



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

Modelo YOLO (You Only Look Once) para el diagnóstico preliminar de la enfermedad de Huanglongbing en la planta de limón en la Región Piura, 2023

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTORES:

Carreño Chero, Irvind Dario (orcid.org/0000-0002-6596-6573)

Obregon Mauricio, Alejandro Junior (orcid.org/0000-0003-1307-2698)

ASESOR:

Mg. More Valencia, Ruben Alexander (orcid.org/0000-0002-7496-3702)

LÍNEA DE INVESTIGACIÓN:

Sistemas de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Promoción de la Salud, nutrición y salud alimentaria

PIURA – PERÚ
2023

DEDICATORIA

A nuestros padres, por brindarnos su apoyo incondicional para poder desarrollarnos profesionalmente y por ser la razón de que se pueda continuar y lograr las metas planteadas, también a todas las personas que estuvieron brindándonos su apoyo incondicional siempre.

AGRADECIMIENTOS

Agradecer a Dios primeramente por brindarnos salud y sabiduría en el transcurso de la carrera. Agradecer también a nuestros amigos cercanos que nos brindaron su apoyo para motivarnos y que no nos dejaron abandonar, son quienes nos han apoyado siempre, inculcando valores y ser un ejemplo a seguir. Agradecemos también a los profesionales docentes por sus conocimientos y apoyo en el desarrollo de este proyecto de investigación, quienes gracias a sus pautas y consejos se pudo culminar correctamente el presente proyecto.

ÍNDICE DE CONTENIDOS

CARÁTULA	i
DEDICATORIA.....	ii
AGRADECIMIENTOS	iii
RESUMEN	vii
ABSTRACT	viii
I. INTRODUCCIÓN.....	1
II. MARCO TEÓRICO	7
III. METODOLOGÍA.....	15
3.1 Tipo y diseño de la investigación	15
3.1.1 Tipo de Investigación.....	15
3.1.2 Diseño de Investigación.....	15
3.2 Variable y operacionalización.....	16
3.3 Población, muestra y muestreo.....	18
3.3.1 Población.....	18
3.3.2 Muestra.....	18
3.3.3 Muestreo.....	19
3.3.4 Unidad de análisis	19
3.4 Técnicas e instrumentos de recolección de datos.....	20
3.5 Procedimientos	22
3.6 Método de análisis de datos.....	23
3.7 Aspectos éticos.....	25
IV. RESULTADOS	27
V. DISCUSIÓN.....	47
VI. CONCLUSIONES.....	58
VII. RECOMENDACIONES	60
REFERENCIAS BIBLIOGRÁFICAS	61
ANEXOS	67

Índice de tablas

Tabla 1. Unidad de análisis	18
Tabla 2. Variable 1: Modelo YOLO (you only look once).....	21
Tabla 3. Variable 2: Diagnóstico preliminar de la enfermedad de huanglongbing	21
Tabla 4. Matriz de Confusión.....	23
Tabla 5. Descripción del dataset propio	28
Tabla 6. Descripción del dataset deteccion_huanglongbing	28
Tabla 7. Descripción del dataset enfermedades_limón_persa	29
Tabla 8. Especificaciones de la PC para correr el Modelo Yolov5	30
Tabla 9. Tabla de experimentos por épocas	33
Tabla 10. Prueba diagnóstica positiva exp4.....	43
Tabla 11. Prueba diagnóstica negativa del exp4.....	44
Tabla 12. Especificación de la Curva ROC	46

Índice de gráficos y figuras

Figura 1. Curva ROC.....	24
Figura 2. Dataset propio ambiente natural	27
Figura 3. Dataset propio ambiente controlado.....	28
Figura 4. Caja y Bigotes de la división de conjunto de datos	30
Figura 5. Cuadros delimitadores según la etiqueta huanglongbing 1	31
Figura 6. Cuadros delimitadores según la etiqueta huanglongbing 2.....	31
Figura 7. Cuadros delimitadores según la etiqueta huanglongbing 3.....	32
Figura 8. Métricas de entrenamiento del exp4	35
Figura 9. Curva de precisión – recuperación (recall) del exp4	36
Figura 10. Curva confianza de la precisión del exp4	37
Figura 11. Curva confianza de recuperación (recall) del exp4.....	38
Figura 12. F1 Score - Curva de confianza del exp4	38
Figura 13. Matriz de confusión del exp4.....	39
Figura 14. Distribución de labels del exp4.....	41
Figura 15. Correlograma de labels del exp4	42
Figura 16. Dispersión de la sensibilidad en el diagnóstico preliminar del HLB.....	44
Figura 17. Dispersión de la especificidad en el diagnóstico preliminar del HLB...	45
Figura 18. Curva ROC (Receiver Operating Characteristic).....	46

RESUMEN

Se sostuvo como objetivo general analizar la precisión del modelo YOLO para el diagnóstico preliminar de enfermedad huanglongbing en la planta de limón, como objetivos específicos evaluar las predicciones sobre las instancias totales en cuanto a la precisión y recuperación, clasificar por patrones positivos utilizando el modelo YOLO la sensibilidad en el diagnóstico preliminar de la enfermedad Huanglongbing en la planta de limón y clasificar por patrones negativos utilizando el modelo YOLO la especificidad en el diagnóstico preliminar de la enfermedad Huanglongbing en la planta de limón. La investigación fue aplicada tecnológica de nivel descriptiva, el diseño de investigación que se utilizó es no experimental de corte transversal con enfoque cuantitativo además la escala fue de razón. La población fue conformada por un dataset de 5475 imágenes de hojas sanas y enfermas, la muestra usada fue por conveniencia, el muestreo se dividió en un 70% para entrenamiento, 15% para evaluación y 15% pruebas. Los resultados de precisión se obtuvieron 67% y recuperación se obtuvo 33.3%, sensibilidad la media fue de 52.86% y la especificidad la media fue de 63.57%. En conclusión, YOLOv5m clasifica correctamente los patrones positivos y negativos, además de demostrar una alta precisión de la enfermedad del Huanglongbing.

Palabras clave: Huanglongbing, YOLOv5m, Precisión, Especificidad, Sensibilidad.

ABSTRACT

The general objective was to analyze the precision of the YOLO model for the preliminary diagnosis of Huanglongbing disease in lemon plants, the specific objectives were to evaluate the predictions on the total instances in terms of precision and recovery, to classify by positive patterns using the YOLO model the sensitivity in the preliminary diagnosis of Huanglongbing disease in lemon plants and to classify by negative patterns using the YOLO model the specificity in the preliminary diagnosis of Huanglongbing disease in lemon plants. The research was applied technologically of descriptive level, the research design used was non-experimental of transversal cut with quantitative approach and the scale was of ratio. The population consisted of a dataset of 5475 images of healthy and diseased leaves, the sample used was by convenience, the sampling was divided into 70% for training, 15% for evaluation and 15% for testing. The results for accuracy were 67% and recall was 33.3%, sensitivity the mean was 52.86% and specificity the mean was 63.57%. In conclusion, YOLOv5m correctly classifies positive and negative patterns and demonstrates high accuracy for Huanglongbing disease.

Keywords: Huanglongbing, YOLOv5m, Accuracy, Specificity, Sensitivity.

I. INTRODUCCIÓN

Se conoce que el limón y otras variedades de cítricos aportan mucho valor económico en la mayor parte del mundo, reconocido por su popularidad, valor nutritivo, disponibilidad, y además ayuda a la contribución de cambio de divisas e industrialización (Donkersley et al. 2018). En la zona norte del Perú, en departamentos y regiones como Tumbes, Piura, Lambayeque, poseen condiciones agradables como el clima y suelos muy fértiles para el cultivo del limón Sutil, cuyas cosechas son constantes durante todo el año ya que produce un movimiento constante en lo económico y social. A pesar de ello, últimamente los cultivos de limón se encuentran siendo atacados por enfermedades y plagas que son muy determinantes en la pérdida progresiva de producción, calidad y pérdida de las plantas, lo que influye en la baja productividad del cultivo durante todo el año (Vegas y Narrea 2011).

Sin embargo, la realidad muestra que hay una enfermedad preocupante que pone en modo alerta al sector e industria cítrica, según Senasa, 2019 afirma que existe y además alerta a los agricultores que hay una enfermedad muy peligrosa llamada Huanglongbing o dragón amarillo en español, es una enfermedad que ataca a los cultivos y afecta indirectamente a los agricultores en el Perú, puesto que esta enfermedad se encuentra propagada en países sudamericanos debido al comercio informal nacional e internacional, el insecto portador de la bacteria fue detectado en la Región Piura en el cual fue controlado y deteriorado a tiempo por especialistas expertos en sanidad agraria (Senasa 2019).

Según Thakuria, 2023 comenta que una detección temprana del Huanglongbing es sumamente importante, crucial y determinante para evitar que se propague o siga propagando (Thakuria et al. 2023). Así mismo, sigue siendo un desafío en la identificación porque los primeros síntomas del Huanglongbing se pueden diferir con la deficiencia nutricional. Además, existen variaciones de la misma bacteria dependiendo el continente como lo es el *Candidatus Liberibacter americanus*, *africanus*, *asiaticus* asociadas con el Huanglongbing, ya que aún no se puede cultivar para hacer los análisis

correspondientes en los laboratorios, por ende, la detección del Huanglongbing se hace más difícil para su estudio, el patógeno se alimenta a base hojas, afectando la forma, tamaño, rasgos, maduración y calidad del cítrico.

De igual manera se observa en otras realidades como “[...] El mercado de fruta fresca disminuyó de 2.10 a 1.70 mil millones de toneladas (20.5 % de reducción) durante el mismo intervalo de tiempo. Este mercado se vio menos afectado que el resto de la industria citrícola porque en Estados Unidos alrededor del 90% de las naranjas producidas en Florida que es el estado con mayor prevalencia del Huanglongbing (HLB), mientras que California suministra naranjas para el mercado fresco. Según Sigerman y Useche, 2017 reportaron un aumento de \$2.89 a \$9.34 (3.2 veces) del precio de una caja de naranja desde que el HLB fue detectado en los Estados Unidos (Singerman, Lence y Useche 2017). No obstante, los productores de cítricos en muchos países enfrentan serios problemas con la aparición del HLB, responsable de la disminución en la producción de naranjas para procesamiento en los Estados Unidos de 7.89 a 2.22 mil millones de toneladas (reducción del 72.2%) de 2007-08 a 2017-18” (Dala-Paula et al. 2019).

En nuestra realidad se observa que el Huanglongbing, ha causado la erradicación de más de 100 millones de plantas de cítricos en todos los continentes del mundo en los últimos años, y logrando reemplazar las costumbres tradicionales de la citricultura. Sin embargo, está forzando a los pequeños agricultores a buscar otras formas de cultivos, al no tener los suficientes medios económicos para poder combatir dicha enfermedad (Senasa 2019).

La planta de limón también es afectada por la plaga Hemiptera: Pseudococcidae que forma parte de una familia de insectos como chinches o normalmente conocido como cochinillas. Esta especie de plaga es difícil para un control adecuado en el campo y requiere un control continuo durante el todo el año. Estas especies son difíciles de distinguir porque son muy similares morfológicamente y su identificación taxonómica se basa en claves que tratan varias estructuras cutáneas. Además, en algunas especies pueden existir

variaciones fenotípicas entre individuos, dependiendo de las condiciones climáticas o del sustrato sobre el que crecen (Correa et al. 2012).

Estas enfermedades y plagas que atacan a los cultivos de limón, en las temporadas de 2008/09 fueron devueltas un total de 158.986 cajas de frutas desde los mercados de Asia, América y Oceanía en la detección como uvas, peras, limones, nectarinas, duraznos, entre otros exportados desde nuestro país (Larrain 2010).

De igual manera se observa que habitualmente los agricultores utilizan dos métodos tradicionales para determinar enfermedades que afectan directamente al cultivo como: Análisis de expertos en laboratorios e inspección visual por humanos (Lowe, Harrison y French 2017). Estos métodos generan desventajas en su efectividad e infructuosidad. En el primer método se toma la evidencia en campos de cultivos y es transportado a laboratorios para los análisis correspondientes obteniendo un diagnóstico preliminar previo, cabe destacar que durante este proceso de análisis la evidencia empieza a evolucionar como también ocurre la propagación eventual de esta, de no tomar las medidas necesarias para prevenirla y evitar que existan cambios en los cultivos puede ocurrir que el insecto mute y, así mismo, la tenacidad de la forma utilizada se ve limitada debido a su precio y laboratorios escasos. Al analizar el segundo método inspección visual por humanos, debido a sus precios elevados por servicios y el problema de contratar a un experto para transportarlo al lugar donde se encuentra el cultivo reducen los medios de ejecución, además de la subjetividad y la deficiencia de garantías en el diagnóstico de las plagas y enfermedades, lo cual le quitan credibilidad.

Para ofrecer ayuda a los agricultores y a las personas expertas en lo agrario, según Lit y otros, 2014 comenta que se han creado diferentes tipos de tecnologías para diagnósticos en los cultivos, como cámaras fotográficas hiper-espectrales equipadas con sensores para detectar objetos insólitos en los parámetros de las plantas, otra porción, se centra en el uso de imágenes dentro del rango visible, primordialmente en los frutos y hojas; estos recursos tecnológicos constan de un método clásico como una etapa de procesamiento,

una excavación de características y toma de decisiones, en la que utilizan métodos como inteligencia artificial (Li et al. 2014).

Actualmente, se han venido presentando proyectos orientados en el aprendizaje profundo para abordar el diagnóstico mediante imágenes en un rango visible. En comparación con el método anterior, el aprendizaje profundo disminuye la etapa de preprocesamiento y evita el desarrollo de algoritmos. La mayoría de estas aplicaciones exigen una mayor capacidad de cálculo para ejecutar los modelos (Jones et al. 2017). Sin embargo, en varios de los casos es difícil disponer de dicha capacidad o conexión a un servidor que la posea, dada la ubicación de los cultivos. Por otro lado, estos desarrollos sirven como prueba de concepto, pero carecen de una etapa de despliegue y validación (QA) a gran escala para evaluar su impacto en el trabajo de los agricultores.

Dado se justifica teóricamente, el desarrollo de un diagnóstico preliminar como una herramienta para que los métodos sean más efectivos en la detección del Huanglongbing ante la falta de métodos precisos, ya que es muy beneficioso tanto en los productores, agricultores y expertos en el sector cítrico. Los cultivos de cítricos como el limón son amenazados por esta peligrosa enfermedad HLB que ha sido considerada como una de las amenazas más importantes para la producción mundial de cítricos comerciales y sostenibles. Se estima que a nivel mundial más de 60 millones de árboles habían sido destruidos por la enfermedad a principios de la década de los 1990 (CABI 2019). Para combatir esta enfermedad, según Pandey y otros, 2019 afirman que requiere un diagnóstico precoz y sólido del HLB para la extracción del inóculo, a fin de evitar la adquisición y propagación de esta enfermedad (Pandey y Wang 2019). Por esta razón, es de gran importancia disponer de mecanismos de diagnóstico rápidos y eficaces para detectar a tiempo los síntomas de la enfermedad y poder combatirla.

Dado se justifica de manera social, debido a que se basa en un sistema realice un diagnóstico preliminar sobre la enfermedad de huanglongbing en la planta de limón, brindando ayuda a los expertos y agricultores concernientes al agro específicamente en el sector cítrico que pueda manifestar si existe contagio preliminar de huanglongbing o no exista contagio, ya que son importante para

mover la economía y gastronomía del país. Por consiguiente, al producir limones de baja calidad generaría una pérdida económica significativa al país como a los pequeños y grandes agricultores o relacionados al sector citrícola.

Dado se justifica de forma práctica, debido al modelo YOLO tiene una gran utilidad práctica en la capacidad de detección y clasificación de objetos sin retrasos significativos y, así mismo, con respuestas inmediatas con una precisión en tiempo real lo que beneficia a la construcción de proyectos como diagnósticos preliminares o pre-diagnósticos en el sector citrícola en lo que demanda la detección de elementos u objetos tanto en video o imágenes, el modelo YOLO además de la clasificación de objetos superpuestos se adapta a cualquier tipo de entorno, la cual posee una eficiencia en el impacto de los sistemas y cuenta con una facilidad de implementación.

Dado se justifica de manera tecnológica, porque mediante un sistema tecnológico se da una solución a una problemática, lo cual consiste en la detección de la enfermedad de huanglongbing mediante un diagnóstico preliminar para contribuir a la toma de decisiones que pueden cambiar un rumbo en el sector citrícola usando inteligencia artificial (IA). Por tal motivo, se apoyó en sistemas o modelos existentes basados en IA.

Dado se justifica metodológicamente, debido a que se propuso instrumentos como guías de observación para medir el comportamiento de los indicadores como el Incrementó Precisión de la detección, Incrementó de Recall de detección, Incrementó de la sensibilidad e Incrementó de la especificidad de las cuales ante la necesidad para la obtención de datos se crearon dos guías de observación para cada una de las dimensiones Modelo YOLO y diagnóstico preliminar para la enfermedad de Huanglongbing (HLB). Además, se usó herramientas para gestionar las investigaciones de autores que ayudaron a formular, seleccionar y categorizar tanto antecedentes como referencias. Por ello, también se hizo uso del método científico en los estudios mencionados ya que permite conceptualizar las variables “Modelo YOLO” y “Diagnóstico Preliminar”, además de conocer en qué nivel de investigación científica y de ciencia se encuentra el tema propuesto. Además de analizar los paradigmas

como cambia ese concepto, estudio o investigación con respecto al tiempo transcurrido.

Por tal fin se va a desarrollar un modelo basado en YOLO, con el uso de tecnología como el reconocimiento de imágenes, evaluación de precisión para el diagnóstico preliminar del Huanglongbing (HLB) donde se definirán conceptos de aprendizaje profundo (Deep Learning) para brindar ayuda en los campos de los agricultores limoneros y/o expertos en el sector agrario. Esto requiere que el modelo a implementar se pueda desplegar en dispositivos con recursos mínimos de capacidad de memoria y computacional.

En relación a lo descrito, surge la idea de investigar ¿Cómo un modelo YOLO para el diagnóstico preliminar analiza la detección de la enfermedad del Huanglongbing en la planta de limón en la Región Piura?, además de las preguntas específicas ¿Cómo evaluar las predicciones sobre las instancias totales en cuanto a la precisión y recuperación?, ¿Cómo clasificar patrones positivos utilizando el modelo YOLO la sensibilidad en el diagnóstico preliminar de la enfermedad Huanglongbing en la planta de limón?, ¿Cómo clasificar patrones negativos utilizando el modelo YOLO la especificidad en el diagnóstico preliminar de la enfermedad Huanglongbing en la planta de limón?

El objetivo principal de la investigación se planteó: Analizar la precisión del modelo YOLO (You Only Look Once) para el diagnóstico preliminar de la enfermedad de huanglongbing en la planta de limón.

Los objetivos específicos de la investigación se plantearon, Evaluar las predicciones sobre las instancias totales en cuanto a la precisión y recuperación, Clasificar por patrones positivos utilizando el modelo YOLO la sensibilidad en el diagnóstico preliminar de la enfermedad Huanglongbing en la planta de limón, Clasificar por patrones negativos utilizando el modelo YOLO la especificidad en el diagnóstico preliminar de la enfermedad Huanglongbing en la planta de limón.

II. MARCO TEÓRICO

En la investigación se han tomado en consideración varios estudios por investigadores y autores a nivel internacional y nacional, bases teóricas que fueron un apoyo para determinar el marco referencial de dimensiones e indicadores correspondientes a la investigación.

En Corea, Xing y Lee, 2019 en su artículo de investigación "Modelo de reconocimiento de plagas y enfermedades de los cítricos utilizando una red de convolución conectada débilmente densa", como objetivo sostuvo desarrollar un modelo CNN efectivo para aumentar la complejidad entre canales y frecuencia para la reutilización de características. La red propuesta se diseñó desde el punto de vista de la eficiencia de los parámetros. El conjunto de datos se incluye 17 especies de plaga que afecta a los cítricos y de 7 enfermedades, las imágenes de las plagas fueron extraídas desde internet que fueron un muestra de 9051 imágenes mientras que las imágenes de enfermedades fueron tomadas desde un huerto con una cámara de alta resolución la muestra fue de 3510 imágenes, el dataset original se dividió en 3 entrenamiento, validación y pruebas, la proporción fue de 4:1:1, la imágenes fueron redimensionadas a 224 x 224 con un enfoque de interpolación bilineal para evaluar la efectividad de Weakly DenseNet se comparó con varias redes de referencia en diferentes aspecto. Los resultados fueron que DenseNet-16 obtuvo una precisión de 93.33%. En conclusión, los resultados del experimento mostraron que Weakly DenseNet-16 obtuvo la mayor precisión de clasificación con menos parámetros. Debido a que esta red es ligera, puede utilizarse en dispositivos móviles (Xing, Lee y Lee 2019).

En China, Qiu y otros, 2022 en su artículo científico "Un sistema de identificación automática para la enfermedad de ecologización de cítricos (Huanglongbing) utilizando una red neuronal convolucional YOLO" sostuvo como objetivo detección de Huanglongbing a partir de imágenes digitales basado en tres modelos, su muestra fue en 12 huertos en varias ciudades de Fujian, China en varias especies de plantas la construcción de su conjunto de datos fue 7516 imágenes, incluyendo imágenes de frutas y hojas saludables el total fue 1413, para hojas y frutos relacionadas con los síntomas del HLB fue

de 3017 imágenes y otros síntomas relacionados fue de 3086 imágenes, se usó un tasa positiva de PCR de *Candidatus Liberibacter Asiaticus* para examinar la viabilidad nos dice que al reconocer síntomas de HLB fue 83.33%, 85.19% y 83.73% incluyendo imágenes similares a HLB ayudó a mejorar la precisión. Se realizaron variaciones de modelo YOLO a evaluar como lo es Yolov5I-HLB1 / Yolov5I-HLB2 / Yolov5I-HLB3, los parámetros que se utilizaron con respecto al tamaño del lote fue 20 en 200 épocas, los experimentos se realizaron con una GPU GeForce RTX 3090, la división porcentual se realizó con un 80 % para el entrenamiento, el 10 % para la validación y el 10 % pruebas se consideraron métricas de entrenamiento incluyendo precisión, recall, puntuaciones F1 y la matriz de confusión, el umbral de nivel de confianza igual a 0,4. En cuanto a los resultados la precisión obtuvo de 90.07% el recall obtuvo 81.84% y el f1 85.19% qué diferencia de los demás variaciones de modelos, la matriz de confusión del modelo HLB2 aplicado el reconocimiento en 12 enfermedades de los cítricos fue adquirida en diferentes condiciones de adquisición. No obstante, incluir imágenes con síntomas irrelevantes no mejoró la precisión. En conclusión, el segundo modelo basado en YOLO superó a todos los modelos con F1 del 85,19%, sin embargo, aún tiene limitaciones por mejorar en la precisión (Qiu et al. 2022).

En Lima, Gamboa, 2018 en su proyecto de tesis “Aplicación móvil para el diagnóstico preliminar de micosis superficiales a través de fotografías digitales”, presenta una investigación detallada sobre la funcionalidad del desarrollo y diseño de una app móvil, cuyo objetivo es el diagnóstico preliminar de micosis superficiales, mediante el análisis de fotografías digitales. Este proyecto de investigación es de tipo experimental ya que se manipula las variables y se divide en dos grupos de población, un grupo de personas enfermas (Grupo experimental) y el otro grupo de personas sanas (Grupo de control), para medir se utilizó los indicadores sensibilidad y especificidad, se desarrolló con un framework llamado IONIC, que se compila para dispositivos móviles tanto para Android y para IOS. En conclusión esta aplicación es de vital importancia para las personas sobre todo porque está relacionada a diagnósticos de enfermedades dermatológicas, ya que de acuerdo a los resultados de la investigación en los indicadores obtenidos si tiene un incremento porcentual en

cuanto a la sensibilidad de 21.1605% y en la especificidad de 33.1361% (Gamboa Jara 2018).

En Puerto Maldonado, Diaz y Vilcas, 2019 nos dice en su tesis "Aplicativo Móvil Para La Detección De La Enfermedad Del Mosaico Común En La Yuca (Manihot Esculenta) Utilizando Redes Neuronales Convolucionales, 2021" de la Universidad Nacional Amazónica de Madre de Dios, su investigación fue de tipo aplicada diseño experimental pre-experimental de nivel descriptiva , como objetivo sostuvo la implementación de un aplicativo móvil para medir el nivel exactitud y mejorar la eficiencia en la detección del mosaico mediante un modelo de CNN, como población se usó los participantes como pacientes, siendo su muestra no probabilística, como resultado según a las encuestas realizadas con la norma ISO 9126 para dimensión de funcionalidad obtuvo 4.5(requisitos esperados), para la dimensión de eficiencia obtuvo 4.2 (requisitos esperados), usabilidad obtuvo 4.7 (requisitos esperados), portabilidad obtuvo 4.3 (requisitos esperados) todo lo que engloba la variable independiente aplicativo móvil, y con respecto a la variable interviniente red neuronal convolucional para el entrenamiento y validación de las 58 mediciones la exactitud para la detección fue mayor al 90%, concluyendo que el app móvil influye significativamente para el diagnóstico del mosaico (Diaz Holgado y Vilcas Villalba 2021).

En Piura, Martinez y Morocho, 2021 nos dice en su tesis "Diagnóstico del estado emocional, a través de reconocimiento facial para una empresa del Sector Público Piura 2021" de la Universidad César Vallejo, como objetivo sostuvo demostrar o evidenciar la efectividad mediante un diagnóstico a trabajadores a través de un reconocimiento facial el estado emocional en una institución educativa, su investigación fue de tipo aplicada de diseño experimental pre-experimental, un enfoque mixto y longitudinal, su población fue 15 colaboradores de distintas áreas, su muestra fue de 8 colaboradores, su muestreo fue no probabilístico por conveniencia, como técnica fue observación e instrumento guía de observación como resultado fue que la precisión fue positiva y exitosa obtuvo un 75%, se concluye que se identifica exitosamente las emociones en los colaboradores mediante la implementación de una

herramienta que usa machine learning (Martinez Ojeda y Morocho Febres 2021).

En Piura, Cordova y Quinde, 2022 nos dice en su tesis "Sistema De Reconocimiento De Imágenes Para Pre-Diagnósticos De Onicomicosis (Hongos En Las Uñas) en el centro podológico DECAPIE – Piura 2022" de la Universidad César Vallejo, como objetivo sostuvo la implementación de un sistema de reconocimiento para determinar la precisión por reconocimiento de imágenes determinando el incremento de la especificidad y sensibilidad. Su población fue de 100 personas que incluyen hombres y mujeres pacientes de un centro podológico que rondan las edades entre 20 y 60 años de edad, su muestra fue de 73 imágenes, para determinar diagnósticos positivos y negativos utilizando 37 pacientes con onicomicosis para los totales positivos y 36 pacientes sin onicomicosis para los totales negativos, como resultado para el indicador incremento de regularización, tomaron en cuenta 212 imágenes y 200 una con onicomicosis y la otra sin onicomicosis tuvieron en cuenta la muestra representativa el 85.40% se reconoció como onicomicosis y 13.60% presentó error en tal identificación de onicomicosis y para decremento de error de entrenamiento (RMSI) en las 20 épocas la asertividad el algoritmo CNN con 98.4% hasta 100%, la pérdida fue de 5.60% con un mínimo de 0.01%, para la sensibilidad fueron 10 y 15 iteraciones con un mínimo de 89.19% y como máximo 97.30% con 20 épocas. La sensibilidad fue 92.57%, la especificidad fue 92.36%. En conclusión, la sensibilidad y especificidad ya que mientras avanzan los entrenamientos la precisión incrementa y el error disminuye (Cordova Zapata y Quinde Ramos 2022).

Para un mejor entendimiento es mejor teorizar en afinidad a la variable **modelo YOLO**, se determina que detecta objetivos de diferentes escalas y predice múltiples categorías con gran precisión (Lyu et al. 2023), La serie YOLO representa algoritmos de una etapa, que son más adecuados para aplicaciones prácticas que los algoritmos de dos etapas (como Faster R-CNN) debido a su mejor equilibrio entre precisión, velocidad y recuperación (Recall).

Por otro lado Zaidi y otros, 2022 menciona que los detectores de dos etapas resuelven la detección de objetos como un problema de clasificación, mientras

que YOLO lo reformula como un problema de regresión, prediciendo los píxeles de los objetos y sus atributos de cuadro delimitador (Zaidi et al. 2022).

El modelo se evalúa utilizando diferentes métricas, como la precisión (P), la recuperación (R), las puntuaciones F1 (F1) y las matrices de confusión.

La precisión (P), donde TP es positivo verdadero: se predice que el objetivo positivo es positivo; FP es falso positivo: se predice erróneamente que el objetivo negativo es positivo; FN es falso negativo: se predice erróneamente que el objetivo positivo es negativo (Lawal 2021).

$$\text{Precisión } (P) = \frac{TP}{TP + FP} \times 100\%$$

La recuperación (R), donde TP es positivo verdadero: se predice que el objetivo positivo es positivo; FP es falso positivo: se predice erróneamente que el objetivo negativo es positivo; FN es falso negativo: se predice erróneamente que el objetivo positivo es negativo (Lawal 2021).

$$\text{Recall } (R) = \frac{TP}{TP + FN} \times 100\%$$

las puntuaciones F1 (F1), se realizó como un compromiso entre Recall y Precisión para mostrar el rendimiento global de los modelos entrenados, definido en la ecuación (Lawal 2021).

$$F1 = \frac{2 * \text{Precisión} * \text{Recall}}{\text{Precisión} + \text{Recall}}$$

La enfermedad de **huanglongbing**, se define Huanglongbing también conocida como HLB o ecologización de cítricos, posee síntomas complejos ya que es la enfermedad más devastadora en el sector citrícola asociada con varios patógenos de acuerdo al continente africanus, asiaticus, americanus (Qiu et al. 2022). El patógeno que produce HLB provoca trastornos tapando los tejidos que transportan los nutrientes para la transportación de nutrientes al floema y debido a la acumulación de almidón (Dong et al. 2021). También se le conoce como el cáncer de los cítricos los síntomas son similares a la deficiencia nutricional (Yang et al. 2021). Una de las mejores estrategias para la

intervención del HLB es plantar plántones de cítricos libres de enfermedades ya que esto ayudará a evitar la propagación del Huanglongbing (Zhang et al. 2022).

Con respecto al **diagnóstico preliminar** se define que ayuda a analizar un fenómeno o situación ya investigado para conocer su inclinación a lo largo del tiempo ya que permite la toma de decisiones para potenciar las ventajas existentes y la solución de detección de problemas a corregir (Obregón et al. 2018). También son evaluaciones mediante pruebas refiriéndose a propósitos para determinar mediante métodos confiables un diagnóstico sobre un estudio en particular (Han et al. 2018). Un diagnóstico es un punto de atención aplicable en distintos sectores ya que facilitará detecciones de forma rápida y precisa para poder abordarse ante un fenómeno (Thakuria et al. 2023).

La prueba diagnóstica sirve para descartar o confirmar que hay una enfermedad, además de poder confirmar ausencia e incluso la presencia de una enfermedad (Sancho-Insenser y González-Castillo 2022).

La **sensibilidad** también conocida como (TRP) verdaderos positivos es probabilidad de cómo se clasifican las etiquetas de manera positiva en la realidad como en la predicción ya que indica que el algoritmo pueda identificar los casos positivos de todos los casos positivos que existen (Guzmán y Yineth 2021). También es la disposición para detectar la cantidad de individuos con la enfermedad que logran un resultado positivo en pocas palabras se clasifica adecuadamente como individuo enfermo (Sancho-Insenser y González-Castillo 2022).

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

La **especificidad** también conocida como (FRP) falsos positivos es como se clasifican las etiquetas de manera negativa. También es la disposición para detectar la cantidad de individuos que se encuentren sanos que logran un resultado negativo en pocas palabras se clasifica como individuo que no cuenta con la enfermedad o no enfermo (Sancho-Insenser y González-Castillo 2022).

$$\text{Especificidad} = \frac{TN}{TN+FP}$$

El **aprendizaje profundo** su tiempo de ejecución en el entrenamiento tienden hacer un poco largo, pero más corto que otros, además que proporciona resultados precisos (Lee et al. 2022). Además, es un subconjunto de la IA ayuda realizar tareas de manera específica a través de sistemas, existen dos métodos o tipos de aprendizaje supervisados y la sin supervisión (Gao et al. 2020).

Redes neuronales convolucionales la arquitectura CNN comparte una red cuando opera convoluciones de una imagen o imágenes de entrada. Por ello, ya no se requiere un detector para el mismo objeto ya que se reduce los parámetros por aprender (Gao et al. 2020). En comparación con la tecnología tradicional de procesamiento de imágenes, las CNN profundas tienen mayores capacidades para expresar características abstractas y pueden obtener información semántica de imágenes complejas.

Los algoritmos de detección de objetivos basados en el aprendizaje profundo pueden dividirse en dos categorías, algoritmos de detección de una etapa (como la serie YOLOv5) (Jocher et al. 2021) y algoritmos de detección de dos etapas (como FasterR-CNN). Las velocidades de procesamiento de los primeros son más rápidas que las de los segundos, lo que los hace más adecuados para la detección en tiempo real de enfermedades de las plantas en entornos de campo complejos (Chen et al. 2022).

El modelo YOLOv5 (Jocher et al. 2021), en función de las diferencias en la profundidad y anchura de la red, se divide en cuatro versiones de modelo de red: YOLOv5s, YOLOv5m, YOLOv5l y YOLOv5x. Entre ellas, la red YOLOv5s tiene la velocidad de cálculo más rápida, pero la precisión media es la más baja, mientras que la red YOLOv5x tiene las características opuestas. El tamaño del modelo de la red YOLOv5 es aproximadamente una décima parte del de la red YOLOv4. Tiene velocidades de reconocimiento y posicionamiento más rápidas, y la precisión no es inferior a la de YOLOv4. La red YOLOv5 consta de cuatro componentes principales: Input, Backbone, Neck and Prediction.

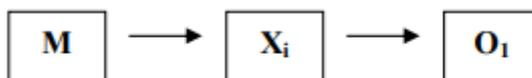
La red YOLOv5s utiliza residuos para evitar la desaparición/explosión de gradientes, integra mapas de características. Esto permite detectar objetivos de diferentes escalas y predecir múltiples categorías con gran precisión (Lyu et al. 2023). Chen y otros, 2022 resalta que YOLOv5s tiene un rendimiento subóptimo en la detección de grupos de objetivo pequeños, por ende, las detecciones de pérdidas son una limitación grave (Chen et al. 2022). Por otro lado, Spatial Pyramid Pooling (SPP) puede aumentar la agrupación multiescala de características de alto nivel mediante el algoritmo YOLOv5s para aumentar el campo receptivo. La estructura general de una SE-Net, F_{sq} se refiere a la operación de compresión, F_{ex} se refiere a la operación de excitación y F_{scale} se refiere a la operación de escala. La capa de cuello de botella es una convolución 1×1 seguida de una convolución 3×3 (**Ver Anexos 5, 6, 7 y 8**).

III. METODOLOGÍA

3.1 Tipo y diseño de la investigación

3.1.1 Tipo de Investigación

El tipo de investigación que se empleo fue aplicada tecnológica de nivel descriptiva, porque se implementó un modelo YOLO para diagnosticar la enfermedad de Huanglongbing o Dragón Amarillo, ya que el Huanglongbing es una enfermedad que ataca principalmente a los cítricos y se propaga de manera rápida, consecuente de su paso destructivo que causa pérdidas económicas en el sector citrícola en el mundo. Esteban Nieto, 2018 afirmó que la investigación aplicada o también conocida por tecnológica se centra en resolver un problema de la sociedad u organización de manera tecnológica, sin un conocimiento netamente puro ya que están orientadas a optimización y soluciones (Esteban Nieto 2018).



M1: Total de imágenes reconocimiento de la enfermedad de Huanglongbing.

Xi: Imágenes de test de reconocimiento de la enfermedad de Huanglongbing.

O1: Guías de observación para el registro de categorización y clasificación de imágenes de reconocimiento para diagnóstico preliminar de la enfermedad de Huanglongbing.

3.1.2 Diseño de Investigación

En la investigación, el diseño que se utilizó es no experimental de corte transversal y de enfoque cuantitativo, porque se usaron técnicas de estadística y se examinó de forma numérica para la ayuda de detección de la enfermedad del Huanglongbing. Según

Flores y Anselmo, 2019 comentó que la investigación de enfoque cuantitativo es la predominancia de acuerdo a la problemática por abordar, donde se pueda cuantificar-controlar la ocurrencia del fenómeno estableciendo relaciones de causa-efecto, examinando variables medibles y numéricas (Flores y Anselmo 2019). También Bencardino, 2019 afirmó que son caracteres de estilo numéricos y mensurables, ya que se puede denominar en los caracteres de enfoque cuantitativo como variables que son cuantificables (Bencardino 2019).

3.2 Variable y operacionalización

- Definición conceptual

V1: Modelo YOLO: Extracción automáticamente de las características según los objetos del conjunto de datos en el entrenamiento y aprender una representación más avanzada de datos, características y resolviendo problemas como la detección de objetos basados en el aprendizaje profundo y la visión por computadora (Lyu et al. 2023).

V2: Diagnóstico Preliminar de la enfermedad de Huanglongbing: Identificación de la presencia de la enfermedad Huanglongbing basada en los resultados de la detección y reconocimiento de síntomas o características específicas (Nehela y Killiny 2020).

- Definición operacional

V1: Modelo YOLO: Ajuste y optimización de los parámetros del modelo YOLO, en el tamaño de las capas, las tasas de aprendizaje y los umbrales de confianza, para mejorar su Precisión, Recall, F1-score y AP (Average Precisión) en la detección de síntomas de la enfermedad (Lawal 2021). En cuanto al procedimiento realizado en el Modelo YOLO es la precisión y recall, se utilizó una guía de observación mediante esta se obtienen datos correspondientes a los indicadores a medir.

V2: Diagnóstico Preliminar de la enfermedad de Huanglongbing: Procesamiento de las imágenes utilizando el modelo YOLO previamente entrenado para detectar y reconocer síntomas o características asociadas a la enfermedad de Huanglongbing (Qiu et al. 2022). En cuanto al procedimiento en el diagnóstico preliminar de la enfermedad de Huanglongbing es el incremento de la especificidad y sensibilidad en la clasificación de imágenes, se utilizó una guía de observación para el entrenamiento y validación del sistema para acentuar la confiabilidad de ello.

- Indicadores

Incrementó precisión de detección se utilizó para medir los patrones positivos que se predicen correctamente del total de patrones predichos en una clase positiva (Lawal 2021; Hossin y M.N 2015).

Incrementó de recuperación de detección se utilizó para medir la fracción de patrones positivos que se clasifican correctamente (Lawal 2021; Hossin y M.N 2015).

Sensibilidad se utilizó para medir la fracción de patrones positivos que se clasifican correctamente (Khan et al. 2021).

Especificidad se utilizó para medir la fracción de patrones negativos que se clasifican correctamente (González-Garza 2017).

- Escala de medición

Se utilizó la escala de razón o también conocida como escala proporcional, la cual posee un cero absoluto o un carácter de origen lo que permite aplicar estadística inferencial y descriptiva, brindando la cantidad de posibilidades de medir la magnitud o tamaño de una circunstancia (Velázquez 2018).

3.3 Población, muestra y muestreo

3.3.1 Población

La población fue conformada por un Dataset de 5475 imágenes (Greening/Healthy) traducido al español como enfermas y sanas. Según Bencardino, 2019 comentó que la población que se emplea para describir un conjunto o agrupación de medidas de elementos o unidades que pueden ser observadas de la cual se logra extraer la muestra, que también se puede llamar población muestreada, ya que puede corresponder no solo a personas sino también a entidades u objetos. En las cuales las entidades pueden ser simples o complejas, además se considera si es finita o infinita (Bencardino 2019).

Tabla 1. *Unidad de análisis*

Unidad de análisis	Cantidad
Imágenes hojas de la planta de limón sutil (<i>Citrus Aurantifolia Swingle</i>) e Imágenes hojas de la planta de limón Tahití o persa (<i>Citrus Latifolia Tanaka</i>) (Greening/Healthy with HLB)	5475

Fuente: elaboración propia

3.3.2 Muestra

La muestra que se usó fue la población finita, en campos de estudio con características donde se presentó los síntomas de la enfermedad HLB, como hojas afectadas por HLB siendo medianas o pequeñas, desarrollando un patrón de áreas amarillas y verdes sin límites claros entre los colores, dando la apariencia de moteado-manchado, también con síntomas HLB que aparecen en una sola rama y posteriormente se extienden de forma aleatoria por la copa del árbol durante varios años, por ende la deficiencia de zinc suele aparecer de forma uniforme por toda la copa (Nehela y Killiny 2020). Cabe precisar que también se tomó parte de la muestra en cultivos de Plantones de limón sutil (*Citrus Aurantifolia Swingle*)

completamente sanos, es importante resaltar que se aplicó la muestra por conveniencia. Según Bencardino, 2019 afirmó que la muestra es parte perteneciente y concerniente a la población, además que la muestra es conjunto de medidas en que los elementos se eligen de forma al azar y que todos tienen el mismo porcentaje o posibilidad de ser elegidos para que la muestra sea representativa (Bencardino 2019).

Con respecto a la muestra por conveniencia Ortega, 2018 comentó que el muestreo por conveniencia, se llevó a cabo para realizar muestras de acuerdo a la disponibilidad, facilidad de las personas o elementos que son parte de la muestra y se considera no aleatorio y no probabilístico, ya que posee ventajas en la recopilación de datos y es muy utilizada cuando hay situaciones de poblaciones grandes a ser evaluadas además se utiliza cuando no hay criterios por considerar para que un elemento o persona sea parte de la muestra (Ortega 2018).

3.3.3 Muestreo

El muestreo se dividió en un 70% para entrenamiento (train), 15% para evaluación (eval) y 15% pruebas (test). Según Otzen y Manterola, 2017 comenta que el muestreo tiende a estudiar las relaciones entre distribuciones de una variable en la muestra de estudio y en la población en blanco, donde se debe tomar ciertos criterios de inclusión y también de exclusión, además nos indica que existen técnicas de muestreo de dos tipos probabilística y no probabilística (Otzen y Manterola 2017).

3.3.4 Unidad de análisis

Se utilizaron imágenes hojas de la planta de limón sutil (*Citrus Aurantifolia Swingle*) y hojas de limón Tahití o persa (*Citrus Latifolia Tanaka*) (Greening/Healthy with HLB). Se conoció que la producción del *Citrus aurantifolia swingle* en el norte del Perú es de

todo el año brindando una actividad fluida en el ámbito económico y social, para este cultivo es de vital importancia realizar un estudio de terreno para determinar características físicas, químicas y biológicas obteniendo un análisis de vocación y fertilidad natural de los suelos. Cabe resaltar que, en su ciclo de vida requiere del recurso hídrico del agua dulce y fertilización. Es importante mencionar que *Citrus Aurantifolia* es rico en vitamina C y aceites esenciales (Vidal-Villacorta y Villacorta-González 2021).

3.4 Técnicas e instrumentos de recolección de datos

Técnica

Las técnicas de recolección de datos consisten en actividades y procedimientos que permiten recopilar información necesaria para permitir dar una respuesta a la pregunta de investigación (Mendoza y Avila 2020). Existen diferentes herramientas para la recolección de información de manera eficaz con distintos fines como de investigación y análisis, ya que existen varias técnicas para la recopilación de datos para demostrar la confiabilidad y validez de una investigación (Narvaez 2023). En la investigación, se empleó la técnica de la **observación**, que consiste en registros tomados para observar el desarrollo o comportamiento de un fenómeno en un entorno específicamente y analizarlo para registrar lo observado.

Instrumento

Los instrumentos se conocen como herramientas que se utilizan para la obtención de datos e información de una realidad. Como instrumentos se realizaron **Guías de Observación**, ya que precisa de manera sistemática distintos aspectos en general a observar el objeto a estudiar para identificar elementos y variables, ya que es medio para la obtención y recolección de datos de un hecho a observar del mismo (Useche et al. 2019).

Tabla 2. Variable 1: Modelo YOLO (you only look once)

Indicadores	Técnica	Instrumento
Incrementó Precisión de detección	Observación	Guía de Observación 01
Incrementó de Recuperación de detección		

Fuente: elaboración propia

Tabla 3. Variable 2: Diagnóstico preliminar de la enfermedad de huanglongbing

Indicadores	Técnica	Instrumento
Sensibilidad	Observación	Guía de Observación 02
Especificidad		

Fuente: elaboración propia

Validez del Instrumento: Se determinó el grado en que los instrumentos miden las variables, se consideró la validez de constructo y criterios de jueces expertos (Terán-Bustamante et al. 2020; López Fernández et al. 2019).

Validación de constructo: Se verificó la validez del constructo se midió y analizó las dimensiones que componen la variable 1, Modelo YOLO (You Only Look Once) y la variable 2, Diagnóstico Preliminar de la enfermedad de Huanglongbing. El análisis de validez de contenido se realizó por el procedimiento de la observación de todos los indicadores basados en la clasificación de imágenes de estudios sobre el HLB.

Aplicación del Instrumento: Consistió en un primer momento en la aplicación de los instrumentos, guías de observación para observar el comportamiento de los indicadores como incrementó precisión de detección, incrementó de recuperación de detección sensibilidad, especificidad de acuerdo a lo establecido en los objetivos en dos guías de observación.

3.5 Procedimientos

Como primera actividad se desarrolló la recolección de datos, lo cual requirió un dataset con fotografías de hojas sanas sin Huanglongbing y enfermas con Huanglongbing (Healthy/Greening). Como técnica de procesamiento se usó imágenes de hojas de la planta del limón sutil (*Citrus Aurantifolia Swingle*) e imágenes hojas de la planta de limón Tahití o persa (*Citrus Latifolia Tanaka*) para el entrenamiento se empleó un 70%, un 15% para evaluación y 15% para pruebas que se ingresan al modelo YOLOv5m. Se aplicó la técnica de la observación con el instrumento de recolección como son las dos guías de observación. Se empezó analizar e implementar la construcción del instrumento para observar el comportamiento de los indicadores mediante una guía de observación que se planteó de acuerdo a las necesidades, llenando las guía de observación con información relevante y las fórmulas a ejecutar para obtener esos datos, se validó con un experto los instrumentos dando su aprobación de validez y confiabilidad, con respecto a la variable 1: Modelo YOLO, para las dimensiones Precisión y Recall lo cual se determinó en los indicadores como el Incrementó Precisión de detección lo cual se obtuvo como resultado mediante el modelo YOLO para obtener las predicciones correctas sobre las instancias totales evaluadas con la siguiente fórmula:

$$\text{Precisión (P)} = \frac{TP}{TP + FP} \times 100\%$$

y el incrementó de recuperación de detección mediante el modelo YOLO para obtener la medición de la fracción de patrones pronosticados positivos que se clasifican correctamente con la siguiente fórmula:

$$\text{Recall (R)} = \frac{TP}{TP + FN} \times 100\%$$

lo cual por consiguiente se utilizó solo una guía observación para los indicadores de la variable 2: Diagnóstico preliminar de la enfermedad de Huanglongbing, para la dimensión Prueba Diagnóstica para determinar con sus indicadores la sensibilidad mediante el diagnóstico del modelo

medir la fracción de patrones positivos que se clasifican correctamente con su fórmula:

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

y con la especificidad mediante el diagnóstico del modelo medir la fracción de patrones negativos que se clasifican correctamente con su fórmula:

$$\text{Especificidad} = \frac{TN}{TN+FP}$$

3.6 Método de análisis de datos

Para el análisis de datos recopilados de los indicadores de incrementó precisión de detección y el incrementó de recuperación de detección de la variable 1, por ello el instrumento que se aplicó fue la guía de observación por ende la técnica de procesamiento fue la clasificación de imágenes y para su análisis se utilizara la Matriz de Confusión, es una matriz de confianza (también conocida como matriz de error) es una tabla de contingencia que se utilizó para describir el rendimiento de un sistema clasificador / sistema de clasificación cuando se conoce la verdad (Yang y Berdine 2017). Cabe resaltar que cada columna (o fila) indica los números de una clase pronosticada, por ejemplo, el número de enfermedades pronosticadas o normales pronosticadas, mientras que cada fila (o columna) indica los números de una clase verdadera, por ejemplo, el número de enfermedades verdaderas o normales verdaderas.

Tabla 4. *Matriz de Confusión*

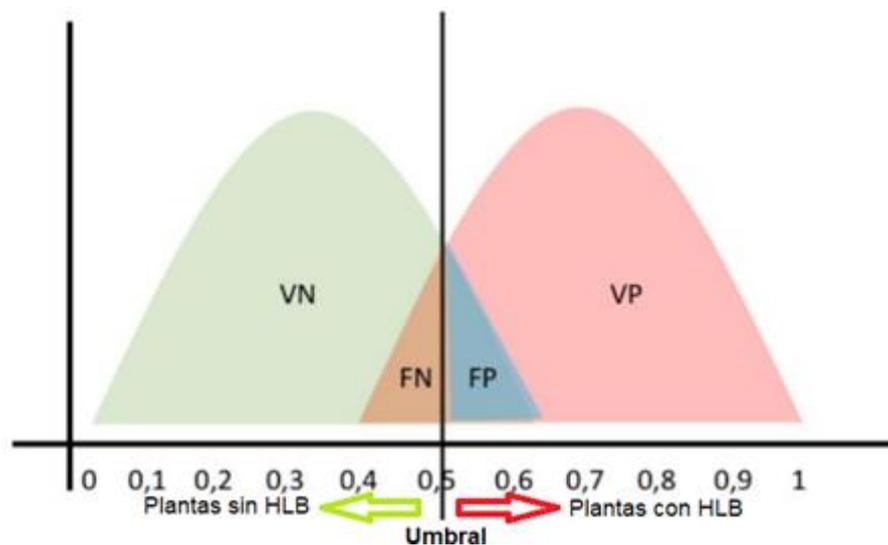
		Condición prevista	
		Enfermedad	Normal
Condición verdadera	Enfermedad	Positivo verdadero (TP)	Negativo falso (FN)
	Normal	Positivo falso (FP)	Negativo verdadero (TN)

Fuente: elaboración propia

- 1) Verdadero Positivo (VP; también llamado sensibilidad; una medida de la proporción de positivos, que se predicen correctamente dado que es realmente positivo),
- 2) Falso negativo (FN; una medida de la proporción de negativos predichos, dado que es realmente positivo),
- 3) Falso Positivo (FP; una medida de la proporción de positivos predichos, dado que es realmente negativo), y
- 4) Verdadero negativo (TN; también llamado especificidad; una medida de la proporción de negativos predichos, dado que es realmente negativo).

Para los datos recopilados de los indicadores sensibilidad y especificidad de la variable 2, el instrumento que se aplicó fue la guía de observación por ende la técnica de procesamiento fue la clasificación de imágenes y para su análisis se utilizó la Curva ROC ya que aborda los problemas que se enfrentó durante la clasificación entre dos grupos (Yang y Berdine 2017). Por ejemplo, Plantas con HLB y Plantas sin HLB y establecer un umbral o punto de quiebre donde se define:

Figura 1. Curva ROC



Fuente: Adaptado de curvas ROC y área bajo la curva (Gonzalez 2019).

- 1) Verdadero positivo (VP; Plantas con HLB clasificados correctamente),
- 2) Falso negativo (FN; Plantas con HLB clasificados en forma incorrectas),
- 3) Falso positivo (FP; Plantas sin HLB clasificados en forma incorrecta), y
- 4) Verdadero negativo (VN; Plantas sin HLB clasificados correctamente).

Analizando el comportamiento de la variable Modelo YOLO y variable Diagnóstico preliminar de la enfermedad de Huanglongbing de la investigación, se usó la estadística descriptiva que representa información que se puede utilizar como base para comparar cómo difieren las series de datos (Green et al. 2023), donde utilizó gráficos estadísticos, correlogramas y cuadros de comparación con la finalidad de analizar los datos y obtener las conclusiones sobre la población establecida.

3.7 Aspectos éticos

Se tienen en consideración los siguientes aspectos éticos en la ejecución de la investigación:

- Respeto por las teorías, conceptos de los autores aplicando la norma internacional ISO 690 requerida por la Universidad César Vallejo al citar y referenciar implica una lista para la bibliografía ordenada y con elementos de referenciación a un entorno formal como lo es la identificación de documentos en internet (DOI) dando acceso a las referencias que se tiene en base al proyecto ya que es un identificador único que se le asigna a un objeto de forma digital que contiene una confiabilidad y credibilidad además de ser utilizada y reconocida en el ámbito académico.
- Se defiende la originalidad de la investigación, la formalidad y veracidad de los resultados obtenidos por los dos investigadores y rechazando el plagio alguno.

Con respecto al artículo 64, Código de Ética en Investigación de la Universidad César Vallejo, la investigación ejecutada cumple con los requisitos que dictan el artículo en mención la cual dice específicamente en el capítulo 9 de la ética en la investigación, que toda investigación realizada en Universidad César Vallejo tiene que contener integridad científica, cumplir estándares que dictan el rigor científico que amerita, honestidad, responsabilidad protegiendo el bienestar y los derechos de los participantes.

Para el entrenamiento del Modelo YOLO, se usó imágenes de dos Dataset respetando los derechos de los autores agregando el bib en la información y el proyecto de investigación requirió la construcción de un dataset por lo que se recurrió a un campo donde la enfermedad de Huanglongbing (HLB) se sabía que existía como lo es la frontera Perú-Ecuador respetando la anonimidad.

IV. RESULTADOS

Primer objetivo específico: Evaluar las predicciones sobre las instancias totales en cuanto a la precisión y recuperación

Se realizaron 6 experimentos de prueba para evaluar el modelo YOLO en su versión 5 y a su vez también la detección preliminar de la enfermedad de Huanglongbing (Dragón Amarillo) correspondiente a las métricas de rendimiento que arrojo YOLOv5, como la precisión (Accuracy), recuperación (Recall), AP, mAP, mAP@50, mAP@[0.5:0.95], F1-Score y la clasificación de la Sensibilidad, Especificidad con los 3 tipos de datasets:

El primer dataset propio que se tomó en campo en la frontera de Perú y Ecuador, comprendió de 752 imágenes de hojas del limón sutil (*Citrus Aurantifolia Swingle*) en la cuales sus características de las imágenes que se lograron tomar, hay dos tipos de tomas de imágenes, una en campo natural/ambiente natural sin intervención y otra porción fue en espacio controlado solo hojas con fondo uniforme con intervención:

Hojas del Dataset Propio: Entorno/Ambiente natural (Enfermas y Sanas)



Figura 2. Dataset propio ambiente natural

Hojas del Dataset Propio: Entorno/Ambiente Controlado (Enfermas y Sanas)



Figura 3. Dataset propio ambiente controlado

Tabla 5. Descripción del dataset propio

Dataset	Cantidad
propio	752

Fuente: elaboración propia

El segundo dataset llamado detección_huanglongbing que se tomó fue de la página Roboflow que comprendió de 1354 imágenes de hojas de limón persa (*Citrus Latifolia Tanaka*) en entorno/ambiente natural:

Tabla 6. Descripción del dataset *deteccion_huanglongbing*

Dataset	Cantidad
detección_huanglongbing	1354
Bib (Autorización/Referencia de uso del Dataset al Autor)	
<pre>@misc{ enfermedades-en-limón-persa_dataset, title = { Enfermedades en limón persa Dataset }, type = { Open Source Dataset }, author = { mafer }, howpublished = { \url{ https://universe.roboflow.com/mafer-6tca8/enfermedades-en-limón-persa } }, url = { https://universe.roboflow.com/mafer-6tca8/enfermedades-en-limón-persa }, journal = { Roboflow Universe },</pre>	

<pre> publisher = { Roboflow }, year = { 2023 }, month = { sep }, note = { visited on 2023-11-03 }, } </pre>
Link
https://universe.roboflow.com/fase-1-loeq2/deteccion-de-enfermedades-en-limón-persa

Fuente: elaboración propia

El tercer dataset llamado enfermedades_limón_persa que se tomó fue de la página Roboflow que comprendió de 3369 imágenes de hojas de limón persa (Citrus Latifolia Tanaka) en entorno/ambiente natural:

Tabla 7. Descripción del dataset enfermedades_limón_persa

Dataset	Cantidad
enfermedades_limón_persa	3369
Bib (Autorización/Referencia de uso del Dataset al Autor)	
<pre> @misc{ enfermedades-en-limón-persa_dataset, title = { Enfermedades en limón persa Dataset }, type = { Open Source Dataset }, author = { mafer }, howpublished = { \url{ https://universe.roboflow.com/mafer-6tca8/enfermedades-en-limón-persa } }, url = { https://universe.roboflow.com/mafer-6tca8/enfermedades-en-limón-persa }, journal = { Roboflow Universe }, publisher = { Roboflow }, year = { 2023 }, month = { sep }, note = { visited on 2023-11-03 }, } </pre>	
Link	
https://universe.roboflow.com/mafer-6tca8/enfermedades-en-limón-persa	

Fuente: elaboración propia

Para el entrenamiento en los 6 experimentos se usó una división de 70% para entrenamiento (train), 15% evaluación (eval) y 15% pruebas (test).

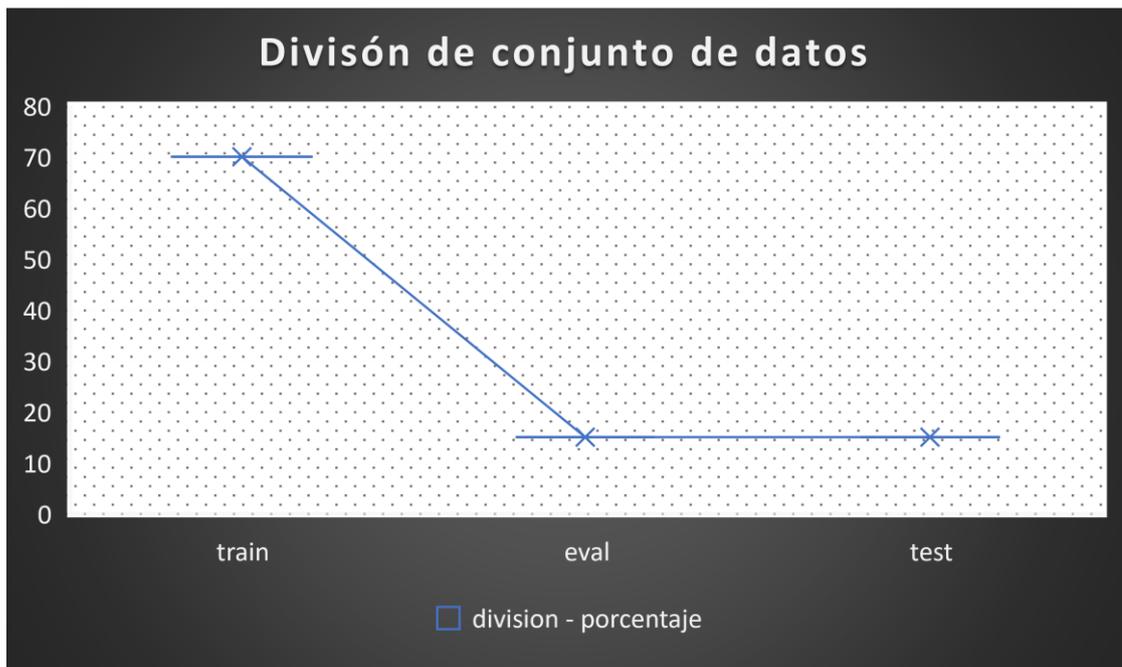


Figura 4. Caja y Bigotes de la división de conjunto de datos

La computadora que se realizaron los experimentos fue de las siguientes especificaciones:

Tabla 8. Especificaciones de la PC para correr el Modelo Yolov5

Especificaciones	Valores
CPU	11th Gen Intel(R) Core (TM) i5-11400 4.400GHz
RAM	16GB
GPU	NVIDIA GeForce RTX 3060 Lite Hash Rate
GPU RAM	12 GB
Disco	500 GB SSD / 1 TB HDD
OS	Arch Linux x86_64

Fuente: elaboración propia

Para el etiquetado se usó la herramienta **Yolo Mark RS** para dibujar cuadros delimitadores conocidos como **Bounding Box** alrededor de las hojas que contiene **Huanglongbing** para asignarle el tipo de clase según la etiqueta al que pertenece, la herramienta es de código abierto en GitHub: https://github.com/MarcusDunn/yolo_mark_rs

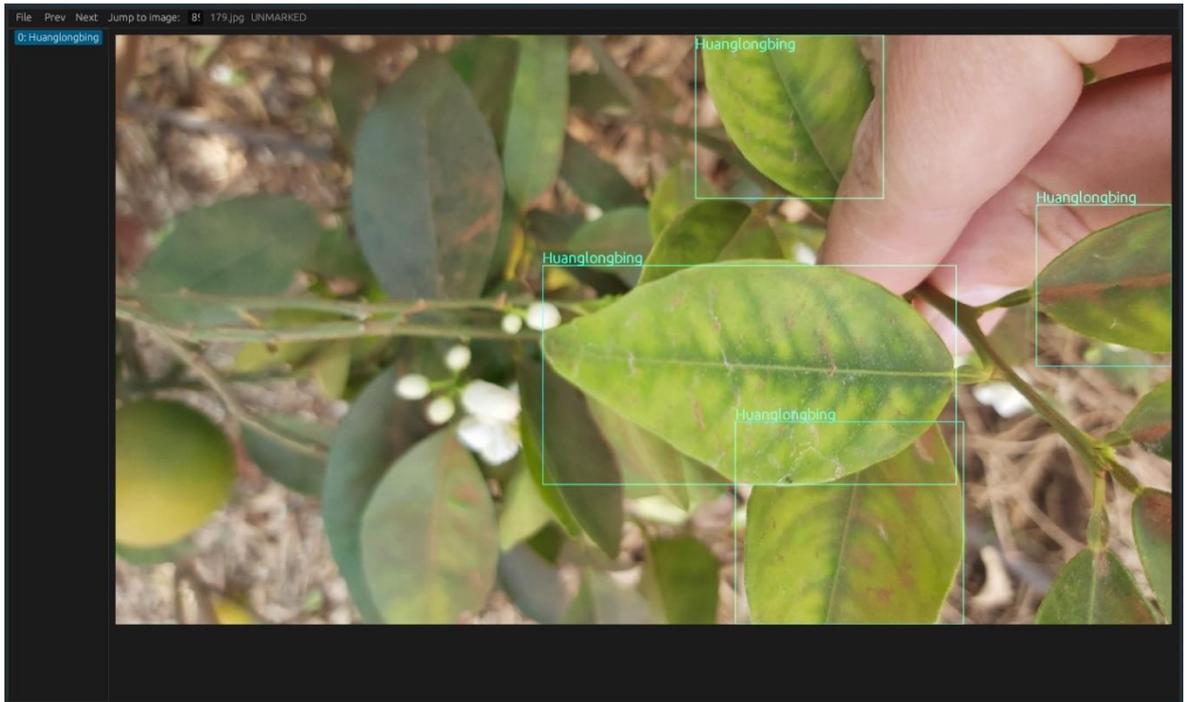


Figura 5. Cuadros delimitadores según la etiqueta huanglongbing 1

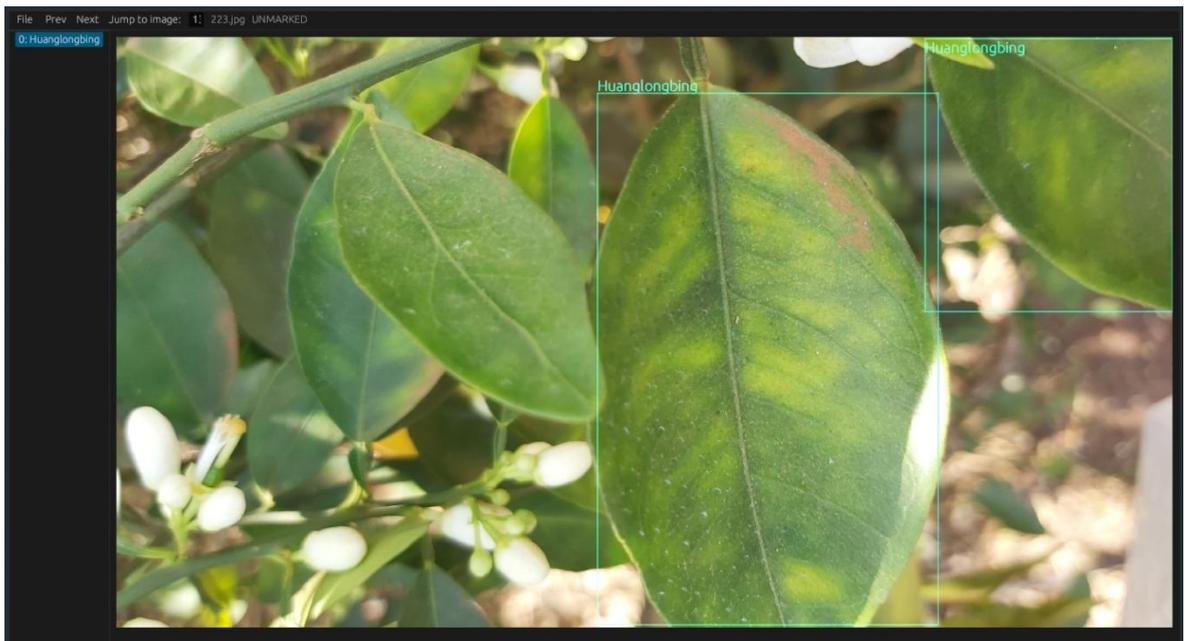


Figura 6. Cuadros delimitadores según la etiqueta huanglongbing 2



Figura 7. Cuadros delimitadores según la etiqueta huanglongbing 3

El IDE de desarrollo fue Visual Studio Code con las extensiones necesarias para que Python en su versión 3.8 logre ejecutarse.

Como se muestra en la **tabla 9**, cada uno de los experimentos fue con el mismo número de épocas salvo el primer experimento que se probó con un punto de partida de 500 épocas por ser un modelo más pequeño, en cuanto al tamaño del lote (**batch size**) se eligió 16 el más pequeño de los lotes por las limitaciones técnicas al principio ya que es mejor cuando se cuenta con memoria y GPU limitado, en cuanto a los demás experimentos 2 - 6 si se realizaron en 300 épocas con el tamaño del lote en automático (**Auto-Batching**) ya que después se obtuvieron los recursos necesarios para una ejecución óptima de Yolov5m y el entrenamiento es más acelerado, el tamaño de entrada de la imagen fue 640x640 pixeles para el arquetipo Yolov5m ya que es adecuado para la detección del Huanglongbing en regiones de interés (**ROI**), en caso de que el conjunto de datos (Dataset) sea pequeño sirve como transferencia de conocimiento ya que se puede aprovechar el modelo que se pre-entreno para complementarlo con un dataset mucho más grande para un óptimo rendimiento reduciendo las épocas. Los mejores experimentos con mejores resultados fueron el Exp4 con una precisión constante y el Exp5 con una precisión con saltos bruscos.

Tabla 9. *Tabla de experimentos por épocas*

Exp	Epoch	Batch size (Tamaño del lote)	Dataset	Image Size	Model base	Precisión	Recall	mAP_0.5	mAP_0.5:0.95	Error	Tiempo de Entrenamiento
Exp	500	16	Own (propio)	640px	Yolov5m	1%	0.2%	1.6%	0.1%	99%	7 h y 15 m
Exp2	300	automático	Own (propio)	640px	Yolov5m	20%	8%	6.6%	2.5	80%	3 h y 25 m
Exp3	-	-	-	-	-	-	-	-	-	-	-
Exp4	300	automático	Detección-de-enfermedades-en-limón-persa_dataset	640px	Yolov5m	67%	55.5%	58.8%	49%	33.3%	3 h y 51 m
Exp5	300	automático	Own+Exp4	640px	Exp4	55%	29%	30%	15%	45%	4 h y 1 m
Exp6	300	automático	Enfermedades-hojas-limón-persa	640px	Exp4	26%	14%	9%	6%	74%	3 h y 58 m

Fuente: elaboración propia

En la **figura 8** las métricas se observan los valores por cada época que el modelo fue aprendiendo, en cuanto a la precisión y la recuperación (Recall) del Epx4 alcanzan un pico estable en las 300 épocas alcanzando la mejor estabilidad en cuanto a la precisión en la época 200 y en la recuperación mantiene una estabilidad desde la época 100.

Con la métrica de precisión es de 0.67 que es muy buena con un porcentaje de 67%, métrica de recall (recuperación) es de 0.55 en porcentaje fue de 55%, en cuanto a la métrica mAP_0.5 es de 0.558 que se adecua para la investigación porque hay un equilibrio 55.8% ya que combina la precisión y el recall para obtener la métrica de mAP_0.5 y con la métrica de mAP_0.5:0.95 es de 0.49 en porcentaje fue de 49%.

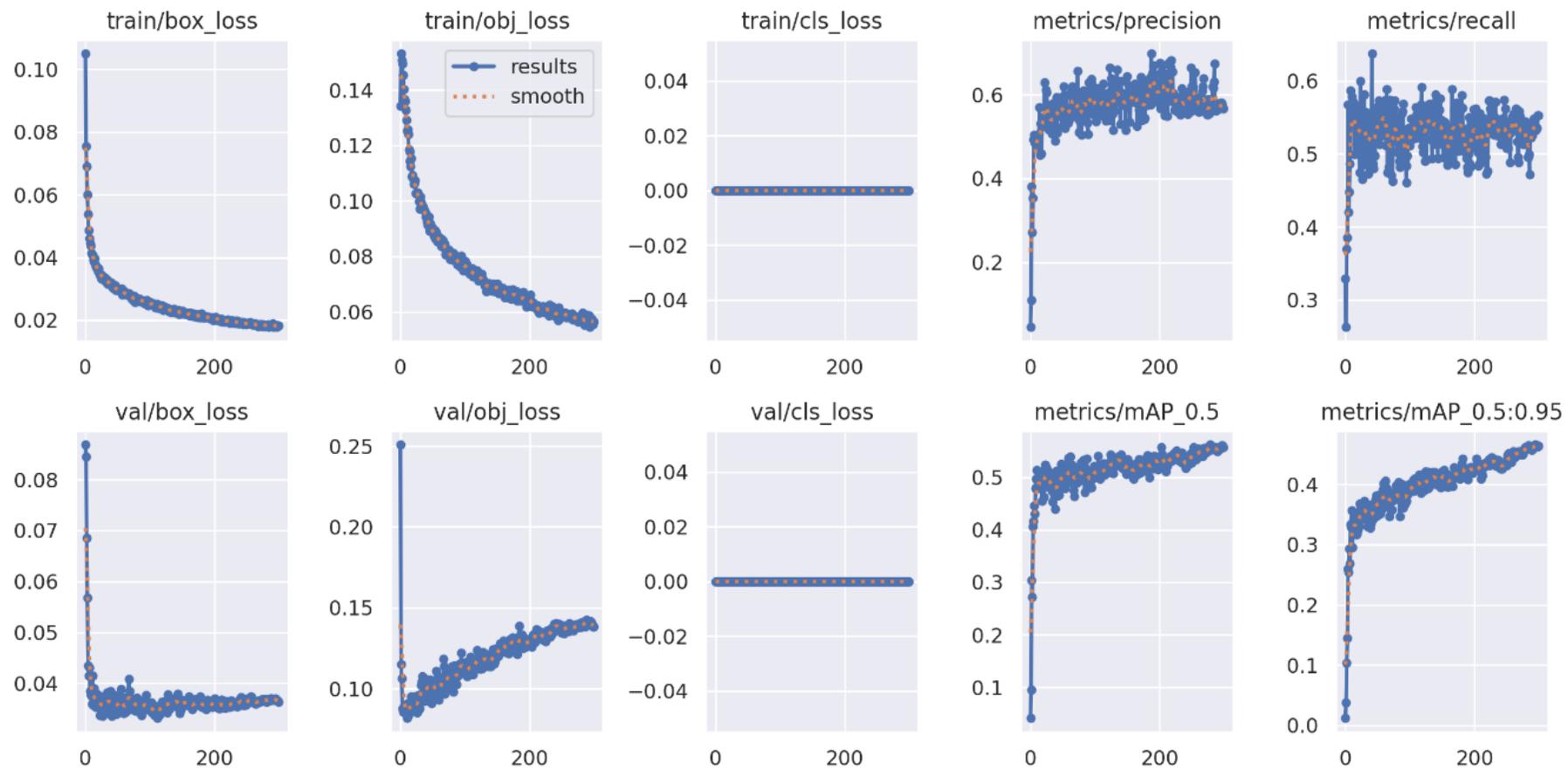


Figura 8. Métricas de entrenamiento del exp4

En la figura 9 en la curva de Precisión y Recall, permitió observar el punto de la recuperación (recall) comienza a sufrir una degradación de precisión y a la inversa ya que requiere mantener un equilibrio, es donde se calcula la métrica de la media de precisión $mAP@0.5$ obteniendo 0.558 ya que es una de las forma de obtener el área bajo la curva (PR AUC o Curva P-R) como resultado de integrar la curva, ya que es muy útil para la evolución del rendimiento de Yolov5m, el mejor de lo casos es que la curva esté cerca de esquina límite superior derecho, cuanto más cerca sea del valor 1 es mejor el modelo.

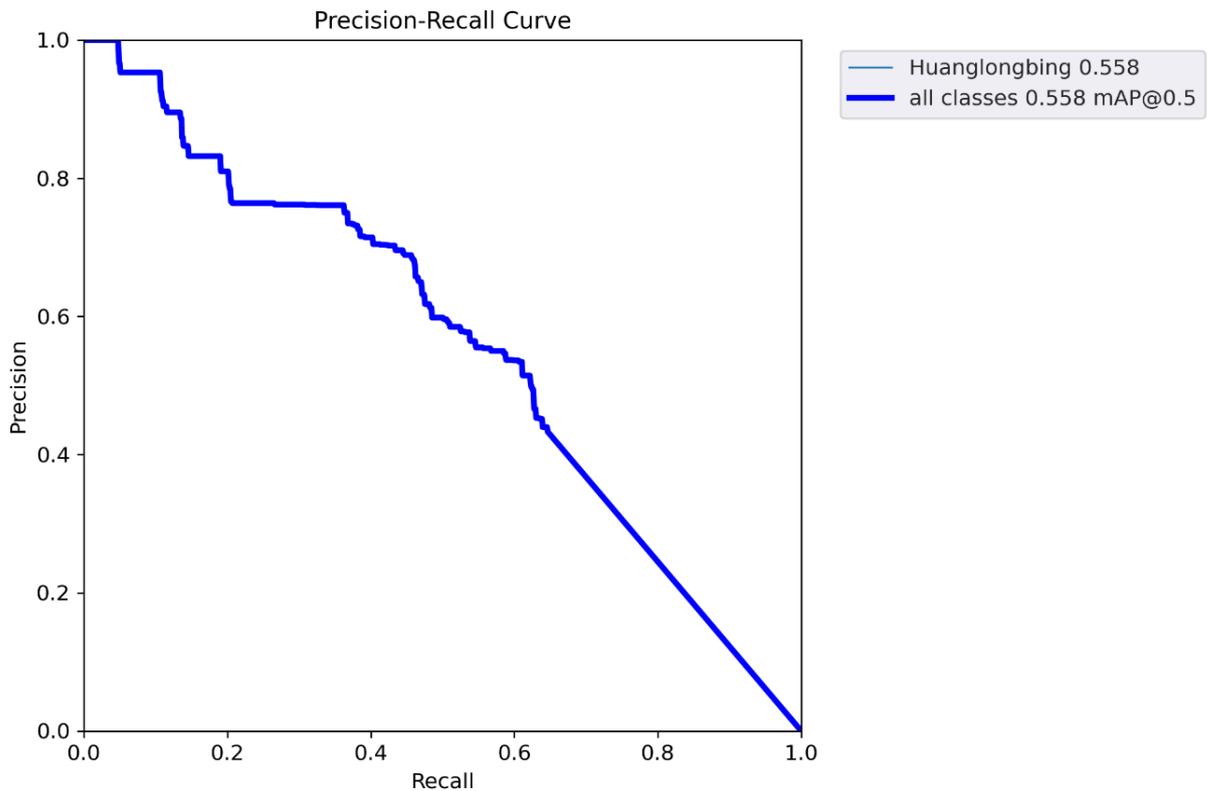


Figura 9. Curva de precisión – recuperación (recall) del exp4

En la figura 10 se necesita la figura 11 y 12 para F1-Confidence Curve o F1-Score se centra en la evaluación del rendimiento en cuanto a la confianza del modelo en las predicciones teniendo en cuenta la combinación de Recall y Precisión en un contexto diferente para obtener una sola puntuación, ya que es el equilibrio a optimizar entre recall y precisión, el grado de confianza que se le dio a optimizar fue de 0.029 obtenido 0.56 de F1-Score en porcentaje fue el F1-Score 56%, en casos diferentes el grado de confianza se puede aumentar para a su vez aumentar el F1-Score alto teniendo en cuenta que la precisión aumenta como lo demuestra el gráfico de la figura 10 y el recall disminuirá como lo muestra el gráfico de la figura 11, aunque subiendo el valor de confianza el F1-Score puede tender a la baja es lo que no se desea en la investigación.

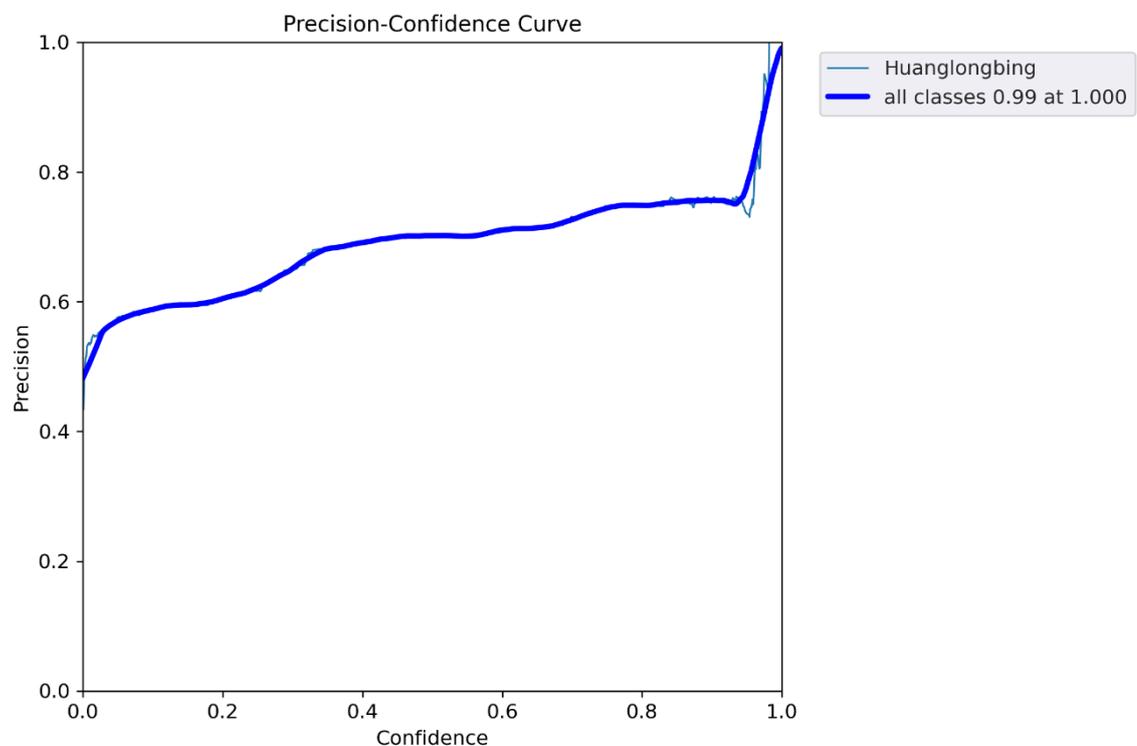


Figura 10. Curva confidencia de la precisión del exp4

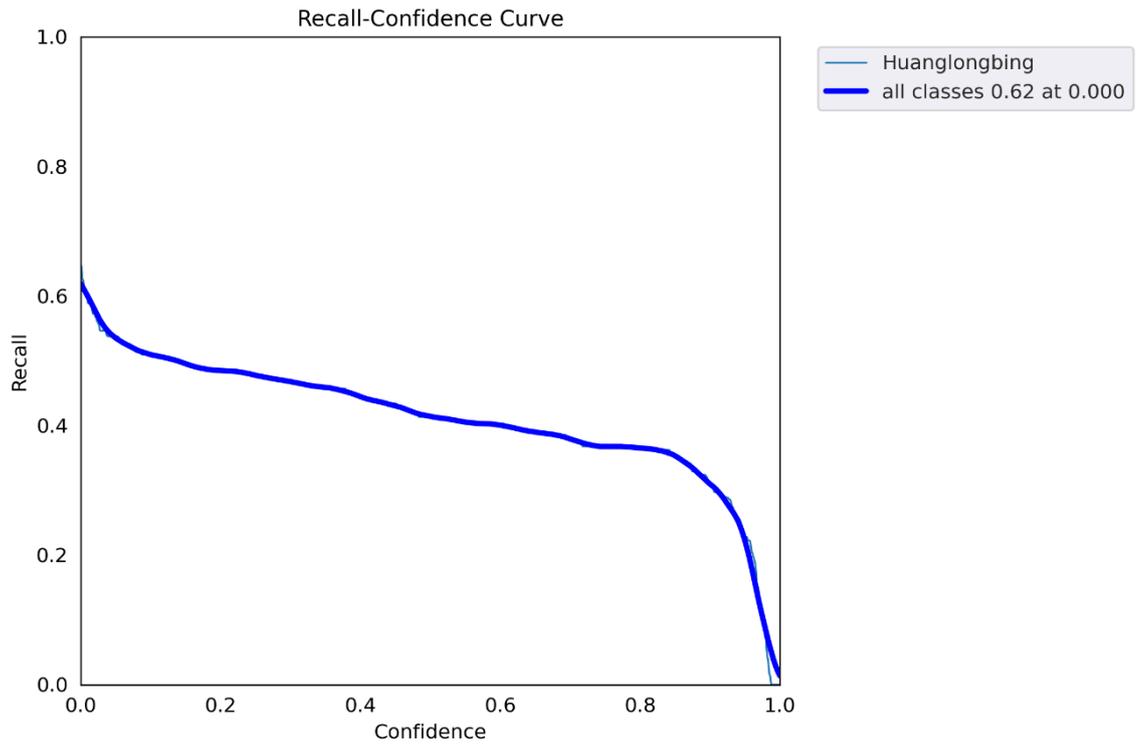


Figura 11. Curva confianza de recuperación (recall) del exp4

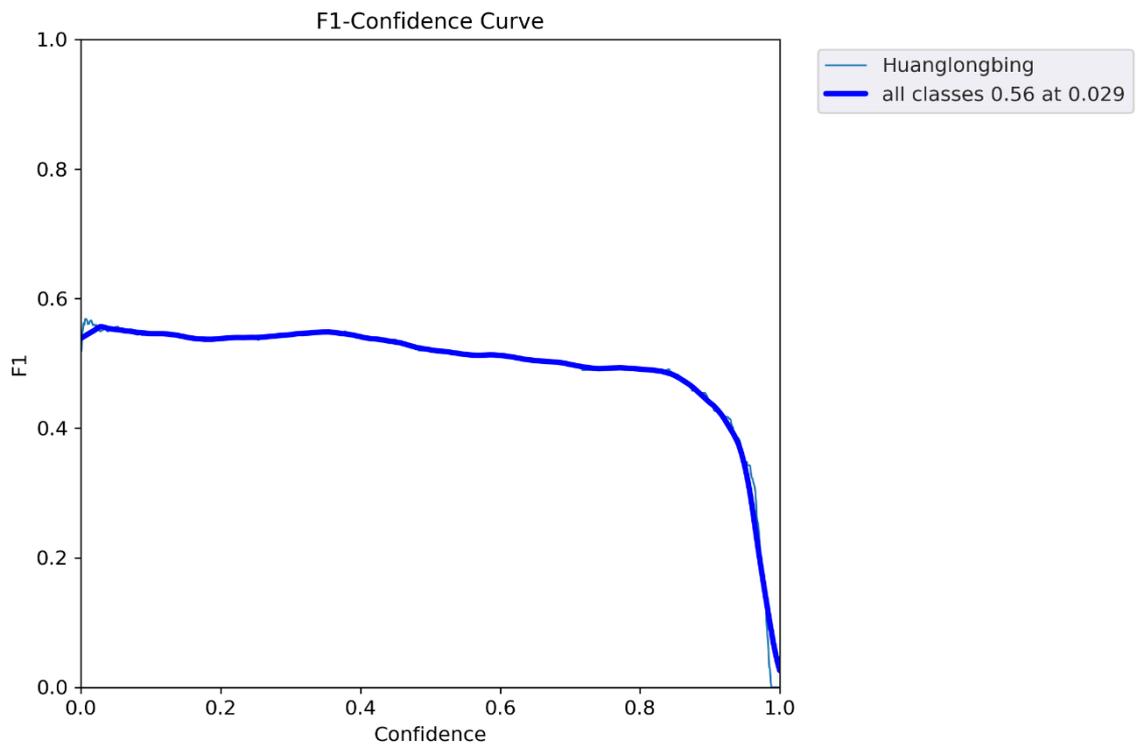


Figura 12. F1 Score - Curva de confianza del exp4

En la figura 13 se observa la matriz de confusión como se trabajó con una clase que es “Huanglongbing”, en la parte superior a la izquierda Huanglongbing/Huanglongbing se muestra la precisión que es 0.67 son los casos reales (Verdaderos Positivos), en la parte inferior a la izquierda Background/Huanglongbing es el error de 0.33, en la parte superior a la derecha Huanglongbing/Background es de 1.00, el Background es la parte que no está detectando y que no debería detectar pero es parte del contexto que está afuera del área a detectar, el valor es 1 porque es la sumatoria de la columna del Background, además si hubiera más clases ese valor que es 1.00 varía pero para la investigación se consignó una clase como lo es “Huanglongbing”, cabe resaltar que la matriz de confusión está normalizada porque todo los valores se llevan de 0 a 1.

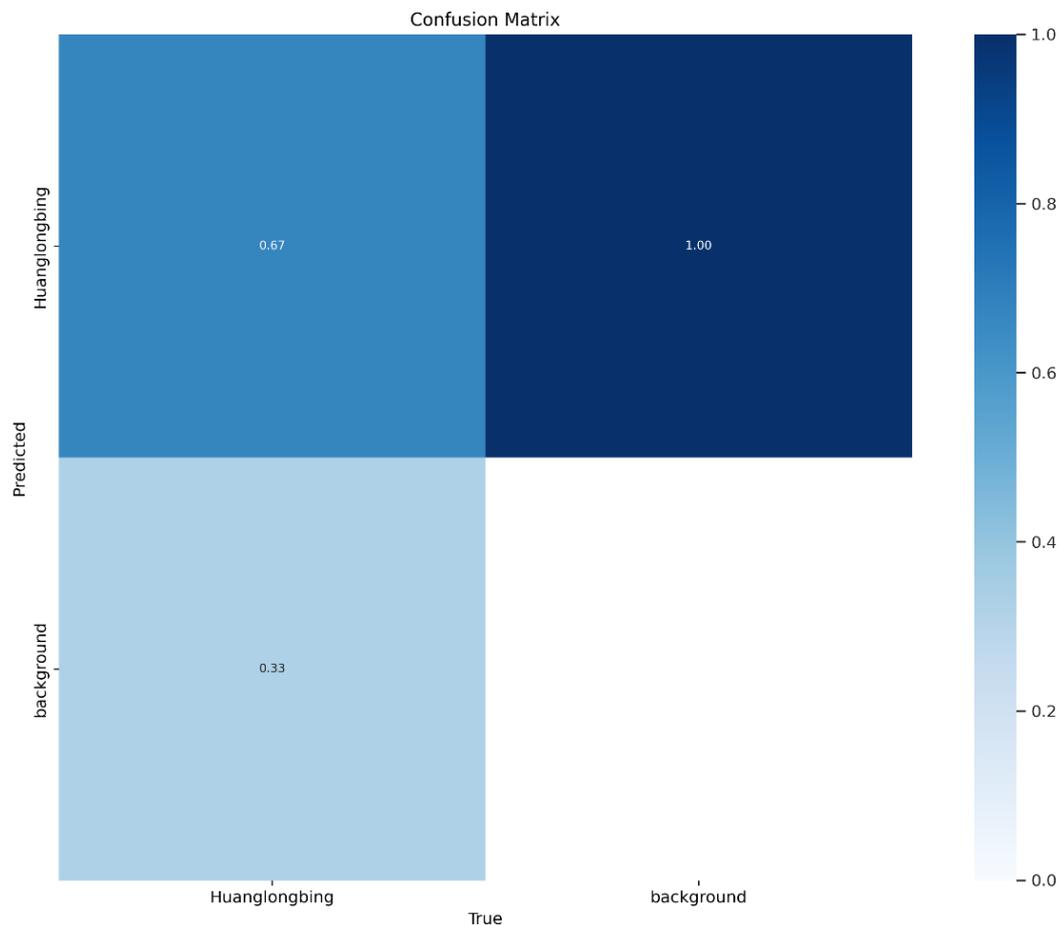


Figura 13. Matriz de confusión del exp4

En la figura 14 se aprecia en la parte superior a la izquierda son las instancias de las imágenes de hojas lo que muestra la barra roja cuando lo labels se tiene el objeto que se va a detectar cuando se hace el etiquetado es una instancia de un objeto si esa imagen aparece una vez, 10 veces esa instancia o pueden aparecer muchas instancias de ese objeto/hoja en el cuadro rojo de la imagen en la detección en las 300 épocas con un resultado de 17,500 instancias del objeto en el conjunto de datos que se entrenó sabiendo que es solo una clase “Huanglongbing”.

En la parte inferior izquierda se muestran puntos “Outliers” juntos de color azul y grises, donde se agrupan puntos azules con un color intenso en medio como lo muestra el gráfico es que las instancias las ha detectado correctamente y los puntos que están dispersos color gris son la detecciones que no se asemejan a las instancias mientras más alejado de los puntos acumulados azules la detección de esa instancia es errónea o que no corresponde, en la parte superior a la derecha son los cuadros delimitadores que se etiquetaron con Yolo Mark RS (w,h,x,y) son las etiquetas de cada una de las instancias de lo que fue en el entrenamiento.

En la parte inferior a la derecha son los tamaños de los insides de la imágenes, cada imagen en Yolov5m se normaliza tomando el ancho y altura del box de la instancia por cada tamaño se encuentra una cajita en una imagen de la hoja que se mandó en el dataset siendo pequeñas en su entorno natural representando el montón de puntos azules juntos bordeando el 0.1 del eje “x” y el 0.1 del eje “y”, algunas grandes en entorno controlado que están en el dataset ahí sí aumentaría el tamaño como se ve en el gráfico la dispersión de grises porque solo un porcentaje de imágenes están cerca y las otras no.

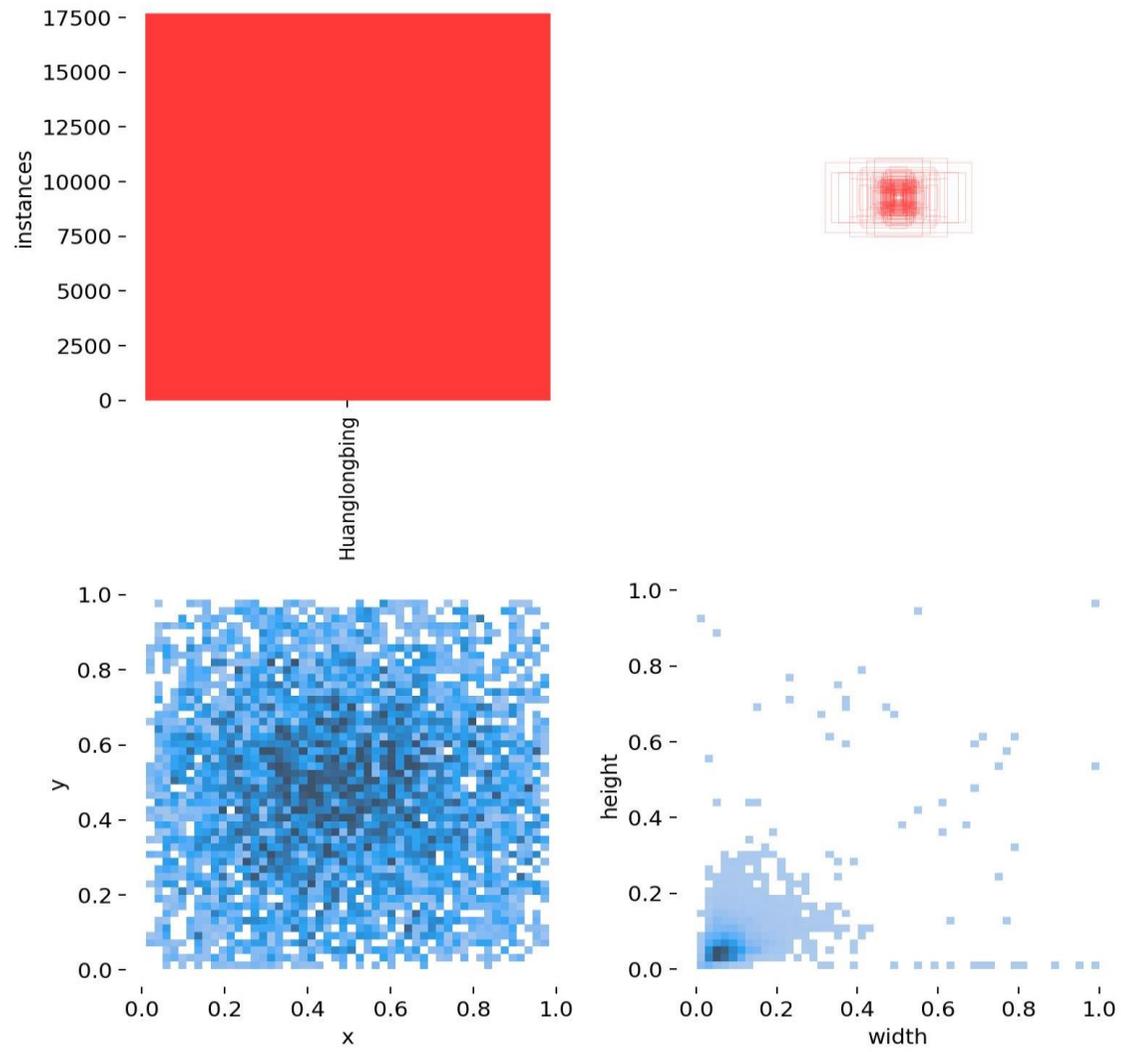


Figura 14. Distribución de labels del exp4

En la figura 15 es la parte discriminativa del correlograma de etiquetas, donde se analizó e identificó las relaciones o patrones entre las etiquetas (labels) en una representación visual de una distribución espacial, indicando en los eje x-y la cuadrícula mientras que el ancho y alto proporciona la confianza de detecciones en esa región, los gráficos de barras al costado muestran la frecuencia de las detecciones del Huanglongbing, representando la cantidad de veces que la clase Huanglongbing aparece en las imágenes.

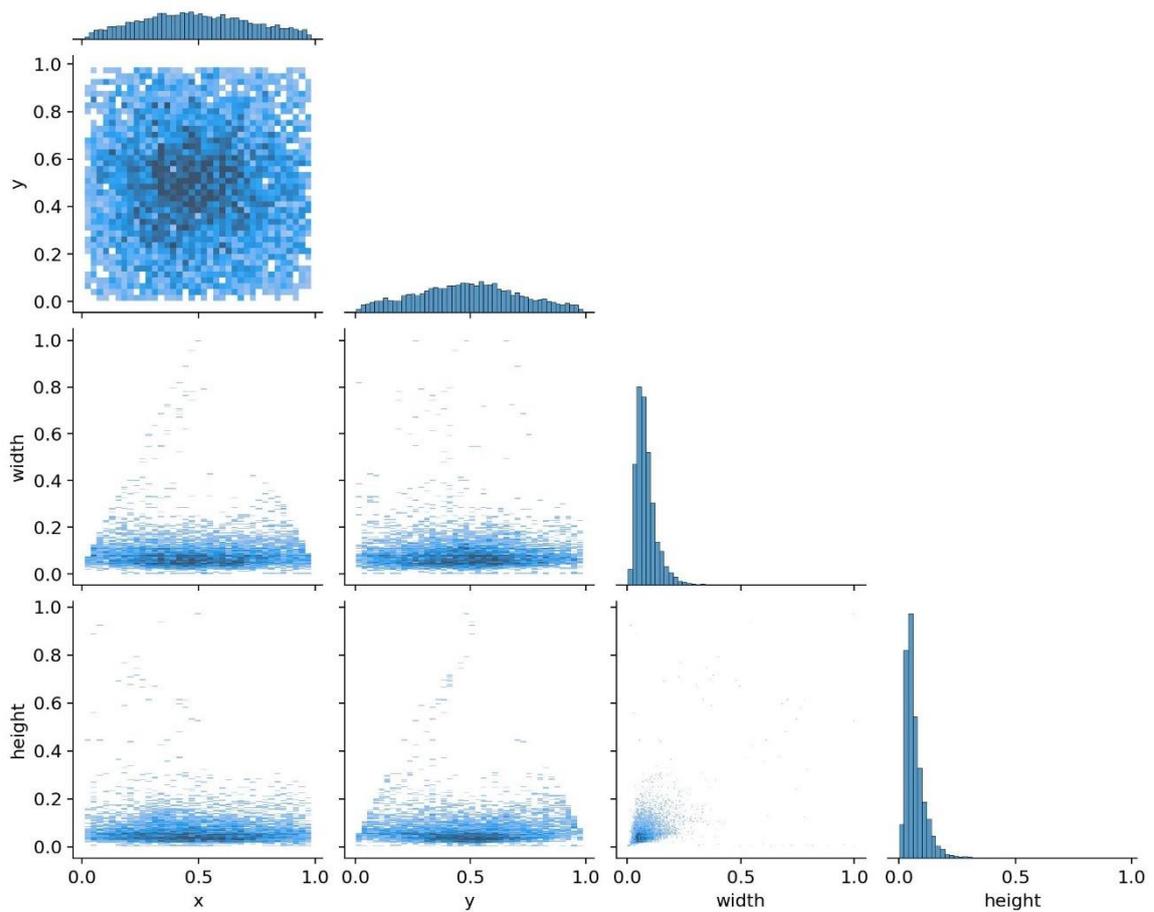


Figura 15. Correlograma de labels del exp4

Segundo objetivo específico: Clasificar por patrones positivos utilizando el modelo YOLO la sensibilidad en el diagnóstico preliminar de la enfermedad del Huanglongbing en la planta de limón.

En el indicador de sensibilidad se resume en el resultado positivo o que están infectados de huanglongbing clasificando por 300 épocas, obteniendo el mejor resultado entre las 200/300 épocas.

Tabla 10. Prueba diagnóstica positiva exp4

Iteración	Enfermos (con Huanglongbing)		Total Positivos
	Verdadero Positivo (VP)	Falso Negativo (FN)	
50/300	309	262	571
100/300	292	279	571
150/300	293	278	571
200/300	320	251	571
250/300	290	281	571
300/300	307	264	571

Fuente: elaboración propia

Fórmula Sensibilidad:

$$VP/VP+FN$$

En la figura 16 se aprecia el gráfico de la sensibilidad sobre el diagnóstico preliminar de la enfermedad del Huanglongbing en la que se iteró en 300 épocas, en la época 50 se obtuvo un 54.12% de media, en la época 100 la media decreció a 51.14%, en la época 150 la media incrementó a 51.31% con respecto a su antecesora, en la época 200 se obtuvo el mejor pico de media más alta con 56.04% con respecto a las 300 épocas, en la época 250 fue la peor media 50.79% de todas las épocas, en la época 300 se recuperó y se mantuvo estable con la media 53.77%. La sensibilidad media concluyente fue de 52.86% con una desviación típica de 2.11%.

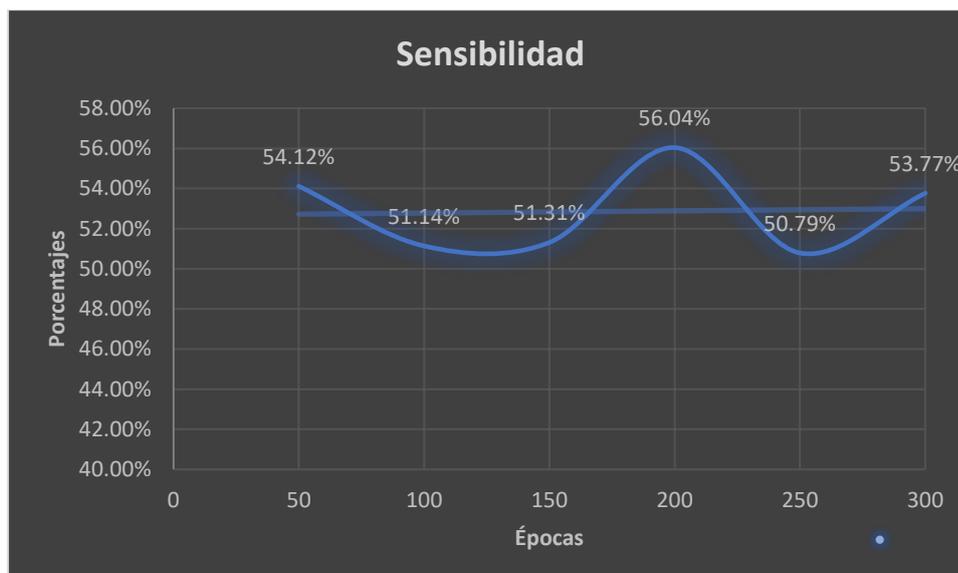


Figura 16. Dispersión de la sensibilidad en el diagnóstico preliminar del HLB

Tercer objetivo específico: Clasificar por patrones negativos utilizando el modelo YOLO la especificidad en el diagnóstico preliminar de la enfermedad del Huanglongbing en la planta de limón.

Este indicador de especificidad se resume en el resultado negativo o que no están infectados de huanglongbing clasificando por 300 épocas, obteniendo los mejores resultados desde la época 150 hasta la época 250.

Tabla 11. Prueba diagnóstica negativa del exp4

Iteración	No enfermos (Sin Huanglongbing)		Total Negativos	Precisión	Error
	Falso Positivo (FP)	Verdadero Negativo (VN)			
50/300	253	318	571	0.55014	0.44986
100/300	218	353	571	0.57299	0.42701
150/300	162	409	571	0.64418	0.35582
200/300	212	359	571	0.60158	0.39842
250/300	178	393	571	0.61932	0.38068
300/300	225	346	571	0.56651	0.43349

Fuente: elaboración propia

Fórmula Especificidad:

$$VN/(VN+FP) *100$$

En la figura 17 se aprecia el gráfico de la especificidad sobre diagnóstico preliminar de la enfermedad del Huanglongbing en la que se iteró en 300 épocas, en la época 50 se obtuvo un 55.69% de media, en la época 100 fue incrementando la media fue de 61.82%, en la época 150 fue incrementando la media fue de 71.63% obteniendo el mejor pico de la media más alta con respecto a las 300 épocas, posteriormente en la época 200 la media bajó a 62.87% manteniendo superioridad a la época 50 y 100, en la época 250 la media aumentó a 68.83% y en la época 300 la media bajó considerable a un 60.60% pero manteniendo superioridad con la época inicial. La especificidad de su media concluyente fue de 63.57% con una desviación típica de 5.78%.



Figura 17. Dispersión de la especificidad en el diagnóstico preliminar del HLB

En el modelo Yolov5m no es muy común sacar la curva ROC ya que existe métricas que dan información precisa, aunque investigaciones si la manejan para la investigación se sacó la curva ROC que no es igual a la Curva de Precisión-Recall considerando que cambian valores además que es solo una clase la cual es "Huanglongbing", con un umbral de 0.45, se ha obtenido con la librería sklearn (Python) obteniendo la métrica de ROC_CURVE, el resultado se calculó mediante el resultado de F1-Score, lo que interesa en la curva es

que este cerca de esquina límite superior izquierdo para que Yolov5m no incluya más falsos positivos cuando se aumenta la sensibilidad.

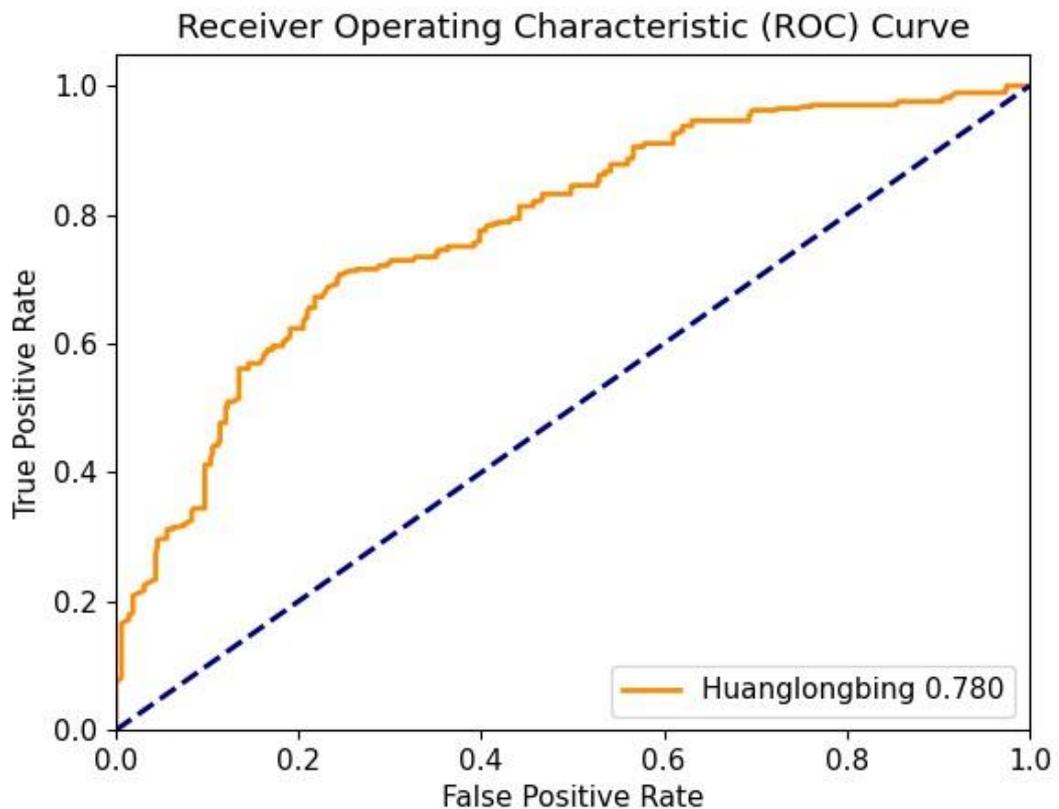


Figura 18. Curva ROC (Receiver Operating Characteristic)

En la curva ROC se muestra que en la clase “Huanglongbing” tiene la capacidad que en las imágenes de hojas se distinguen entre las que cuentan o no cuentan con Huanglongbing, como se muestra en la **tabla 12** el área bajo la curva es de 0.780.

Tabla 12. Especificación de la Curva ROC

Área bajo la curva	
Clases	Área bajo la curva
Huanglongbing	0.780

Fuente: elaboración propia

V. DISCUSIÓN

De acuerdo con el primer objetivo específico: Evaluar las predicciones sobre las instancias totales en cuanto a la precisión y recuperación, se tomó en cuenta los indicadores incrementó precisión de detección e incrementó de recuperación de detección.

La propuesta de Xing y otros, centrada en el indicador de precisión de detección, se basa en un trabajo realizado en Corea con un modelo CNN. Su conjunto de datos abarcó 17 especies de plagas y 7 enfermedades que afectan a los cítricos, con 9051 imágenes de plagas y 3510 imágenes de enfermedades tomadas con una cámara de alta resolución desde Internet. La distribución porcentual fue de 4:1:1 para entrenamiento, validación y pruebas, con imágenes redimensionadas a 224 x 224 píxeles. En la comparación de redes, **Xing** destacó DenseNet-16, logrando una precisión de detección del 93.33%. Su argumento se centra en la eficiencia, resaltando que esta red liviana es apta para dispositivos móviles, subrayando la importancia de considerar no sólo la precisión, sino también la eficiencia en términos de recursos (Xing, Lee y Lee 2019).

El estudio de Qiu y otros en China, financiado por una entidad privada, se centró en la evaluación de varias variantes del modelo YOLO (Yolov5I-HLB1, Yolov5I-HLB2 y Yolov5I-HLB3). Se recolectaron datos de 12 huertos en la ciudad de Fujian, utilizando una tarjeta gráfica GPU GeForce RTX 3090 para el entrenamiento y fijando parámetros como un tamaño de lote de 20 durante 200 épocas. La distribución porcentual incluyó un 80% para entrenamiento, 10% para validación y otro 10% para métricas, con una resolución de 640 píxeles. Se estableció un umbral de confianza de 0,4.

En cuanto a los resultados, se destacan cifras importantes, con una precisión del 90.07%, recall del 81.84%, y F1 del 85.19%, mostrando notables diferencias con otras variantes de modelos. En la matriz de confusión del modelo HLB2 fue aplicado en el reconocimiento de 12 enfermedades en cítricos bajo diversas condiciones de adquisición, reveló que la inclusión de imágenes con síntomas irrelevantes no contribuyó a mejorar la precisión del modelo (Qiu et al. 2022).

En el proyecto de investigación se logró medir el indicador precisión de detección y recuperación de detección. Se usó el modelo YOLO al igual que **Qiu y otros** ahora mismo está a la vanguardia de la tecnología de detección de objetos como lo es el modelo Yolov5, a diferencia de **Xing y otros** que uso DensetNet-16. En el proyecto de investigación la precisión se obtuvo dentro de las 300 épocas, alcanzando la mejor precisión en la época 200 y una estabilidad desde la época 100, ya que más épocas más es el tiempo de entrenamiento, en la investigación el proyecto se tardó de 3 horas y 51 minutos en entrenar, configurando el batch size o tamaño de lote en automático para aprovechar los recursos de la tarjeta gráfica.

Qiu sus experimentos los realizó con 200 épocas y un batch size de 20 mejorando el tiempo de entrenamiento ya que utiliza una gráfica potente que le permitió manipular el batch size a placer junto con los demás parámetros.

El proyecto de investigación sobre el diagnóstico preliminar del Huanglongbing se realizó con 6 experimentos, obteniendo los mejores resultados en los experimentos 4 y 5, siendo el mejor el 4. En cuanto a **Xing y otros** tuvo un dataset de 7 enfermedades y 17 plagas, comprendiendo 12,561 imágenes entre plagas y enfermedades. Mientras que **Qiu** su muestra y el lugar donde se recogió la unidad de análisis fue más amplia en 12 huertos diferentes de la ciudad de Fujian en China, comprendiendo su dataset de 11,946 de imágenes de frutas-hojas sanas tanto sanas como enfermas con y sin Huanglongbing.

En el dataset del proyecto de investigación se consideró solo una clase que va a ser una plaga “Huanglongbing” a estudiar obteniendo 751 imágenes de hojas de Limón Sutil en campo que se tomó en la frontera de Perú y Ecuador ya que la enfermedad no existe en el Perú, porque la entidad SENASA la exterminó en su momento que fue detectado, por ende se recurre a tal lugar para la obtención de la unidad de análisis que comprendió de 751 imágenes en ambiente/entorno natural y controlado a diferencia de **Xing y Qiu** que obtuvieron sus unidades de análisis en entornos naturales con variaciones de no solo hojas sino combinadas con frutos.

En la matriz de confusión de la investigación **figura 13**, se enfocó exclusivamente en la enfermedad de Huanglongbing, limitándose a una sola clase. En contraste, **Qiu y otros** ampliaron su estudio incluyendo varias enfermedades relacionadas con la planta de limón, como la antracnosis, entre otras. Esta diferencia en el enfoque se refleja en la considerable disparidad en el tamaño de sus matrices de confusión, siendo la de **Qiu** más extensa debido a la diversidad de muestras recolectadas en áreas más extensas y variadas, respaldada por mayores recursos económicos para financiar su investigación. Este contraste destaca la relevancia de considerar los alcances y recursos disponibles al diseñar y comparar proyectos de investigación.

Una de las limitaciones del modelo Yolo ocurrió en su documentación recomendando tener un mínimo de ≥ 1000 hojas por clase, el dataset al principio quedó muy corto, ya que en los entrenamientos como se obtuvo en el experimento 1 fueron muy bajos en tema de precisión y recuperación (recall) con respecto a los experimentos 2, 4, 5, 6. Las imágenes se recogieron de la página RoboFlow ya que se necesitaban aumentar la unidad de análisis comprendiendo de 2 dataset de hojas más conseguidos en esa página y la unidad de análisis subió a 5,475 imágenes de la planta limón sutil (*Citrus Aurantifolia Swingle*) e imágenes de limón Tahití o persa (*Citrus Latifolia*).

En cuanto a la proporción que **Xing y otros** uso fue 4:1:1 mientras que **Qiu y otros** uso división de porcentaje de 80%,10%,10%, esto se usa comúnmente en el aprendizaje automático en tema de etiquetado para evitar un sobreajuste o conocido normalmente como Overfitting ya que las proporciones son relativas más no absolutas dependiendo de la naturaleza y el tamaño del dataset como es más grande que el dataset de **Xing** y del proyecto de investigación, su porcentaje de entrenamiento es mucho más alta.

En el proyecto de investigación, se utilizó una proporción de entrenamiento de un 70%, 15% de evaluación y 15% de pruebas, el tamaño de las imágenes fue de 640 píxeles ese es el tamaño estándar que se usa en la mayoría de los proyectos de deep learning y adecuando a la investigación. En cuanto a **Xing y otros** usó la redimensión de imágenes de 224 x 224 píxeles con un enfoque bilineal que consiste en la modificación basada en la selección de píxeles.

Mientras que **Qiu y otros** también usó la redimensión estándar de 640x640px, siendo inicialmente sus imágenes con un tamaño procesado a JPG desde Photoshop con la resolución 72px/pulgada.

En el estudio de **Xing y otros** se usó el modelo DenseNet-16 es una versión liviana ya que se está usando 16 capas de profundidad de red mientras que **Qiu y otros** usa YOLO, pero una versión más grande YOLOv5l que contiene más capas que Yolov5m, consumiendo más recursos computacionales como memoria y configurando su batch size manualmente en 20, ya que posee financiación privada lo que le permite tener recursos casi ilimitados.

Cómo aporta **Jocher** desde la teoría existen varios tamaños del modelo o arquetipos de Yolo como Yolov5s, Yolov5m, Yolov5l, y Yolov5x para ver la profundidad y anchura de la red, también comenta que Yolov5s es más rápido que sus otras familias. El proyecto de investigación se ha considerado el modelo base o el arquetipo YOLOv5m, manteniendo el equilibrio entre precisión y velocidad lo cual se decantó por tal arquetipo configurándose el batch size o el tamaño de lote en automático para que aproveche todos los recursos de la computadora mientras el batch size abarque más espacio en la RAM ya que automáticamente hace el cálculo (Jocher et al. 2021).

En un principio se tuvo limitaciones técnicas en el momento de ejecutar el modelo, ya que necesita buena memoria RAM, GPU y sobre todo tarjeta gráfica, comenzando usar el sistema operativo Arch Linux para optimizar los recursos al ejecutar el modelo ya que demandaba recursos económicos y técnicos en la parte del Hardware.

Asimismo Lawal y otros desde la teoría hacen saber que el total de los patrones positivos se predicen correctamente obteniendo los patrones predichos en una clase positiva (Lawal 2021; Hossin y M.N 2015; López-Cabrera et al. 2022), eso vendría a ser el incrementó de la precisión de detección lo cual se obtuvo en el proyecto de investigación la precisión de los experimentos: en el exp1 se obtuvo 1%, en el exp2 se obtuvo 20%, exp5 se obtuvo 55%, exp6 se obtuvo 26%, siendo mejor el experimento 4 la cual se obtuvo 67% de precisión en la clasificación total de los patrones predichos como verdaderos.

En cuanto a **Xing y otros** la precisión fue de 93.33% ya que las capas de profundidad solo fueron 16 obteniendo más precisión que es lo que su investigación lo demandaba en la clasificación con menos parámetros, mientras que **Qiu y otros** la precisión que se obtuvo fue 90.07% ya que la diferencia estuvo en las variaciones de modelos ya que se está usando 3 variaciones del modelo Yolov5l y su dataset es mucho más amplio.

Desde la teoría **Lawal y otros** comenta que la fracción de los patrones positivos que se clasifican correctamente (Lawal 2021; Hossin y M.N 2015; López-Cabrera et al. 2022), eso vendría a ser la recuperación de detección **en la investigación** en el exp1 se obtuvo 0.2%, en el exp2 se obtuvo 8%, en el exp5 se obtuvo 29%, en el exp6 se obtuvo 14%, siendo el mejor exp4 de recall se obtuvo 55.5% de recuperación de detección (recall) en la clasificación de la fracción de los patrones positivos del experimento 4, mientras **Qiu y otros** obtuvo 81.84% de recall. Lawal, 2021 desde la teoría comenta que F1 es el compromiso entre recall y precisión para observar el rendimiento del modelo entrenado y comparar (Lawal 2021), pero en el proyecto de investigación solo se usa 1 clase y solo un arquetipo, el F1 del proyecto de investigación es de 56% con un grado de confianza de 0.029 del exp4, ya que solo se está buscando el equilibrio en la precisión y recall. Mientras que **Qiu y otros** obtuvieron el F1 85.19% ya que hay si está comparando 3 modelos.

En cuanto al segundo objetivo específico: Clasificar por patrones positivos utilizando el modelo YOLO la sensibilidad en el diagnóstico preliminar de la enfermedad del Huanglongbing en la planta de limón.

En cuanto al estudio de Gamboa, 2018 en Lima, usando como proyecto una investigación experimental que se dividió en dos grupos de población, grupo de personas enfermas en Grupo experimental y otro grupo de personas sanas en Grupo control, midiendo el indicador sensibilidad relacionado con el diagnóstico de enfermedades dermatológicas obteniendo en la sensibilidad un 83.6244% aumentando 21.1605% de 62.4639% (Gamboa Jara 2018). Asimismo, concuerdan con el estudio de Cordova y Quinde, 2022 en Piura, su población fue de 100 personas que incluyen hombres y mujeres, en un centro podológico, con una muestra de 73 pacientes, usando 37 pacientes. En la

enfermedad de onicomiosis y 36 pacientes sin onicomiosis, usó 20 épocas en el algoritmo CNN para el entrenamiento, su sensibilidad fue de 92.57% (Cordova Zapata y Quinde Ramos 2022).

En el proyecto de investigación se evaluó el **incrementó de sensibilidad** se traduce en la detección positiva de hojas infectadas con la enfermedad de Huanglongbing. Este proceso abarcó un total de 300 épocas, destacando un rendimiento superior en la época 200, donde se logró una sensibilidad del 56.04%. Este resultado se obtuvo específicamente en el experimento 4, identificado como el más exitoso entre los seis experimentos realizados, todos enfocados en verificar la presencia de la enfermedad. La medida de sensibilidad se define en este contexto como la capacidad del modelo para identificar correctamente las hojas infectadas, es decir, los verdaderos positivos. Este proceso se llevó a cabo utilizando el modelo YOLOv5m y recolectando datos a través de una guía de observación en un total de 300 épocas. Es fundamental destacar que el énfasis en la sensibilidad implica una atención especial a los verdaderos positivos y falsos negativos en las imágenes de hojas afectadas por la enfermedad de Huanglongbing.

El procedimiento y la metodología empleada en la investigación revelaron que la sensibilidad media fue del 52.86%, con una desviación típica de 2.11%. Este análisis estadístico proporciona una visión más completa de la variabilidad en los resultados de la enfermedad a lo largo de las épocas evaluadas. **Mientras Gamboa** en su estudio obtuvo 83.6244% de sensibilidad aumentando la sensibilidad en un 21.1605% de 62.4639% ya que lo realizó en un proyecto experimental contando de Pre-test y Post-test aumentando su sensibilidad y con solo una muestra de 18 pacientes usando distintos algoritmos. **En cuanto a Cordova Zapata y otros**, su muestra fue 73 personas particionado en 37 y 36 con y sin onicomiosis las imágenes por individuo, usando CNN como algoritmo y realizado en 20 épocas en 5 en 5 siguiendo la serie (5|10|15|20) obteniendo un 92.57% de sensibilidad.

Los resultados obtenidos en este proyecto de investigación se compararon con los hallazgos previos de **Gamboa y Zapata**. Sin embargo, es esencial destacar que la muestra utilizada en el proyecto fue significativamente más amplia, ya

que no se limitó a imágenes de un solo tipo de hoja de limón, sino que abarcó dos variedades: Limón Sutil y Limón Persa. Esta decisión se tomó estratégicamente para evitar el riesgo de Underfitting, que se refiere a la posibilidad de que el modelo YOLOv5m entrene exclusivamente con un tipo de hojas, lo que podría resultar en un fallo al detectar otro tipo de hojas de limón. Esta ampliación en la diversidad de la muestra no sólo mitiga el riesgo de subajuste sino que también enriquece la capacidad del modelo para generalizar su aprendizaje a diferentes contextos. Además, la elección de plasmar las épocas de entrenamiento en incrementos de 50 (50|100|150|200|250|300) se presenta como una estrategia efectiva para mejorar la visualización de resultados del entrenamiento. Este enfoque secuencial permite un ajuste más detallado del modelo a medida que avanza en las etapas de entrenamiento.

A pesar de estas mejoras sustanciales, es crucial reconocer las limitaciones que surgieron, tanto técnicas como económicas. No obstante, es alentador observar que, a pesar de estas restricciones, el modelo demostró un rendimiento prometedor. El proyecto sugiere que aún hay espacio para mejoras adicionales. Por ejemplo, la incorporación de más clases de hojas o plagas podría enriquecer aún más el aprendizaje del modelo y aumentar su capacidad de detección en diversos escenarios. Esta expansión, sin embargo, debe considerarse cuidadosamente en términos de recursos y objetivos específicos del proyecto.

No obstante, la interpretación crítica de estos resultados podría centrarse en la necesidad de analizar la variabilidad en la sensibilidad y considerar posibles factores externos que podrían haber afectado la precisión del modelo. Además, la elección de la época 200 como el punto óptimo para la sensibilidad merece un análisis más profundo, ya que podría indicar un equilibrio efectivo entre el tiempo de entrenamiento y la capacidad del modelo para detectar la enfermedad. Asimismo, se respalda en la teoría sustentada de Sancho y Gonzales, 2022 argumentando la disposición de detectar la cantidad de individuos con la enfermedad logrando un resultado positivo, en pocas palabras, clasificando como un individuo enfermo (Sancho-Insenser y González-Castillo 2022).

En el tercer objetivo específico: Clasificar por patrones negativos utilizando el modelo YOLO la especificidad en el diagnóstico preliminar de la enfermedad del Huanglongbing en la planta de limón.

En cuanto al estudio de Gamboa, 2018 se obtuvo como media en la especificidad 83.3328% aumentando en un 33.1361% de especificidad de 50.1967%, con una muestra de 18 pacientes de micosis superficiales sobre un diagnóstico preliminar de (Gamboa Jara 2018). Asimismo concuerda con el estudio de Cordova y Quinde, 2022 se obtuvo como media en la especificidad un 92.36% comprendiendo en 20 épocas que se plasmó de 5 en 5 con la serie de (5|10|15|20) épocas para visualizar cómo evoluciona el algoritmo CNN (Cordova Zapata y Quinde Ramos 2022).

En el proyecto de investigación se evaluó el **incrementó de especificidad** se enfoca en la identificación negativa, es decir, la capacidad del modelo para reconocer hojas libres de la enfermedad de Huanglongbing. Este proceso se extendió a lo largo de 300 épocas, correspondientes a iteraciones completas en las que el modelo YOLOv5m fue perfeccionando su desempeño. Es digno de destacar que este proceso se llevó a cabo evitando caer en sobreajuste (overfitting), un aspecto crucial para la generalización efectiva del modelo.

Durante este extenso proceso de aprendizaje, se observó un progreso constante desde la época 150 hasta la 250, siendo este intervalo el más óptimo en términos de especificidad. El pico máximo se alcanzó en la época 150, registrando un destacado 71.63% de especificidad. Este rendimiento sobresaliente se asoció con el experimento cuatro, identificado como el más exitoso entre los seis experimentos totales del proyecto.

La medición de especificidad se basó en una prueba negativa, enfocándose en la capacidad del modelo para identificar correctamente las hojas libres de la enfermedad. Este proceso también abarcó las 300 épocas, y los resultados se recolectaron utilizando una guía de observación como instrumento. El análisis detallado de falsos positivos y verdaderos negativos en las imágenes de hojas sin la enfermedad permitió determinar la especificidad global del modelo. El proceso arrojó una especificidad media del 63.57%, con una desviación típica

de 5.78%. **Mientras Gamboa** en su estudio obtuvo 83.3328% de sensibilidad aumentando la sensibilidad en un 33.1361% de 50.1967%. **En cuanto a Córdova y otros**, usando su algoritmo CNN y realizado en 20 épocas en 5 en 5 siguiendo la serie (5|10|15|20) obteniendo un 92.36% de especificidad.

Los resultados obtenidos en este proyecto de investigación se compararon con los hallazgos previos de **Gamboa y Zapata** en cuanto a las épocas es muy poco en un modelo o algoritmo como los es Yolo sea cualquier arquetipo pero como en Inteligencia Artificial es dependiendo el contexto donde se usa o aplica de acuerdo a los objetivos, pero en un entorno de detección de plagas son muy pocas épocas y muestras mientras que en Yolo como se discutió antes la limitación es que requiere más de 1000 imágenes por clase en la cual se tiene una que es “Huanglongbing” lo que da un mejor resultado para el modelo aprenda y beneficio para aun seguir perfeccionando el modelo sin caer en un sobreajuste (Overfitting) en el entrenamiento ya que puede causar pérdidas económicas en la toma de decisiones sobre un finca/huerto que se sospeche de contaminación del Huanglongbing por eso se entrenó con lo recomendado que son 300 épocas como lo recomiendo la propia documentación de Yolo, usando la serie de 50 en 50 con la serie (50|100|150|200|250), obteniendo el pico de especificidad en la época 150 clasificando las pruebas negativas, en cuanto a los instrumentos fueron similares como lo es la guía de observación con respecto a **Zapata** ya que se usa la misma fórmula para determinar la especificidad esa es fortaleza de Yolo por lo que ahora es lo mejor de lo mejor en la detección de objetos.

A pesar de los resultados positivos y consistentes, es crucial reflexionar sobre la interpretación de la especificidad y su impacto en el contexto de detección de enfermedades. La elección de la época 150 como el punto álgido puede indicar no sólo un rendimiento excepcional, sino también una capacidad efectiva del modelo para discernir entre casos positivos y negativos. No obstante, la reflexión crítica podría dirigirse hacia la comprensión de la variabilidad en la especificidad a lo largo de las épocas, así como la ponderación de la desviación típica. Este análisis podría proporcionar insights adicionales sobre la robustez y la confiabilidad de la propuesta de investigación.

Se respalda en la teoría sustentada de Sancho y González, 2022 argumenta la clasificación de las etiquetas de manera negativa de individuos que se encuentran sanos logrando un resultado negativo (Sancho-Insenser y González-Castillo 2022). En pocas palabras, se clasifica como un individuo que no cuenta la enfermedad o no está enfermo.

En el proyecto de investigación se requirió un balanceo equilibrio entre la sensibilidad y la especificidad, ya que no existe un valor específico que considere bueno o malo para la especificidad y sensibilidad al determinar los valores depende de las características específicas de su aplicación ya que el rendimiento actual de modelo satisface los objetivos y los requisitos en contexto de su aplicación en la detección de la enfermedad del Huanglongbing en la región Piura.

En el desarrollo de la investigación, se incorporó la **curva ROC** para analizar la tasa de verdaderos positivos y falsos positivos, obteniendo un área bajo la curva (AUC) de 0.780. Este análisis se realizó considerando únicamente la clase de Huanglongbing, con un umbral de confianza establecido en 0.45. La observación detallada del gráfico de la **figura 18** revela que la curva en la esquina límite superior izquierda se alinea estratégicamente, evitando la inclusión de más falsos positivos a medida que se aumenta la sensibilidad. Este enfoque es crucial, ya que busca optimizar la capacidad del modelo YOLOv5m para distinguir entre clases positivas y negativas.

La interpretación de los valores de la curva ROC, que varían de 0 a 1, es esencial para comprender el rendimiento del modelo. En este caso, el valor de 0.780 se sitúa en un rango razonable, indicando que el modelo tiene una capacidad decente en la distinción entre clases positivas y negativas. Este resultado se alinea con los criterios estándar de evaluación de la curva ROC, donde un valor entre 0.7 y 0.8 se considera bueno. La clasificación del rendimiento según rangos establecidos (**0.5-0.6: Pobre, 0.6-0.7: Aceptable, 0.7-0.8: Bueno, 0.8-0.9: Muy bueno, 0.9-1.0: Excelente**) permite contextualizar aún más el resultado obtenido.

En este caso, al caer en la categoría de "bueno", el rendimiento de la curva ROC satisface las expectativas y se ajusta a los requisitos y objetivos del proyecto. Asimismo, se da soporte al teórico Yang y Berdine, 2017 abordando sobre la Curva ROC los problemas que se enfrenta durante la clasificación entre dos grupos, Por ejemplo, Plantas con HLB y Plantas sin HLB y establecer un umbral de confianza o punto de quiebre, por lo que se confirma la capacidad que en las imágenes de hojas se distinguen entre las que cuentan o no cuentan con Huanglongbing de manera discriminatoria (Yang y Berdine 2017).

En la justificación tecnológica delineada en el proyecto de investigación, se estableció inicialmente el desarrollo del modelo Yolov5m adaptado para la detección de la enfermedad del Huanglongbing (HLB). El propósito fundamental de este modelo radica en brindar apoyo a la toma de decisiones mediante la inteligencia artificial, con el fin de implementar medidas necesarias contra el HLB. Sin embargo, es crucial destacar que el modelo Yolov5m, en su estado actual, se limita a ser un algoritmo que cumple con sus características de detección de hojas enfermas con HLB o no. Esta limitación se atribuye a restricciones temporales de tiempo, impidiendo la implementación del modelo en una aplicación móvil, un avance logrado por **Xing y otros**, que podría potenciar aún más la utilidad práctica del modelo en entornos aplicados.

VI. CONCLUSIONES

La evaluación de predicciones a lo largo de 300 épocas reveló variaciones notables en la precisión del modelo YOLOv5m. Destacando el experimento 4 con un 67% de precisión, se optó por su estrategia de distribución de datos (70% train, 15% eval, 15% test) como la más efectiva. Aunque la cantidad inicial de imágenes fue aceptable pero no óptima, la inclusión de dos datasets adicionales compensa esta limitación, resultando en mejoras significativas en el rendimiento. En última instancia, se concluye que el YOLOv5m demuestra una alta precisión en la detección de la enfermedad de Huanglongbing.

La evaluación de seis experimentos a lo largo de 300 épocas reveló variaciones en la recuperación (recall). Aunque los primeros experimentos mostraron resultados modestos (exp1: 0.2%, exp2: 8%), el experimento 4 se destacó con un recall del 55.5%. Este resultado sugiere que YOLOv5m demuestra una capacidad significativa para identificar correctamente los patrones positivos asociados con la enfermedad de Huanglongbing. La adaptabilidad del modelo al ajustar automáticamente el batch size refuerza la robustez de su desempeño, consolidando su utilidad en la detección efectiva de esta enfermedad.

La clasificación de patrones positivos en cuanto a la sensibilidad abarcó 300 épocas, destacando un rendimiento superior en la época 200, donde se logró una sensibilidad del 56.04%. Este resultado sobresalió en el experimento 4, identificado como el más exitoso entre los seis realizados para verificar la presencia de la enfermedad de Huanglongbing. La sensibilidad media fue del 52.86%, con una desviación típica de 2.11%. Al abarcar dos variedades, Limón Sutil y Limón Persa, estratégicamente para evitar el riesgo de Underfitting, se concluye que el modelo yolov5m demostró un rendimiento prometedor en la discriminación precisa de casos positivos de los negativos.

La clasificación de patrones negativos en cuanto a la especificidad se llevó a cabo a lo largo de 300 épocas, con el modelo YOLOv5m perfeccionando su desempeño. Evitando sobreajuste, se destacó un progreso constante desde la época 150 hasta la 250, siendo este intervalo óptimo en especificidad con un pico máximo del 71.63% en la época 150. La media de especificidad fue del

63.57%, con una desviación típica de 5.78%. En conclusión, se afirma que el modelo yolov5m demuestra una capacidad efectiva para discriminar entre casos negativos de los positivos.

En conclusión, se evidencia que el modelo yolov5m mejora su precisión al aumentar el número de imágenes y la diversidad de enfermedades en las plantas de limón. El incremento de épocas resulta beneficioso en ciertos aspectos como que en otros no, abarcando más recursos técnicos y alargando el tiempo de entrenamiento.

En el proyecto de investigación, se logró un equilibrio entre sensibilidad y especificidad, satisfaciendo los objetivos planteados para la detección de huanglongbing. La sensibilidad se valoró especialmente para la detección temprana, previniendo su propagación, mientras que la especificidad evitar acciones innecesarias con posibles costos ambientales y económicos por ello se planteó un equilibrio entre los dos, el modelo YOLOv5m desarrollado sirve como base para diagnósticos preliminares, aunque se destaca que su implementación debe complementarse con análisis más específicos en laboratorio, con la asesoría de expertos en plagas.

VII. RECOMENDACIONES

Se recomienda el desarrollo e implementación de un Aplicativo Móvil para la detección en tiempo real de enfermedades en plantas como lo es el Huanglongbing y otras enfermedades relacionadas con la planta del limón utilizando un arquetipo como lo es YOLOv5l integrándose con una interfaz de usuario buscando ayudar a los profesionales y agricultores una herramienta más accesible y efectiva al alcance de la mano.

Se recomienda a los próximos investigadores la integración de sensores remotos como drones ya que ayudan en la mejora de la precisión del modelo a usar, permitiendo en las áreas de cultivo un mejor control y monitoreo más amplio para la ayuda de toma de decisiones.

Se recomienda en futuras investigaciones el enfoque multiespectral para las detecciones tempranas sobre el Huanglongbing ya que ayudan en la detección a diferencia de que en imágenes convencionales no se logra apreciar.

Se recomienda llevar el enfoque de la investigación a un nivel más elevado como es el nivel de investigación relacional con un análisis de estadística bivariado.

Se recomienda el desarrollo y comparativa de métodos sobre la detección de enfermedades en plántones de limón con Yolov5 (cualquier arquetipo superior a Yolov5m) y técnicas convencionales.

Se recomienda datasets o conjuntos de datos más grandes y la incorporación de más clases de hojas o plagas para enriquecer aún más el aprendizaje del modelo y aumentar su capacidad de detección en diversos escenarios para mejorar su detección y no caer en Underfitting.

REFERENCIAS BIBLIOGRÁFICAS

- BENCARDINO, C.M., 2019. *Estadística básica aplicada*. S.I.: Ecoe Ediciones. ISBN 978-958-771-748-8.
- CABI, 2019. citrus huanglongbing (greening) disease (citrus greening). *CABI Compendium*, vol. CABI Compendium, DOI 10.1079/cabicompendium.16567.
- CHEN, Z., WU, R., LIN, Y., LI, C., CHEN, Siyu, YUAN, Z., CHEN, Shiwei y ZOU, X., 2022. Plant Disease Recognition Model Based on Improved YOLOv5. *Agronomy*, vol. 12, no. 2, ISSN 2073-4395. DOI 10.3390/agronomy12020365.
- CORDOVA ZAPATA, A.B. y QUINDE RAMOS, M.J., 2022. Sistema de reconocimiento de imágenes para pre-diagnósticos de onicomicosis (Hongos en las uñas) en el centro podológico Decapie – Piura 2022. En: Accepted: 2022-10-19T22:20:23Z, *Repositorio Institucional - UCV* [en línea], [consulta: 8 junio 2023]. Disponible en: <https://repositorio.ucv.edu.pe/handle/20.500.12692/99173>.
- CORREA, M.C.G., GERMAIN, J.-F., MALAUSA, T. y ZAVIEZO, T., 2012. Molecular and morphological characterization of mealybugs (Hemiptera: Pseudococcidae) from Chilean vineyards. *Bulletin of Entomological Research*, vol. 102, no. 5, ISSN 1475-2670. DOI 10.1017/S0007485312000053.
- DALA-PAULA, B.M., PLOTTO, A., BAI, J., MANTHEY, J.A., BALDWIN, E.A., FERRAREZI, R.S. y GLORIA, M.B.A., 2019. Effect of Huanglongbing or Greening Disease on Orange Juice Quality, a Review. *Frontiers in Plant Science* [en línea], vol. 9, [consulta: 13 mayo 2023]. ISSN 1664-462X. Disponible en: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01976>.
- DIAZ HOLGADO, R. y VILCAS VILLALBA, D.R., 2021. Aplicativo móvil para la detección de la enfermedad del Mosaico común en la yuca (Manihot esculenta) utilizando redes neuronales convolucionales, 2021. En: Accepted: 2022-11-21T13:54:46Z, *Universidad Nacional Amazónica de Madre de Dios - UNAMAD* [en línea], [consulta: 13 mayo 2023]. Disponible en: <http://repositorio.unamad.edu.pe/handle/20.500.14070/889>.
- DONG, Z., SRIVASTAVA, A.K., LIU, X., RIAZ, M., GAO, Y., LIANG, X., TAN, Q., SUN, X., WU, S. y HU, C., 2021. Interactions between nutrient and Huanglongbing pathogen in citrus: An overview and implications. *Scientia Horticulturae*, vol. 290, ISSN 0304-4238. DOI 10.1016/j.scienta.2021.110511.
- DONKERSLEY, P., SILVA, F.W.S., CARVALHO, C.M., AL-SADI, A.M. y ELLIOT, S.L., 2018. Biological, environmental and socioeconomic threats to citrus lime production. *Journal of Plant Diseases and Protection*, vol. 125, no. 4, ISSN 1861-3837. DOI 10.1007/s41348-018-0160-x.

- ESTEBAN NIETO, N., 2018. Tipos de Investigación. En: Accepted: 2018-07-02T01:44:39Z, *Universidad Santo Domingo de Guzmán* [en línea], [consulta: 8 junio 2023]. Disponible en: <http://repositorio.usdg.edu.pe/handle/USDG/34>.
- FLORES, S. y ANSELMO, F., 2019. Fundamentos epistémicos de la investigación cualitativa y cuantitativa: consensos y disensos. *Revista Digital de Investigación en Docencia Universitaria*, vol. 13, no. 1, ISSN 2223-2516. DOI 10.19083/ridu.2019.644.
- GAMBOA JARA, K.B., 2018. Aplicación móvil para el diagnóstico preliminar de micosis superficiales a través de fotografías digitales. En: Accepted: 2019-01-29T20:21:21Z, *Universidad César Vallejo* [en línea], [consulta: 13 mayo 2023]. Disponible en: <https://repositorio.ucv.edu.pe/handle/20.500.12692/25620>.
- GAO, Z., LUO, Z., ZHANG, W., LV, Z. y XU, Y., 2020. Deep Learning Application in Plant Stress Imaging: A Review. *AgriEngineering*, vol. 2, no. 3, ISSN 2624-7402. DOI 10.3390/agriengineering2030029.
- GONZALEZ, L., 2019. Curvas ROC y Área bajo la curva (AUC).  *Aprende IA* [en línea]. [consulta: 7 julio 2023]. Disponible en: <https://aprendeia.com/curvas-roc-y-area-bajo-la-curva-auc-machine-learning/>.
- GONZÁLEZ-GARZA, R., 2017. Evolución de técnicas de diagnóstico de virus fitopatógenos. *Revista mexicana de fitopatología*, vol. 35, no. 3, ISSN 0185-3309. DOI 10.18781/r.mex.fit.1706-1.
- GREEN, J.L., MANSKI, S.E., HANSEN, T.A. y BROATCH, J.E., 2023. Descriptive statistics. En: R.J. TIERNEY, F. RIZVI y K. ERCIKAN (eds.), *International Encyclopedia of Education (Fourth Edition)* [en línea]. Oxford: Elsevier, pp. 723-733. [consulta: 6 julio 2023]. ISBN 978-0-12-818629-9. Disponible en: <https://www.sciencedirect.com/science/article/pii/B9780128186305100831>.
- GUZMÁN, S. y YINETH, V., 2021. Algoritmo de detección de mosca blanca por medio de inteligencia artificial en las hojas de plátano. En: Accepted: 2021-09-21T23:24:20Z [en línea], [consulta: 13 mayo 2023]. Disponible en: <http://repository.unad.edu.co/handle/10596/42374>.
- HAN, S.S., KIM, M.S., LIM, W., PARK, G.H., PARK, I. y CHANG, S.E., 2018. Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm. *The Journal of Investigative Dermatology*, vol. 138, no. 7, ISSN 1523-1747. DOI 10.1016/j.jid.2018.01.028.
- HOSSIN, M. y M.N, S., 2015. A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, vol. 5, DOI 10.5121/ijdkp.2015.5201.

- JOCHER, G., STOKEN, A., BOROVEC, J., NANOCODE012, CHRISTOPHERSTAN, CHANGYU, L., LAUGHING, TKIANAI, YXNONG, HOGAN, A., LORENZOMAMMANA, ALEXWANG1900, CHAURASIA, A., DIACONU, L., MARC, WANGHAOYANG0106, ML5AH, DOUG, DURGESH, INGHAM, F., FREDERIK, GUILHEN, COLMAGRO, A., YE, H., JACOBSOLAWETZ, POZNANSKI, J., FANG, J., KIM, J., DOAN, K. y 于力军 L.Y., 2021. ultralytics/yolov5: v4.0 - nn.SiLU() activations, Weights & Biases logging, PyTorch Hub integration. En: ADS Bibcode: 2021zndo...4418161J, *Zenodo* [en línea], [consulta: 6 julio 2023]. DOI 10.5281/zenodo.4418161. Disponible en: <https://ui.adsabs.harvard.edu/abs/2021zndo...4418161J>.
- JONES, J.W., ANTLE, J.M., BASSO, B., BOOTE, K.J., CONANT, R.T., FOSTER, I., GODFRAY, H.C.J., HERRERO, M., HOWITT, R.E., JANSSEN, S., KEATING, B.A., MUNOZ-CARPENA, R., PORTER, C.H., ROSENZWEIG, C. y WHEELER, T.R., 2017. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agricultural Systems*, vol. 155, ISSN 0308-521X. DOI 10.1016/j.agsy.2016.09.021.
- KHAN, S., TUFAIL, M., KHAN, M.T., KHAN, Z.A., IQBAL, J. y WASIM, A., 2021. Real-time recognition of spraying area for UAV sprayers using a deep learning approach. *PLoS One* [en línea], vol. 16, no. 4, [consulta: 18 mayo 2023]. DOI 10.1371/journal.pone.0249436. Disponible en: <https://www.proquest.com/docview/2507957035/abstract/FB84BA4F8EE14792PQ/19>.
- LARRAIN, P., 2010. Biología y comportamiento de chanchitos blancos (*Pseudococcus* spp.). En: Accepted: 2020-12-15T01:51:57Z [en línea], [consulta: 2 junio 2023]. Disponible en: <https://biblioteca.inia.cl/handle/20.500.14001/7350>.
- LAWAL, M.O., 2021. Tomato detection based on modified YOLOv3 framework. *Scientific Reports*, vol. 11, ISSN 2045-2322. DOI 10.1038/s41598-021-81216-5.
- LEE, S., NUEVA, E. a sitio externo E. enlace se abrirá en una ventana, CHOI, G., NUEVA, E. a sitio externo E. enlace se abrirá en una ventana, PARK, H.-C., NUEVA, E. a sitio externo E. enlace se abrirá en una ventana, CHOI, C. y NUEVA, E. a sitio externo E. enlace se abrirá en una ventana, 2022. Automatic Classification Service System for Citrus Pest Recognition Based on Deep Learning. *Sensors* [en línea], vol. 22, no. 22, [consulta: 18 mayo 2023]. DOI 10.3390/s22228911. Disponible en: <https://www.proquest.com/docview/2739457209/abstract/1F1D39CCFB56459DPQ/1>.
- LI, H., LEE, W.S., WANG, K., EHSANI, R. y YANG, C., 2014. 'Extended spectral angle mapping (ESAM)' for citrus greening disease detection using airborne hyperspectral imaging. *Precision agriculture* [en línea], [consulta: 2 junio

2023]. ISSN 1385-2256. Disponible en: <https://doi.org/10.1007/s11119-013-9325-6>.

LÓPEZ FERNÁNDEZ, R., AVELLO MARTÍNEZ, R., PALMERO URQUIZA, D.E., SÁNCHEZ GÁLVEZ, S., QUINTANA ÁLVAREZ, M., LÓPEZ FERNÁNDEZ, R., AVELLO MARTÍNEZ, R., PALMERO URQUIZA, D.E., SÁNCHEZ GÁLVEZ, S. y QUINTANA ÁLVAREZ, M., 2019. Validación de instrumentos como garantía de la credibilidad en las investigaciones científicas. *Revista Cubana de Medicina Militar* [en línea], vol. 48, [consulta: 2 julio 2023]. ISSN 0138-6557. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S0138-65572019000500011&lng=es&nrm=iso&tlng=es.

LÓPEZ-CABRERA, J.D., RUIZ-GONZALEZ, Y., DÍAZ-AMADOR, R. y TABOADA-CRISPI, A., 2022. Estrategias de fusión para la clasificación automática de imágenes de úlceras de pie diabético usando técnicas de visión por computadora. *Revista Cubana de Ciencias Informáticas*, vol. 16, no. 1,

LOWE, A., HARRISON, N. y FRENCH, A.P., 2017. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods*, vol. 13, no. 1, ISSN 1746-4811. DOI 10.1186/s13007-017-0233-z.

LYU, S., KE, Z., LI, Z., XIE, J., ZHOU, X. y LIU, Y., 2023. Accurate Detection Algorithm of Citrus Psyllid Using the YOLOv5s-BC Model. *Agronomy* [en línea], vol. 13, no. 3, [consulta: 27 mayo 2023]. DOI 10.3390/agronomy13030896. Disponible en: <https://www.proquest.com/docview/2791560757/abstract/4599696FDFA24943PQ/8>.

MARTINEZ OJEDA, B.S. y MOROCHO FEBRES, L.E., 2021. Diagnóstico del estado emocional, a través de reconocimiento facial para una empresa del sector público Piura 2021. En: Accepted: 2022-10-29T04:45:02Z, *Repositorio Institucional - UCV* [en línea], [consulta: 8 junio 2023]. Disponible en: <https://repositorio.ucv.edu.pe/handle/20.500.12692/100114>.

MENDOZA, S.H. y AVILA, D.D., 2020. Técnicas e instrumentos de recolección de datos. *Boletín Científico de las Ciencias Económico Administrativas del ICEA*, vol. 9, no. 17, ISSN 2007-4913. DOI 10.29057/icea.v9i17.6019.

NARVAEZ, M., 2023. Técnicas de recolección de datos: Qué son y cuáles existen. *QuestionPro* [en línea]. [consulta: 23 junio 2023]. Disponible en: <https://www.questionpro.com/blog/es/tecnicas-de-recoleccion-de-datos/>.

NEHELA, Y. y KILLINY, N., 2020. Revisiting the Complex Pathosystem of Huanglongbing: Deciphering the Role of Citrus Metabolites in Symptom Development. *Metabolites*, vol. 10, no. 10, ISSN 2218-1989. DOI 10.3390/metabo10100409.

- OBREGÓN, R.R.S., SINCHE, J.C.L., ZARUMA, G.A.M., BADILLO, P.L.L., ALTAMIRANO, H.R.S. y VITE, G.A.O., 2018. Diagnóstico preliminar para diseñar actividades físico-recreativas para promocionar la salud en estudiantes de la ESPOCH. *Lecturas: Educación Física y Deportes*, vol. 23, no. 247, ISSN 1514-3465.
- ORTEGA, C., 2018. ¿Qué es el muestreo por conveniencia? *QuestionPro* [en línea]. [consulta: 30 junio 2023]. Disponible en: <https://www.questionpro.com/blog/es/muestreo-por-conveniencia/>.
- OTZEN, T. y MANTEROLA, C., 2017. Técnicas de Muestreo sobre una Población a Estudio. *International Journal of Morphology*, vol. 35, no. 1, ISSN 0717-9502. DOI 10.4067/S0717-95022017000100037.
- PANDEY, S.S. y WANG, N., 2019. Targeted Early Detection of Citrus Huanglongbing Causal Agent 'Candidatus Liberibacter asiaticus' Before Symptom Expression. *Phytopathology*®, vol. 109, no. 6, ISSN 0031-949X. DOI 10.1094/PHYTO-11-18-0432-R.
- QIU, R.-Z., CHEN, S.-P., CHI, M.-X., WANG, R.-B., HUANG, T., FAN, G.-C., ZHAO, J. y WENG, Q.-Y., 2022. An automatic identification system for citrus greening disease (Huanglongbing) using a YOLO convolutional neural network. *Frontiers in Plant Science*, vol. 13, ISSN 1664-462X. DOI 10.3389/fpls.2022.1002606.
- SANCHO-INSENER, J.-J. y GONZÁLEZ-CASTILLO, A.M., 2022. Pruebas diagnósticas. ¿Cómo describir su validez? *Cirugía Española*, vol. 100, no. 9, ISSN 0009-739X. DOI 10.1016/j.ciresp.2022.02.006.
- SENASA, 2019. Piura: Prevención mantiene fuera de Perú al Dragón Amarillo. *SENASA al día* [en línea]. [consulta: 13 mayo 2023]. Disponible en: <https://www.senasa.gob.pe/senasacontigo/piura-prevencion-mantiene-fuera-de-peru-al-dragon-amarillo/>.
- SINGERMAN, A., LENCE, S. y USECHE, P., 2017. Is Area-Wide Pest Management Useful? The Case of Citrus Greening. *Applied Economic Perspectives and Policy*, vol. 39, no. 4, ISSN 2040-5790.
- TERÁN-BUSTAMANTE, A., RAMÍREZ-CASTILLO, C.E., MARTÍNEZ-VELASCO, A., TERÁN-BUSTAMANTE, A., RAMÍREZ-CASTILLO, C.E. y MARTÍNEZ-VELASCO, A., 2020. Confiabilidad y validez de un instrumento de selección de capital humano. *Revista mexicana de economía y finanzas*, vol. 15, no. 3, ISSN 1665-5346. DOI 10.21919/remef.v15i3.516.
- THAKURIA, D., CHALIHA, C., DUTTA, P., SINHA, S., UZIR, P., SINGH, S.B., HAZARIKA, S., SAHOO, L., KHARBIKAR, L.L. y SINGH, D., 2023. Citrus Huanglongbing (HLB): Diagnostic and management options. *Physiological and Molecular Plant Pathology*, vol. 125, ISSN 0885-5765. DOI 10.1016/j.pmp.2023.102016.

- USECHE, M.C., ARTIGAS, W., QUEIPO, B. y PEROZO, É., 2019. *Técnicas e instrumentos de recolección de datos cuali-cuantitativos*. [en línea]. S.l.: Universidad de la Guajira. [consulta: 24 junio 2023]. ISBN 978-956-603-704-0. Disponible en: <https://repositoryinst.uniguajira.edu.co/handle/uniguajira/467>.
- VEGAS, U. y NARREA, M., 2011. Manejo Integrado del cultivo de limón. *Universidad Nacional Agraria la Molina*,
- VELÁZQUEZ, A., 2018. Escala proporcional y escala de intervalo ¿En qué difieren? *QuestionPro* [en línea]. [consulta: 24 junio 2023]. Disponible en: <https://www.questionpro.com/blog/es/escala-proporcional-y-escala-de-intervalo/>.
- VIDAL-VILLACORTA, E.P. y VILLACORTA-GONZÁLEZ, M.Y., 2021. Eficacia del aceite esencial del *Citrus aurantifolia* y de *Ricinus communis* sobre larvas de *Aedes aegypti* comparado con temephos, in vitro. *UCV Scientia Biomédica*, vol. 4, no. 3, ISSN 2663-614X. DOI 10.18050/ucvscientiabiomedica.v4i3.02.
- XING, S., LEE, M. y LEE, K., 2019. Citrus Pests and Diseases Recognition Model Using Weakly Dense Connected Convolution Network. *Sensors*, vol. 19, no. 14, ISSN 1424-8220. DOI 10.3390/s19143195.
- YANG, D., WANG, F., HU, Y., LAN, Y. y DENG, X., 2021. Citrus Huanglongbing Detection Based on Multi-Modal Feature Fusion Learning. *Frontiers in Plant Science*, vol. 12, ISSN 1664-462X. DOI 10.3389/fpls.2021.809506.
- YANG, S. y BERDINE, G., 2017. The receiver operating characteristic (ROC) curve. *The Southwest Respiratory and Critical Care Chronicles*, vol. 5, DOI 10.12746/swrccc.v5i19.391.
- ZAIDI, S.S.A., ANSARI, M.S., ASLAM, A., KANWAL, N., ASGHAR, M. y LEE, B., 2022. A survey of modern deep learning based object detection models. *Digital Signal Processing*, vol. 126, ISSN 1051-2004. DOI 10.1016/j.dsp.2022.103514.
- ZHANG, F., QIU, Z., FENG, T., DAI, Y. y FAN, G., 2022. Modeling the Importation and Local Transmission of Huanglongbing Disease: Bifurcation and Sensitivity Analysis. *International Journal of Bifurcation & Chaos in Applied Sciences & Engineering*, vol. 32, no. 8, ISSN 02181274. DOI 10.1142/S0218127422501176.

ANEXOS

Anexo 1. Tabla de operacionalización de variables

Matriz de Operacionalización de Variables						
Variable	Definición conceptual	Definición Operacional	Dimensión	Indicador	Instrumento	Escala de Medición
V1: Modelo YOLO (You Only Look Once)	Modelo que permite extraer automáticamente características de objetivos de conjuntos de datos de entrenamiento, aprender una representación más avanzada de datos y resolver problemas como la detección de objetivos basados en el aprendizaje profundo y la visión por computadora. (Lyu et al. 2023)	Ajuste y optimización de los parámetros del modelo YOLO, como el tamaño de las capas, las tasas de aprendizaje y los umbrales de confianza, para mejorar su Precisión, Recall, F1-score y AP (Average Precisión) en la detección de síntomas de la enfermedad. (Lawal 2021). En cuanto al procedimiento realizado en el Modelo YOLO es la precisión y recall, se utilizará una guía de observación mediante esta se obtienen datos correspondientes a los indicadores a medir.	Precisión (Lawal 2021; López-Cabrera et al. 2022)	Incrementó Precisión de detección $Precisión (P) = \frac{TP}{TP+FP} \times 100\%$ TP = Verdadero Positivo FP = Falso positivo	Guía de Observación 01 (Useche et al. 2019)	Razón / Continua (Velázquez 2018)
			Recall (Lawal 2021; López-Cabrera et al. 2022)	Incrementó de Recuperación de detección $Recall (R) = \frac{TP}{TP+FN} \times 100\%$ TP = Verdadero positivo FN = Falso negativo		
V2: Diagnóstico Preliminar de la enfermedad de Huanglongbing	Identificación de la presencia de la enfermedad Huanglongbing basada en los resultados de la detección y reconocimiento de síntomas o características específicas. (Nehela y Killiny 2020)	Procesamiento de las imágenes utilizando el modelo YOLO previamente entrenado para detectar y reconocer síntomas o características asociadas a la enfermedad de Huanglongbing. (Qiu et al. 2022). En cuanto al procedimiento en el diagnóstico preliminar de la enfermedad de Huanglongbing es el incremento de la especificidad y sensibilidad en la clasificación de imágenes, se utilizará una guía de	Prueba Diagnóstica (Sancho-Insenser y González-Castillo 2022)	Incrementó de la Sensibilidad (Khan et al. 2021; Zhang et al. 2022; Guzmán y Yineth 2021; López-Cabrera et al. 2022; Sancho-Insenser y González-Castillo 2022) $Sensibilidad = \frac{TP}{TP+FN}$	Guía de Observación 02 (Useche et al. 2019)	Razón / Continua (Velázquez 2018)

		observación para el entrenamiento y validación del sistema para acentuar la confiabilidad de ello.		TP= Verdadero positivo FN = Falso Negativo Incrementó de la Especificidad (González-Garza 2017; Sancho-Insenser y González-Castillo 2022) $Especificidad = \frac{TN}{TN+FP}$ TN = Verdadero negativo FP = Falso Positivo		
--	--	--	--	---	--	--

Fuente: Elaboración propia

Anexo 2. Matriz de consistencia

MATRIZ DE CONSISTENCIA											
Título	Pregunta General	Objetivo General	Preguntas específicas	Objetivos específicos	VARIABLE	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Escala de medición	Diseño metodológico
Modelo YOLO (You Only Look Once) para el diagnóstico preliminar de la enfermedad de Huanglongbing en la planta de limón en la Región Piura, 2023	¿Cómo un modelo YOLO para el diagnóstico preliminar analiza la detección de la enfermedad del Huanglongbing en la planta de limón en la Región Piura?	Analizar la precisión del modelo YOLO (You Only Look Once) para el diagnóstico preliminar de la enfermedad de Huanglongbing en la planta de limón.	¿Cómo evaluar las predicciones sobre las instancias totales en cuanto a la precisión y recuperación?	Evaluar las predicciones sobre las instancias totales en cuanto a la precisión y recuperación	Modelo YOLO (You Only Look Once)	Modelo que permite extraer automáticamente características de objetivos de conjuntos de datos de entrenamiento, aprender una representación más avanzada de datos y características y resolver problemas como la detección de objetivos basados en el aprendizaje profundo y la visión por computadora. (Lyu et al. 2023)	Ajuste y optimización de los parámetros del modelo YOLO, como el tamaño de las capas, las tasas de aprendizaje y los umbrales de confianza, para mejorar su Precisión, Recall, F1-score y AP (Average Precisión) en la detección de síntomas de la enfermedad (Lawal 2021). En cuanto al procedimiento realizado en el Modelo YOLO es la precisión y recall, se utilizará una guía de observación mediante esta se obtienen datos correspondientes a los indicadores a medir.	Precisión (Lawal 2021; López-Cabrera et al. 2022)	Incrementó Precisión de detección Precisión (P) = $TP/TP+FP * 100\%$ TP = Verdadero Positivo FP = Falso positivo	Razón / Continua (Velázquez 2018)	Tipo de Investigación: Aplicada - Tecnológica Diseño de Investigación: Descriptiva – No experimental – Transversal Enfoque de la Investigación: Cuantitativo Población: 5475 imágenes hojas de la planta de limón sutil (Citrus Aurantifolia Swingle) e imágenes hojas de la planta de limón Tahití o persa (Citrus Latifolia Tanaka) (Greening/Healthy with HLB) Muestra:
			¿Cómo clasificar patrones positivos utilizando el modelo YOLO la sensibilidad en el diagnóstico preliminar de la enfermedad Huanglongbing en la planta de limón?	Clasificar por patrones positivos utilizando el modelo YOLO la sensibilidad en el diagnóstico preliminar de la enfermedad Huanglongbing en la planta de limón.		Identificación de la presencia de la enfermedad Huanglongbing basada en los resultados de la detección y reconocimiento de	Procesamiento de las imágenes utilizando el modelo YOLO previamente entrenado para detectar y reconocer síntomas o características asociadas a la	Recall (Lawal 2021; López-Cabrera et al. 2022)	Incrementó de Recuperación de detección Recall(R) = $TP/TP+FN * 100\%$ TP = Verdadero positivo FN = Falso negativo	Razón / Continua (Velázquez 2018)	
			¿Cómo clasificar patrones negativos utilizando el modelo YOLO la especificidad en el diagnóstico preliminar de la enfermedad Huanglongbing	Clasificar por patrones negativos utilizando el modelo YOLO la especificidad en el diagnóstico preliminar de				Prueba Diagnóstica (Sancho-Insenser y González	Incrementó de la Sensibilidad (Khan et al. 2021; Zhang et al. 2022; Guzmán y Yineth 2021;	Razón / Continua (Velázquez 2018)	

			en la planta de limón?	la enfermedad Huanglongbing en la planta de limón.	Diagnóstico Preliminar de la enfermedad de Huanglongbing	síntomas o características específicas. (Nehela y Killiny 2020)	enfermedad de Huanglongbing (Qiu et al. 2022). En cuanto al procedimiento en el diagnóstico preliminar de la enfermedad de Huanglongbing es el incrementó de la especificidad y sensibilidad en la clasificación de imágenes, se utilizará una guía de observación para el entrenamiento y validación del sistema para acentuar la confiabilidad de ello.	-Castillo 2022)	López-Cabrera et al. 2022; Sancho-Insenser y González-Castillo 2022) Sensibilidad = TP/TP+FN TP= Verdadero positivo FN = Falso Negativo Incrementó de la Especificidad (González-Garza 2017; Sancho-Insenser y González-Castillo 2022) Especificidad= TN/TN+FP TN = Verdadero negativo FP = Falso Positivo	Razón / Continua (Velázquez 2018)	Toda la población finita por conveniencia Muestreo: El muestreo se usó un 70% para entrenamiento, 15% para validación y 15% para pruebas (Healthy/ Greening with HLB)
--	--	--	------------------------	--	--	---	---	-----------------	---	-----------------------------------	---

Fuente: Elaboración propia

Anexo 3. Guía de observación 01 – Modelo YOLO

Guía de Observación 01 – Precisión y Recall							
Investigadores	Carreño Chero Irving Dario Obregon Mauricio Alejandro Junior					Tipo de Prueba	Descriptiva
Institución	Universidad César Vallejo						
Dimensión de estudio del Modelo YOLO (You Only Look Once)	<ul style="list-style-type: none"> Precisión Recall 						
Objetivos	<ul style="list-style-type: none"> Evaluar las predicciones sobre las instancias totales en cuanto a la precisión y recuperación 			Fecha de Inicio	27/09/23	Fecha de término	28/09/23
Variable	Indicador			Medida	Fórmulas		
Modelo YOLO (You Only Look Once)	<ul style="list-style-type: none"> Incrementó de Precisión de detección Incrementó de Recuperación de detección 			Categorías por Clasificación	Precisión: $(P) = TP/TP+FP*100\%$	Recall: $(R) = TP/TP+FN*100\%$	
#	Fecha de Revisión	Citrus Aurantifolia (Greening/Enfermo) con HLB		Citrus Aurantifolia (Healthy/ No Enfermo) sin HLB		Resultado	Resultado
		Verdadero Positivo (TP)	Falso Negativo (FN)	Falso Positivo (FP)	Verdadero Negativo (TN)		
1							

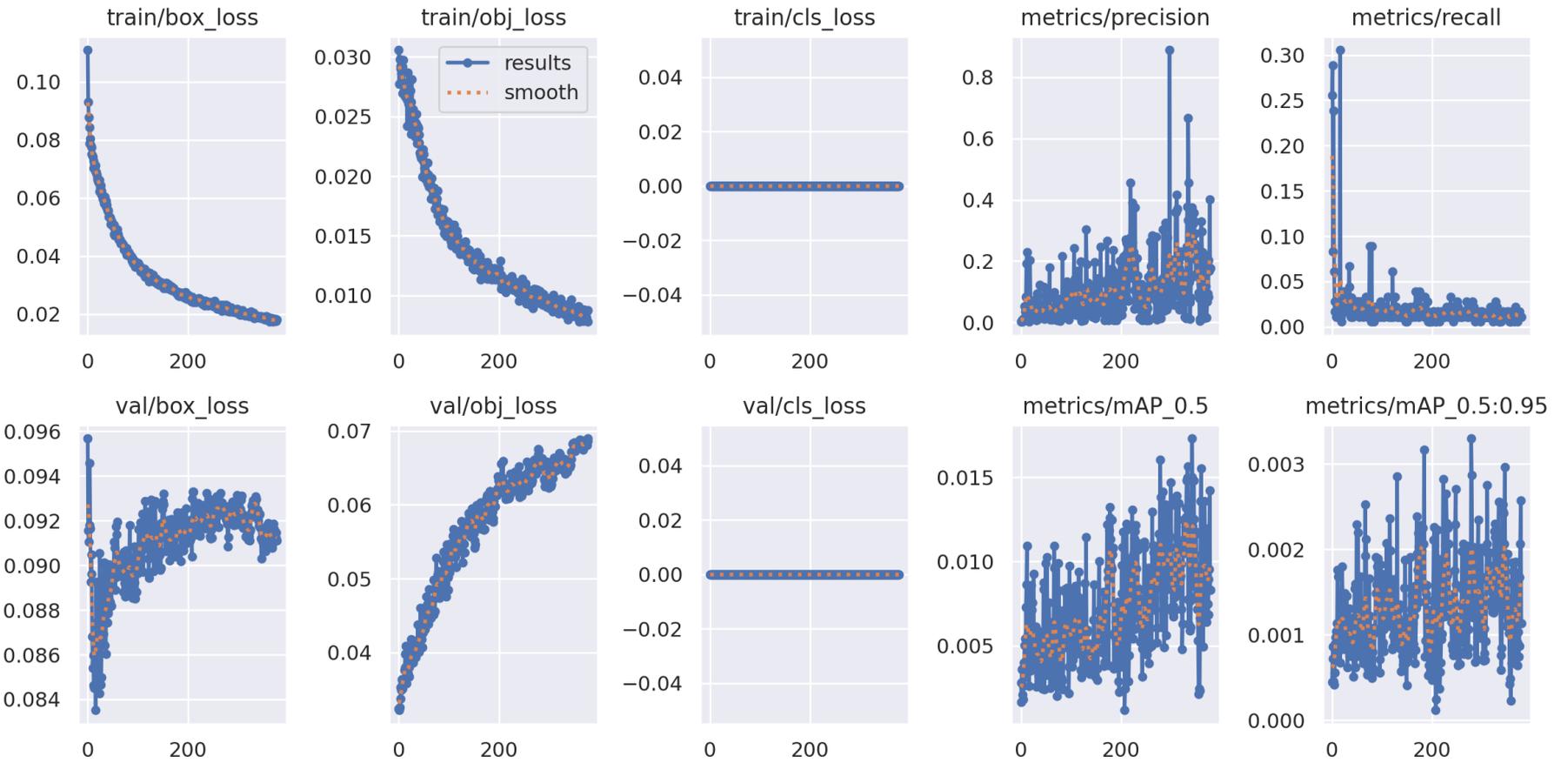
Fuente: Elaboración propia

Anexo 4. Guía de observación 02 - Prueba diagnóstica

Guía de Observación 02 – Prueba diagnóstica								
Investigadores	Carreño Chero Irving Dario Obregon Mauricio Alejandro Junior					Tipo de Prueba	Descriptiva	
Institución	Universidad César Vallejo							
Dimensión de estudio del diagnóstico Preliminar de la enfermedad de Huanglongbing	<ul style="list-style-type: none"> Prueba diagnóstica 							
Objetivos	Clasificar por patrones negativos & positivos utilizando el modelo YOLO la especificidad & sensibilidad en el diagnóstico preliminar de la enfermedad Huanglongbing en la planta de limón			Fecha de Inicio	29/09/23	Fecha de término	30/09/23	
Variable	Indicador			Medida	Fórmulas			
Diagnóstico Preliminar de la enfermedad de Huanglongbing	<ul style="list-style-type: none"> Incrementó de la Sensibilidad Incrementó de la Especificidad 			Categorías por Clasificación	Precisión de Clasificación: $AP + AN / AP + AN + FP + FN$ Error: $FP + FN / n$			
# Iteración (n)	Enfermos (Greening) con HLB		Total, Positivos (AP)	No enfermos (Healthy) sin HLB		Total, Negativos (AN)	Precisión	Error
	Verdadero Positivo (TP)	Falso Negativo (FN)		Falso Positivo (FP)	Verdadero Negativo (TN)			
1								

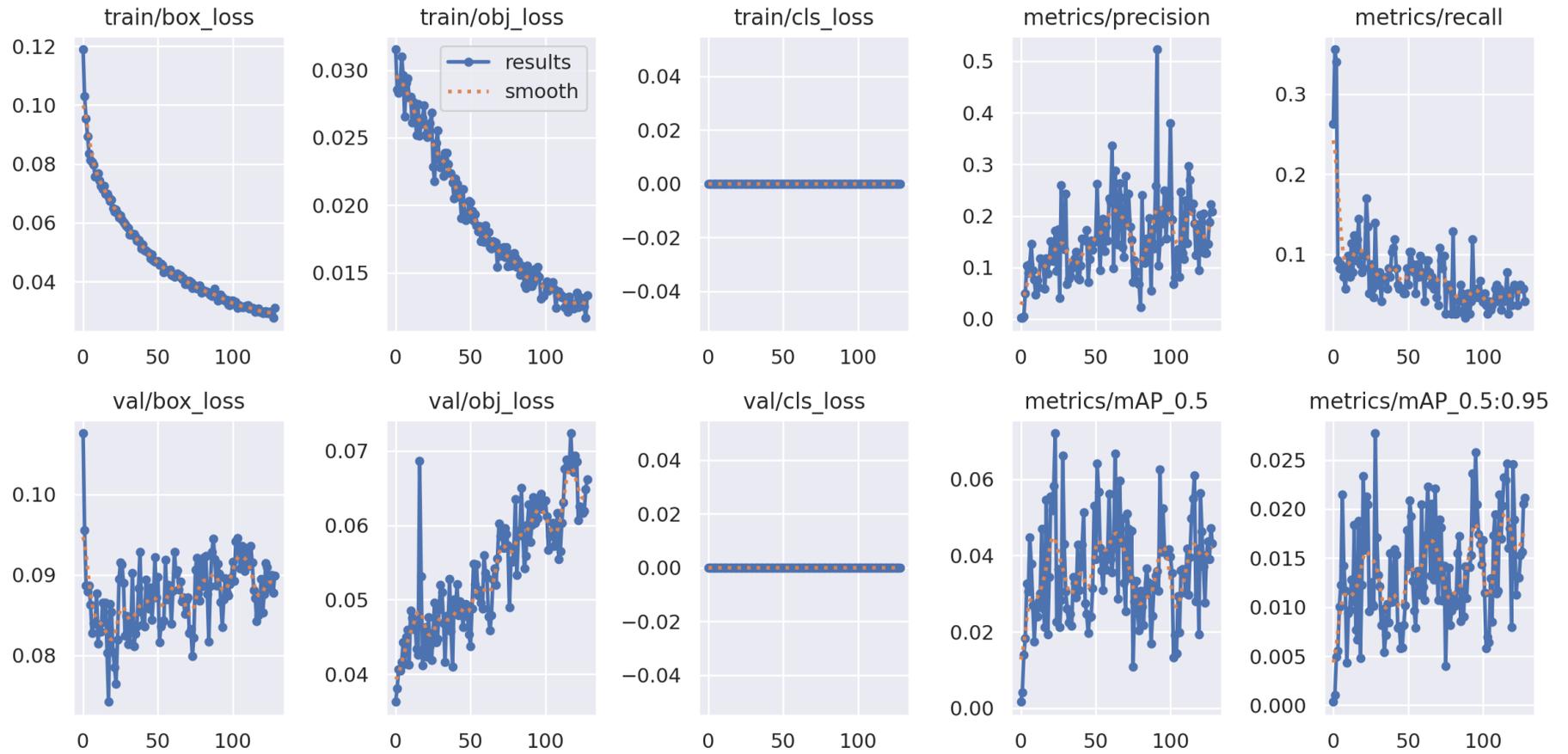
Fuente: Elaboración propia

Anexo 5. Métricas de entrenamiento exp1



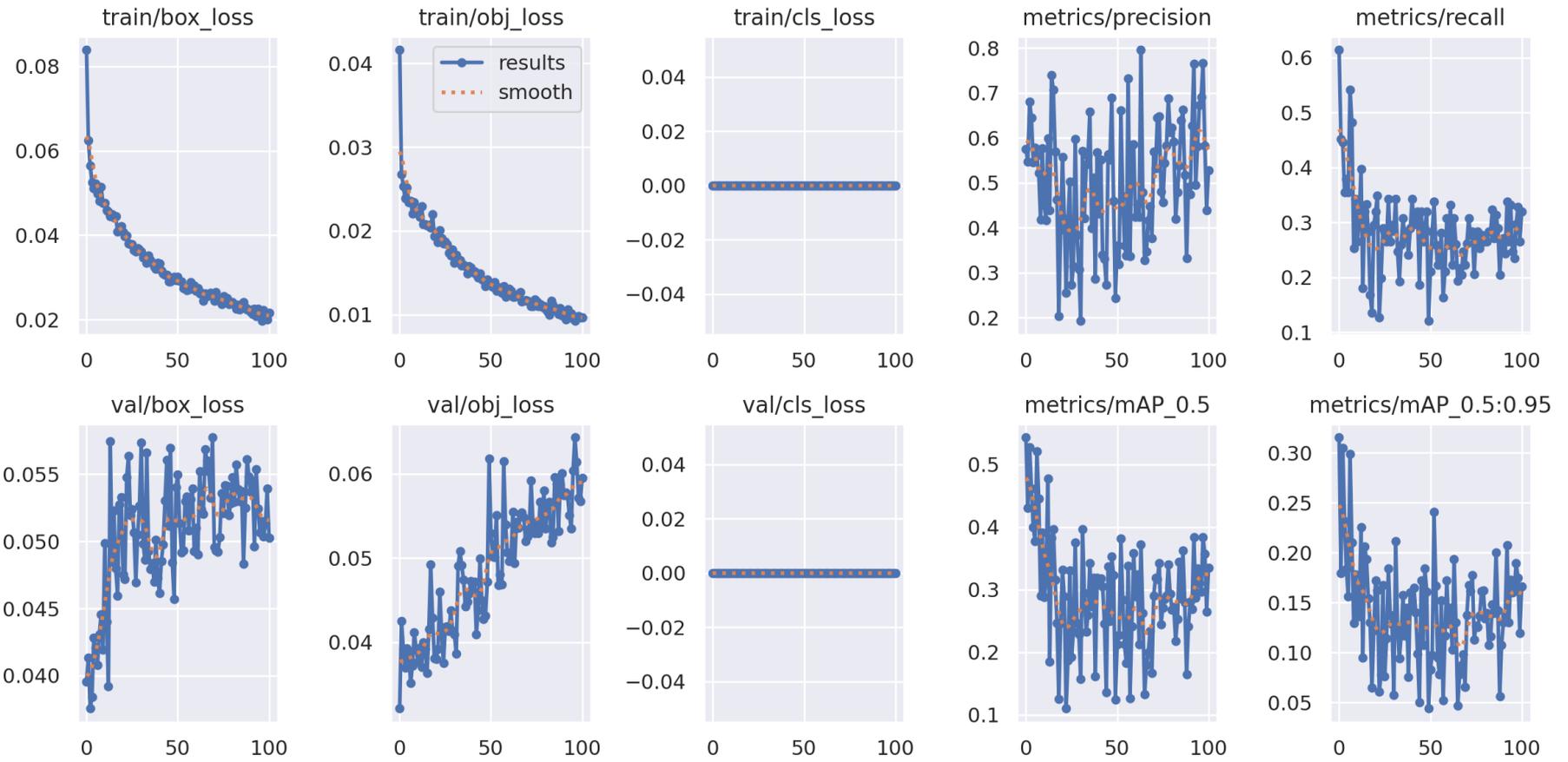
Fuente: Elaboración propia

Anexo 6. Métricas de entrenamiento exp2



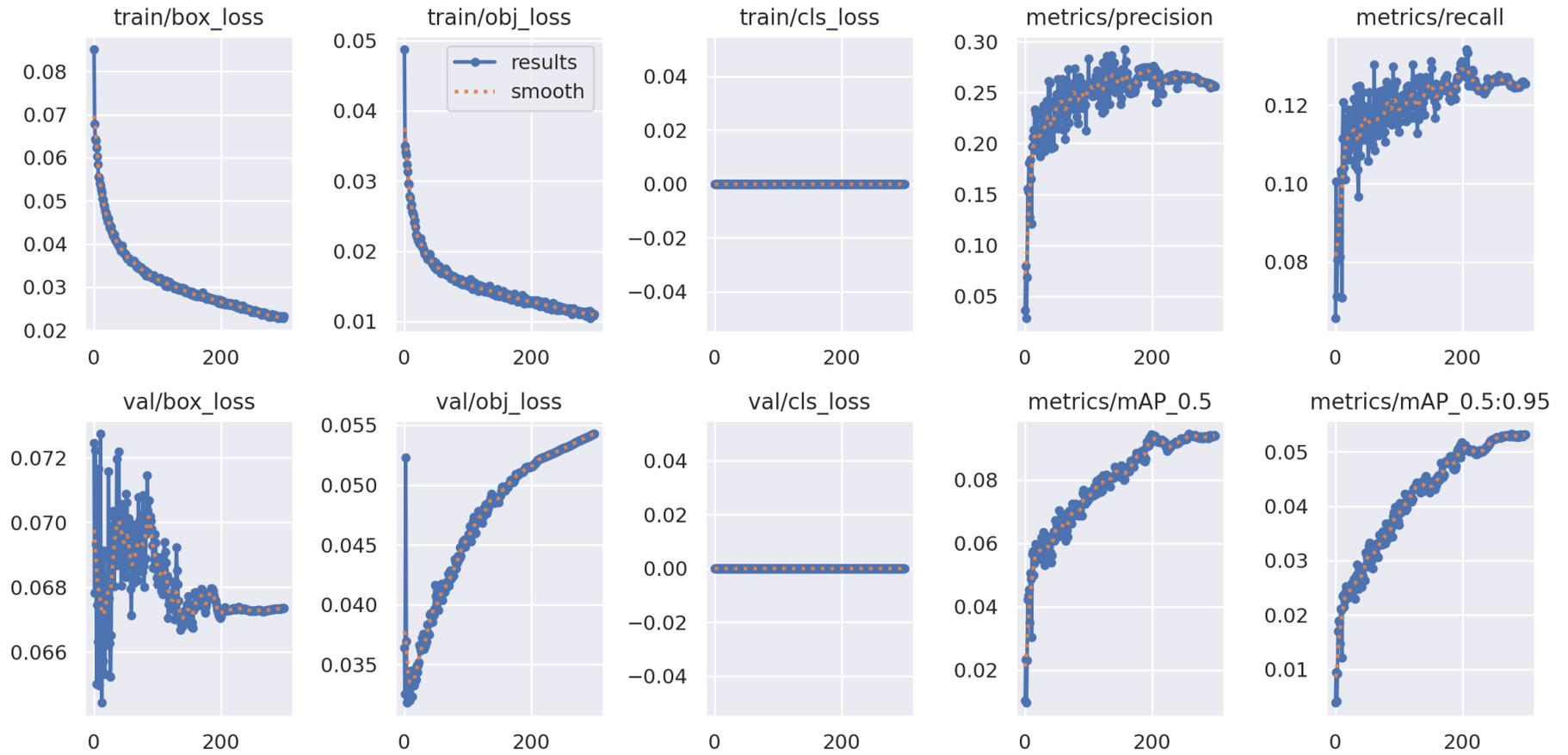
Fuente: Elaboración propia

Anexo 7. Métricas de entrenamiento exp5



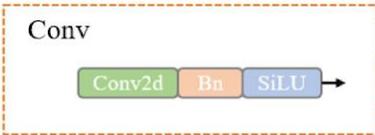
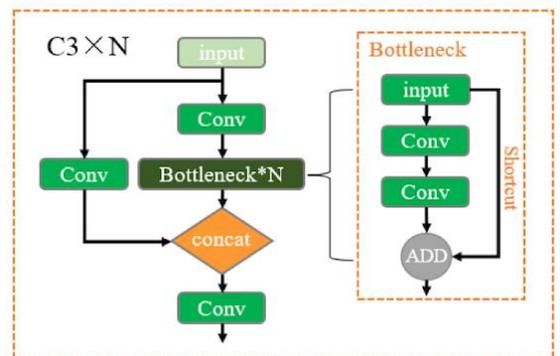
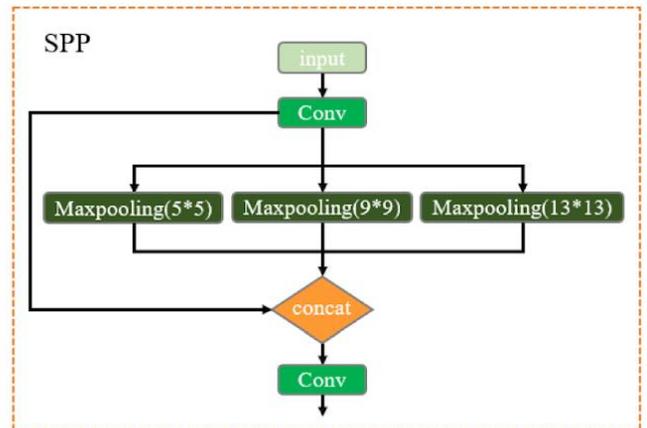
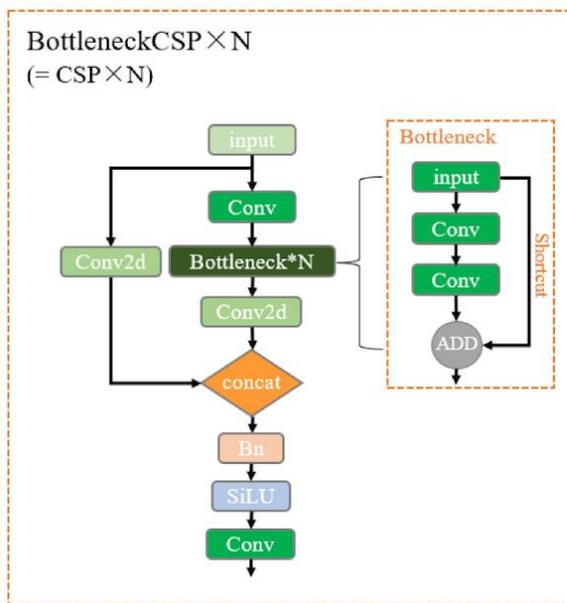
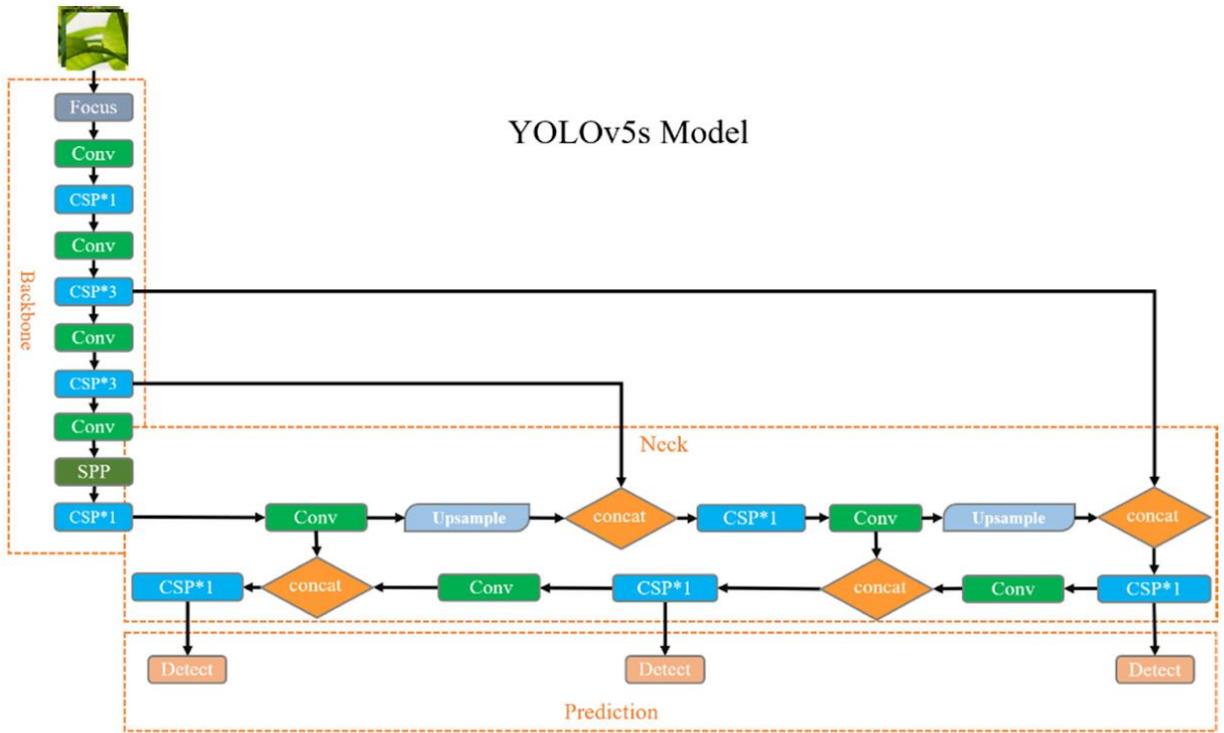
Fuente: Elaboración propia

Anexo 8. Métricas de entrenamiento exp6



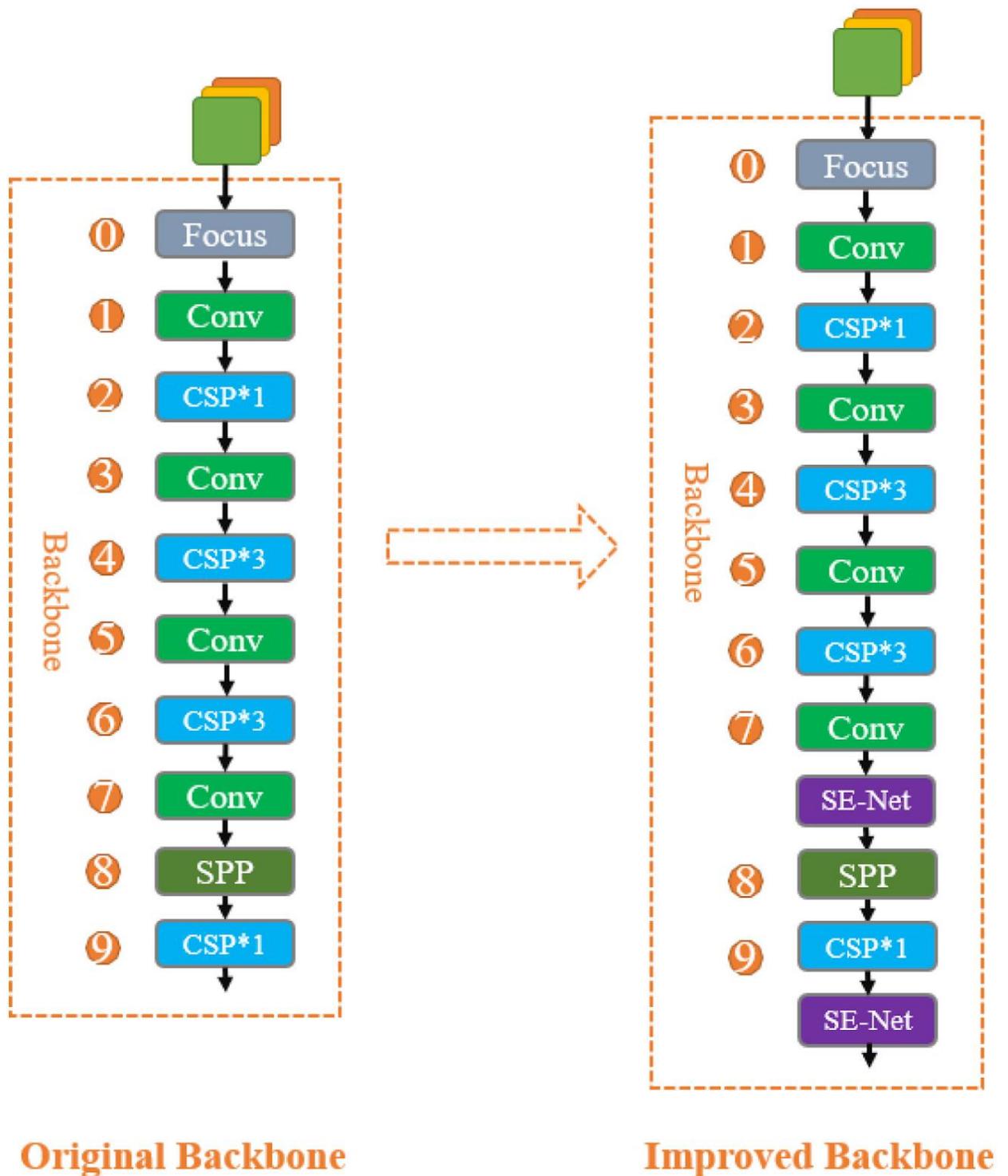
Fuente: Elaboración propia

Anexo 9. Estructura de YOLOv5s



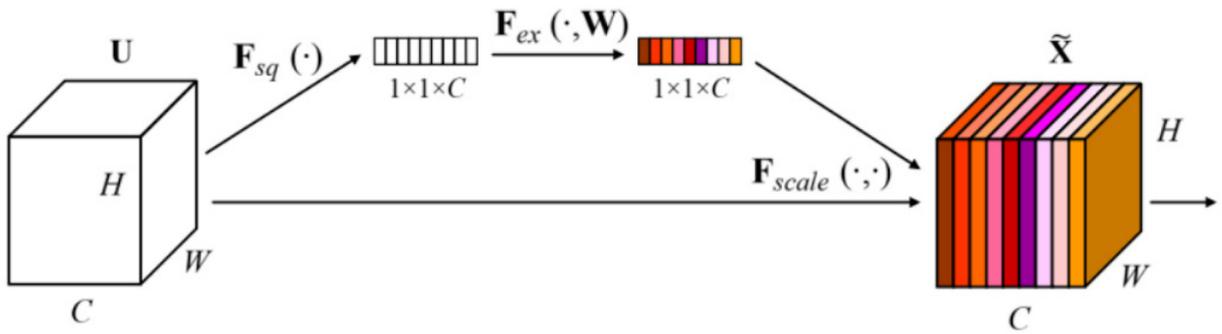
Fuente: Estructura de YOLOv5s (Lyu et al. 2023)

Anexo 10. Comparación de las columnas vertebrales antes y después de la mejora



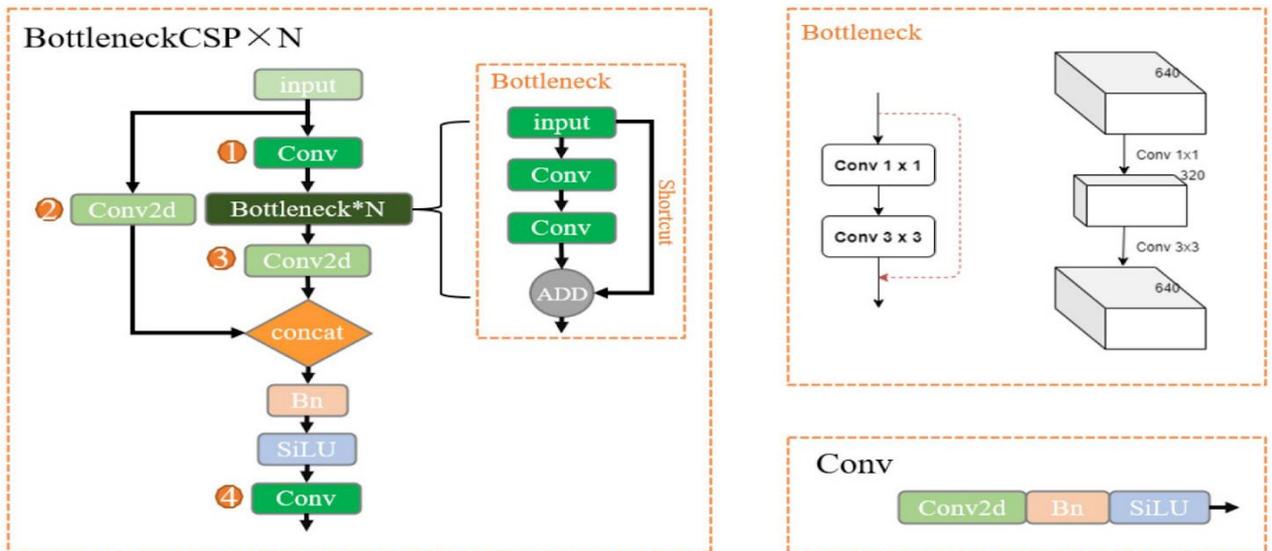
Fuente: Columnas vertebrales antes y después (Lyu et al. 2023)

Anexo 11. Estructura de la SE-Net



Fuente: Estructura de SE-Net (Lyu et al. 2023)

Anexo 12. Estructura de BottleneckCSP



Fuente: Estructura de BottleneckCSP (Lyu et al. 2023)

Anexo 13. Tabla de abreviaturas

Sigla	Significado	Página
HLB	Huanglongbing (Enfermedad Dragón amarillo)	12 - 13 - 15 - 16 - 17 - 21 - 25 - 30 - 34 - 35 - 37 - 39 - 40 - 41
QA	Quality Assurance (Control de calidad)	15
IA	Inteligencia artificial	17 - 27
CNN	Convolutional neural networks (Redes neuronales convolucionales)	19 - 20 - 22 - 27
PCR	Reacción en cadena de la polimerasa	21
AUP	Metodología Agile Unified Process	21
RMSI	Decremento de error de entrenamiento	23
NMS	Non-maximum Suppression (Algoritmo de supresión no máxima)	24
(P)	Precisión	24
(R)	Recall	24 - 25
(F1)	Puntuación F1 (F1-score)	21 - 24 - 25
TP	Positivo verdadero	24 - 25 - 26 - 38 - 39 - 40
FP	Falso positivo	24 - 25 - 26 - 38 - 39 - 40
FN	Falso Negativo	24 - 25 - 26 - 38 - 39 - 40
TN	Verdadero negativo	26 - 38 - 39
TRP	Potencial receptor transitorio (Transient receptor potential)	25 - 26
FRP	Fenómeno de flexión-relajación	26
PCA	Análisis de componentes principales (Principal component analysis)	27
Conv	Módulo de convolución	28 - 29

SPP	Módulo de agrupación de pirámides espaciales (Spatial Pyramid Pooling)	28 - 29
FPN	Red de pirámide de características	28 -
Fsq	Operación de compresión	29
F ex	Operación de excitación	29
Fscale	Operación de escala	29
Conv2d	Capa de convolución	29
BN	Capa de normalización por lote	29
PW	Convolución puntual	29
sRGB	Espacio de color que por defecto se utiliza en imágenes	30
AP	Average Precision (Precisión media)	32
DOI	Identificación de documentos en internet	41
ISO	International Organization for Standardization (Organización Internacional de Normalización)	22 - 40
JPEG, JPG	Joint Photographic Experts Group (Grupo Mixto de Expertos en Fotografía)	30

Fuente: Elaboración propia

Anexo 14. Código para generar la Curva ROC

```
289     if len(stats) and stats[0].any():
290         import sklearn.metrics
291         from utils.metrics import plot_auroc_curve
292         y_true, y_score = stats[0][:,0].astype(int), stats[1]
293         fpr, tpr, thresholds = sklearn.metrics.roc_curve(y_true, y_score, pos_label=1)
294         auc = sklearn.metrics.auc(fpr, tpr)
295         file = open(Path(save_dir)/ "roc_curve_data.txt", "a")
296         file.write(f'{fpr} {tpr}\n')
297         file.close()
298         plt.figure()
299         plt.plot(fpr,tpr, color='darkorange', lw=2, label=f'Huanglongbing {auc:.3f}')
300         plt.plot([0, 1], [0, 1], color='navy', lw=2, linestyle='--')
301         plt.xlim([0.0, 1.0])
302         plt.ylim([0.0, 1.05])
303         plt.xlabel('False Positive Rate')
304         plt.ylabel('True Positive Rate')
305         plt.title('Receiver Operating Characteristic (ROC) Curve')
306         plt.legend(loc='lower right')
307         plt.savefig(Path(save_dir)/ 'ROC_curve_from_val.png')
308         plot_auroc_curve(fpr, tpr, auc, save_dir=Path(save_dir)/ f'ROC_curve_GITHUB.png')
```

Fuente: Elaboración propia

Anexo 15. Código para generar la Matriz de confusión

```
segment > val.py > ...
22
23 import argparse
24 import json
25 import os
26 import subprocess
27 import sys
28 from multiprocessing.pool import ThreadPool
29 from pathlib import Path
30
31 import numpy as np
32 import torch
33 from tqdm import tqdm
34
35 FILE = Path(__file__).resolve()
36 ROOT = FILE.parents[1] # YOLOv5 root directory
37 if str(ROOT) not in sys.path:
38     sys.path.append(str(ROOT)) # add ROOT to PATH
39 ROOT = Path(os.path.relpath(ROOT, Path.cwd())) # relative
40
41 import torch.nn.functional as F
42
43 from models.common import DetectMultiBackend
44 from models.yolo import SegmentationModel
45 from utils.callbacks import callbacks
46 from utils.general import (LOGGER, NUM_THREADS, TQDM_BAR_FORMAT, Profile, check_dataset, check_img_size,
47                             check_requirements, check_yaml, coco80_to_coco91_class, colorstr, increment_path,
48                             non_max_suppression, print_args, scale_boxes, xywh2xyxy, xyxy2xywh)
49 from utils.metrics import ConfusionMatrix, box_iou
50 from utils.plots import output_to_target, plot_val_study
51 from utils.segment.dataloaders import create_dataloader
52 from utils.segment.general import mask_iou, process_mask, process_mask_native, scale_image
53 from utils.segment.metrics import Metrics, ap_per_class_box_and_mask
54 from utils.segment.plots import plot_images_and_masks
55 from utils.torch_utils import de_parallel, select_device, smart_inference_mode
56
57
58 def save_one_txt(predn, save_conf, shape, file):
59     # Save one txt result
```

Fuente: Elaboración propia

Anexo 16. Código para generar la Matriz de confusión 2

```
segment > val.py > run > 00 confusion_matrix
229 confusion_matrix = ConfusionMatrix(nc=nc)
230 names = model.names if hasattr(model, 'names') else model.module.names # get class names
231 if isinstance(names, (list, tuple)): # old format
232     names = dict(enumerate(names))
233 class_map = coco80_to_coco91_class() if is_coco else list(range(1000))
234 s = ('%22s' * '%11s' * 10) % ('Class', 'Images', 'Instances', 'Box(P', 'R', 'mAP50', 'mAP50-95)', 'Mask(P', 'R',
235                             'mAP50', 'mAP50-95)')
236 dt = Profile(), Profile(), Profile()
237 metrics = Metrics()
238 loss = torch.zeros(4, device=device)
239 jdict, stats = [], []
240 # callbacks.run('on_val_start')
241 pbar = tqdm(data_loader, desc=s, bar_format=TQDM_BAR_FORMAT) # progress bar
242 for batch_i, (im, targets, paths, shapes, masks) in enumerate(pbar):
243     # callbacks.run('on_val_batch_start')
244     with dt[0]:
245         if cuda:
246             im = im.to(device, non_blocking=True)
247             targets = targets.to(device)
248             masks = masks.to(device)
249             masks = masks.float()
250             im = im.half() if half else im.float() # uint8 to fp16/32
251             im /= 255 # 0 - 255 to 0.0 - 1.0
252             nb, _, height, width = im.shape # batch size, channels, height, width
253
254     # Inference
255     with dt[1]:
256         preds, protos, train_out = model(im) if compute_loss else (*model(im, augment=augment)[:2], None)
257
258     # Loss
259     if compute_loss:
260         loss += compute_loss((train_out, protos), targets, masks)[1] # box, obj, cls
261
262     # NMS
263     targets[:, 2:] *= torch.tensor((width, height, width, height), device=device) # to pixels
264     lb = [targets[targets[:, 0] == i, 1:] for i in range(nb)] if save_hybrid else [] # for autolabelling
265     with dt[2]:
```

Fuente: Elaboración propia

Anexo 17. Código para generar el Correlograma de labels

```
utils > plots.py > plot_labels
270
271 @TryExcept() # known issue https://github.com/ultralytics/yolov5/issues/5395
272 def plot_labels(labels, names=(), save_dir=Path('')):
273     # plot dataset labels
274     LOGGER.info(f"Plotting labels to {save_dir / 'labels.jpg'}... ")
275     c, b = labels[:, 0], labels[:, 1:].transpose() # classes, boxes
276     nc = int(c.max() + 1) # number of classes
277     x = pd.DataFrame(b.transpose(), columns=['x', 'y', 'width', 'height'])
278
279     # seaborn correlogram
280     sn.pairplot(x, corner=True, diag_kind='auto', kind='hist', diag_kws=dict(bins=50), plot_kws=dict(pmax=0.9))
281     plt.savefig(save_dir / 'labels_correlogram.jpg', dpi=200)
282     plt.close()
283
284     # matplotlib labels
285     matplotlib.use('svg') # faster
286     ax = plt.subplots(2, 2, figsize=(8, 8), tight_layout=True)[1].ravel()
287     y = ax[0].hist(c, bins=np.linspace(0, nc, nc + 1) - 0.5, rwidth=0.8)
288     with contextlib.suppress(Exception): # color histogram bars by class
289         [y[2].patches[i].set_color([x / 255, (module) np
290         ax[0].set_ylabel('instances')
291         if 0 < len(names) < 30:
292             ax[0].set_xticks(range(len(names)))
293             ax[0].set_xticklabels(list(names.values()))
294         else:
295             ax[0].set_xlabel('classes')
296     sn.histplot(x, x='x', y='y', ax=ax[2], bins=50, rwidth=0.8)
297     sn.histplot(x, x='width', y='height', ax=ax[3], bins=50, rwidth=0.8)
298
299     # rectangles
300     labels[:, 1:3] = 0.5 # center
301     labels[:, 1:] = xywh2xyxy(labels[:, 1:])
302     img = Image.fromarray(np.ones((2000, 2000, 3), dtype=np.uint8) * 255)
303     for cls, *box in labels[:1000]:
304         ImageDraw.Draw(img).rectangle(box, width=1, outline=colors(cls)) # plot
305     ax[1].imshow(img)
306     ax[1].axis('off')
```

Fuente: Elaboración propia

Anexo 18. Código para generar la Curva de Precisión-Recall

```
utils > metrics.py > ap_per_class
53 # Create Precision-Recall curve and compute AP for each class
54 px, py = np.linspace(0, 1, 1000), [] # for plotting
55 ap, p, r = np.zeros((nc, tp.shape[1])), np.zeros((nc, 1000)), np.zeros((nc, 1000))
56 for ci, c in enumerate(unique_classes):
57     i = pred_cls == c
58     n_l = nt[ci] # number of labels
59     n_p = i.sum() # number of predictions
60     if n_p == 0 or n_l == 0:
61         continue
62
63     # Accumulate FPs and TPs
64     fpc = (1 - tp[i]).cumsum(0)
65     tpc = tp[i].cumsum(0)
66
67     # Recall
68     recall = tpc / (n_l + eps) # recall curve
69     r[ci] = np.interp(-px, -conf[i], recall[:, 0], left=0) # negative x, xp because xp decreases
70
71     # Precision
72     precision = tpc / (tpc + fpc) # precision curve
73     p[ci] = np.interp(-px, -conf[i], precision[:, 0], left=1) # p at pr_score
74
75     # AP from recall-precision curve
76     for j in range(tp.shape[1]):
77         ap[ci, j], mpre, mrec = compute_ap(recall[:, j], precision[:, j])
78         if plot and j == 0:
79             py.append(np.interp(px, mrec, mpre)) # precision at mAP@0.5
80
81 # Compute F1 (harmonic mean of precision and recall)
82 f1 = 2 * p * r / (p + r + eps)
83 names = [v for k, v in names.items() if k in unique_classes] # list: only classes that have data
84 names = dict(enumerate(names)) # to dict
85 if plot:
86     plot_pr_curve(px, py, ap, Path(save_dir) / f'{prefix}PR_curve.png', names)
87     plot_mc_curve(px, f1, Path(save_dir) / f'{prefix}F1_curve.png', names, ylabel='F1')
88     plot_mc_curve(px, p, Path(save_dir) / f'{prefix}P_curve.png', names, ylabel='Precision')
89     plot_mc_curve(px, r, Path(save_dir) / f'{prefix}R_curve.png', names, ylabel='Recall')
```

Fuente: Elaboración propia

Anexo 19. Código para generar las métricas de entrenamiento

```
utils > loggers > __init__.py > Loggers > on_train_end
267
268 def on_train_end(self, last, best, epoch, results):
269     # Callback runs on training end, i.e. saving best model
270     if self.plots:
271         plot_results(file=self.save_dir / 'results.csv') # save results.png
272         files = ['results.png', 'confusion matrix.png', *f'{x} Curve.png' for x in ('F1', 'PR', 'P', 'R')]
273         files = [self.save_dir / f for f in files if (self.save_dir / f).exists()] # filter
274         self.logger.info(f"Results saved to {colorstr('bold', self.save_dir)}")
275
276     if self.tb and not self.clearml: # These images are already captured by ClearML by now, we don't want doubles
277         for f in files:
278             self.tb.add_image(f.stem, cv2.imread(str(f))[..., ::-1], epoch, dataformats='HWC')
279
280     if self.wandb:
281         self.wandb.log(dict(zip(self.keys[3:10], results)))
282         self.wandb.log({'Results': [wandb.Image(str(f), caption=f.name) for f in files]})
283         # Calling wandb.log. TODO: Refactor this into WandbLogger.log_model
284         if not self.opt.evolve:
285             wandb.log_artifact(str(best if best.exists() else last),
286                               type='model',
287                               name=f'run_{self.wandb.wandb_run.id}_model',
288                               aliases=['latest', 'best', 'stripped'])
289             self.wandb.finish_run()
290
291     if self.clearml and not self.opt.evolve:
292         self.clearml.task.update_output_model(model_path=str(best if best.exists() else last),
293                                               name='Best Model',
294                                               auto_delete_file=False)
295
296     if self.comet_logger:
297         final_results = dict(zip(self.keys[3:10], results))
298         self.comet_logger.on_train_end(files, self.save_dir, last, best, epoch, final_results)
299
300 def on_params_update(self, params: dict):
301     # Update hyperparams or configs of the experiment
302     if self.wandb:
303         self.wandb.wandb_run.config.update(params, allow_val_change=True)
304     if self.comet_logger:
```

Fuente: Elaboración propia

Anexo 20. Código para generar la media y desviación típica de la sensibilidad

```
1 # HALLAR LA MEDIA DE LA SENSIBILIDAD Y DESV. TÍPICA
2 # Crear un vector con tus valores
3 valores <- c(54.12, 51.14, 51.31, 56.04, 50.79, 53.77)
4 media <- mean(valores) # Calcular la media
5 # Calcular la desviación típica (muestral) utilizando la función sd()
6 desviacion_tipica <- sd(valores)
7 # Redondear la media y la desviación típica a dos decimales
8 media <- round(media, 2)
9 desviacion_tipica <- round(desviacion_tipica, 2)
10 # Imprimir los resultados
11 cat("Media:", media, "%\n")
12 cat("Desviación Típica:", desviacion_tipica, "%\n")
13
14
15
```

Console output:

```
> # HALLAR LA MEDIA DE LA SENSIBILIDAD Y DESV. TÍPICA
> # Crear un vector con tus valores
> valores <- c(54.12, 51.14, 51.31, 56.04, 50.79, 53.77)
> media <- mean(valores) # Calcular la media
> # Calcular la desviación típica (muestral) utilizando la función sd()
> desviacion_tipica <- sd(valores)
> # Redondear la media y la desviación típica a dos decimales
> media <- round(media, 2)
> desviacion_tipica <- round(desviacion_tipica, 2)
> # Imprimir los resultados
> cat("Media:", media, "%\n")
Media: 52.86 %
> cat("Desviación Típica:", desviacion_tipica, "%\n")
Desviación Típica: 2.11 %
```

Fuente: Elaboración propia

Anexo 21. Código para generar la media y desviación típica de la especificidad

```
14 # HALLAR LA MEDIA DE LA ESPECIFICIDAD Y DESV. TÍPICA
15 # Crear un vector con tus valores
16 valores <- c(55.69, 61.82, 71.63, 62.87, 68.83, 60.60)
17 media <- mean(valores) # Calcular la media
18 # Calcular la desviación típica (muestral) utilizando la función sd()
19 desviacion_tipica <- sd(valores)
20 # Redondear la media y la desviación típica a dos decimales
21 media <- round(media, 2)
22 desviacion_tipica <- round(desviacion_tipica, 2)
23 # Imprimir los resultados
24 cat("Media:", media, "%\n")
25 cat("Desviación Típica:", desviacion_tipica, "%\n")
26
27
```

Console output:

```
> # HALLAR LA MEDIA DE LA ESPECIFICIDAD Y DESV. TÍPICA
> # Crear un vector con tus valores
> valores <- c(55.69, 61.82, 71.63, 62.87, 68.83, 60.60)
> media <- mean(valores) # Calcular la media
> # Calcular la desviación típica (muestral) utilizando la función sd()
> desviacion_tipica <- sd(valores)
> # Redondear la media y la desviación típica a dos decimales
> media <- round(media, 2)
> desviacion_tipica <- round(desviacion_tipica, 2)
> # Imprimir los resultados
> cat("Media:", media, "%\n")
Media: 63.57 %
> cat("Desviación Típica:", desviacion_tipica, "%\n")
Desviación Típica: 5.78 %
```

Fuente: Elaboración propia

Anexo 22. Carta de presentación para validación de instrumentos por juicio de expertos



CARTA DE PRESENTACIÓN

Mg. Correa Calle, Teófilo Roberto

Presente

Asunto: VALIDACIÓN DE INSTRUMENTOS A TRAVÉS DE JUICIO DE EXPERTO.

Nos es muy grato comunicarnos con usted para expresarle nuestros saludos y así mismo, hacer de su conocimiento que, siendo estudiante de la Escuela profesional de Ingeniería de Sistemas de la Universidad César Vallejo, en la sede Piura, requerimos validar los instrumentos con los cuales recogeremos la información necesaria para poder desarrollar nuestra investigación.

El título de nuestro proyecto de investigación es: **“Modelo YOLO (You Only Look Once) para el diagnóstico preliminar de la enfermedad de Huanglongbing en la planta de limón en la Región Piura, 2023”** y siendo imprescindible contar con la aprobación de docentes especializados para poder aplicar los instrumentos en mención, hemos considerado conveniente recurrir a usted, ante su connotada experiencia en temas de investigación educativa.

El expediente de validación, que le hacemos llegar contiene:

- Carta de presentación.
- Definiciones conceptuales de las variables y dimensiones.
- Matriz de consistencia
- Certificado de validez de contenido de los instrumentos.
- Instrumento de validación de la metodología de desarrollo
- Instrumento de validación de los indicadores

Expresándole nuestros sentimientos de respeto y consideración nos despedimos de usted, no sin antes agradecerle por la atención que dispense a la presente.

Atentamente.

Firma
Apellidos y nombre:
Carreño Chero, Irving Darío
D.N.I: 74628463

Firma
Apellidos y nombre:
Obregón Mauricio, Alejandro Junior
D.N. I: 71078410

Fuente: Elaboración propia

Anexo 24. Certificado de validez de contenido del indicador incremento de recuperación

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:

Nº	DIMENSIONES / ítems	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		Si	No	Si	No	Si	No	
	DIMENSIÓN 2: Recall							
1	INDICADOR 1: Incremento de recuperación de detección							
a	Es formulado con lenguaje apropiado.	X		X		X		
b	Es adecuado el avance, la ciencia y tecnología.	X		X		X		
c	Existe una organización lógica.	X		X		X		
d	Comprende los aspectos de cantidad y calidad.	X		X		X		
e	Adecuado para valorar los aspectos del sistema metodológico y científico.	X		X		X		
f	Está basado en aspectos teóricos y científicos.	X		X		X		
g	En los datos respecto al indicador.	X		X		X		
h	Responde al propósito de investigación.	X		X		X		
i	El instrumento es adecuado al tipo de investigación.	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: **Aplicable [X]** **Aplicable después de corregir []** **No aplicable []**

Apellidos y nombres del juez validador: **Mg. Teófilo Roberto Correa Calle** **DNI: 02820231**

Especialidad del validador: **DIRECCION Y GESTION DE LAS TECNOLOGIAS DE LA INFORMACION Y COMUNICACIONES**

¹**Pertinencia:** El ítem corresponde al concepto teórico formulado.

²**Relevancia:** El ítem es apropiado para representar al componente o dimensión específica del constructo

³**Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

10 de agosto del 2023



 TEOFILO ROBERTO
 CORREA CALLE

Firma del Experto Informante.

Fuente: Elaboración propia

Anexo 25. Certificado de validez de contenido del indicador incremento de la sensibilidad

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:

N°	DIMENSIONES / ítems	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		Si	No	Si	No	Si	No	
	DIMENSIÓN 3: Prueba Diagnóstica							
1	INDICADOR 1: Incremento de la sensibilidad							
a	Es formulado con lenguaje apropiado.	X		X		X		
b	Es adecuado el avance, la ciencia y tecnología.	X		X		X		
c	Existe una organización lógica	X		X		X		
d	Comprende los aspectos de cantidad y calidad.	X		X		X		
e	Adecuado para valorar los aspectos del sistema metodológico y científico.	X		X		X		
f	Está basado en aspectos teóricos y científicos.	X		X		X		
g	En los datos respecto al indicador.	X		X		X		
h	Responde al propósito de investigación.	X		X		X		
i	El instrumento es adecuado al tipo de investigación.	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: **Aplicable [X]** **Aplicable después de corregir []** **No aplicable []**

Apellidos y nombres del juez validador. **Mg. Teófilo Roberto Correa Calle** **DNI: 02820231**

Especialidad del validador: **DIRECCION Y GESTION DE LAS TECNOLOGIAS DE LA INFORMACION Y COMUNICACIONES**

¹**Pertinencia:** El ítem corresponde al concepto teórico formulado.

²**Relevancia:** El ítem es apropiado para representar al componente o dimensión específica del constructo

³**Claridad:** Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

10 de agosto del 2023



TEOFILO ROBERTO
CORREA CALLE

Firma del Experto Informante.

Fuente: Elaboración propia

Anexo 26. Certificado de validez de contenido del indicador incrementó de la especificidad

CERTIFICADO DE VALIDEZ DE CONTENIDO DEL INSTRUMENTO QUE MIDE:

Nº	DIMENSIONES / ítems	Pertinencia ¹		Relevancia ²		Claridad ³		Sugerencias
		Si	No	Si	No	Si	No	
	DIMENSION 3: Prueba Diagnóstica							
1	INDICADOR 2: Incremento de la Especificidad							
a	Es formulado con lenguaje apropiado.	X		X		X		
b	Es adecuado el avance, la ciencia y tecnología.	X		X		X		
c	Existe una organización lógica.	X		X		X		
d	Comprende los aspectos de cantidad y calidad.	X		X		X		
e	Adecuado para valorar los aspectos del sistema metodológico y científico.	X		X		X		
f	Está basado en aspectos teóricos y científicos.	X		X		X		
g	En los datos respecto al indicador.	X		X		X		
h	Responde al propósito de investigación.	X		X		X		
i	El instrumento es adecuado al tipo de investigación.	X		X		X		

Observaciones (precisar si hay suficiencia): _____

Opinión de aplicabilidad: Aplicable [X] Aplicable después de corregir [] No aplicable []

Apellidos y nombres del juez validador. Mg. Teófilo Roberto Correa Calle DNI: 02820231

Especialidad del validador: DIRECCION Y GESTION DE LAS TECNOLOGIAS DE LA INFORMACION Y COMUNICACIONES

¹Pertinencia: El ítem corresponde al concepto teórico formulado.

²Relevancia: El ítem es apropiado para representar al componente o dimensión específica del constructo

³Claridad: Se entiende sin dificultad alguna el enunciado del ítem, es conciso, exacto y directo

Nota: Suficiencia, se dice suficiencia cuando los ítems planteados son suficientes para medir la dimensión

10 de agosto del 20



TEOFILO ROBERTO
CORREA CALLE

Firma del Experto Informante.

Fuente: Elaboración propia

Anexo 27. Tabla de validación de instrumento de expertos de los indicadores incrementó de precisión y de recuperación

TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Incremento de la precisión de detección e Incremento de la recuperación de la detección	
I. DATOS GENERALES	
Apellidos y Nombres del Experto:	CORREA CALLE TEÓFILO ROBERTO
Título y/o Grado Académico:	DIRECCION Y GESTION DE LAS TIC
Doctor () Magister (X) Ingeniero () Licenciado () Otro ()	
Universidad que labora:	UNIVERSIDAD CESAR VALLEJO
Fecha :10/08/23	10/08/23
TESIS : Modelo YOLO (You Only Look Once) para el diagnóstico preliminar de la enfermedad de Huanglongbing en la planta de limón en la Región Piura, 2023	

Autores: Carreño Chero, Irving Darío y Obregón Mauricio, Alejandro Junior

Deficiente (0-20%) Regular (21-50%) Bueno (51-70%) Muy Bueno (71-80%) Excelente (81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucrado mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.					85
OBJETIVIDAD	Esta expresado en conducta observable.					85
ACTUALIDAD	Es adecuado el avance, la ciencia y tecnología.					90
ORGANIZACIÓN	Existe una organización lógica.					85
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					85
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.					85
CONSISTENCIA	Está basado en aspectos teóricos y científicos.					85
COHERENCIA	En los datos respecto al indicador.					85
METODOLOGÍA	Responde al propósito de investigación.					85
PERTENENCIA	El instrumento es adecuado al tipo de investigación.					85
TOTAL						855

III. PROMEDIO DE VALIDACIÓN

85.5

IV. OPCIÓN DE APLICABILIDAD

- (X) El instrumento puede ser aplicado, tal como está elaborado
 () El instrumento debe ser mejorado antes de ser aplicado

FIRMA DEL EXPERTO



TEÓFILO ROBERTO
CORREA CALLE

Fuente: Elaboración propia

Anexo 28. Tabla de validación de instrumento de expertos de los indicadores incrementó de sensibilidad y de especificidad

TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Incremento de sensibilidad e Incremento de la especificidad

I. DATOS GENERALES

Apellidos y Nombres del Experto: **CORREA CALLE TEOFILO ROBERTO**
 Título y/o Grado Académico: **DIRECCION Y GESTION DE LAS TIC**

Doctor () Magister (X) Ingeniero () Licenciado () Otro ().....

Universidad que labora: **UNIVERSIDAD CESAR VALLEJO**
 Fecha: **10/08/23**

TESIS : Modelo YOLO (You Only Look Once) para el diagnóstico preliminar de la enfermedad de Huanglongbing en la planta de limón en la Región Piura, 2023

Autores: Carreño Chero, Irving Darío y Obregón Mauricio Alejandro Junior

Deficiente (0-20%) Regular (21-50%) Bueno (51-70%) Muy Bueno (71-80%) Excelente (81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucrado mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.					85
OBJETIVIDAD	Esta expresado en conducta observable.					85
ACTUALIDAD	Es adecuado el avance, la ciencia y tecnología.					90
ORGANIZACIÓN	Existe una organización lógica.					85
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					85
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.					85
CONSISTENCIA	Está basado en aspectos teóricos y científicos.					85
COHERENCIA	En los datos respecto al indicador.					85
METODOLOGIA	Responde al propósito de investigación.					85
PERTENENCIA	El instrumento es adecuado al tipo de investigación.					85
TOTAL						855

III. PROMEDIO DE VALIDACIÓN

85,5

IV. OPCIÓN DE APLICABILIDAD

- (X) El instrumento puede ser aplicado, tal como está elaborado
 () El instrumento debe ser mejorado antes de ser aplicado

FIRMA DEL EXPERTO



TEOFILO ROBERTO
CORREA CALLE

Fuente: Elaboración propia



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Autenticidad del Asesor

Yo, MORE VALENCIA RUBEN ALEXANDER, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - PIURA, asesor de Tesis titulada: "Modelo YOLO (You Only Look Once) para el diagnóstico preliminar de la enfermedad del Huanglongbing en la planta de limón en la Región Piura, 2023", cuyos autores son CARREÑO CHERO IRVIND DARIO, OBREGON MAURICIO ALEJANDRO JUNIOR, constato que la investigación tiene un índice de similitud de 8.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

PIURA, 08 de Diciembre del 2023

Apellidos y Nombres del Asesor:	Firma
MORE VALENCIA RUBEN ALEXANDER DNI: 02897931 ORCID: 0000-0002-7496-3702	Firmado electrónicamente por: RMOREV el 08-12- 2023 23:55:51

Código documento Trilce: TRI - 0689247