío), ya que aparecen aislamientos que pueden resistir diferentes fungicidas, también pueden infectar plantas resistentes y muchas razones hacen que este desafío sea más relevante.

2.2. Bases Teóricas

2.2.1. Inteligencia Artificial

Durante la conferencia de Dartmouth en 1956, el informático John McCarthy presentó como primera persona el término "Inteligencia Artificial".^a al mundo. McCarthy basó su concepto en los fundamentos teóricos publicados por Turing en 1950, donde se planteaba la posibilidad de que las máquinas pudieran pensar. En este evento, diversos investigadores y científicos expusieron las metas y la visión de la IA. Esta conferencia es abiertamente tomada como el inicio de la inteligencia artificial según se sabe en la actualidad. (**teamredac2022**). Asimismo, hoy en día, no tenemos un concepto exacto acerca de la inteligencia artificial, debido a que es un tema complejo, por lo tanto, es posible hallar distintos conceptos acerca de ella. No obstante, la definición que se usará para términos de la investigación es que la inteligencia artificial es el poder de un ordenador de utilizar algoritmos, recibir datos, procesarlos y en base a esos pasos ser capaz de tomar decisiones similares a las de un ser humano. A diferencia de los humanos, la IA a través del procesamiento de conjuntos de información es capaz de crear máquinas y sistemas para resolver problemas que usualmente necesitan de inteligencia humana para resolverse. Muchos de los algoritmos de la IA se entrena constituyéndose de datos para mejorar su rendimiento y optimizar las reglas establecidas, lo que se conoce como aprendizaje profundo (**rouhiainen2018inteligenciaricardo2021inteligenciacajahuanca2021inteligencia**). Por otro lado, hoy podemos encontrar diferentes tipos de IA, cada una con sus diferentes propósitos y características, de las más conocidas y aplicables en distintos campos académicos y profesionales, como la agricultura e ingeniería, son Deep learning, Vision Computer, este mismo contempla técnicas como Vision Transformer, Redes Neuronales Convolucionales, You Only Look One y técnicas para clasificación y regresión como Support Vector Machine.

2.2.2. Deep Learning

El Aprendizaje profundo, que particularmente se usa en los contextos donde la data es compleja y donde hay enormes cantidades de datos disponibles. Este subcampo de la inteligencia artificial se desarrolla mediante el uso de redes neuronales, las cuales se estructuran en niveles de procesamiento para identificar patrones y estructuras en conjuntos de datos extensos. Cada capa aprende un concepto de los datos sobre el que se basan las capas siguientes; cuanto más alto el nivel (capa), más abstractos son los conceptos aprendidos. El aprendizaje profundo no requiere un procesamiento previo de los datos y es capaz de extraer características de forma automática. Por poner una utilización sencilla, una red neuronal encargada de descifrar

figuras aprendería a reconocer bordes simples en la primera capa y luego añadiría el reconocimiento de las figuras más complejas compuestas por esos bordes en las capas siguientes. No hay una regla fija sobre cuántas capas son necesarias para constituir una red neuronal profunda (**rusk2016deeprouhiainen2018inteligencia**). Hoy en día muchas de las compañías online y grandes consumidoras de tecnología usan Inteligencia profunda. Por citar a una, Facebook usa esta tecnología para analizar los textos de las conversaciones. Otras compañías como Google, Baidu, y Microsoft usan inteligencia profunda para búsqueda de imágenes, y también traslación de máquinas. Todos los teléfonos inteligentes poseen sistemas de inteligencia profunda corriendo en ellos. Ahora, inteligencia profunda es el estándar para tecnología para reconocimiento del habla, y también para la detección de rostros por cámaras digitales. La inteligencia profunda, también es el centro de los autos que se manejan por sí mismos, donde es usual la localización y el mapeo, la percepción del entorno, planificación y dirección del movimiento, así como el seguimiento del estado del conductor. La inteligencia profunda está revolucionando la agricultura al proporcionar herramientas avanzadas para la clasificación y el análisis, lo que resulta en una mayor eficiencia, precisión y sostenibilidad en las prácticas agrícolas (**kelleher2019deep**).

2.2.3. Support Vector Machine: A comprehensive survey on support vector machine classification: Applications, challenges and trends (tecnica3)

En los últimos años, se ha realizado una gran cantidad de investigación sobre las máquinas de soporte vectorial (SVM) y sus aplicaciones en varios campos de la ciencia. Las SVM son algoritmos de clasificación y regresión muy poderosos y robustos, especialmente en el reconocimiento de patrones. Aunque en algunos campos las SVM no funcionan bien, se han desarrollado aplicaciones para grandes conjuntos de datos, clasificación múltiple y datos desbalanceados. Además, las SVM se han integrado con métodos avanzados para mejorar su capacidad de clasificación y optimización de parámetros. Este artículo ofrece una introducción a las SVM, describe sus aplicaciones, resume los desafíos y tendencias, identifica limitaciones y discute su futuro y posibles nuevas aplicaciones.

2.2.3.1. Introducción

El aprendizaje automático es un campo multidisciplinario que integra conceptos de la ciencia cognitiva, la informática, la estadística y la optimización, entre otros.. En este campo, la clasificación es un enfoque supervisado que analiza un conjunto de datos y construye un modelo para separar los datos en clases distintas. Existen diversas técnicas de clasificación como el redes neuronales artificiales, k-vecinos más cercanos,, árboles de decisión. redes bayesianas, y

SVM.

- **k-vecinos más cercanos:** Fácil de implementar, pero lento con conjuntos de datos grandes y sensible a parámetros irrelevantes.
- **Árboles de decisión:** Rápidos en la fase de entrenamiento, pero menos flexibles para modelar parámetros.
- **Redes neuronales:** Ampliamente utilizadas y universales, pero sensibles al ruido en los datos de entrenamiento y requieren considerar muchos factores al construirlas.

De estas técnicas, las SVM son conocidas por su capacidad de optimización y generalización. Introducidas por Vapnik, las SVM son modelos de aprendizaje automático basados en kernel para tareas de clasificación y regresión. Las SVM se destacan por su capacidad discriminativa y han demostrado ser superiores a otros métodos de aprendizaje supervisado, convirtiéndose en una de las técnicas de clasificación más usadas.

Las funciones de decisión en SVM se determinan directamente a partir de los datos de entrenamiento, maximizando la separación entre los bordes de decisión en un espacio de características de alta dimensión, lo que minimiza los errores de clasificación y mejora la capacidad de generalización. Una ventaja notable de las SVM es que obtienen un subconjunto de vectores de soporte durante la fase de aprendizaje, lo cual representa una tarea de clasificación y suele ser una pequeña parte del conjunto de datos original.

El documento se organiza en secciones que presentan las bases teóricas, características, ventajas y desventajas de las SVM, sus debilidades, implementaciones y aplicaciones en problemas del mundo real, finalizando con tendencias y desafíos futuros.

2.2.3.2. Bases teóricas de SVM

El objetivo principal en la ordenación de patrones es obtener una técnica la cual maximice el rendimiento en los datos de preparación. Las maneras de preparación tradicionales determinan las técnicas de tal fin que cada pareja entrada-salida se clasifique bien en la clase que debe estar. Por otro lado, si el clasificador se ajusta demasiado a los datos de preparación, comienza a memorizar los datos en lugar de aprender a generalizar, degradando su capacidad de generalización. Preparar una SVM requiere un grupo de n ejemplos, cada uno tiene que ver con ser un par, un vector de entrada x_i y la etiqueta asociada y_i . Supongamos que se da un grupo de entrenamiento X como:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (\text{Ecuación 2.1})$$

Dado un conjunto $X = \{(x_i, y_i)\}_{i=1}^n$ donde $x_i \in \mathbb{R}^d$ y $y_i \in \{+1, -1\}$, consideremos el caso de una entrada bidimensional, es decir, $x \in \mathbb{R}^2$. Los datos son linealmente separables y existen numerosos hiperplanos que pueden realizar esta tarea. En la Figura 1 se presentan varios hiperplanos que dividen el conjunto de datos de entrada de manera perfecta. Es evidente que hay una cantidad infinita de hiperplanos capaces de lograr esto. No obstante, la capacidad de generalización depende de la posición del hiperplano de separación óptimo asociado con el margen máximo. El plano de decisión, o el hiperplano que divide el espacio de entrada, se define mediante la ecuación central. $w^T x_i + b = 0$.

El caso más sencillo de SVM es el caso linealmente separable en el espacio de características. Optimizamos el margen geométrico configurando el margen funcional $\kappa_i = 1$ (también conocido como Hiperplano Canónico).

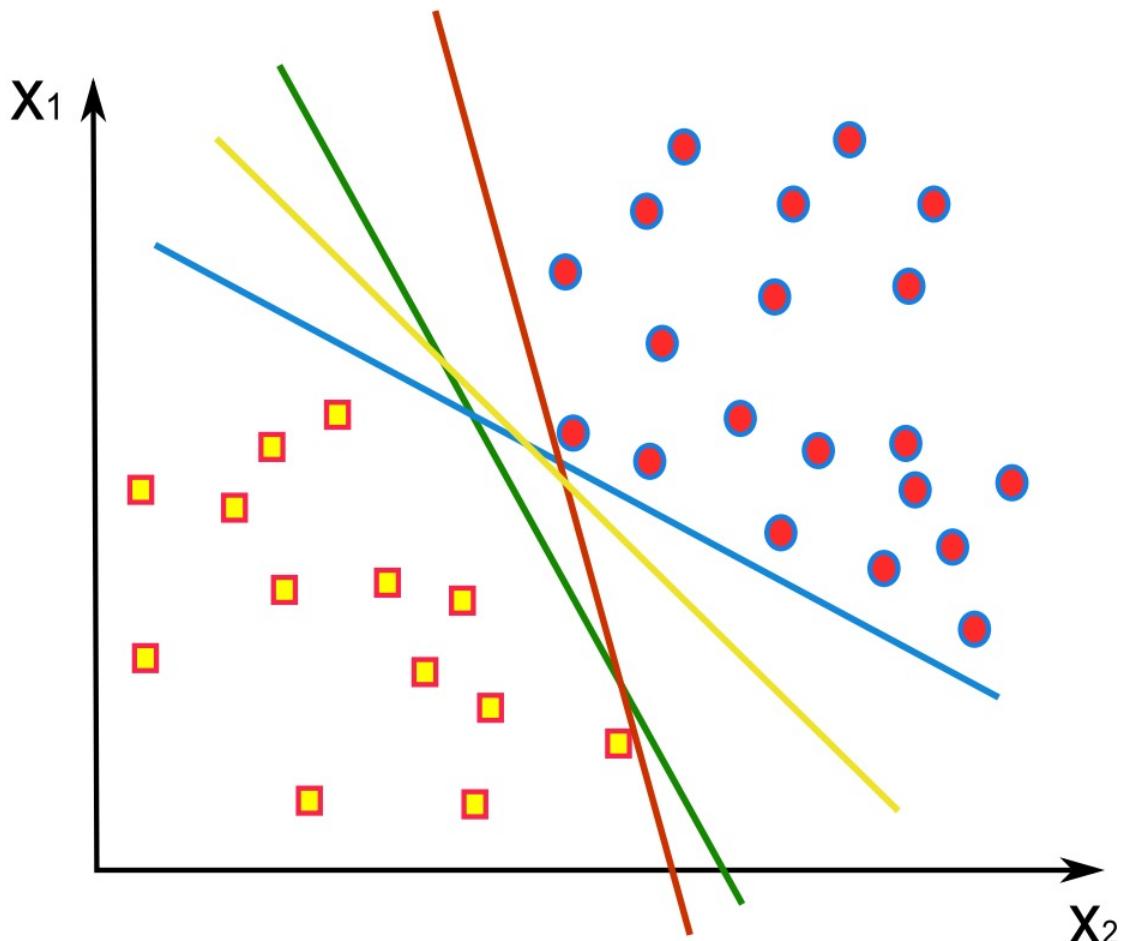


Figura 2.15: Hiperplanos de separación(técnica3)

■ **Margen Geométrico:**

- El margen geométrico se define como:

$$\gamma = \frac{1}{\|w\|}$$

- El hiperplano de separación óptimo maximiza este margen, mejorando la capacidad de generalización del modelo.

■ Optimización del Margen:

- La optimización del margen geométrico implica minimizar la norma del vector de pesos w .
- Esto se resuelve mediante un problema de programación cuadrática para encontrar el hiperplano óptimo y dos hiperplanos paralelos (H_1 y H_2) que maximicen la distancia entre ellos sin que haya datos entre estos hiperplanos.

■ Vectores de Soporte:

- Los puntos de datos más cercanos al hiperplano de separación, llamados vectores de soporte, definen este hiperplano. Los demás puntos no afectan la solución del SVM.

■ Formulación Dual:

- Se utiliza la formulación dual mediante multiplicadores de Lagrange para simplificar el problema.
- Esto permite que los datos de entrenamiento aparezcan solo como productos punto entre vectores, lo cual es fundamental para generalizar el procedimiento a casos no lineales.
- La Lagrangiana es:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i [y_i (w \cdot x_i + b) - 1]$$

■ Solución Dual:

- La solución del problema dual implica derivar con respecto a w y b y luego sustituir estas derivadas en la Lagrangiana original:

$$\frac{\partial L(w, b, \alpha)}{\partial w} = w - \sum_{i=1}^l \alpha_i y_i x_i = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i y_i x_i$$

$$\frac{\partial L(w, b, \alpha)}{\partial b} = - \sum_{i=1}^l \alpha_i y_i = 0 \Rightarrow \sum_{i=1}^l \alpha_i y_i = 0$$

- Los vectores de soporte son aquellos con $\alpha_i > 0$, y son cruciales para definir los hiperplanos H_1 y H_2 .

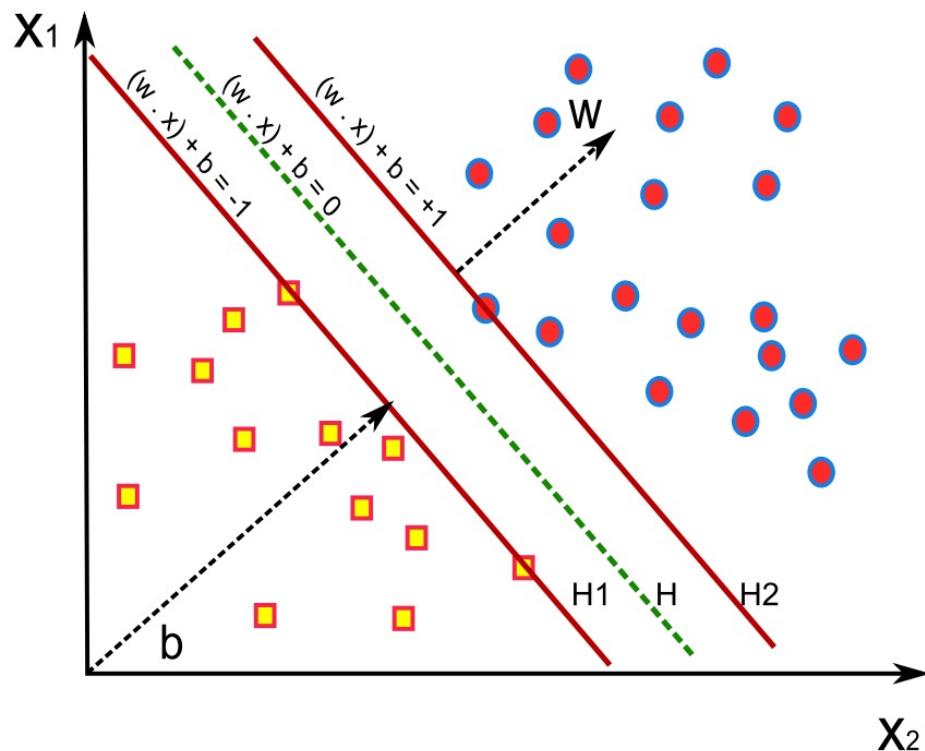


Figura 2.16: Clasificador óptimo(**tecnica3**)

- El problema de aprendizaje presentado anteriormente es válido solo para datos linealmente separables, lo cual es raro en la vida real.
- En muchos casos, los datos de entrenamiento tienen intersecciones y no pueden ser separados linealmente sin errores.
- Los métodos de programación cuadrática anteriores no pueden ser usados en estos casos porque la condición $y_i(w \cdot x_i + b) \geq 1$ no puede ser satisfecha para todos los puntos de datos.
- En casos de intersección, algunos puntos de datos no pueden ser clasificados correctamente, y los valores correspondientes de α_i tienden a infinito.
- Para manejar esto, se introduce el concepto de "soft margin", permitiendo una cierta cantidad de errores de clasificación.

- Se agregan variables de holgura no negativas ξ_i (slack variables) en la ecuación de separación:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad \text{con} \quad \xi_i \geq 0$$

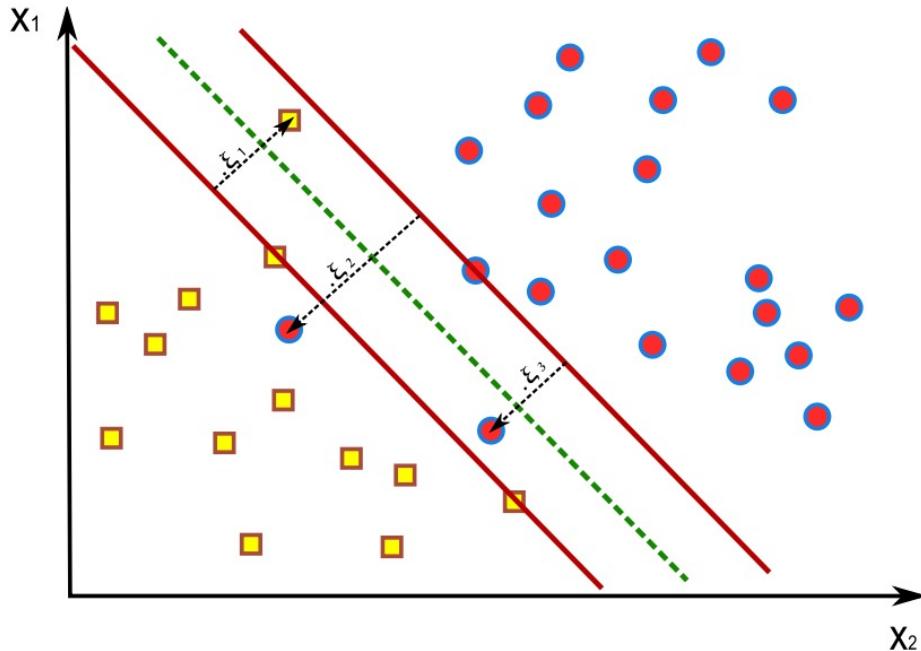


Figura 2.17: Hiperplanos de margen suave.(tecnica3)

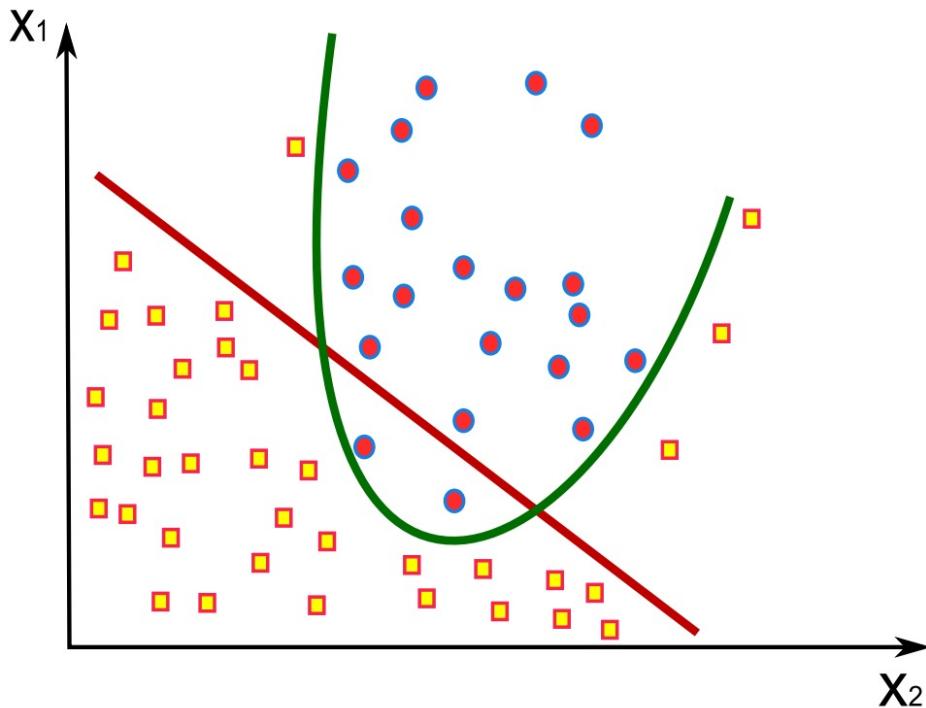


Figura 2.18: Clasificación no lineal.(técnica3)

En un SVM, el hiperplano óptimo se determina para maximizar la capacidad de generalización del modelo. Sin embargo, si la data de preparación son linealmente unibles, el clasificador que se tuvo podría no tener una alta capacidad de generalización, incluso si los hiperplanos se determinan de manera óptima. Es decir, para engrandecer el hueco entre clases, el espacio de entrada original se transforma en un espacio de características altamente dimensional llamado “espacio de características”.

La idea simple en la arquitectura de SVM no lineales es convertir los vectores que reciben $x \in \mathbb{R}^n$ en vectores $U(x)$ de un cardumen de características altísimamente dimensional F (donde U representa la asignación: $\mathbb{R}^n \rightarrow \mathbb{R}^f$) y terminar la dificultad de clasificación lineal en este espacio de características. El conjunto de hipótesis consideradas será de la forma:

$$f(x) = \sum_{i=1}^l w_i \phi_i(x) + b$$

donde $\phi : X \rightarrow F$ es una asignación no lineal de un espacio de entrada a un espacio de características.

Una característica de los componentes de aprendizaje lineales es que pueden expresarse en una vista dual, lo que significa que la ecuación anterior se puede expresar como una combinación lineal de los puntos de datos de entrenamiento. Por lo tanto, la regla de decisión puede

evaluarse utilizando productos punto:

$$f(x) = \sum_{i=1}^l \alpha_i y_i K(x_i, x) + b$$

donde $K(x_i, x)$ es una función kernel que representa el producto punto en el espacio de características.

Los kernel son funciones que satisfacen ciertas propiedades y permiten calcular eficientemente la función de decisión. Algunos de los kernel más utilizados son:

- Kernel lineal: $K(x_i, x_j) = (x_i \cdot x_j)$
- Kernel Gaussiano: $K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$
- Kernel RBF: $K(x_i, x_j) = e^{-c\|x_i - x_j\|^2}$
- Kernel sigmoide: $K(x_i, x_j) = \tanh(g(x_i \cdot x_j) + m)$

La elección del kernel depende de las características de los datos y es necesario determinar los parámetros óptimos del kernel utilizado para obtener buenos resultados.

2.2.3.3. Debilidades de SVM

A pesar de la capacidad de generalización y las muchas ventajas de SVM, tienen algunas debilidades muy marcadas, entre las que se encuentran: la selección de parámetros, la complejidad algorítmica que afecta el tiempo de entrenamiento del clasificador en conjuntos de datos grandes, el desarrollo de clasificadores óptimos para problemas multiclas y el rendimiento de SVM en conjuntos de datos desequilibrados.

Complejidad algorítmica Limitación principal: alto costo computacional en conjuntos de datos voluminosos. Las SVM, si bien son herramientas poderosas, presentan una desventaja significativa: su elevado costo computacional al trabajar con conjuntos de datos de gran tamaño. La razón de esto reside en el crecimiento cuadrático de la matriz del kernel de entrenamiento con respecto al tamaño del conjunto de datos. Esta dependencia cuadrática deriva en un proceso de entrenamiento sumamente lento para conjuntos de datos voluminosos.

Los métodos de entrenamiento para SVM se pueden categorizar en selección de datos, descomposición, implementaciones geométricas, implementaciones paralelas y heurísticas. Sus ideas centrales y los algoritmos más representativos se presentan en esta sección.

Métodos de selección de datos para SVM: intentan disminuir el tamaño de los conjuntos de datos eliminando las instancias que no contribuyen a la definición del hiperplano separador óptimo. Estos métodos se basan en las instancias que están más cerca del límite de separación y se denominan vectores de soporte (SVs).

Métodos de descomposición: se basan en que el tiempo de entrenamiento se puede reducir si solo se tienen en cuenta las restricciones activas del problema QP. Un método similar a los métodos de conjunto activo para la optimización se aplica en estos métodos de descomposición.

Métodos de implementación paralela: dividen el conjunto de entrenamiento en subconjuntos independientes para entrenar SVM en diferentes procesadores.

Métodos geométricos: están basados en que calcular el hiperplano separador óptimo es equivalente a encontrar el par de puntos más cercanos que pertenecen a envolventes convexas.

Métodos heurísticos: incluyen técnicas como la inicialización de valores de parámetros para el inicio del problema QP, entre otros.

2.2.3.4. Implementaciones de SVM

Resumen de Implementaciones SVM:

- Actualmente existen varias implementaciones de Máquinas de Vectores de Soporte (SVM) en la literatura.
- El tiempo computacional varía entre las implementaciones debido a diferentes heurísticas utilizadas para resolver el problema de programación cuadrática.
- En conjuntos de datos grandes, las SVM enfrentan tiempos de entrenamiento enormes debido a su complejidad computacional casi cúbica.

Enfoques para Mejorar el Tiempo de Entrenamiento de SVM:

- **Reducción de Datos:** Se centra en entrenar SVM con un subconjunto de datos probablemente vectores de soporte.
- **Fragmentación (Chunking):** Divide el problema en fragmentos más pequeños, resolviendo cada uno de forma iterativa.

- **Descomposición:** Resuelve una secuencia de problemas de optimización más pequeños para reducir la complejidad.
- **Optimización Secuencial Mínima (SMO):** Optimiza un subconjunto mínimo de dos puntos en cada iteración, escalando bien con conjuntos de datos grandes.
- **Reducción (Shrinking):** Acelera la optimización al reducir el número de valores del kernel necesarios.
- **Selección de Trabajo (Working Selection):** Selecciona un conjunto inicial de variables para optimizar SMO de manera eficiente.

Beneficios de las Implementaciones SVM:

1. Pueden manejar conjuntos de datos grandes de manera eficiente.
2. Soportan funciones de kernel estándar y personalizadas.
3. Eficientes para clasificación multiclas y validación cruzada.
4. Pueden manejar SVM ponderadas para datos desbalanceados.
5. Proporcionan estimaciones de probabilidad.

Implementaciones SVM Populares:

- **SVMLight:** Utiliza técnicas de Selección de Trabajo y Reducción, eficiente para conjuntos de datos grandes.
- **SVM Torch:** Utiliza Selección de Trabajo y Reducción para problemas de regresión a gran escala.
- **Pegasos:** Implementa métodos de descomposición para reducción del tiempo de entrenamiento, adecuado para SVM no lineales.
- **LIBSVM:** Basado en SMO con algoritmo avanzado de selección de conjunto de trabajo, eficiente para conjuntos de datos grandes.
- **SVM Incremental:** Marco para aprendizaje incremental y adaptación de clasificadores SVM.

Tabla 2.1: Implementaciones SVM

Implementación	Desarrollador	Código Fuente y Universidad
SVM Torch	Ronan Collobert and Samy Bengio	C++ - Université de Montréal
Pegasos	Shai Shalev-Shwartz	C++ - The Hebrew University of Jerusalem
LibSVM	Chih-Chung Chang and Chih-Jen Lin	C and Java - National Taiwan University
SVMLight	Thorsten Joachims	C - Cornell University
Incremental	Chris Diehl	M - Carnegie Mellon
SVM		

2.2.3.5. Aplicaciones en problemas del mundo real en Clasificación de imágenes

Clasificación de Cálculos Renales: En **svmm1**, se emplean filtros de mediana y Gaussiano, y se aplica un enmascaramiento no agudo para mejorar las imágenes. Se utilizan operaciones morfológicas y segmentación basada en entropía para encontrar la región de interés, y luego se emplean técnicas de clasificación KNN y SVM para el análisis de imágenes de cálculos renales.

Detección Temprana de Melanoma: En **svmm2**, se presenta un dispositivo de manejo con bajo costo y alto rendimiento para mejorar la detección temprana de melanoma en atención primaria. Se propone un sistema de hardware dinámico para implementar un clasificador SVM en cascada en FPGA para la detección temprana de melanoma.

Reconocimiento de Expresiones Faciales: En **svmm3**, se presenta un marco para el reconocimiento de expresiones independiente de la persona mediante el aprendizaje de múltiples tipos de características faciales a través del aprendizaje de múltiples núcleos en SVM multiclas.

Reconocimiento de Gestos de Mano: Se aborda el reconocimiento de signos indios basado en técnicas de reconocimiento de gestos de mano dinámicos en escenarios en tiempo real. Se utilizan momentos de Hu y trayectorias de movimiento para la extracción de características y la clasificación de gestos mediante SVM.

Clasificación de Imágenes Hiperespectrales: En **svmm4**, se propone una técnica para la clasificación espectral-espacial de imágenes hiperespectrales que combina la clasificación pixel a pixel con SVM y la información contextual espacial con un enfoque de Campo Aleatorio de Markov para refinar los resultados de la clasificación.

Selección de Características para la Clasificación de Masas Mammográficas: En **svmm5**, se abordan métodos de selección de características para la clasificación de masas en mamografías, integrando un procedimiento basado en eliminación recursiva de características SVM con una selección de características de información mutua normalizada.

Clasificación de Imágenes mediante SVM con Kernel de Intersección de Histogramas: En **svmm6**, se propone un método para la clasificación de imágenes mediante SVM con kernel de intersección de histogramas. Se presenta un solver de kernel de intersección de histogramas determinista y escalable.

Decodificación de Códigos QR a Color: En **svmm7**, se proponen dos enfoques para resolver problemas de decodificación de códigos QR a color. LSVM-CMI y QDA-CMI, que modelan conjuntamente diferentes tipos de distorsión cromática.

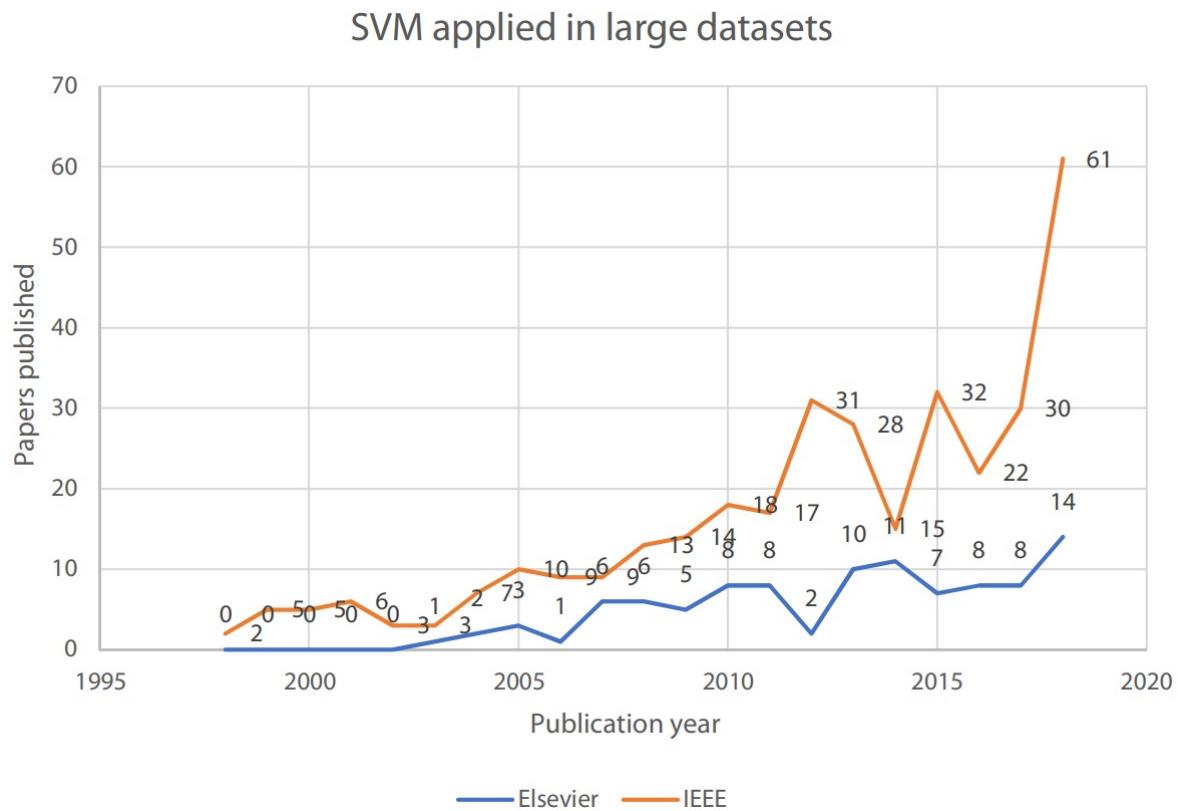


Figura 2.19: Número de publicaciones, en capítulos de libros y revistas, por año que contienen los términos de búsqueda SVM y Grandes conjuntos de datos.(tecnica3)

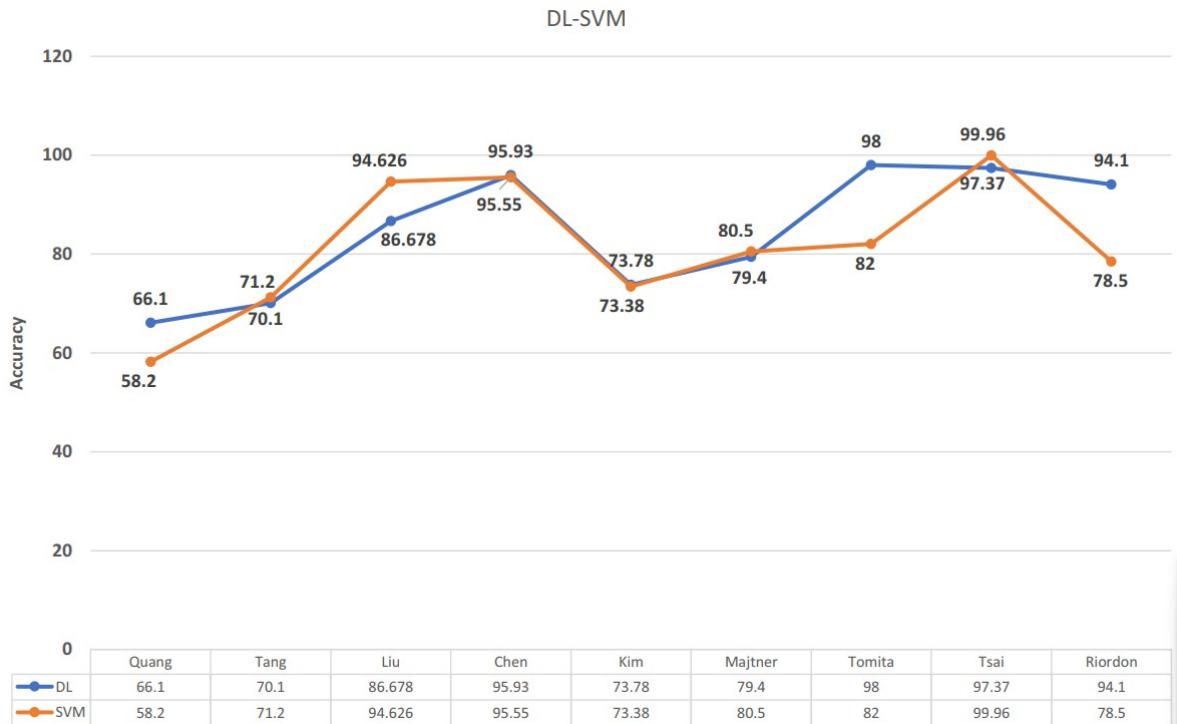


Figura 2.20: Rendimiento de SVM frente a Aprendizaje profundo(tecnica3)

2.2.3.6. Conclusiones

Debido a sus sólidos fundamentos teóricos y capacidad de generalización, entre otras ventajas, las SVM han sido implementadas en muchas aplicaciones del mundo real. Los algoritmos SVM han sido implementados en diversos campos de investigación como: categorización de texto (e hipertexto), detección de pliegues de proteínas y homología remota, clasificación de imágenes, bioinformática (clasificación de proteínas y cáncer), reconocimiento de caracteres escritos a mano, detección de rostros, control predictivo generalizado y muchos más. Muchos investigadores han demostrado que las SVM son mejores que otras técnicas de clasificación actuales. Sin embargo, a pesar de que las SVM tienen algunas limitaciones relacionadas con la selección de parámetros, la complejidad algorítmica, conjuntos de datos multiclas y conjuntos de datos desequilibrados, han sido implementadas en muchos problemas de clasificación de la vida real debido a sus buenos fundamentos teóricos y rendimiento de generalización.

Es importante mencionar que las SVM no son tan populares cuando los conjuntos de datos son muy grandes porque algunas implementaciones de SVM requieren un tiempo de entrenamiento enorme o, en otros casos, cuando los conjuntos de datos están desequilibrados, la precisión de las SVM es pobre. Hemos presentado algunas técnicas para enfrentar estos desequilibrios en los conjuntos de datos. Este artículo describe detalladamente las principales desventajas de las SVM y muchos algoritmos implementados para enfrentar estas desventajas, y cita los trabajos de investigadores que han enfrentado estas desventajas

2.2.4. Redes neuronales convolucionales: Review of Image Classification Algorithms Based on Convolutional Neural Networks (técnica2)

Este artículo destaca el papel crucial de las redes neuronales convolucionales (CNN) en la clasificación de imágenes desde 2012. Explora su evolución desde modelos básicos hasta arquitecturas avanzadas y su influencia en áreas de reconocimiento visual. Compara métodos de clasificación y analiza el impacto de CNN en diversas aplicaciones, subrayando tendencias actuales en el campo. La introducción aborda la importancia de la clasificación de imágenes y su evolución desde la extracción manual de características hasta el uso de CNN, inspiradas en la biología. Modelos como LeNet-5, AlexNet y ZFNet han mejorado significativamente el rendimiento en este ámbito.

A pesar de que las CNN se diseñaron originalmente para la visión por computadora, su aplicación exitosa en el análisis de imágenes de teledetección ha llevado a la necesidad de un análisis sistemático de los métodos de clasificación de imágenes basados en CNN. Este artículo se dedica a detallar el desarrollo de casi todas las arquitecturas de CNN típicas en

tareas de clasificación de imágenes, con el objetivo de proporcionar inspiración para el diseño de modelos CNN en el campo del análisis de imágenes de teledetección.

2.2.4.1. Visión general de las CNN

La CNN tiene una estructura principal compuesta por capas convolucionales, de agrupación, de activación no lineal y completamente conectadas. Después de preprocesar la imagen, se introduce en la red a través de la capa de entrada, se procesa mediante capas convolucionales y de agrupación alternadas, y finalmente se clasifica mediante la capa completamente conectada. En comparación con MLP, la CNN agrega capas convolucionales y de agrupación, lo que mejora el rendimiento en términos de tamaño del modelo y capacidad de procesamiento. La capa convolucional identifica eficazmente la correlación entre las características de los píxeles de la imagen, mientras que la capa de agrupación reduce la carga computacional y la sensibilidad excesiva a la posición. La CNN asegura cierto grado de invarianza ante cambios en la imagen como desplazamientos, escalados y distorsiones.

Para CNNs con una cierta profundidad, la operación de convolución de múltiples capas convolucionales puede extraer diferentes características de la entrada. La capa convolucional inferior generalmente extrae características comunes como textura, líneas y bordes, mientras que la capa superior extrae características más abstractas. La capa convolucional tiene varios núcleos de convolución con parámetros aprendibles, que son matrices compuestas por pesos aprendibles. Estas matrices de pesos suelen ser de tamaño 3×3 , 5×5 y 7×7 , con una longitud y ancho iguales y un número impar. Normalmente, la capa convolucional ingresará a los mapas de características. La matriz de pesos del núcleo de convolución corresponde al área local del mapa de características de conexión, y el núcleo de convolución realiza operaciones de convolución secuenciales en el área del mapa de características mediante deslizamiento.

La fórmula para la salida de una superficie de características en la capa convolucional puede expresarse aproximadamente como:

$$\text{feature_surface_out} = \sum_{i=1}^N M_i * W_i + B$$

Donde M_i representa una superficie de características de los mapas de características de entrada, W_i es la matriz de pesos del núcleo de convolución, la matriz de sesgo es B , $f(\cdot)$ es la función de activación no lineal y $\text{feature_surface_out}$ es una superficie de características de salida.

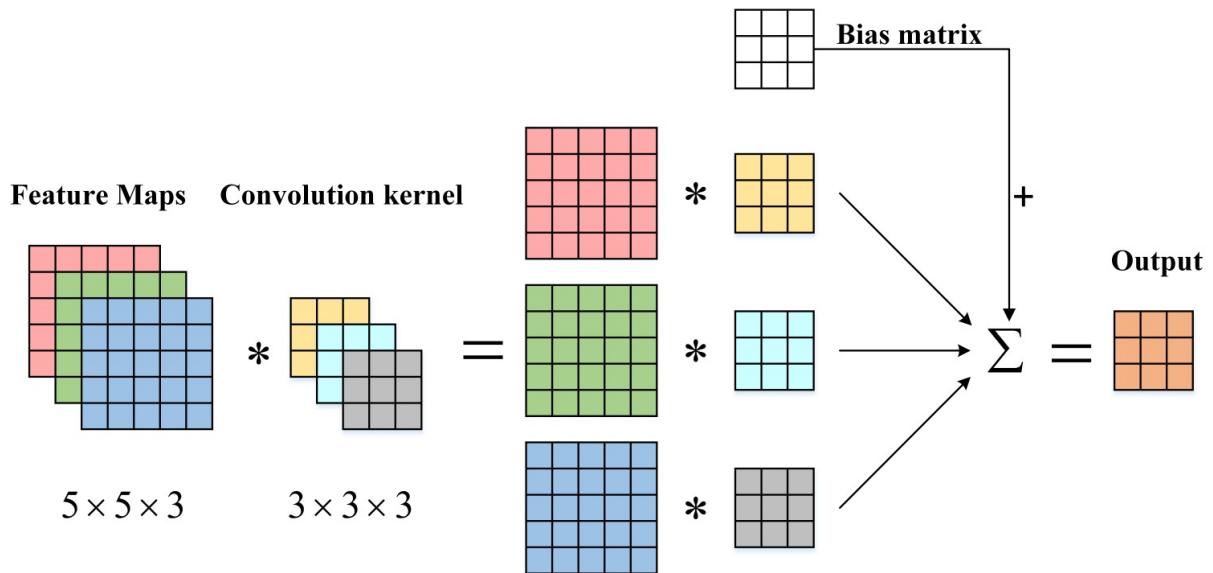


Figura 2.21: Diagrama esquemático del proceso de convolución(**técnica2**)

La capa de agrupación generalmente sigue a la capa convolucional. Las principales razones para usar la capa de agrupación son: realizar un procesamiento de submuestreo y reducción de dimensionalidad en la imagen de entrada para reducir el número de conexiones de la capa convolucional, reduciendo así la carga computacional de la red; lograr invariancia de escala, invariancia de traslación e invariancia de rotación de la imagen de entrada; hacer que el mapa de características de salida sea más robusto ante la distorsión y el error de una sola neurona.

Los métodos de agrupación más utilizados son el promedio y el máximo. Aunque existen otras formas de agrupación que pueden mitigar de manera más efectiva el sobreajuste de las redes neuronales convolucionales, como la Agrupación Lp, Agrupación Mixta, Agrupación Estocástica, Agrupación de Pirámide Espacial (SPP) y Agrupación Ordenada Multiescala, entre otros. Sin embargo, para el modelo clásico de red neuronal convolucional, aunque la mejor operación de agrupación no es la agrupación promedio o la agrupación máxima, son los dos métodos más clásicos. En [72], Bourbeau realizó principalmente un análisis teórico sobre el rendimiento de la agrupación promedio y la agrupación máxima. La operación de agrupación satisface la siguiente relación general entre el tamaño de las matrices de entrada y salida:

$$o = \left(\frac{i-k}{s} \right) + 1$$

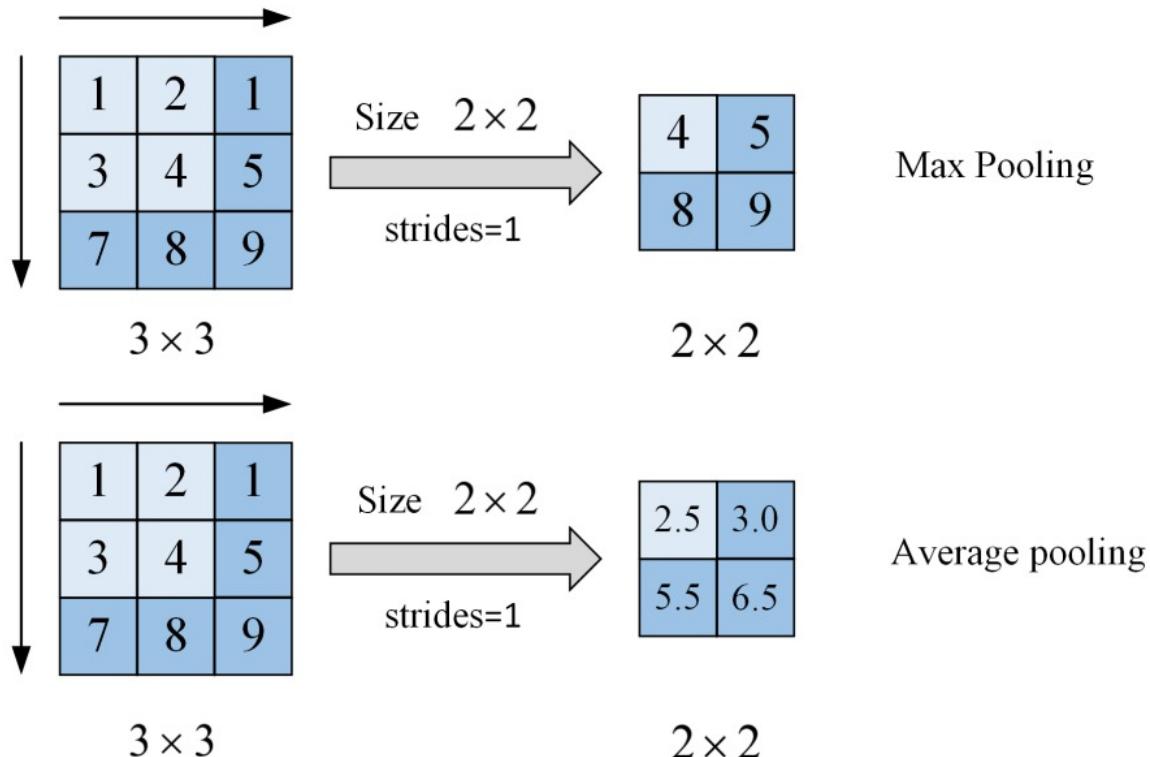


Figura 2.22: Agrupación máxima y agrupación promedio, no implica relleno cero.(técnica2)

El propósito de la función de activación es establecer una relación funcional entre la entrada y la salida, introduciendo así un componente no lineal en la red neural. Una función de activación adecuada puede mejorar significativamente el rendimiento de la red. Algunas funciones comunes incluyen sigmoides y Tanh, que son no linealidades saturantes. Sin embargo, para abordar los problemas asociados con estas no linealidades saturantes, se han propuesto funciones no saturantes como ReLU, Leaky ReLU, PReLU, RReLU y ELU. Estas funciones no saturantes son mucho más rápidas para el entrenamiento por descenso de gradiente, lo que resulta en redes mucho más eficientes. Las neuronas con funciones no saturantes se conocen como Unidades Lineales Rectificadas (ReLUs).

En la arquitectura de las CNN, la clasificación final se logra desde la capa de salida, generalmente la última capa de la capa completamente conectada (FC layer). Las diferentes funciones de pérdida también afectan el rendimiento de la arquitectura de la CNN y se aplican a diferentes tareas visuales (como la clasificación de imágenes, reconocimiento facial y reconocimiento de objetos). La función de pérdida más comúnmente utilizada es Softmax+Entropía Cruzada, con muchas versiones mejoradas basadas en ella, como center-loss, L-Softmax, A-Softmax, AM-Softmax, PEDCC-loss, entre otras, que desempeñan un papel importante en diferentes tareas visuales.

Tabla 2.2: Funciones de pérdida comunes para modelos CNN.

Función de Pérdida	Ecuación	Característica
L1 (MAE)	$\text{Pérdida}(y, y^*) = \frac{1}{m} \sum_{i=1}^m y_i^* - y_i $	Ampliamente utilizado en problemas de regresión. La pérdida L1 se denomina error absoluto medio (MAE).
L2 (MSE)	$\text{Pérdida}(y, y^*) = \frac{1}{m} \sum_{i=1}^m (y_i^* - y_i)^2$	Ampliamente utilizado en problemas de regresión. La pérdida L2 se denomina error cuadrático medio (MSE).
Softmax + Entropía Cruzada	$\text{Pérdida}(y, y^*) = -\sum_i y_i \log \left(\frac{\sum_{j=1}^{i-1} y_j}{\sum_i y_i} \right)$	= Esta función suele emplearse como sustitución del MSE en problemas de clasificación multi-clase. También se utiliza comúnmente en modelos CNN.

El flujo de datos en la arquitectura CNN se introduce básicamente en la sección anterior. Entendemos claramente que el entrenamiento de la red se basa en el paso fundamental de la actualización del gradiente, es decir, necesita calcular el gradiente de la función objetivo (función de pérdida) aplicando una derivada de primer orden con respecto a los parámetros de la red, y luego la información del gradiente se transfiere a la capa de red anterior en forma de cálculo de derivada parcial para lograr la actualización de los parámetros de aprendizaje de cada capa de red. La función del optimizador es proporcionar una manera de hacer que la actualización del gradiente sea más razonable, es decir, la macroscópica es que toda la red puede converger más rápido, un valor óptimo local más pequeño (pérdida más pequeña), cálculos más baratos, etc.

Tabla 2.3: Métodos de optimización para modelos CNN.

Nombre	Método	Características
Batch Gradient Descent (BGD)	Calcula el gradiente de todo el conjunto de entrenamiento y posteriormente utiliza este gradiente para actualizar los parámetros.	1. Para un conjunto de datos de pequeño tamaño, el modelo CNN converge más rápido y crea un gradiente extra estable utilizando BGD. 2. Generalmente no es adecuado para un conjunto de entrenamiento grande. 3. Requiere una cantidad sustancial de recursos.
Stochastic Gradient Descent (SGD)	Muestrea seleccionando arbitrariamente parte del conjunto de entrenamiento.	1. Para un conjunto de entrenamiento de gran tamaño, esta técnica es más eficiente en memoria y mucho más rápida que BGD. 2. Se introduce aleatoriedad y ruido debido a sus actualizaciones frecuentes. Su convergencia no es estable, pero la expectativa sigue siendo igual al descenso de gradiente correcto.
Mini-batch Gradient Descent	Divide las muestras de entrenamiento en varios mini lotes, y luego se realiza la actualización de parámetros siguiendo el cálculo del gradiente en cada mini lote.	Este método combina las ventajas técnicas de SGD y SGD, que tienen una convergencia estable, más eficiencia computacional y una eficacia de memoria extra.
Momentum	Introduce un parámetro de momento en SGD que acumula información de gradiente histórica.	Cuando el entrenamiento cae en un mínimo local, la información del gradiente con momento puede ayudar a la red a escapar y encontrar el mínimo global.
Adaptive Moment Estimation (Adam)	Calcula una tasa de aprendizaje adaptativa para cada parámetro en el modelo.	Se combinan las ventajas del momento y RMSprop. Es ampliamente utilizado en el aprendizaje profundo y representa la última tendencia de optimización.

2.2.4.2. Clasificación de imágenes basada en CNN

En general, la clasificación de imágenes implica extraer características de la imagen manualmente o mediante métodos de aprendizaje de características, y luego usar un clasificador para identificar la categoría del objeto. Antes del aprendizaje profundo, se utilizaba ampliamente el modelo de Bag of Words, que involucraba tres procesos: extracción de características de bajo nivel, codificación de características y diseño de clasificador. Este método tradicional fue común hasta 2012, siendo utilizado en competiciones como PASCAL VOC y ILSVRC 2010. Sin embargo, la aparición de las CNNs revolucionó la clasificación de imágenes, permitiendo un aprendizaje de características más eficaz y un proceso de clasificación de extremo a extremo que no requiere la extracción manual de características, superando las limitaciones de los métodos tradicionales. Esta sección presenta los modelos clásicos de clasificación de imágenes basados en CNN a lo largo del tiempo.

- **LeNet Network:** En 1998, Lecun desarrolló el modelo LeNet-5 para clasificar imágenes digitalmente, superando a otros métodos de la época y utilizando retropropagación por primera vez en CNNs. LeNet-5, con 7 capas y 60k parámetros, se divide en un área de convolución y un área FC, utilizando la función de activación sigmoid y el clasificador softmax. Aunque tuvo buenos resultados en MNIST, su rendimiento en conjuntos de datos más grandes fue limitado debido a la complejidad computacional y la falta de investigación en inicialización de parámetros y algoritmos de optimización. Eventualmente, fue superado por otros métodos de aprendizaje automático como SVM.
- **AlexNet Network:** En 2012, AlexNet, desarrollado por Krizhevsky et al., ganó la ILSVRC 2012 con gran ventaja, demostrando que las características aprendidas superan a las diseñadas manualmente. Utilizó procesamiento paralelo entre GPUs debido a las limitaciones de la GPU GTX 580. AlexNet tiene 8 capas y 60M parámetros, con 5 capas de convolución y 3 capas FC. Requiere un tamaño de convolución mayor para las imágenes más grandes de ImageNet. Mejoras clave incluyen: ReLU: Aceleró la convergencia y redujo la desaparición del gradiente. Dropout: Controló la complejidad del modelo y mitigó el sobreajuste. Aumento de datos: Usó técnicas como volteo, recorte y cambios de color para ampliar conjuntos de datos y reducir el sobreajuste. Pooling superpuesto: Mejoró la precisión del modelo y alivió el sobreajuste.

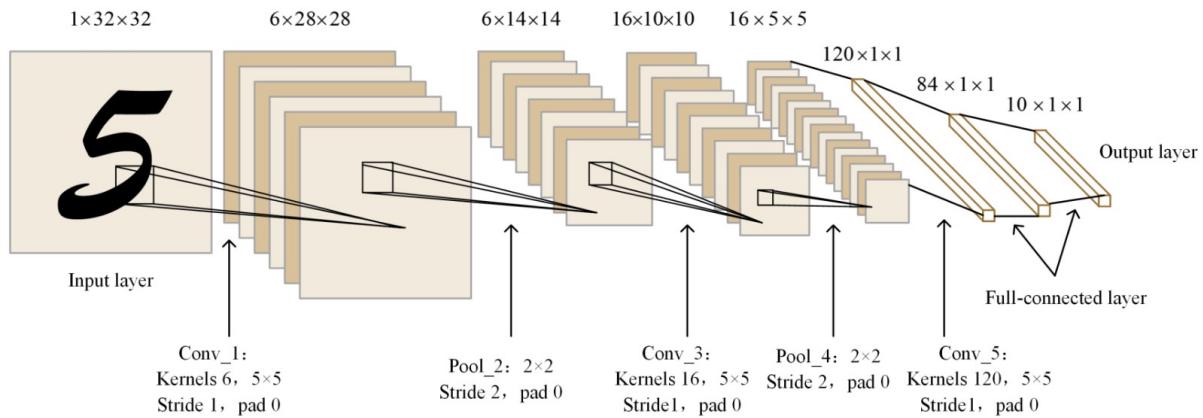


Figura 2.23: La arquitectura de la red LeNet-5. La forma de salida es canal × altura × anchura. Cada capa convolucional Utiliza tallas 5×5 , relleno 0, zancadas 1. Cada capa de agrupación tiene un tamaño de 2×2 y zancadas 2.(técnica2)

- **VGGNet:** En 2014, Simonyan et al. propusieron el modelo VGG, obteniendo el segundo lugar en ILSVRC 2014. Similar a AlexNet, VGG usa una estructura de área de convolución seguida de área FC. Utiliza varias capas de convolución idénticas seguidas de una capa de pooling que reduce la altura y anchura a la mitad. VGG-16, una variante con 16 capas, conecta cinco bloques en serie y termina con dos capas FC de 4096 neuronas y una capa de salida de 1000 clasificaciones.

Mejoras respecto a AlexNet:

Red modular: Uso de módulos básicos para construir el modelo. Convoluciones más pequeñas: Filtros de 3×3 que aumentan la profundidad y reducen parámetros comparado con filtros más grandes. Entrenamiento multi-escala: Escala la imagen de entrada y la recorta aleatoriamente, logrando aumento de datos y previniendo el sobreajuste.

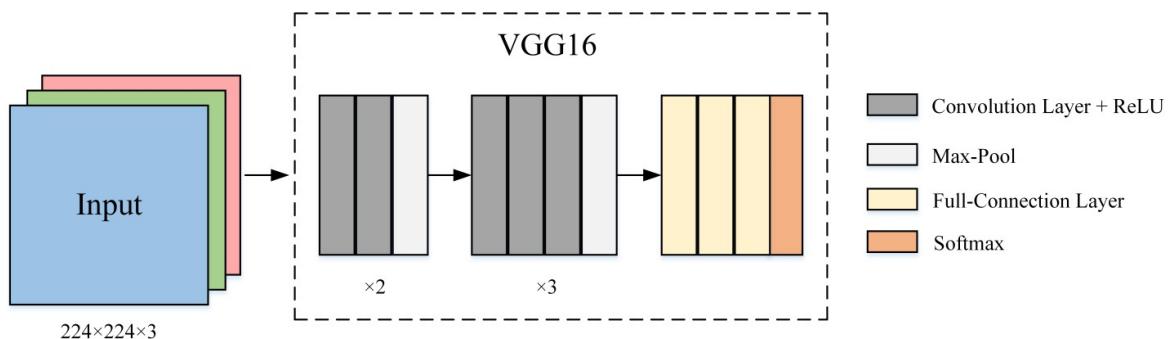


Figura 2.24: La arquitectura de la red VGG-16. Conv: tamaño = 3×3 , zancada = 1, relleno = 1. Piscina: tamaño = 3×3 , zancada = 2. (técnica2)

- **GoogLeNet/InceptionV1 to V4:** En 2014, GoogLeNet, propuesta por Christian Szegedy et al., ganó el campeonato de ILSVR 2014 introduciendo el módulo Inception.

InceptionV1:

Tiene 22 capas y alrededor de 6 millones de parámetros. El módulo Inception contiene 4 ramas paralelas: tres con capas de convolución de diferentes tamaños (1x1, 3x3, 5x5) y una con una capa de max-pooling, seguido de convolución 1x1. Mejora la adaptabilidad y la fusión multi-escala al aumentar el ancho del modelo.

InceptionV2:

Reemplaza convoluciones de 5x5 con dos convoluciones de 3x3 para reducir el tiempo de cálculo. Introduce Batch Normalization (BN) para estabilizar y acelerar el entrenamiento mediante la normalización de las entradas de las capas.

InceptionV3:

Utiliza convoluciones factoradas y asimétricas, reemplazando una convolución 3x3 por una 1x3 seguida de una 3x1 para reducir los parámetros y mejorar la eficiencia computacional. Implementa reducción del tamaño de la cuadrícula mediante operaciones de pooling y convoluciones con stride 2. Introduce técnicas adicionales para optimizar la eficiencia computacional y prevenir el sobreajuste, como el uso de múltiples tamaños de escala para las imágenes de entrada.

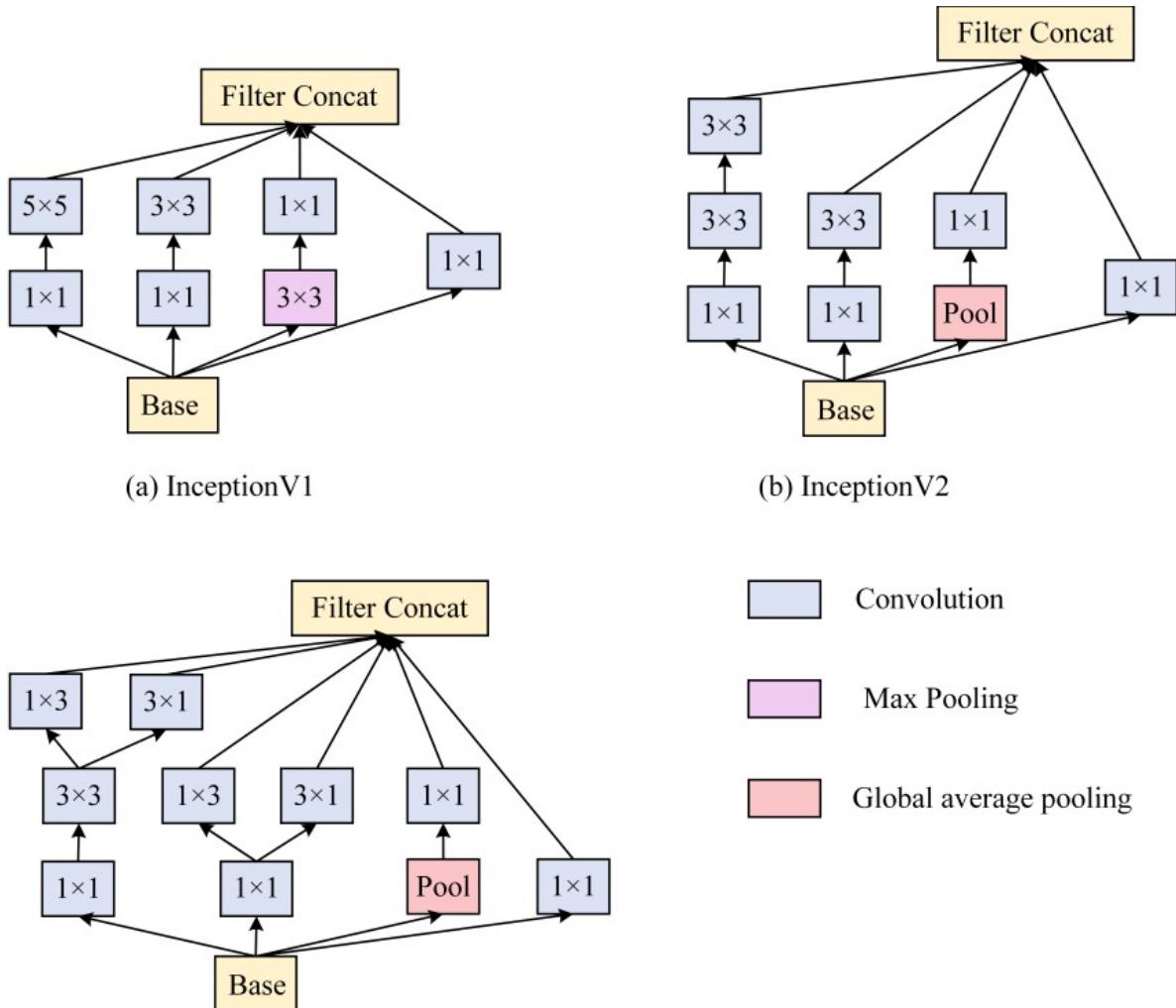


Figura 2.25: Módulo InceptionV1 a V3 (técnica2)

InceptionV4:

Reduce la complejidad de InceptionV3 mediante la unificación de las opciones de cada bloque de Inception. Los bloques de Inception y de Reducción son utilizados para simplificar y optimizar la arquitectura. El "stem" de la arquitectura se modifica para ser más uniforme y mejorar la eficiencia. Se optimiza el uso de memoria durante la retropropagación, permitiendo entrenar el modelo sin particionar réplicas.

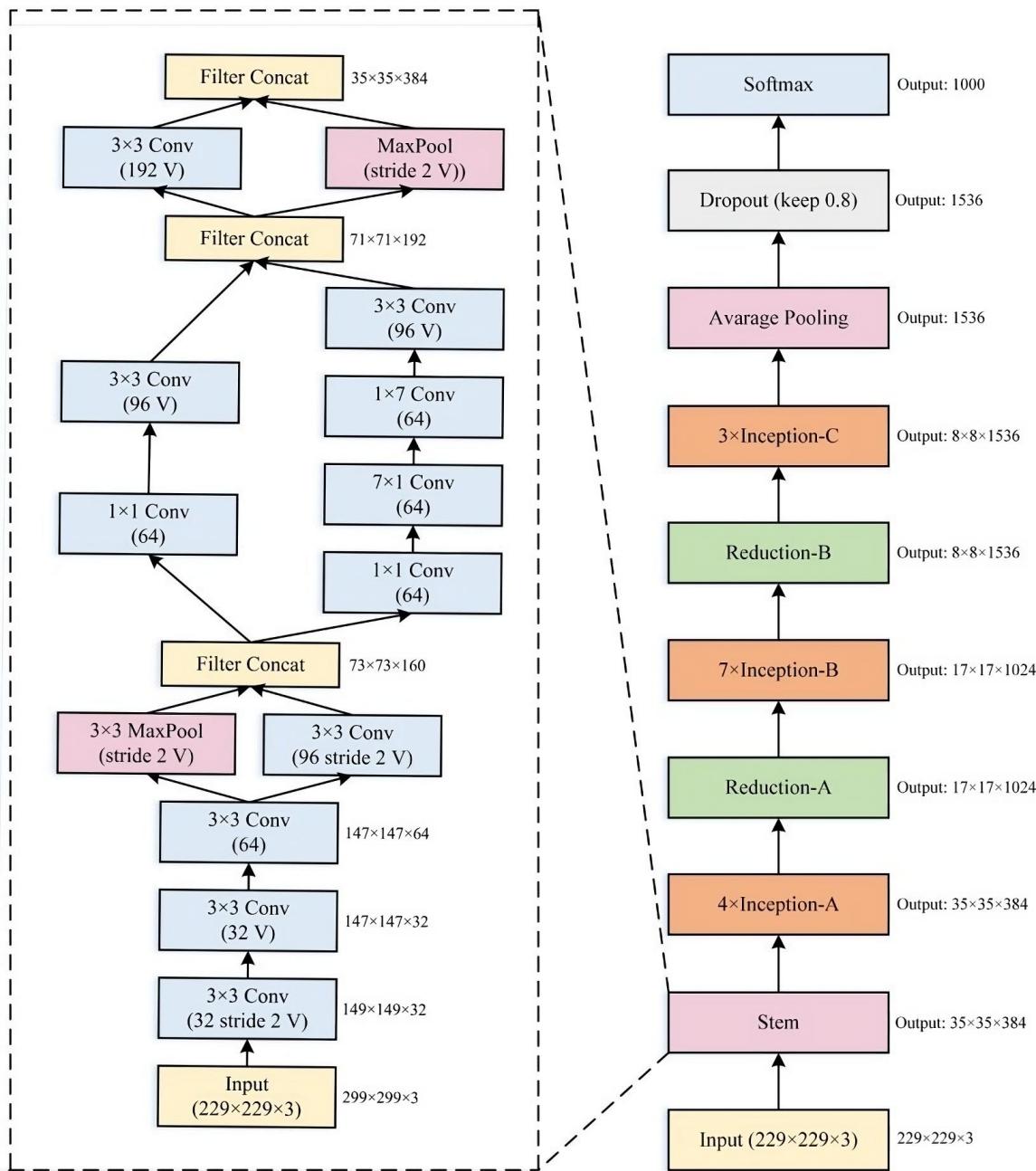


Figura 2.26: Arquitectura general de InceptionV4. La parte superior de la imagen es la estructura general, la parte inferior de la imagen es el tallo de la arquitectura. (técnica2)

- **Residual Learning Networks (ResNet):** En 2015, la red residual profunda ResNet, propuesta por Kaiming He et al., ganó el primer premio en ILSVRC2015. A medida que las redes se vuelven más profundas, aumenta el rendimiento de la red, pero simplemente aumentar la profundidad de la red no mejora efectivamente el rendimiento. ResNet aborda este problema mediante conexiones residuales, que permiten que las redes profundas

alcancen una alta precisión. En lugar de aprender la asignación deseada directamente, las conexiones residuales ajustan una asignación residual relacionada con la identidad, lo que facilita la optimización y el aprendizaje. Los bloques residuales en ResNet contienen dos capas convolucionales 3x3, seguidas de normalización y activación ReLU, con una conexión directa de entrada al bloque final. Para profundidades mayores, se puede añadir una capa convolucional 1x1 para controlar el número de canales. ResNet ha demostrado ser una contribución significativa al lograr precisión incluso con profundidades mayores, superando a las redes previas. En 2015, ResNet, una red residual profunda propuesta por Kaiming He et al., ganó el primer premio en ILSVR2015. A medida que las redes se vuelven más profundas, el rendimiento aumenta, pero simplemente aumentar la profundidad no garantiza una mejora efectiva. ResNet aborda este desafío mediante conexiones residuales, que permiten a las redes profundas alcanzar altos niveles de precisión. En lugar de aprender directamente la asignación deseada, las conexiones residuales ajustan una asignación residual relacionada con la identidad, lo que facilita la optimización y el aprendizaje. Los bloques residuales en ResNet consisten en dos capas convolucionales 3x3, seguidas de normalización y activación ReLU, con una conexión directa de entrada al bloque final. Para profundidades mayores, se puede añadir una capa convolucional 1x1 para controlar el número de canales. ResNet ha demostrado ser una contribución significativa al lograr una alta precisión incluso con profundidades mayores, superando a las redes anteriores.

- **Mejoras de ResNet:** ResNet con preactivación: He et al. introdujeron una estructura de preactivación que mejora el rendimiento de ResNet al preactivar la BN y ReLU. Este enfoque permite el entrenamiento exitoso de ResNet con más de 1000 capas, enfatizando la importancia del mapeo de identidad.

Profundidad estocástica: El método de profundidad estocástica, propuesto por Huang et al., elimina ciertas capas durante el entrenamiento, reduciendo significativamente el tiempo de entrenamiento y aumentando la profundidad de ResNet. Este método resulta efectivo incluso con más de 1200 capas, mejorando el error de prueba y el tiempo de entrenamiento en los conjuntos de datos CIFAR-10/100.

Redes Residuales Amplias (WRNs): Para contrarrestar la desaceleración del entrenamiento debido a la disminución de la reutilización de características en redes más profundas, las WRNs introducen bloques de eliminación amplia, ampliando las capas de peso de las unidades residuales originales y agregando eliminación entre ellas. Este enfoque, con menos capas que las ResNets más profundas, reduce el tiempo de entrenamiento y tiene un mejor rendimiento en los conjuntos de datos CIFAR e ImageNet.

ResNeXt: ResNeXt introduce el concepto de “Cardinalidad C”, el número de rutas en

un bloque, para superar los desafíos de adaptación del conjunto de datos. Aumentar la cardinalidad resulta más efectivo que aumentar la profundidad o el ancho. Las estructuras de convolución agrupada, como en la Figura 18c, son más rápidas y eficientes, formando la base de ResNeXt.

Redes Residuales Dilatadas (DRN): Yu et al. abordan la pérdida de resolución y la reducción de la información de características en el submuestreo al introducir convoluciones dilatadas. Estas convoluciones aumentan el campo receptivo de las capas superiores, compensando la reducción de campos receptivos inducida por la eliminación del submuestreo. DRN logra una precisión significativamente mejorada en la clasificación de imágenes en comparación con ResNet.

Otros modelos: Variantes como la ResNet reducida de Veit et al., Resnet en Resnet (RiR), la técnica DropBlock, Big Transfer (BiT) y NFNet, ofrecen enfoques únicos para mejorar el rendimiento de ResNet. Estos métodos incluyen la eliminación de capas, arquitecturas de doble flujo, eliminación de características y estructuras sin BN, logrando mejoras notables en velocidad de entrenamiento y precisión.

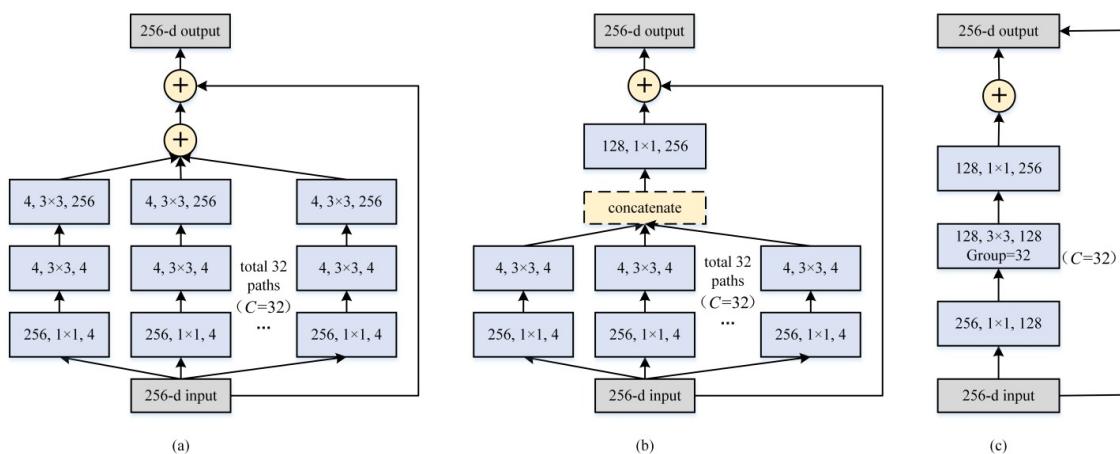


Figura 2.27: Bloques de construcción equivalentes de ResNeXt. (técnica2)

- **MobileNet V1 to V3:** En 2017, Google presentó MobileNetV1, una red liviana diseñada para dispositivos móviles y embebidos. Utiliza convolución separable en profundidad, que consta de convolución en profundidad y convolución puntual, en lugar de convolución estándar, reduciendo significativamente el costo computacional y los parámetros. MobileNetV1 ofrece dos hiperparámetros, el multiplicador de ancho x y el multiplicador de resolución x, para equilibrar eficazmente el cálculo y la precisión.

En 2018, MobileNetV2 abordó el problema de los parámetros cero en la convolución separable en profundidad al introducir Residuos Invertidos y Cuellos de Botella Lineales.

A diferencia del bloque residual estándar, el bloque residual invertido de MobileNetV2 sigue una secuencia de 1×1 (expansión) $\rightarrow 3 \times 3 \rightarrow 1 \times 1$ (compresión). Además, reemplaza el último ReLU con una transformación lineal para evitar la pérdida de información.

En 2019, MobileNetV3 mejoró la eficiencia y precisión al incorporar el bloque SE para atención por canal y la tecnología de Búsqueda de Arquitectura Neural (NAS). El enfoque NAS consciente de la plataforma se utiliza para la búsqueda por bloques para encontrar estructuras globales de red, mientras que NetAdapt ajusta individualmente las capas de manera secuencial. MobileNetV3 también adopta la función de activación h-swish, modificando el sigmoidal de la función swish para mejorar la precisión.

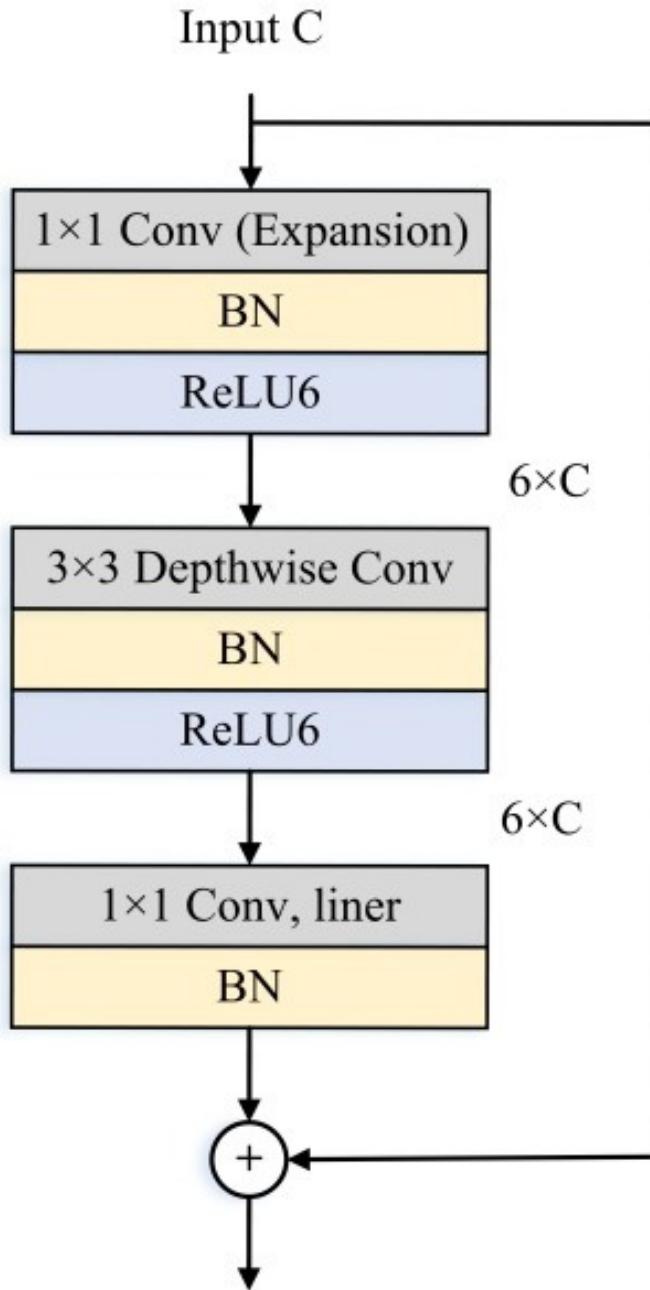


Figura 2.28: El bloque residual de cuello de botella de MobileNetV2. C es el número de canales y las relaciones de expansión son 6. (**tecnica2**)

- **ShuffleNet V1 to V2:**

En 2017, ShuffleNetV1, desarrollado por Face++, se enfocó en plataformas móviles como drones, robots y teléfonos inteligentes. Utiliza convolución Pointwise Group y Channel Shuffle para mejorar el bloque residual. La convolución Pointwise Group resuelve el problema de los canales limitados debido a convoluciones costosas, mientras que Chan-

nel Shuffle aborda la pérdida de información entre grupos de canales en bloques de convolución de grupo. La unidad ShuffleNet (stride = 1) reemplaza la primera convolución 1×1 con convolución de grupo pointwise seguida de una operación de mezcla de canales. La unidad ShuffleNet (stride = 2) agrega un GAP de 3×3 a la ruta de acceso directo y reemplaza la concatenación de canales con una suma de elementos. MobileNetV1 experimentó con diferentes números de grupos para convoluciones y escalas para el número de filtros.

En 2018, ShuffleNetV2 optimizó la velocidad y precisión considerando la complejidad computacional, el costo de acceso a la memoria y las características de la plataforma. Cuatro directrices principales surgieron de los experimentos: el ancho de canal igual minimiza el costo de acceso a la memoria; la convolución de grupo excesiva aumenta este costo; la fragmentación de la red reduce el paralelismo y las operaciones elemento a elemento tienen una alta relación MAC/FLOPs. La unidad ShuffleNetV2 evita violar estas directrices para mejorar la eficiencia.

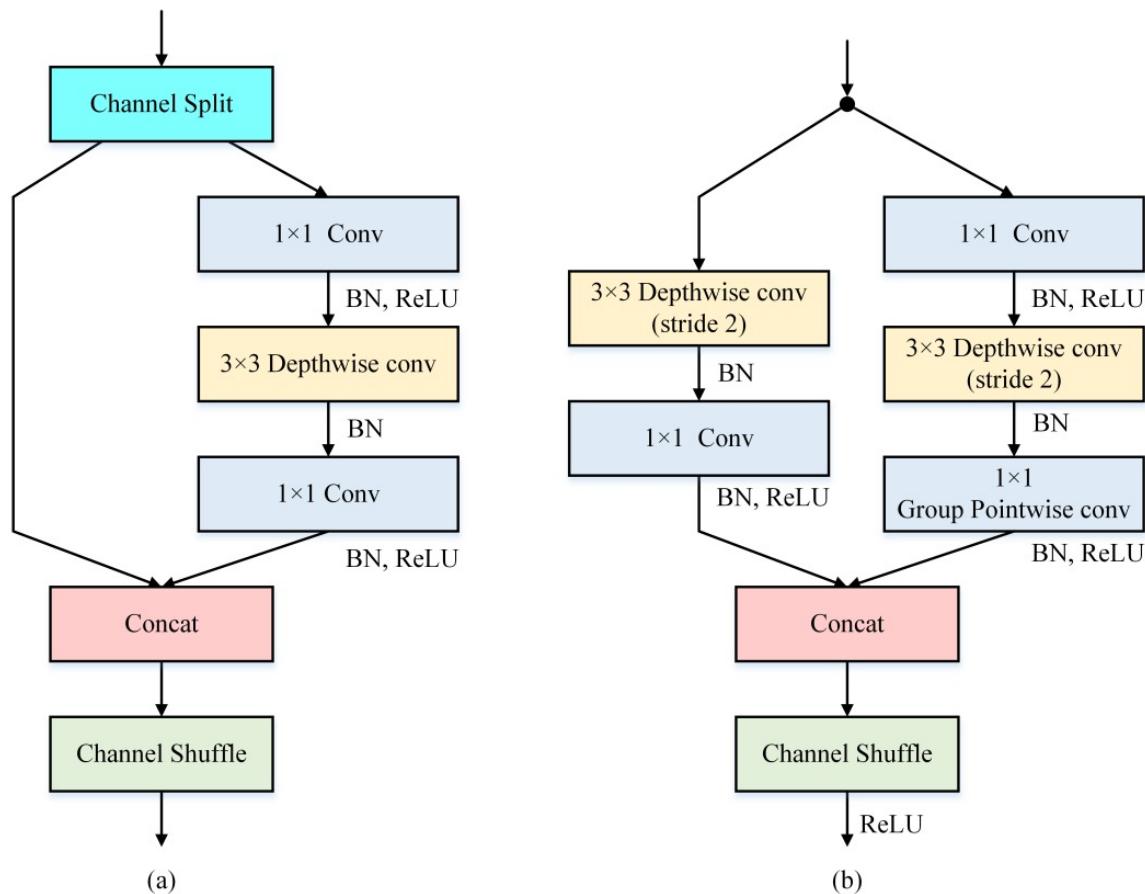


Figura 2.29: El bloque residual de cuello de botella de MobileNetV2. C es el número de canales y las relaciones de expansión son 6. (**tecnica2**)

■ **Data Augmentation:** as CNN a menudo enfrentan el riesgo de sobreajuste debido a datos limitados. Las técnicas tradicionales de aumento de datos incluyen una colección de métodos, como voltear, cambiar el espacio de color, recortar, rotar, trasladar e inyectar ruido, que pueden mejorar los atributos y el tamaño del conjunto de datos de entrenamiento. Además, tienen el potencial de mejorar significativamente la generalización de los modelos de DL. El aumento de datos automatizado tiene el potencial de abordar algunas debilidades de los métodos tradicionales de aumento de datos, ya que entrenar un modelo CNN con una política de aumento de datos aprendida puede mejorar significativamente la precisión, la robustez del modelo y el rendimiento en el aprendizaje semi-supervisado para la clasificación de imágenes. La técnica de Aumento de Datos de Muestra Mixta (MSDA) consiste en mezclar aleatoriamente dos muestras de entrenamiento y sus etiquetas según una cierta proporción, lo que no solo puede reducir la identificación errónea de algunas muestras difíciles, sino también mejorar la robustez del modelo y hacerlo más estable durante el entrenamiento.

2.2.4.3. Conclusiones

Esta revisión abarca no solo los modelos convencionales de CNN, sino también métodos mixtos y estrategias de entrenamiento, destacando puntos clave en la clasificación de imágenes.

Los modelos clásicos de CNN (2012-2017) sentaron las bases para el diseño estructural actual. La integración del mecanismo de atención en las CNN, como los bloques SE, mejora su rendimiento. Las redes para plataformas móviles son más pequeñas y eficientes, aprovechando al máximo los recursos limitados. La elección de hiperparámetros es crucial, y la búsqueda NAS facilita el diseño de redes de alto rendimiento. El aprendizaje por transferencia y el aumento de datos mejoran las predicciones y el rendimiento.

Los modelos livianos sacrifican precisión por eficiencia, y aún se explora su uso en sistemas limitados. El campo de NLP ha avanzado más en el aprendizaje semi-supervisado y no supervisado.

Futuras direcciones incluyen la combinación de convolución y Transformer, como en la red CoAtNet, y la revisión de componentes convencionales de CNN, que podría llevar a avances significativos.

2.2.5. Vision Computer

Según **alonso2016vision**, la Visión por computadora también conocido como visión artificial es otra rama importante de la IA que desempeña la tarea de dotar al computador la función de captar y comprender una imagen con el objetivo de emular el proceso que realizan los humanos. A pesar de tener menos tiempo a comparación con las otras, desempeña una función básica, primordial y compleja del reconocimiento. Esta es capaz de aprender y reconocer formas para luego darles una clasificación correctamente. Posee funciones como el reconocimiento donde en la imagen se indaga un objeto singular o reconocer diferentes instancias de una categoría genérica donde se hace dicho reconocimiento de la instancia. Este reconocimiento está asignado para categorizar diferentes clases a los objetos, el instrumento que ejecuta este procedimiento se llama clasificador. Otro método es la representación de objetos, que se trata acerca de que los objetos sean identificados mediante segmentación de imágenes, pueden ser subdivididos en múltiples agrupaciones, desde el punto de agrupación, se provee de características comunes que tienen los objetos entre sí. El objeto medido según sus características es denominado patrón. Por último, tenemos al seguimiento de objetos en tiempo real, que consiste en que las técnicas o algoritmos de seguimiento estimen el movimiento de objetivo en el plano de la imagen mientras se da un contexto en movimiento, para eso el sistema coloca etiquetas fijas al objeto o los objetos a continuar durante continuidad de imágenes. La dificultad de esta técnica radica en los cambios inesperados en el movimiento, como la alteración en los aspectos de patrones, tal como de la escena, el propio objeto y las occlusiones entre objetos (**alonso2016vision**).

2.2.6. Vision Transformer: Explainability of Vision Transformers: A Comprehensive Review and New Perspectives (tecnica1)

Los transformers han sido importantes en el procesamiento del lenguaje y, recientemente, se han destacado en la visión por computadora. Aunque su funcionamiento interno no se comprende completamente, la explicabilidad es crucial. Este estudio revisa métodos de explicabilidad para transformers visuales, propone una taxonomía y ofrece criterios de evaluación y herramientas. También señala áreas no exploradas que podrían mejorar la explicabilidad y sugiere futuras investigaciones.

2.2.6.1. Introducción

La Inteligencia Artificial (IA) ha experimentado avances notables en los últimos años, principalmente impulsada por el éxito de las Redes Neuronales Profundas (DNNs) en una amplia gama de aplicaciones, como el diagnóstico médico, las aplicaciones financieras, las evaluaciones de riesgos y la generación de imágenes y videos. Estos logros han sido significativos en términos de rendimiento y precisión en diversas tareas, sin embargo, la aplicación práctica de las DNNs sigue siendo limitada debido a su naturaleza opaca y la falta de transparencia en su toma de decisiones.

La opacidad de las DNNs plantea preocupaciones importantes en términos de confiabilidad y seguridad, ya que los usuarios y los responsables de la toma de decisiones pueden tener dificultades para comprender por qué un modelo ha llegado a una determinada conclusión o recomendación. Esto es particularmente problemático en áreas donde las decisiones basadas en IA pueden tener un impacto significativo en la vida de las personas, como la salud y la justicia. La falta de transparencia también puede ocultar sesgos y errores en los modelos, lo que socava aún más la confianza en su uso.

Para abordar estos desafíos, ha surgido el campo de la Inteligencia Artificial Explicable (XAI), cuyo objetivo es mejorar la comprensión y la transparencia de los modelos de IA. La XAI busca proporcionar explicaciones claras y comprensibles sobre cómo se toman las decisiones por parte de los modelos de IA, permitiendo a los usuarios entender y confiar en sus resultados. Al comprender cómo funciona un modelo y por qué toma ciertas decisiones, los usuarios pueden evaluar mejor su fiabilidad y corregir posibles sesgos o errores.

En este contexto, los transformers, modelos basados en atención, han surgido como una alternativa poderosa a las arquitecturas de redes neuronales convolucionales (CNNs) tradicionales en áreas como el Procesamiento del Lenguaje Natural (NLP) y la Visión por Computadora (CV). Los transformers han demostrado ser altamente efectivos para capturar relaciones a largo plazo en datos secuenciales, lo que los hace adecuados para tareas que requieren un procesamiento de contexto complejo.

En particular, los Vision Transformers (ViT) aplican la arquitectura de transformers a la tarea de visión por computadora, lo que ha llevado a avances significativos en tareas como el reconocimiento de imágenes, la detección de objetos y la segmentación de imágenes. Los ViT han demostrado resultados comparables e incluso superiores a los obtenidos con CNNs en algunas aplicaciones.

Sin embargo, a pesar de su éxito, la comprensión y la explicabilidad de los transformers, especialmente en el contexto de la visión por computadora, siguen siendo áreas de investiga-

ción activa. Se necesitan métodos y técnicas robustas para explicar las decisiones tomadas por estos modelos, así como herramientas y marcos para evaluar y comparar sus explicaciones. Al mejorar la explicabilidad de los transformers, podemos fortalecer la confianza en su uso y abrir nuevas oportunidades para aplicaciones prácticas en una variedad de campos.

2.2.6.2. Arquitectura de Vision Transformer

Los Vision Transformers (ViTs) aprovechan los avances de los transformers en NLP para aplicaciones visuales. Utilizan bloques de transformers que permiten integrar información global en toda la imagen mediante auto-atención. Las imágenes se representan como secuencias de tokens, con una capa de incrustación de parches para dividirlas en parches. Estos parches se aplanan y se tratan como tokens, luego se transforman en vectores de características. Para la clasificación, se agrega un token de clasificación. La red ViT también incluye otros componentes como una activación GELU, normalización de capa y conexión residual. Se puede encontrar una ilustración de la arquitectura ViT en la Figura 7.

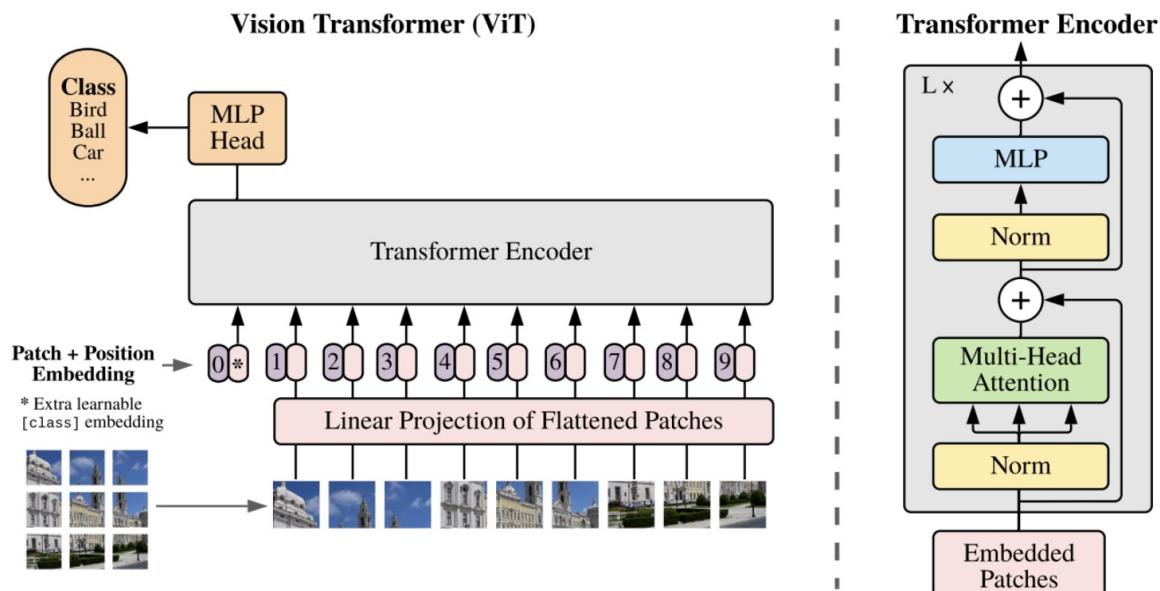


Figura 2.30: Arquitectura de Vision Transformer (técnica1)

2.2.6.3. Explicación y métodos para Visual Transformer

Tras los trabajos innovadores presentados basados en transformers visuales para diversos dominios de visión por computadora, han surgido múltiples enfoques para mejorar la explicabilidad de estas redes. Sin embargo, se necesita una encuesta exhaustiva para comprender mejor estos métodos e identificar áreas de mejora. Con un enfoque en la tarea de clasificación,

esta sección presenta una visión general de las técnicas de explicabilidad existentes para transformers visuales. Para proporcionar una visión clara, categorizamos y resumimos estos métodos en cinco grupos distintos según sus procedimientos de trabajo, motivaciones y características estructurales, como se ilustra en la Figura 8.

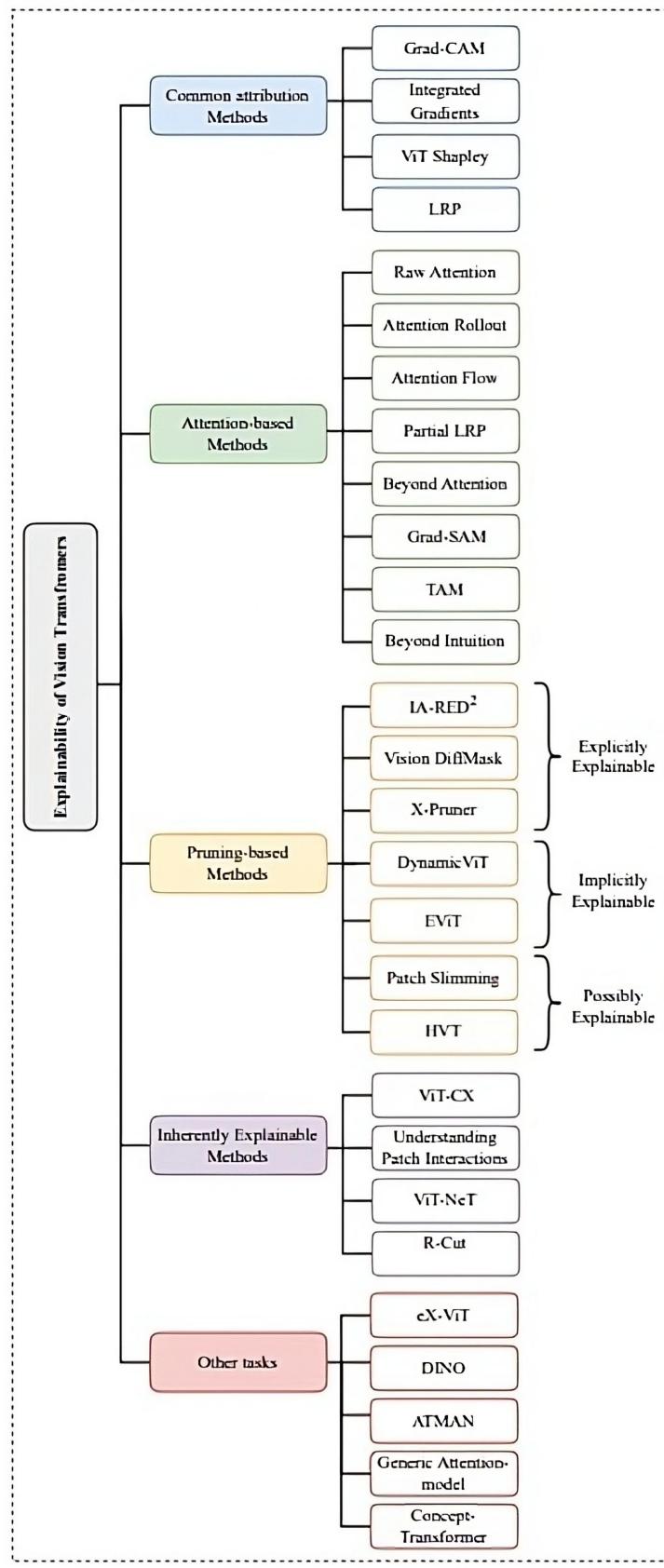


Figura 2.31: Taxonomía de los métodos de explicabilidad para Transformadores de visión (técnica1)

2.2.6.4. Métodos basados en atención

Los métodos basados en atención aprovechan el mecanismo de atención de los modelos para identificar y priorizar las partes más relevantes de una secuencia de entrada. Muchos enfoques existentes se centran en utilizar los pesos de atención o el conocimiento codificado en ellos para explicar el comportamiento del modelo. En aplicaciones basadas en visión, la visualización de los pesos de atención puede ser útil para identificar patrones de atención, aunque puede volverse menos confiable a medida que la red crece más profunda y más compleja. Para superar estos desafíos, se han introducido dos métodos: "Attention Rollout" y "Attention Flow", que cuantifican el flujo de información y aproximan la atención a los tokens de entrada de manera más holística. Estos métodos tienen sus limitaciones, por lo que se han desarrollado enfoques adicionales, como "GradSAM" y "Transition Attention Maps" (TAM), que aplican funciones como gradientes a los pesos de atención para mejorar la explicación de las predicciones del modelo. Además, el marco "Beyond Intuition" propone una aproximación novedosa para aproximarse a las contribuciones de los tokens, operando en dos etapas: percepción de atención y retroalimentación de razonamiento.

Tabla 2.4: Methods for Attention-based Class-specific Multi-modality

Method	Attention	Class-specific	Multi-modality	Backbone	Date
Raw Attention	Yes	No	No	VIT, DEIT	2017
Attention Rollout	Yes	No	No	VIT, DEIT	2020
Attention Flow	Yes	No	No	VIT, DEIT	2020
Partial LRP	Yes	No	No	VIT	2019
Grad-SAM	Yes	Yes	No	VIT	2021
Beyond Attention	Yes	Yes	Yes	VIT	2021
TAM	Yes	Yes	No	VIT, DEIT	2021
Beyond Intuition	Yes	Yes	Yes	BERT, VIT, CLIP	2023

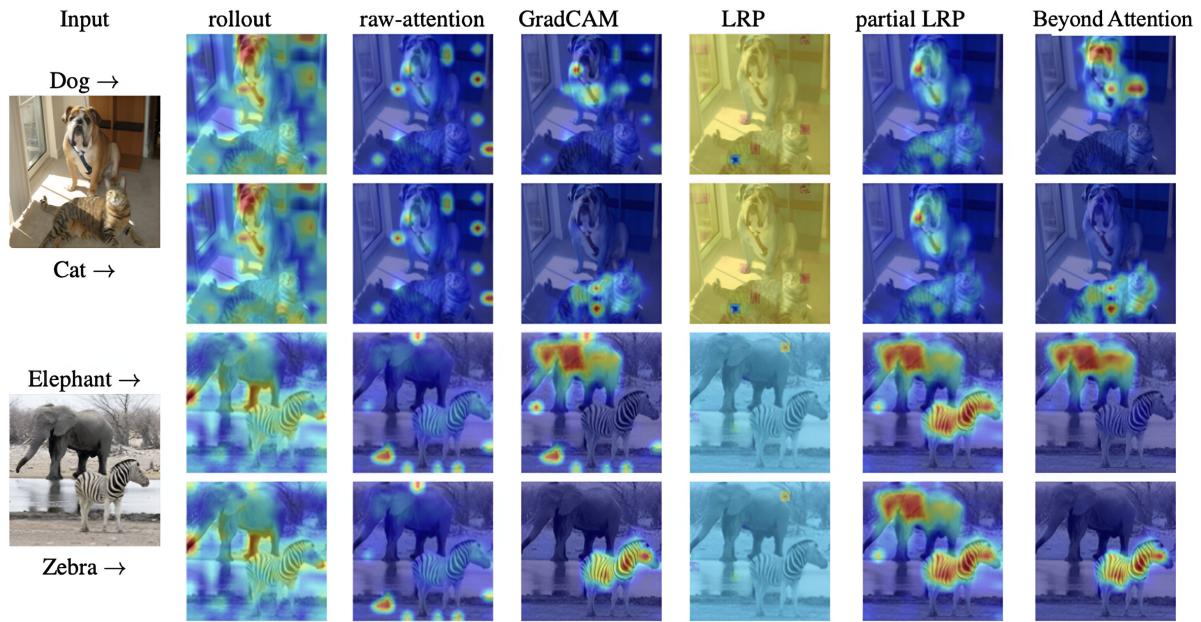


Figura 2.32: Visualizaciones específicas de clase de varios métodos basados en la atención, para cada imagen se pueden ver resultados de dos clases diferentes(**técnica1**)

2.2.6.5. Métodos basados en la poda

Los métodos basados en poda son una poderosa herramienta utilizada para optimizar la eficiencia y complejidad de los transformers. Estos métodos intentan eliminar elementos redundantes o poco informativos como tokens, parches, bloques o cabezas de atención de las redes. Algunos de estos métodos están explícitamente desarrollados con fines de explicabilidad, mientras que otros se centran principalmente en mejorar la eficiencia y no tienen como objetivo específico la explicabilidad. Sin embargo, estudios indican que las técnicas del segundo grupo también pueden impactar positivamente la explicabilidad del modelo. Se pueden categorizar los métodos de poda basados en ViT en tres grupos: métodos explícitamente explicables, implícitamente explicables y posiblemente explicables.

Entre los métodos de poda explícitamente explicables, hay varios enfoques notables que buscan proporcionar modelos menos complejos y más interpretables. Por ejemplo, el método IA-RED2 busca encontrar el equilibrio perfecto entre eficiencia e interpretabilidad al eliminar dinámicamente los parches menos informativos, lo que resulta en una velocidad significativamente mayor con una pérdida mínima de precisión. Otro método, X-Pruner, está diseñado específicamente para podar unidades menos significativas, logrando importantes ahorros computacionales sin perder precisión.

Por otro lado, los métodos de poda implícitamente explicables, como el marco Dy-

namicViT, están diseñados principalmente para mejorar la eficiencia de la red, pero también pueden mejorar la explicabilidad al localizar las partes críticas de la imagen que contribuyen más a la clasificación. EViT es otro enfoque innovador que reorganiza los tokens de una imagen basándose en el concepto de atención, manteniendo los tokens más relevantes mientras fusiona los menos atentos en un solo token, lo que reduce los costos computacionales sin comprometer la precisión del modelo. Estos métodos mejoran la interpretabilidad y ofrecen una comprensión más clara de las decisiones del modelo.

Tabla 2.5: Comparacion de metodos de poda basados en DeiT-S Touvron en 2021 en el conjunto de datos ImageNet

Método de poda	GFLOPs ↓ (%)	TOP-1 Exactitud ↓ (%)	Rendimiento ↑ (%)
IA-RED2	–	0.7	46
DynamicViT	37	0.5	54
EViT	35	0.3	50
HVT	47.8	1.8	–

Los métodos posiblemente explicables son enfoques adicionales de poda que, aunque inicialmente no se diseñaron para mejorar la interpretabilidad de ViT, podrían ofrecer un potencial para investigar su impacto en la explicabilidad de los modelos. Por ejemplo, Patch Slimming es un algoritmo novedoso que acelera ViTs al dirigirse a los parches redundantes en las imágenes de entrada, lo que potencialmente destaca características visuales importantes y mejora la interpretabilidad. Otro enfoque, Hierarchical Visual Transformer (HVT), mejora la escalabilidad y el rendimiento de ViTs al reducir gradualmente la longitud de la secuencia a medida que aumenta la profundidad del modelo. Aunque estos métodos se han evaluado principalmente en términos de eficiencia, existe una brecha significativa en la literatura en cuanto a la evaluación de su explicabilidad. En contraste, los métodos inherentemente explicables, como ViT-CX, se centran en desarrollar modelos que puedan explicarse a sí mismos, utilizando herramientas interpretables como mapas de saliencia para proporcionar explicaciones más significativas.

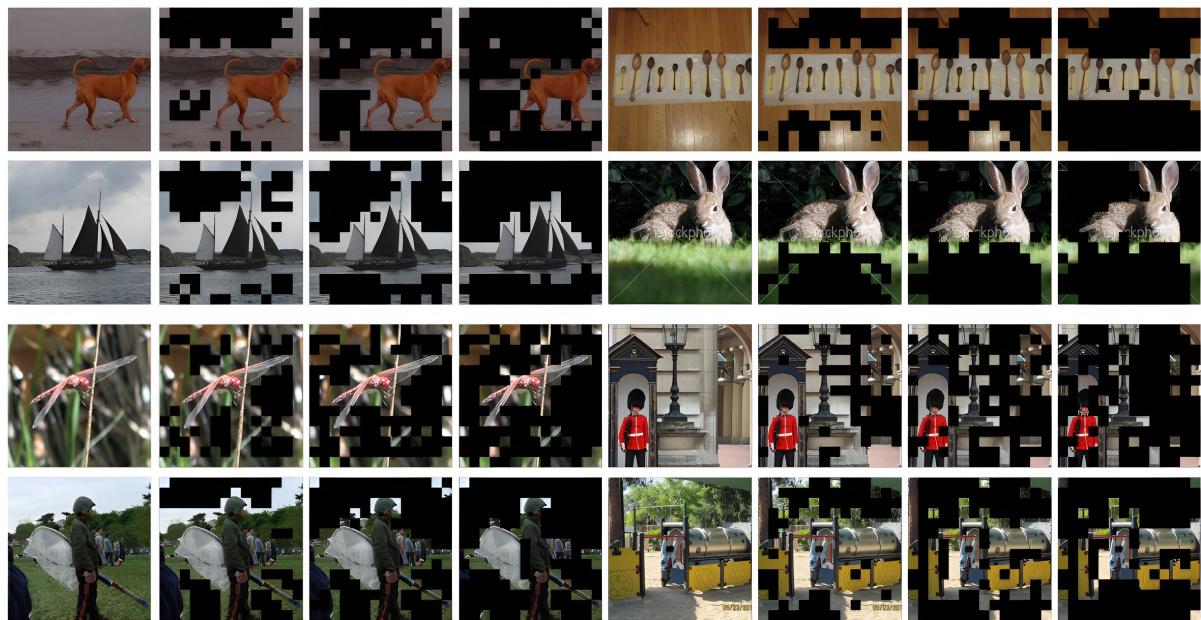


Figura 2.33: Visualización de tokens desatentos en EViT-DeiT-S con 12 capas; Se puede ver que las fichas de falta de atención se fusionan gradualmente (como se representa mediante áreas enmascaradas) o se eliminan, mientras que las fichas más informativas se conservan. Esto permite a los ViT centrarse en tokens específicos de clase en imágenes, lo que conduce a una mejor interpretabilidad(**técnica1**)

2.2.6.6. Evaluación de explicación

En secciones anteriores, presentamos varias técnicas de explicación desarrolladas específicamente para aplicaciones basadas en ViT. Sin embargo, evaluar qué tan bien estas técnicas representan el proceso de razonamiento de un modelo presenta diferentes desafíos. Para abordar esta preocupación, la literatura sugiere una serie de criterios evaluativos, que ayudan a seleccionar y diseñar la técnica de explicabilidad más apropiada. A continuación, se resumen estos criterios:

- **Deletion and Insertion:** Se utilizan para evaluar la fidelidad de un mapa de saliencia al modelo objetivo. Calculan cómo el mapa de saliencia identifica los píxeles más influyentes para la predicción del modelo.
- **Effective Complexity:** Evalúa el número de atribuciones que superan un umbral, indicando la importancia o insignificancia de las características correspondientes.
- **Faithfulness:** Método para evaluar la calidad de las atribuciones de características sin intervención humana, midiendo cuán precisamente las atribuciones de características se alinean o correlacionan con las predicciones del modelo.

- **(In)fidelity:** Se utiliza para evaluar qué tan bien una explicación captura los cambios en las predicciones de un modelo cuando la entrada sufre perturbaciones significativas.
- **Intersection over Union (IoU) test:** Métrica estándar para evaluar el rendimiento de los detectores y seguidores de objetos, que también se ha utilizado para evaluar métodos de explicabilidad midiendo la superposición entre los mapas de explicabilidad predichos y las cajas delimitadoras de la verdad terrenal de los objetos de interés.
- **Perturbation Tests:** Funcionan al enmascarar gradualmente tokens de entrada basándose en las explicaciones proporcionadas por el método de explicabilidad dado.
- **Pointing Game:** Método para evaluar los mapas de saliencia de la explicación en comparación con las cajas delimitadoras anotadas por humanos.
- **Segmentation Tests:** Consideran cada visualización como una segmentación suave de la imagen y las comparan con la verdad terrenal proporcionada en el conjunto de datos.
- **Sensitivity:** Evalúa cómo varían las explicaciones con pequeñas perturbaciones en la entrada.
- **Sensitivity-n:** Propuesto para probar valores de atribución específicos en lugar de considerar solo las clasificaciones de importancia.

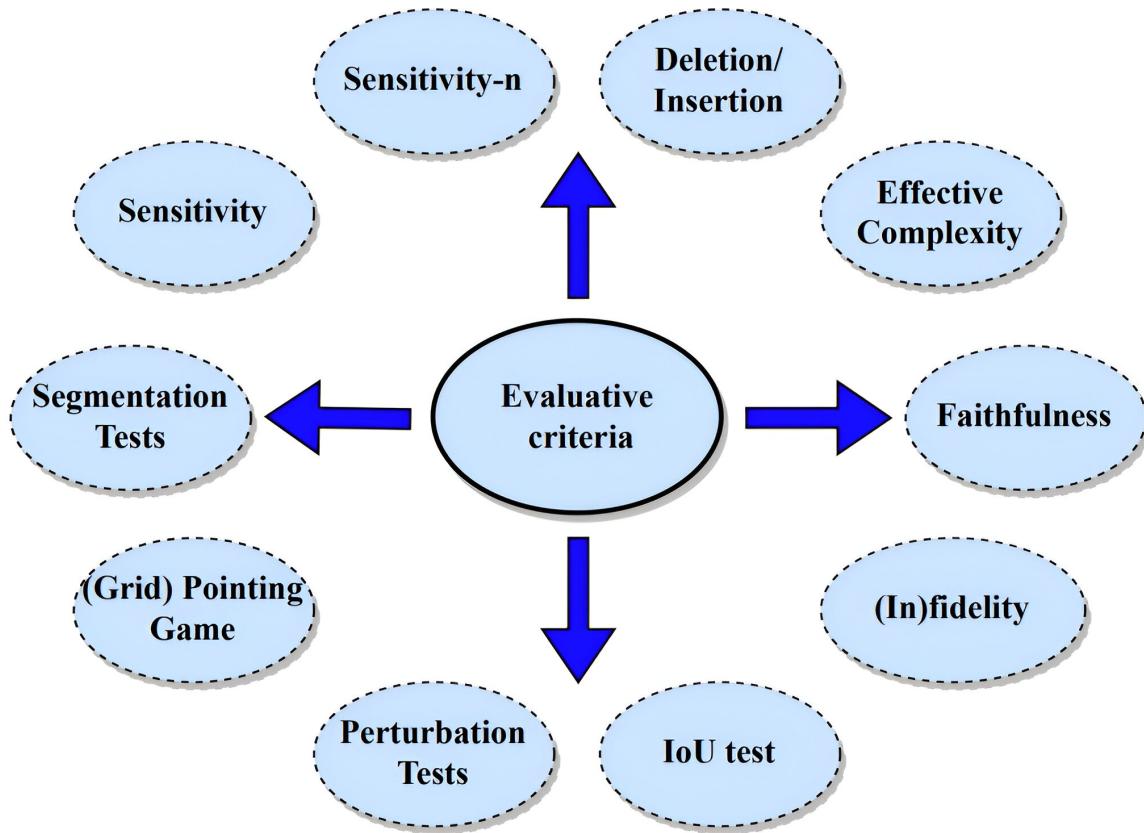


Figura 2.34: Diferentes criterios para evaluar los métodos de explicabilidad en aplicaciones basadas en la visión(**técnica1**)

2.2.6.7. Conclusión

En resumen, este trabajo ofrece una visión completa de las técnicas de explicabilidad propuestas para los transformers visuales. Hemos proporcionado una taxonomía de los métodos basada en sus motivaciones, estructuras y escenarios de aplicación, categorizándolos en cinco grupos. Además, detallamos los criterios de evaluación de la explicabilidad, así como las herramientas y marcos de trabajo utilizados. Por último, discutimos varios problemas esenciales pero poco explorados para mejorar la explicabilidad de los transformers visuales y sugerimos direcciones de investigación potenciales para futuras inversiones. Esperamos que este artículo de revisión ayude a los lectores a comprender mejor los mecanismos internos de los transformers visuales, así como a resaltar problemas abiertos para trabajos futuros.

2.2.7. You Only Look One (YOLO): You Only Look Once: Unified, Real-Time Object Detection (tecnica4)

YOLO trata la detección de objetos como un problema de regresión, prediciendo cuadros delimitadores y probabilidades de clase directamente de imágenes completas en una sola evaluación. Esto permite una arquitectura muy rápida, procesando imágenes en tiempo real a 45 cuadros por segundo, con Fast YOLO alcanzando 155 cuadros por segundo. Aunque YOLO puede cometer más errores de localización, es menos propenso a falsos positivos de fondo y generaliza mejor a diferentes dominios como el arte.

2.2.7.1. Introducción

En la detección de objetos, los sistemas actuales utilizan enfoques complejos que reutilizan clasificadores para detectar objetos, lo que los hace lentos y difíciles de optimizar. En contraste, el enfoque de YOLO reframes la detección como un problema de regresión única, lo que permite predecir objetos y sus ubicaciones directamente desde los píxeles de la imagen en una sola evaluación. Esto ofrece varias ventajas: en primer lugar, YOLO es extremadamente rápido, con una tasa de ejecución de hasta 150 cuadros por segundo, lo que permite el procesamiento de video en tiempo real con una latencia mínima. Además, YOLO considera globalmente la imagen al realizar predicciones, lo que le permite codificar implícitamente información contextual sobre las clases y su apariencia. Por último, YOLO aprende representaciones generalizables de objetos, lo que lo hace menos propenso a descomponerse al aplicarse en nuevos dominios o entradas inesperadas. Sin embargo, YOLO aún se queda atrás en precisión en comparación con otros sistemas de detección de vanguardia, especialmente en la localización precisa de objetos pequeños.

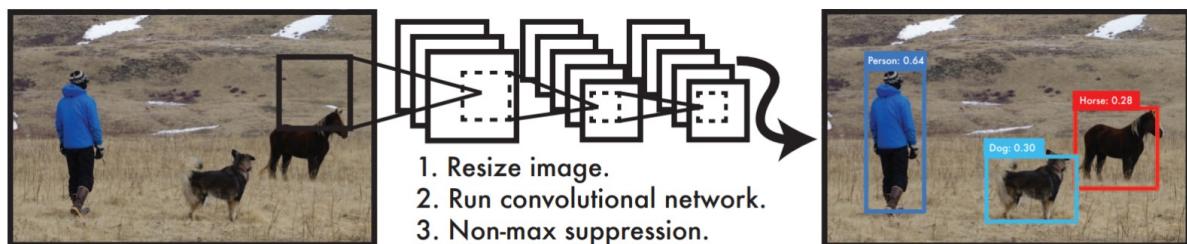


Figura 2.35: El sistema de detección YOLO. Procesamiento de imágenes con YOLO es simple y directo. Nuestro sistema (1) cambia de tamaño la imagen de entrada a 448×448 , (2) ejecuta una única red convolucional en la imagen y (3) establece umbrales para las detecciones resultantes mediante La confianza del modelo.(tecnica4)

2.2.7.2. Detección unificada

El método fusiona todos los elementos de detección de objetos en una única red neuronal. Esta red utiliza características de toda la imagen para predecir simultáneamente las cajas delimitadoras y las clases de objetos, permitiendo un entrenamiento de extremo a extremo y velocidades en tiempo real sin perder precisión. La imagen se divide en una cuadrícula $S \times S$, donde cada celda detecta un objeto si su centro cae en ella. Cada celda predice B cajas delimitadoras y sus puntuaciones de confianza, indicando la certeza del modelo sobre la presencia y precisión del objeto, además de C probabilidades de clase condicionales. Durante la prueba, se combinan estas probabilidades y las predicciones de confianza para obtener puntuaciones específicas de clase para cada caja.

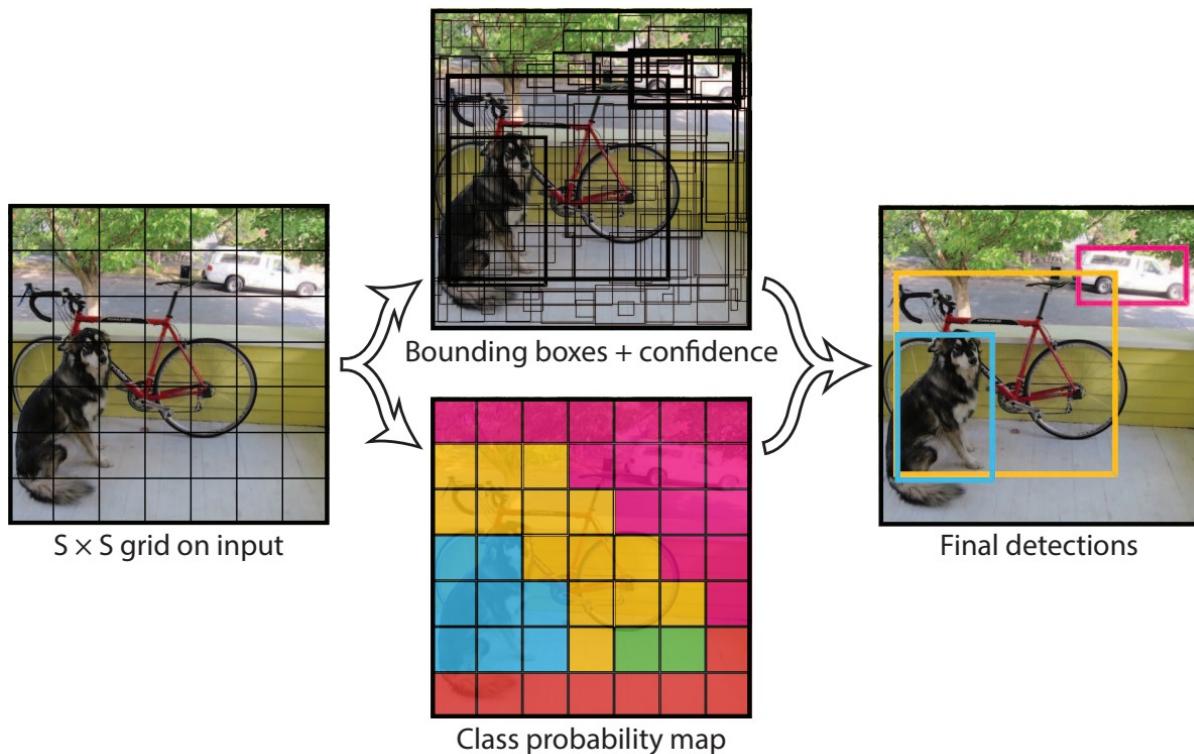


Figura 2.36: El sistema aborda la detección como un problema de regresión, donde la imagen se divide en una cuadrícula $S \times S$. Para cada celda de esta cuadrícula, se predicen cuadros delimitadores, confianza asociada con esos cuadros y probabilidades de clase. Estas predicciones se codifican en un tensor con dimensiones $S \times S \times (B \times 5 + C)$. (**técnica4**)

Este modelo, una red neuronal convolucional, se evalúa en el conjunto de datos de detección PASCAL VOC. Utiliza capas convolucionales para extraer características de la imagen y capas completamente conectadas para predecir las salidas. Basado en GoogLeNet, consta de 24 capas convolucionales y 2 capas completamente conectadas. En lugar de módulos de

inception, emplea capas de reducción de 1×1 seguidas de convoluciones de 3×3 . Además, desarrolla una versión rápida de YOLO con menos capas convolucionales y filtros para mejorar la detección rápida de objetos.

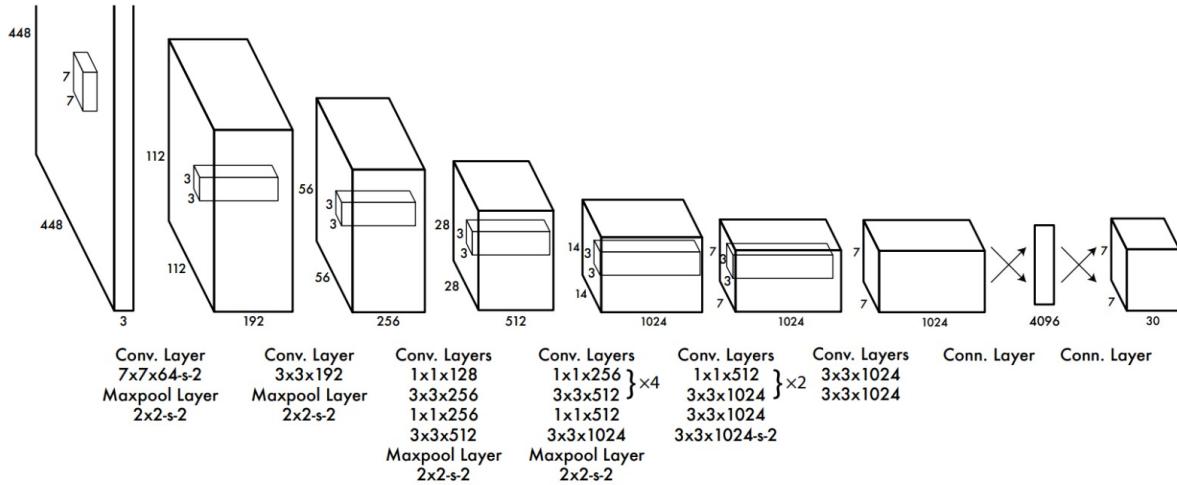


Figura 2.37: La red de detección consta de 24 capas convolucionales y 2 capas completamente conectadas. Se utilizan capas convolucionales de 1×1 para reducir el espacio de características de las capas anteriores y se preentrenan en la tarea de clasificación de ImageNet a la mitad de la resolución (224×224), incrementando la resolución para la detección. (**técnica4**)

Training Las capas convolucionales se preentrenan en el conjunto de datos ImageNet de 1000 clases, logrando una precisión top-5 del 88 %. La red se convierte para la detección, aumentando la resolución de entrada de 224×224 a 448×448 para capturar detalles. La capa final predice las probabilidades de clase y las coordenadas de las cajas delimitadoras, optimizando para el error cuadrático medio. Se entrena durante aproximadamente 135 épocas en los conjuntos de datos de PASCAL VOC 2007 y 2012, utilizando un tamaño de lote de 64 y una tasa de aprendizaje variable. Durante la inferencia, predice detecciones con una sola evaluación de red, empleando supresión no máxima para corregir múltiples detecciones..

Limitaciones de YOLO YOLO impone restricciones espaciales significativas en las predicciones de las cajas delimitadoras. Cada celda de la cuadrícula solo puede predecir dos cajas y asignar una única clase, lo que limita la capacidad del modelo para detectar objetos cercanos, especialmente objetos pequeños que aparecen en grupos, como bandadas de pájaros.

Dado que el modelo aprende de los datos, enfrenta dificultades para generalizar a objetos con relaciones de aspecto nuevas o inusuales. Además, utiliza características relativamente gruesas para predecir las cajas delimitadoras debido a múltiples capas de muestreo descendente

desde la imagen de entrada.

La función de pérdida utilizada durante el entrenamiento no distingue entre errores en cajas delimitadoras pequeñas y grandes, lo que puede llevar a una penalización desproporcionada por errores en cajas pequeñas, que tienen un impacto significativo en la evaluación de la superposición de IOU. En consecuencia, las localizaciones incorrectas son la principal fuente de error en nuestro modelo.

2.2.7.3. Experimentos

Se contrastó YOLO con otros sistemas de detección en tiempo real en PASCAL VOC 2007, revelando que YOLO y Fast R-CNN muestran distintos perfiles de errores. Asimismo, se evidenció la capacidad de YOLO para generalizar mejor en nuevos dominios, como conjuntos de datos de obras de arte, en comparación con otros métodos actuales en VOC 2012.

Dentro del ámbito de la detección de objetos en tiempo real, la mayoría de los esfuerzos de investigación se enfocan en mejorar la velocidad de los procesos estándar de detección. YOLO fue evaluado frente a otros métodos, resaltando su rapidez y precisión. Se introdujo una versión más ágil, Fast YOLO, que sobrepasa considerablemente a los enfoques previos en cuanto a precisión y velocidad.

Además, se examinaron otros métodos como Fastest DPM, R-CNN menos R, Fast R-CNN y Faster R-CNN, destacando sus velocidades y precisión en comparación con YOLO. Sin embargo, ninguno de ellos logra igualar el desempeño en tiempo real de YOLO.

Tabla 2.6: Comparación de Detectores en Tiempo Real y Menos que en Tiempo Real

Detector	Entrenamiento	mAP	FPS
100Hz DPM DPM	2007	16.0	100
30Hz DPM DPM	2007	26.1	30
Fast YOLO	2007+2012	52.7	155
YOLO	2007+2012	63.4	45
Menos que en Tiempo Real			
Fastest DPM DPM	2007	30.4	15
R-CNN Minus R RCNN	2007	53.5	6
Fast R-CNN FastRCNN	2007+2012	70.0	0.5
Faster R-CNN VGG-16 FasterRCNN	2007+2012	73.2	7
YOLO VGG-16	2007+2012	66.4	21

Para examinar las diferencias entre YOLO y detectores de última generación, se realizó un análisis detallado de los resultados en VOC 2007, comparando YOLO con Fast R-CNN. Usando la metodología de Hoiem, se clasificaron las predicciones en categorías basadas en el tipo de error:

- **Correcto:** clase correcta e IOU ≥ 0.5
- **Localización:** clase correcta, $0.1 \leq \text{IOU} < 0.5$
- **Similar:** clase similar, $\text{IOU} \geq 0.1$
- **Otro:** clase incorrecta, $\text{IOU} < 0.1$
- **Fondo:** $\text{IOU} < 0.1$ para cualquier objeto

La Figura 2.38 muestra el desglose promedio de cada tipo de error en las 20 clases. YOLO tiene más errores de localización que cualquier otra fuente combinada, mientras que Fast R-CNN comete menos errores de localización pero muchos más errores de fondo. El 13.6% de las detecciones de Fast R-CNN son falsos positivos sin objetos, siendo casi 3 veces más propenso a predecir detecciones de fondo en comparación con YOLO.

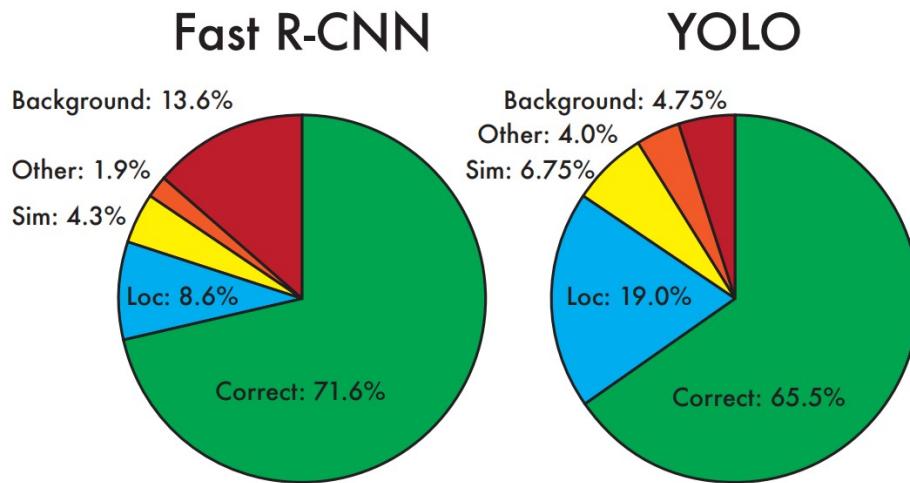


Figura 2.38: Análisis de Errores: Fast R-CNN vs. YOLO. Estos gráficos muestran el porcentaje de errores de localización y de fondo en las principales detecciones de varias categorías.

YOLO comete muchos menos errores de fondo que Fast R-CNN. Al usar YOLO para eliminar las detecciones de fondo de Fast R-CNN, se obtiene una mejora significativa en el rendimiento. Por cada cuadro delimitador que predice R-CNN, se verifica si YOLO predice un cuadro similar. Si es así, se le da un impulso a esa predicción basado en la probabilidad predicha por YOLO y la superposición entre los dos cuadros.

El mejor modelo de Fast R-CNN alcanza un mAP de 71.8 % en el conjunto de prueba de VOC 2007. Cuando se combina con YOLO, su mAP aumenta a 75.0 %, con una ganancia de 3.2 %, como se muestra en la Tabla 2.7.

Modelo	mAP	Combinado	Ganancia
Fast R-CNN	71.8	-	-
Fast R-CNN (2007 data)	66.9	72.4	0.6
Fast R-CNN (VGG-M)	59.2	72.4	0.6
Fast R-CNN (CaffeNet)	57.1	72.1	0.3
YOLO	63.4	75.0	3.2

Tabla 2.7: Experimentos de combinación de modelos en VOC 2007. Se examina el efecto de combinar varios modelos con la mejor versión de Fast R-CNN.

El aumento de rendimiento no es simplemente un subproducto de la combinación de modelos, ya que la combinación de diferentes versiones de Fast R-CNN produce beneficios pequeños (entre 0.3 y 0.6 %). En cambio, la eficacia de YOLO se debe a que comete diferentes tipos de errores en las pruebas, lo que mejora significativamente el rendimiento de Fast R-CNN.

2.2.7.4. Detección en Tiempo Real en el Mundo Real

YOLO es un detector de objetos rápido y preciso, lo que lo hace ideal para aplicaciones de visión por computadora. Se conectó YOLO a una cámara web y se verificó que mantiene el rendimiento en tiempo real, incluyendo el tiempo para capturar imágenes desde la cámara y mostrar las detecciones.

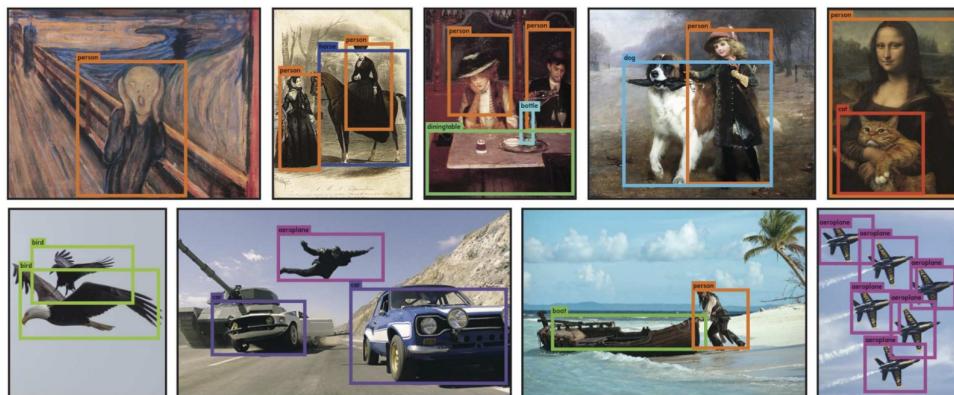


Figura 2.39: Resultados cualitativos. YOLO se ejecuta con obras de arte de muestra e imágenes naturales de Internet. Es mayoritariamente exacto, aunque cree que una persona es un avión.(técnica4)

El sistema resultante es interactivo y atractivo. Aunque YOLO procesa imágenes individualmente, cuando se conecta a una cámara web funciona como un sistema de seguimiento,

detectando objetos a medida que se mueven y cambian de apariencia.

2.2.7.5. Conclusion

YOLO es un detector de objetos rápido y preciso, ideal para aplicaciones de visión por computadora. Al conectarse a una cámara web, mantiene el rendimiento en tiempo real, incluyendo la captura y visualización de imágenes. El sistema es interactivo y funciona como un sistema de seguimiento, detectando objetos en movimiento

2.3. Marco Conceptual

2.3.1. Vision Computacional (vc1)

La visión computacional es un campo de la inteligencia artificial que se enfoca en simular la visión humana y la cognición. Utiliza métodos computacionales y algoritmos para procesar imágenes y videos (incluyendo datos 3D) con el objetivo de:

- **Medir:** Cuantificar características y dimensiones en imágenes.
- **Clasificar:** Identificar y categorizar objetos dentro de las imágenes.
- **Interpretar información visual:** Comprender y derivar significado del contenido visual para crear modelos del mundo real.

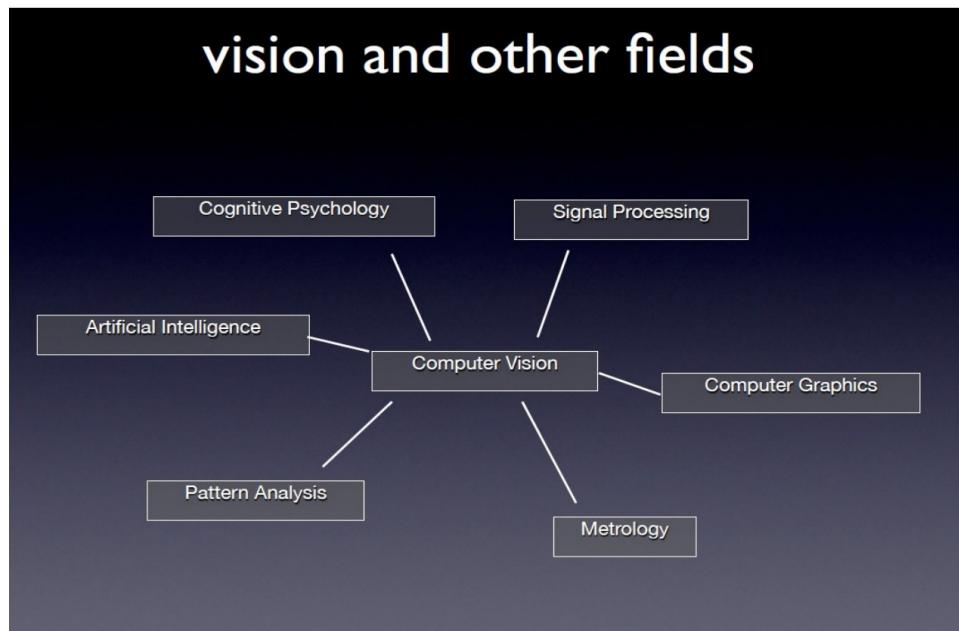


Figura 2.40: Los otros campos de Vision computacional

La visión computacional tiene diversas aplicaciones en múltiples campos:

- **Inteligencia Artificial (IA):** La visión actúa como la etapa de entrada.
- **Medicina:** Estudio de la visión humana y cirugía asistida.
- **Ingeniería y Computación:** Modelado y extracción de modelos.
- **Gráficas:** Generación de contenido y creación.

2.3.1.1. ¿Para qué estudiamos Visión Computacional?

El objetivo de la visión computacional (VC) es hacer juicios prácticos sobre objetos físicos del mundo real (escenas) a partir de imágenes digitales capturadas.

Por lo tanto, la tarea de la VC implica crear descriptores de la escena basados en características relevantes encontradas en una imagen..

2.3.1.2. Modelo de Imagen “Pinhole”

La proyección de perspectiva puede producir imágenes invertidas. En algunas situaciones, es beneficioso tener en cuenta la imagen virtual asociada con el plano situado frente al "pinhole", a la misma distancia que el plano de la imagen, dependiendo del contexto.

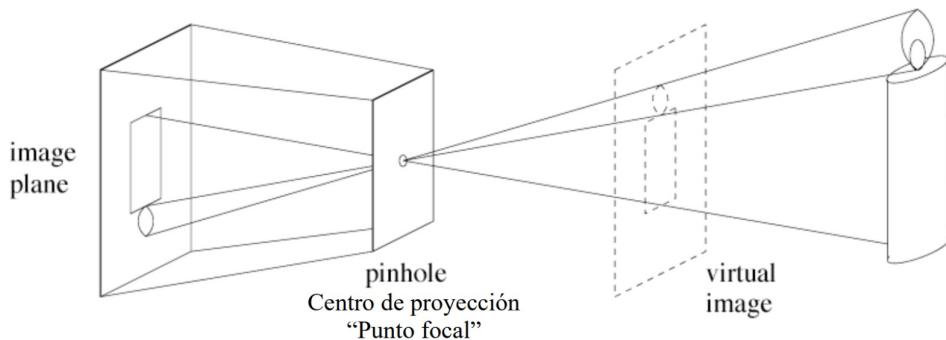


Figura 2.41: Modelo pinhole

2.3.1.3. Efectos de Perspectiva

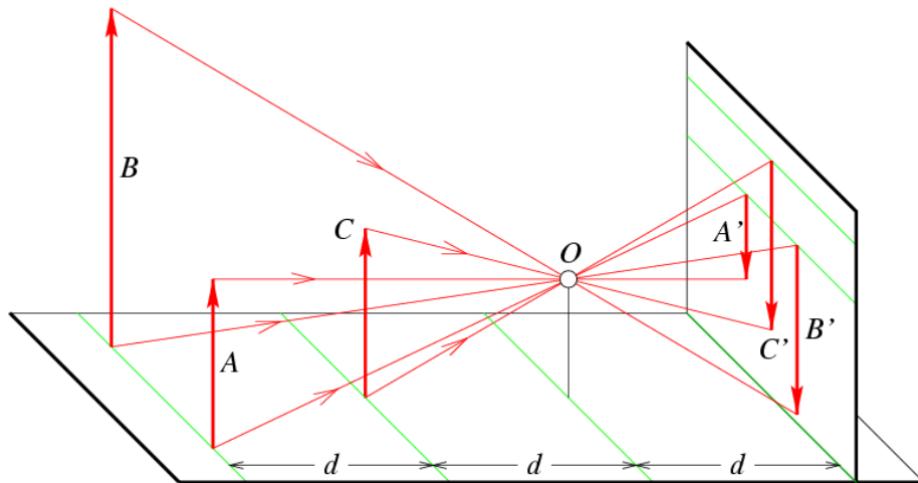


Figura 2.42: Efectos de Perspectiva

- **Tamaño Aparente**: Líneas C' y B' simulan ser similares en cuanto a grandeza, por otro lado, $C = \frac{1}{2} B$ ya que la distancia de B a O es $2d$ y la distancia de C a O es d .
- **Punto de Fuga**: Las líneas paralelas en la escena se encuentran en un objetivo de fuga en la imagen, que es la intersección de las rectas paralelas en el panorama.

2.3.1.4. Metodos de Proyección

Simulacion de la visión humana

- Proyección de perspectiva débil: Esta paráfrasis enfatiza que la proyección en perspectiva

débil da como resultado una distorsión mínima de los tamaños relativos de los objetos en la escena.

- Proyección ortográfica: Se dice que la cámara se encuentra a una distancia continua de la escena, con rayos paralelos al eje ocular y ortogonales al plano de proyección.

2.3.1.5. Aberraciones

- **Geométricas:** Incluyen aberraciones esféricas, astigmatismo y distorsión.
- **Cromáticas:** Dependen de la longitud de onda de la luz.

Tipos de Aberraciones Geométricas

- Aberraciones esféricas: Los rayos exteriores tienden a converger en una posición distinta a los rayos más internos, creando un círculo de confusión.
- Astigmatismo: Diferente distancia focal para rayos inclinados.
- Distorsión: Puede ser corregida si se conocen los parámetros. Se presenta como distorsión en cojín (tele-foto) o en barril (gran-angular).

2.3.1.6. ¿Qué es una imagen?

- **Definición:** Una imagen es una matriz de valores organizada en dos dimensiones, que usualmente representa la intensidad de la radiación electromagnética
- **Características:**
 - **Muestreo espacial**
 - **Cuantización numérica:** Ejemplo, Negro = 0, Blanco = 255.
 - **Cuantización espectral**

2.3.1.7. Representación del Color

- **Motivación:**
 - En análisis automático, el color ayuda en la detección de objetos y en la caracterización de escenas.

- En el análisis humano, el color posibilita la diferenciación de una gama más amplia de matices e intensidades en contraste con los niveles de gris.

■ **Fundamentos del color:**

- La luz blanca se divide en un espectro de color con seis regiones: violeta, azul, verde, amarillo, naranja y rojo.
 - Los colores percibidos están determinados por la naturaleza de la luz reflejada por los objetos.
 - La luz visible es una banda estrecha del espectro electromagnético.
- **Síntesis aditiva:** Los colores se suman (objetos luminosos).
- **Síntesis sustractiva:** Los colores se restan (pigmentos).



Figura 2.43: La luz blanca se divide en un espectro de color que contiene 6 regiones: violeta, azul, verde, amarillo, naranja y rojo.

2.3.2. Rancha "Phytophthora Infestans" Tizón Tardío (prom2)

El cultivo de papa es esencial para la seguridad alimentaria mundial, pero enfrenta desafíos por enfermedades como el tizón tardío, causado por *Phytophthora infestans*. Este patógeno puede causar pérdidas significativas en la producción y afecta a todas las fases de desarrollo de la planta. Se han desarrollado diversas estrategias de manejo, incluyendo el control químico y cultural, pero se requiere un enfoque integrado para su control efectivo. Esta revisión se centra en la prevalencia, ciclo de vida y estrategias de manejo del tizón tardío.

2.3.2.1. Introducción

La papa (*Solanum tuberosum L.*) es un cultivo significativo, especialmente en Europa y Asia, donde más del 80 % de la producción mundial se concentra. China es el mayor productor

mundial de papas, seguido por Rusia e India. Sin embargo, enfermedades como el tizón tardío, causado por *Phytophthora infestans*, representan una amenaza seria para la producción. Esta enfermedad puede causar pérdidas considerables y afecta todas las etapas de crecimiento de la planta, incluyendo los tubérculos, lo que puede llevar a una pérdida del 100 % de la producción. Las estrategias de manejo, como el control químico y cultural, son importantes para su control. La tizón tardío es especialmente devastador en áreas donde la papa es un alimento básico, como en Etiopía, donde puede causar pérdidas del 31 al 100 %. A nivel mundial, se estima que la enfermedad causa pérdidas anuales de hasta 5 mil millones de dólares, y ha tenido un impacto significativo en eventos históricos, como la Gran Hambruna Irlandesa en la década de 1840.

2.3.2.2. Ciclo de vida y epidemiología del tizón tardío

El tizón tardío en los campos de papa comienza con la aparición de lesiones pequeñas, irregulares y de color verde claro a oscuro que exudan agua. Estas generalmente comienzan en las hojas inferiores y se propagan rápidamente en ambientes húmedos, formando áreas marrones deterioradas con bordes irregulares. Un crecimiento micelial blanco y mohoso se desarrolla alrededor de las lesiones en la superficie interna de las hojas. La infección puede matar y dañar hojas enteras, y el micelio crece profusamente entre las células del hospedero, causando lesiones de color verde marrón o amarillento que eventualmente se vuelven negras. La enfermedad puede persistir en plantas de papa, suelo, tubérculos infectados y tubérculos voluntarios que sobreviven el invierno. Los esporangios de *Phytophthora infestans* pueden dispersarse por el viento, la precipitación, el transporte mecánico y los animales, promoviendo la transmisión de la enfermedad. La enfermedad se desarrolla rápidamente en condiciones húmedas y temperaturas entre 15 y 25 °C, y puede causar la muerte de plantas enteras en pocos días o semanas.

2.3.2.3. *Phytophthora infestans* - Síntomas de la enfermedad, Modo de infección y propagación

Phytophthora infestans, un patógeno vegetal diploide, comparte rasgos comunes con los hongos y se reproduce asexualmente mediante esporangios con forma de limón que se propagan a través de mecanismos mecánicos, del viento y de la lluvia. Los zoosporos se forman en condiciones húmedas y frescas, mientras que los esporangios germinan en ambientes secos y cálidos. La reproducción sexual puede ocurrir entre hifas compatibles, lo que resulta en la producción de oosporos, que pueden sobrevivir en el suelo hasta por cuatro años. Tanto la reproducción asexual como la sexual contribuyen a la capacidad del patógeno para infectar nuevas plantas. Los esporangios, zoosporos u oosporos acceden a los tejidos de las plantas huésped a través de

los estomas o áreas lesionadas, donde estructuras especializadas llamadas haustorios ayudan en la absorción de nutrientes. El desarrollo de esporangios en esporangióforos conduce a la esporulación, con esporangios transportados por el aire hacia plantas sanas. La infección exitosa implica interacciones complejas entre el patógeno, el huésped y el medio ambiente, influenciadas por factores como la humedad, la temperatura, la virulencia del patógeno y la resistencia del huésped, lo que conduce a pérdidas agrícolas significativas.

Crop loss %	Country	Reference
10-75	West Bengal-India	[25]
20	Nepal	[26]
50-70	Pakistan	[27]
72	Ethiopia	[28]
75	England	[29]
80	Kenya	[30]

Figura 2.44: Pérdida de cosechas a nivel nacional debido al tizón tardío de la papa

2.3.2.4. Conclusion

El enfoque herbal ha mostrado promesa en la búsqueda de extractos vegetales con acción anti-oomycete, demostrando eficacia contra *P. infestans*. Algunas de estas sustancias han resultado tan efectivas como los fungicidas sintéticos en la prevención del crecimiento de *P. infestans* in vitro o en la reducción de la gravedad del tizón tardío en plantas hospederas. La resistencia de las plantas hospederas al patógeno puede ofrecer beneficios económicos a largo plazo para los agricultores y reducir las alteraciones en la estructura poblacional de *P. infestans*, lo que disminuye el potencial de resistencia a los fungicidas. Sin embargo, dado que no existe un método de manejo único que sea efectivo en todo el mundo debido a la introducción de nuevas cepas, es crucial desarrollar nuevos enfoques para tratar esta enfermedad y superar el problema de la resistencia.

Capítulo 3

METODOLOGÍA DE LA INVESTIGACIÓN

3.1. Diseño de la investigación

En esta sección del documento se explicará cual es el diseño, el tipo y el enfoque del trabajo de investigación, así como también la población y la muestra.

3.1.1. Diseño no experimental

El diseño es no experimental transversal, ya que las variables no serán manipuladas y serán analizadas tal como se encuentran. Es decir, tanto los datos visuales de las hojas de papa peruana afectadas por la enfermedad 'Rancha' (*Phytophthora infestans*) como las técnicas de visión computacional serán utilizados sin modificar ningún aspecto de las condiciones naturales. El objetivo es aplicar técnicas de visión computacional para clasificar tempranamente la presencia de la enfermedad en las hojas de papa. La recolección de datos se llevará a cabo en un periodo de tiempo específico, sin intervenir en el desarrollo natural de la enfermedad en las plantas

3.1.2. Tipo explicativo

El alcance de esta investigación es explicativo, ya que se enfoca en construir un modelo de visión computacional para la detección temprana de la enfermedad 'Rancha' (*Phytophthora infestans*) a partir de imágenes de hojas de papa. Esto busca entender cómo identificar si una

hoja de papa está afectada por el tizón tardío, estableciendo así una relación de causa y efecto. En otras palabras, al identificar un patrón específico en las hojas de papa con Rancha, se podrá determinar si están afectadas por la patología.

3.1.3. Enfoque cuantitativo

El enfoque de esta investigación es cuantitativo, dado que se emplearán técnicas de visión computacional, las cuales implican procesar datos de imagen en valores numéricos (vectores de características). Posteriormente, se usarán técnicas estadísticas para analizar estos datos y determinar la presencia de la enfermedad 'Rancha' (*Phytophthora infestans*) en las hojas de papa peruana.

3.2. Población y Muestra

Tabla 3.1: Poblacion y muestra

Poblacion	La población fueron todas las imágenes de hojas de papa infestadas con Tizón tardío hojas de papa sanas.
Muestra	La muestra se extrajo un conjunto de imágenes de los dataset Plant Diseases Training Dataset, el cual es una recopilación de otras agrupaciones de datos como PlantVillage, Potato Leaf Disease, Cassava Leaf Disease Dataset, etc. Se cuenta con 1777 imágenes de hojas de papa etiquetadas como sanas y 2020 como infestada con Tizón tardío. Adicionalmente, cabe mencionar, que para el propósito de esta investigación se utilizará 80 % del conjunto de imágenes total para Training y 20 % para el Testing de los modelos correspondientes.
Unidad de análisis	En este caso, la unidad de análisis es cada imagen individual de hojas de papa, tanto sanas como infestadas con Tizón tardío.
Variable y tipo de análisis	Variable cuantitativa y discreta, debido a que el presente trabajo de investigación está centrado en variables numéricas y la precisión o exactitud.



Figura 3.1: Hoja de papa infestada con Tizón tardío

3.3. Operacionalización de Variables

Definición de Variables

VARIABLE Y DEFINICIÓN	INDICADOR	FÓRMULA DE INDICADOR
TIZON TARDIO Enfermedad causada por un hongo que afecta a los tomates y papas, provocando manchas oscuras en las hojas y frutos	Hoja de papa	Zona con manchas oscuras en la hoja
VISION COMPUTACIONAL Disciplina científica que incluye métodos para adquirir, procesar, analizar y comprender las imágenes	Resolución de la imagen	Cantidad de pixeles en una imagen
	Preprocesamiento de la imagen	Nivel de nitidez, contraste o ruido de la imagen
MODELO DE CLASIFICACIÓN: Modelo con la capacidad de clasificar según la clase de la variable, evaluado con las metricas mas conocidas y robustas.	Accuracy	$\frac{TP+TN}{TP+FP+FN+TN}$
	Precision	$\frac{TP}{TP+FP}$
	Recall	$\frac{TP}{TP+FN}$

3.4. Técnicas de recolección de Datos

En el contexto de una tesis sobre la Implementación de técnicas de Visión Computacional para la detección temprana de 'Rancha' (*Phytophthora infestans*) en hojas de papa peruana", la técnica de recolección de datos más importante sería la captura y anotación de imágenes de hojas de papa.

3.4.1. Captura de Imágenes

Cámaras de Alta Resolución: Utilizar cámaras de alta resolución para capturar imágenes detalladas de las hojas de papa. Condiciones de Iluminación Controladas: Asegurar que las imágenes se tomen en condiciones de iluminación controladas para evitar sombras y variaciones de color que puedan afectar la precisión de la detección. Diversidad de Condiciones: Capturar imágenes en diferentes condiciones ambientales (luz solar, sombra, humedad) y eta-

pas de crecimiento de la planta para asegurar que el modelo de visión por computadora sea robusto y generalizable.

3.4.2. Anotación de Imágenes

Etiquetado de Datos: Colaborar con expertos en fitopatología para etiquetar las imágenes, identificando áreas afectadas por la 'Rancha' (*Phytophthora infestans*). Software de Anotación: Utilizar herramientas de software de anotación de imágenes que permitan marcar con precisión las áreas afectadas en cada imagen.

3.5. Técnicas para el Procesamiento y Análisis de Información

3.5.1. Metodología de la implementación de la solución

De acuerdo con Szeliski R. (2010), la Visión Computacional se compone de varias etapas que abarcan desde la captura de una imagen hasta su interpretación. Para alcanzar la etapa final, la imagen capturada debe pasar por un proceso que se describirá en detalle en los próximos puntos. La metodología de este trabajo de investigación se ilustra gráficamente en la siguiente Figura.

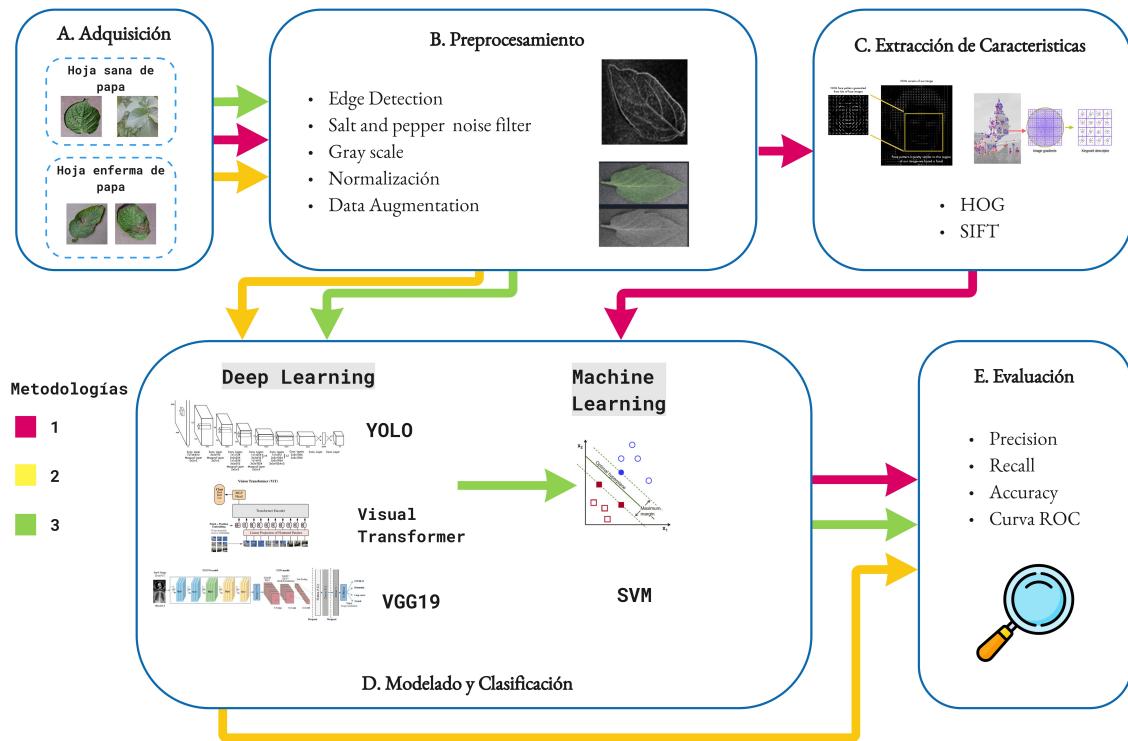


Figura 3.2: Metodología de la Investigación

3.5.2. Metodología para la medición de resultados

El análisis de los resultados de la implementación se llevó a cabo evaluando la relación entre las variables definidas, con el objetivo de medir la efectividad del modelo de clasificación.

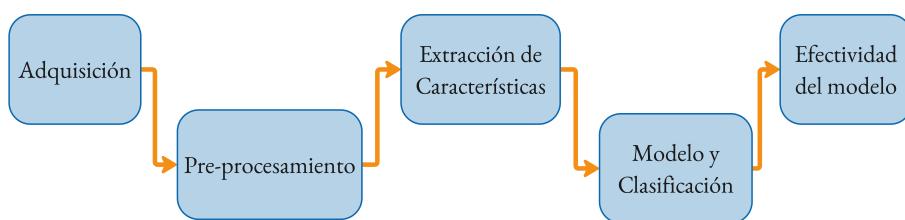


Figura 3.3: Medición de resultados de implementación

Una vez entrenados los modelos de Machine Learning y Deep Learning , se procede a evaluarlos utilizando las siguientes métricas:

Accuracy: El Accuracy es una métrica que indica la proporción de observaciones correctamente predichas en relación con el total de observaciones. Esta métrica es útil cuando el

conjunto de datos tiene una distribución simétrica de falsos positivos y falsos negativos; de lo contrario, puede no ser completamente confiable (Joshi, 2016).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN}$$

Precision: La Precisión es una medida que indica qué proporción de las predicciones positivas son correctas en relación con el total de predicciones positivas realizadas. Una alta precisión implica una baja tasa de falsos positivos (Joshi, 2016).

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall: El Recall es una medida que indica qué proporción de las observaciones positivas reales son identificadas correctamente entre todas las observaciones positivas predichas. (Joshi, 2016)

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1 Score: El F1 Score es un promedio ponderado de Precision y Recall, lo que significa que considera tanto los falsos positivos como los falsos negativos. Esta métrica es especialmente útil cuando se trabaja con conjuntos de datos desbalanceados, es decir, cuando las observaciones de las clases no están distribuidas de manera uniforme. (Joshi, 2016)

$$\text{F1 Score} = 2 \times \frac{TP}{TP+FN}$$

Donde:

TP = True Positives: Son aquellos valores positivos predichos correctamente.

FP = False Positives: Son aquellos valores negativos predichos correctamente.

TN = True Negatives: Son aquellos valores positivos predichos incorrectamente.

FN = False Negatives: Son aquellos valores negativos predichos incorrectamente.

Anexos A

Anexo I: Matriz de Consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS
Problema General	Objetivo General	Hipótesis General
¿Es posible desarrollar un sistema de visión computacional que permita la clasificación temprana de la rancha (<i>Phytophthora infestans</i>) en hojas de papa peruana con alta precisión?	Desarrollar un sistema de visión computacional basado en técnicas de aprendizaje profundo para la clasificación temprana de la rancha (<i>Phytophthora infestans</i>) en hojas de papa peruana.	Se sostiene que mediante el desarrollo de un sistema de clasificación automática de las lesiones de Rancha " <i>Phytophthora infestans</i> " en las hojas de papa peruana, se puede lograr un método de detección la Rancha en las hojas de papa peruana.
Problemas Específicos	Objetivos Específicos	Hipótesis Específicas
¿Cómo podemos desarrollar un sistema de detección automática de Rancha " <i>Phytophthora infestans</i> " en las hojas de papa peruana que sea preciso, eficiente y robusto a la variabilidad de las lesiones y las condiciones ambientales?	Desarrollar un sistema de detección automática de Rancha " <i>Phytophthora infestans</i> " que alcance una precisión alta en la detección de lesiones en las hojas de papa.	La implementación de técnicas de aprendizaje automático y detección de imágenes permitirá desarrollar un sistema de detección automática de Rancha " <i>Phytophthora infestans</i> " con alta precisión.
¿Cómo obtener un conjunto de datos representativo y diverso de imágenes de hojas de papa con Rancha y sin ella?	Recolectar un conjunto de datos amplio y diverso de imágenes de hojas de papa que cubra una variedad de condiciones y escenarios relevantes para la detección de la Rancha.	La recopilación de un conjunto de datos representativo y diverso proporcionará una base sólida para el entrenamiento y la validación de los algoritmos de visión computacional, lo que mejorará su capacidad para detectar con precisión la Rancha en las 95 hojas de papa.
¿Cuáles son las técnicas de vi-	Identificar las técnicas de vi-	Se sostiene que al investigar específicamente técnicas de visión

Anexos B

Anexo II: Resumen de Papers investigados

Tipo	Nº	Título	Autor	Año	País	Fuente
Problema	1	Design and Development of AI-Powered Healthcare WhatsApp Chatbot	Prakasam S and N. Balakrishnan and Kirthickram T R and Ajith Jerom B and Deepak S	2023	India	IEEE
	2	Artificial Intelligence Powered Chatbot for Mental Healthcare based on Sentiment Analysis	Ansh Mehta and Sukhada Virkar and Jay Khatri and Rhutuja Thakur and Ashwini Dalvi	2022	India	IEEE
Propuesta	3	Use of chatbots for customer service in MSMEs	Jorge Cordero, Luis Barba-Guaman and Franco Guamán	2022	Ecuador	Universidad Técnica Particular de Loja
	4	Chatbot: una propuesta viable para la atención al cliente en el centro de soporte de la UCI	Rosbel Caballero Ramírez	2021	Cuba	Revista Cubana de Ciencias Informática
	5	The Science of Detecting LLM-Generated Text	Ruixiang Tang and Yu-Neng Chuang and Xia Hu	2024	China	Communications of the ACM
	6	Review on Implementation Techniques of Chatbot	Nithuna S and Laseena C.A	2020	India	International Conference on Communication and Signal Processing
						International Journal

Anexos C

Anexo III: Árbol del problema

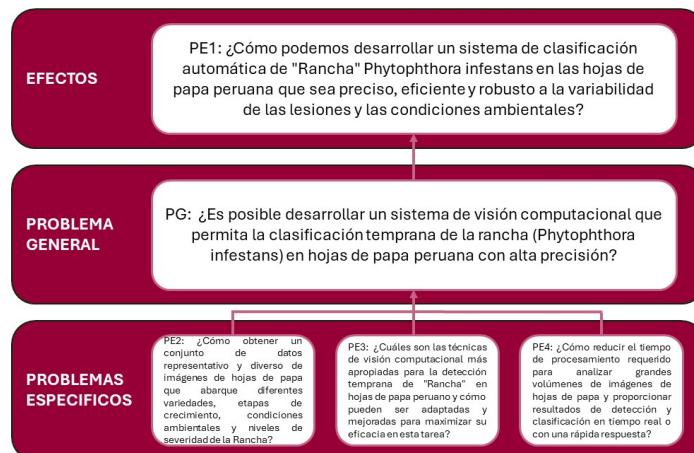


Figura C.1: Arbol de problemas

Fuente: Creación Propia

Anexos D

Anexo III: Árbol de objetivo

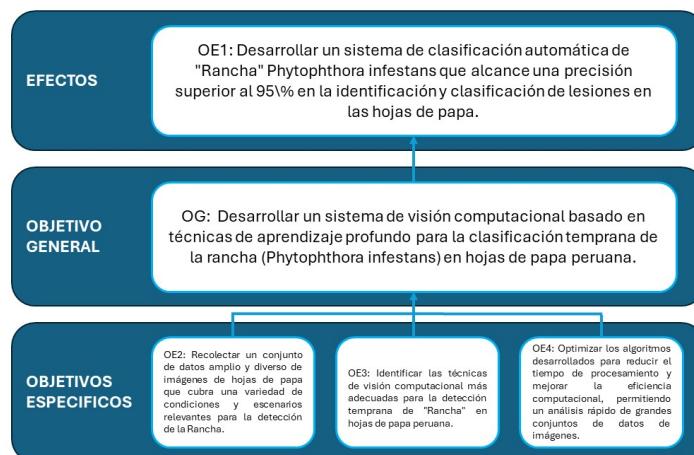


Figura D.1: Arbol de problemas

Fuente: Creación Propia