



Implementación de un modelo computacional predictivo de cosecha con herramientas tecnológicas avanzadas para la producción sostenible del cultivo de cacao

Mariana Sofía Candamil Cortés

Tesis de maestría presentada para optar al título de Magíster en Bioinformática y Biología Computacional

Tutor: Jorge William Arboleda Valencia, Doctor (PhD) en Ciencias Biológicas: Biología Molecular

Cotutora: Anyela Valentina Camargo Rodríguez, Doctora (PhD) en Biología Computacional

Asesores de recursos académicos: Harold Estiven García Álvarez (asesor bibliográfico), Claudia Marcela Cerón Rubio (asesora Centro de Escritura) y Elvia Lucía Sánchez García (asesora de integridad académica)

Universidad de Manizales
Facultad de Ciencias e Ingeniería
Maestría en Bioinformática y Biología Computacional
Manizales, Caldas, Colombia

2025

Cita	(Candamil-Cortés et al., 2025)
Referencia	Candamil-Cortés, M. S., Arboleda-Valencia J. W., Camargo-Rodríguez A. V., (2025). <i>Implementación de un modelo computacional predictivo de cosecha con herramientas tecnológicas avanzadas para la producción sostenible del cultivo de cacao</i> [Tesis de maestría]. Universidad de Manizales. RIDUM: Repositorio Institucional Universidad de Manizales.
Estilo APA 7 (2020)	



Maestría en Bioinformática y Biología Computacional, III

Centro de Investigaciones en Medio Ambiente y Desarrollo

Línea de Investigación Bioinformática.

Centro de Investigaciones en Medio Ambiente y Desarrollo - CIMAD.

Declaración de inteligencia artificial: el o los autores de este trabajo de grado declaran que han utilizado herramientas de inteligencia artificial (IA), tales como ChatGPT, Grammarly y Turnitin, de manera ética y responsable, tal como se establece en el Acuerdo UManizales 002 (julio 26 de 2023) sobre propiedad intelectual e IA. Estas herramientas son empleadas como apoyo en la redacción, revisión gramatical y generación de ideas, pero en ningún caso sustituyen el análisis crítico, la argumentación académica ni la originalidad del trabajo. Asimismo, cualquier contenido generado con asistencia de IA está citado y referenciado adecuadamente, garantizando la integridad académica y el cumplimiento de los principios éticos de la investigación.

Biblioteca y Centro de Recursos: <https://biblioteca.umanizales.edu.co/>

Repositorio Institucional: <http://ridum.umanizales.edu.co/>

Universidad de Manizales: www.umanizales.edu.co

Revistas: <http://revistasum.umanizales.edu.co/>

Fondo Editorial: <https://editorialum.umanizales.edu.co/>

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Manizales ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Dedicatoria

A Dios, por ser mi guía y fortaleza. Gracias por darme la fuerza para superar los obstáculos y por iluminar mi camino en cada paso de este proceso.

A mi abuela Adiela, quien ha sido una inspiración constante en mi vida. Su ejemplo de perseverancia y fuerza ha sido el motor que me ha impulsado a seguir adelante. Esta tesis es un pequeño tributo a su amor y dedicación.

A mis padres, Luz Adriana Cortés y Jhon Jairo Candamil. A Laura, Luis Miguel, Gilma, Socorro, Claudia y Roberto, por su apoyo constante, comprensión y sacrificio. Gracias por creer en mí, por estar siempre a mi lado, y por ser mi mayor motivación en cada momento de este recorrido. Esta dedicación es también para ustedes, que me han dado todo para seguir adelante.

Agradecimientos

Instituciones:

- A la Universidad de Manizales por proporcionar los recursos necesarios para el desarrollo de la tesis. Esta investigación fue apoyada por “Biofábricas: una oportunidad de desarrollo bioeconómico para Caldas a través de la biotecnología” código 1235-903-86957, financiado por el Ministerio de Ciencia y Tecnología - Minciencias
- Al Centro de Bioinformática y Biología Computacional de Colombia - BIOS por poner a disposición los recursos computacionales y su talento humano.
- A la Fundación Luker, quienes me apoyaron con una beca para el estudio de la maestría y por el apoyo económico para realizar una pasantía internacional en la Universidade Católica Dom Bosco, Brasil.

Doctores e investigadores:

- Al Dr. Jorge William Arboleda por su orientación, paciencia y apoyo incondicional durante todo el proceso de investigación. Su experiencia y conocimientos fueron fundamentales para el desarrollo de este trabajo.
- A la Dra. Anyela Camargo por su orientación y guía en la adquisición, entrenamiento y validación del modelo desarrollado. Por su gran apoyo, paciencia y los valiosos comentarios para el desarrollo del trabajo.
- A la Dra. Sara Piedrahita por sus aportes en la construcción del documento y sistematización de la información.

Agradecimientos personales:

- A Dios, amigos y familia.
- A mis padres, Luz Adriana Cortés y Jhon Jairo Candamil, y a mi abuela Adiela, gracias por su amor, apoyo constante y por brindarme la fuerza necesaria para seguir adelante, incluso en los momentos de dificultad. Su confianza en mí ha sido una fuente de motivación inquebrantable.
- A Laura, Luis Miguel, Gilma, Socorro, Claudia y Roberto, gracias por su apoyo incondicional siempre.
- A mis compañeros de la maestría, por siempre tener una actitud positiva y por el excelente trabajo en equipo.
- A Don Eduardo y Doña Francy por su trato amable en cada visita al cultivo.
- A todas las personas que, de una u otra manera me apoyaron para concluir este proceso. Muchas gracias.

Tabla de contenido

Resumen	10
Abstract	12
1. Introducción	14
1.1. Campo temático.....	15
2. Planteamiento del problema.....	17
3. Justificación	20
4. Objetivos	23
4.1. Objetivo general	23
4.2. Objetivos específicos.....	23
5. Referente teórico y Antecedentes	24
5.1. Generalidades	24
5.2. Cadena Productiva de Cacao.....	25
5.2.1. Bioeconomía.....	27
5.2.2. Agricultura sostenible	27
5.3. Agricultura 4.0	28
5.3.1. Inteligencia Artificial	28
5.3.1.1. <i>Aprendizaje Automático (ML)</i>	29
5.3.1.2. <i>Aprendizaje Profundo (DL)</i>	29
6. Materiales y métodos	32
6.1. Especificaciones de Hardware.....	32
6.2. Diseño Metodológico	33
6.3. Diseño Experimental	33
6.3.1. Objetivo Específico 1	33
6.3.2. Objetivo Específico 2	34

6.3.3. Objetivo Específico 3	36
7. Materiales y métodos	38
7.1. Resultados y discusión	38
7.2. Productos	59
8. Conclusiones	60
9. Impactos y perspectivas	61
10. Recomendaciones	63
11. Referencias	64
12. Anexos	72

Lista de tablas

Tabla 1	Parámetros que se ajustan en el modelo computacional	35
----------------	--	----

Lista de figuras

Figura 1 Producción y área cosechada de granos de cacao en el mundo (1994-2022).	24
Figura 2 10 países con mayor producción de cacao (1994-2022).	25
Figura 3 Cadena productiva del Cacao.	26
Figura 4 Esquema de la metodología de DL y ML.	30
Figura 5 Prestaciones del Hardware usadas para el entrenamiento del modelo.	33
Figura 6 Metodología para el cumplimiento del objetivo específico 1.	34
Figura 7 Metodología para el cumplimiento del objetivo específico 2.	36
Figura 8 Metodología para el cumplimiento del objetivo específico 3.	37
Figura 9 Ubicación de la finca La Esmeralda.	38
Figura 10 Finca La Esmeralda.	39
Figura 11 Etiquetado de 30 árboles en la finca La Esmeralda.	39
Figura 12 Registro fotográfico de los árboles seleccionados.	40
Figura 13 A. Desarrollo de aplicación en Kobotoolbox para la adquisición de datos. B. Estructura de la información recopilada en la aplicación	41
Figura 14 Ejemplo de seguimiento a árbol de interés.	41
Figura 15 Uso de herramienta CVAT para el etiquetado manual de los objetos.	42
Figura 16 Etiquetas de los objetos identificados en el árbol de cacao. Etiquetas de A.Flor, B.Progreso, C.Maduración, D.Aborto.	43
Figura 17 Apoyo de productores en el etiquetado de los datos.	43
Figura 18 Ejemplo de la anotación requerida para el entrenamiento del modelo.	44
Figura 19 Conteo de las clases de interés.	45
Figura 20 Resultados del modelo en términos de segmentación y clasificación de las imágenes.	46
Figura 21 Validación del modelo de clasificación.	47

Figura 22 Rendimiento del modelo de clasificación de la cosecha de cacao.....	48
Figura 23 Matriz de confusión del modelo con mejor rendimiento.	49
Figura 24 Curva precisión-confianza para el modelo con el mejor rendimiento.	49
Figura 25 Curva precision-recall con el modelo que presenta mayor rendimiento.	50
Figura 26 Confiabilidad en datos de prueba para cada una de las clases.....	51
Figura 27 Gráfica de Precisión vs Confiabilidad del modelo.	52
Figura 28 Rendimiento del modelo de clasificación de la cosecha de cacao (3 clases)	53
Figura 29 Matriz de confusión del modelo con mejor rendimiento (3 clases).....	53
Figura 30 Curva precisión-confianza para el modelo con el mejor rendimiento (3 clases).....	54
Figura 31 Curva precision-recall con el modelo que presenta mayor rendimiento (3 clases)	55
Figura 32 Confianza en datos de prueba para cada una de las clases (3 clases)	55
Figura 33 Aplicación móvil para supervisar los frutos de cacao en los árboles.	57

Resumen

El incremento de la población y el impacto del cambio climático han ocasionado una fuerte presión sobre los sistemas productivos agrícolas, en búsqueda de la identificación de soluciones innovadoras para generar productos y servicios que suplan las demandas de la sociedad. Específicamente, en Colombia, la agricultura es una de las principales actividades económicas, por lo cual, es necesario contar con estrategias enfocadas en cultivos sostenibles, adaptación climática y agricultura de precisión.

En Colombia, uno de los cultivos de mayor relevancia es el cacao. Alrededor de 65 mil familias (Ministerio de Agricultura y Desarrollo Rural, 2021a) se benefician de este producto. Particularmente, en Caldas se priorizó este cultivo mediante el Plan Departamental de Extensión Agropecuaria (PDEA), en el cual se evidencia el reto relacionado con la poca adopción de paquetes tecnológicos, prácticas culturales, procesos organizativos y sistemas eficientes de comercialización y uso de las Tecnologías de Información y Comunicaciones (TICs), con el objetivo de fortalecer las capacidades técnicas del cultivo. Debido a esto, es importante establecer estrategias que impacten positivamente la cadena productiva. Es así como se propone la implementación de un modelo computacional de la cosecha del cultivo de cacao, en donde se identifiquen patrones de comportamiento en datos de imágenes del cultivo y se puedan emplear para predecir el tiempo óptimo de su cosecha.

La metodología implementada en este estudio se basa en una combinación de tecnologías avanzadas y colaboración con expertos del sector agrícola. Se generó un repositorio de imágenes en campo que documenta de manera longitudinal el crecimiento y desarrollo de los frutos de cacao, considerando las etapas de flor, progreso, cosecha y aborto, con lo cual se crea una base de datos organizada para el análisis posterior. Con esto, se realizó un proceso de etiquetado teniendo en cuenta la asesoría de expertos en el cultivo, lo que garantizó una correcta delimitación de las diferentes categorías de frutos. A partir de esta base de datos, se realizaron experimentos utilizando modelos computacionales para clasificar las imágenes en diferentes etapas de maduración. Los resultados mostraron que, al reducir el número de clases a tres, progreso, cosecha y aborto, se logró una mejora significativa en la precisión del modelo, alcanzando un 71%, frente al 63% obtenido

con la clasificación de cuatro clases. Adicionalmente, se pudo determinar que la función de optimización Adam, ofreció un rendimiento óptimo, lo que permitió mejorar la eficiencia computacional y la precisión de las predicciones.

Estas tecnologías y herramientas bioinformáticas, al estar enfocadas en el uso de ciencia datos, específicamente para información del cultivo de cacao, contribuyen en gran medida a la implementación de alternativas de solución para la mejora del rendimiento de cultivos y aumento de producción, lo cual a su vez genera información de interés que permite el aprovechamiento de productos derivados, el conocimiento y conservación de la biodiversidad, siempre con el objetivo de crear estrategias que permitan mitigar el cambio climático.

Palabras clave: Bioeconomía, modelo computacional, producción sostenible, *Theobroma cacao*, visión por computador

Abstract

Population growth and the impact of climate change have caused strong pressure on agricultural production systems, in search of innovative solutions to generate products and services that meet the demands of society. Specifically, in Colombia, agriculture is one of the main economic activities; therefore, it is necessary to have strategies focused on sustainable crops, climate adaptation and precision agriculture.

In Colombia, one of the most important crops is cocoa; around 65 thousand families (Ministry of Agriculture and Rural Development, 2021) benefit from this product. Particularly in Caldas, this crop was prioritized through the Departmental Agricultural Extension Plan (PDEA), which shows the challenge related to the low adoption of technological packages, cultural practices, organizational processes and efficient marketing systems and the use of ICTs. This plan was defined with the objective of strengthening the technical capabilities of the crop, through the establishment of strategies that positively impact the production chain. Thus, the implementation of a computational model of the cocoa crop harvest is proposed, in which behavioral patterns are identified, both agroclimatic and image data of the crop, and can be used to predict the optimal harvesting time.

The methodology implemented in this study is based on a combination of advanced technologies and collaboration with experts from the agricultural sector. A repository of field images was generated that longitudinally documents the growth and development of cocoa fruits, considering the stages of flower, progress, harvest and abortion, thus creating an organized database for subsequent analysis. A labeling process was carried out considering the advice of experts in the crop, which ensured a correct delimitation of the different categories of fruits. From this database, experiments were carried out using computational models to classify the images at different stages of ripening. The results showed that by reducing the number of classes to three, progress, harvest and abortion, a significant improvement in the accuracy of the model was achieved, reaching 71%, compared to 63% obtained with the four-class classification. In addition, it was determined that the Adam optimization function offered optimal performance, which improved computational efficiency and prediction accuracy.

These bioinformatics technologies and tools, focused on the use of data science, specifically for cocoa crop information, contribute greatly to the implementation of alternative solutions to improve crop yields and increase production, which in turn generates information of interest that allows the use of derived products, knowledge and conservation of biodiversity, always with the aim of creating strategies to mitigate climate change.

Keywords: Bioeconomy, computational model, sustainable production, *Theobroma cacao*, computer vision.

1. Introducción

La producción de cacao es una actividad económica importante a nivel mundial. En el periodo comprendido entre 2019 y 2022, la producción mundial aumentó de 4,7 millones de toneladas (CAF, 2019) a 5,8 millones de toneladas (FAO, 2022), con un área de cosecha de 11.939.436 hectáreas. Dicha producción proviene en su mayoría de Costa de Marfil, Ghana, Indonesia y Nigeria (FAO, 2022). En particular, Colombia ocupa el décimo puesto como productor de cacao en el mundo y el quinto lugar en Latinoamérica, registrando en el 2022 una producción de 51.827 toneladas (FAO, 2022). Aquellas empresas que son procesadoras de cacao en el país han aportado \$782.000 millones al PIB, exportando chocolates, cocoa, licor de cacao, entre otros, a 74 países (ANDI, 2019).

El sector cacaotero representa una oportunidad para más de 38 mil familias en Colombia y gracias a iniciativas nacionales e internacionales como *cacao for peace* propuesta por la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID por sus siglas en inglés), es posible generar desarrollo rural en las comunidades productoras de cacao, a través de proyectos investigativos, educación y acceso a mercados internacionales. Dichas propuestas están orientadas en mejorar el rendimiento de los cultivos y como consecuencia aumentar el ingreso de los agricultores (USAID, 2021). Todo esto ha permitido que durante los últimos cuatro años el área sembrada en cultivos de cacao en Colombia presentara un crecimiento promedio del 3% anual, destinando en el 2020 un total de 188.000 hectáreas para este proceso productivo (Ministerio de Agricultura y Desarrollo Rural, 2021b).

El cacao colombiano es reconocido por la Organización Internacional del Cacao por su finura tanto en sabor como en aroma (Proexport Colombia, 2012), además, considerando las condiciones climáticas y geográficas únicas del país, los granos de cacao colombianos tienen un perfil de sabor complejo y equilibrado (Quintana Fuentes et al., 2018). Estas características únicas son resultado de una combinación de factores, como el clima, la variedad de cacao, las técnicas de cultivo y los procesos de fermentación utilizados por los agricultores colombianos (PROAMAZONIA, 2003).

Sin embargo, la producción de cacao, al ser un cultivo tropical, se ve afectada por diversos factores, incluyendo condiciones climáticas adversas, enfermedades y prácticas agrícolas inadecuadas (CNCH, 2019), motivo por el cuál, se debe considerar el tiempo óptimo para realizar la cosecha del fruto. En este contexto, la agricultura 4.0 ofrece grandes oportunidades para la producción de cacao. Con el uso de tecnologías como la inteligencia artificial, el internet de las cosas y la automatización, los agricultores pueden mejorar la calidad, eficiencia y el rendimiento del cultivo, reduciendo el impacto ambiental y optimizando sus procesos de producción (Ojeda Beltran, 2022).

Una de las estrategias que se utilizan en la agricultura 4.0 es el aprendizaje automático (ML por sus siglas en inglés), esta técnica ha sido ampliamente utilizada para resolver diversos problemas genómicos y de evolución de sistemas biológicos (Larrañaga et al., 2006). Éste se centra en la aplicación de algoritmos computacionales basados en la experiencia previa y como resultado se obtiene una inferencia estadística de los datos ingresados (Mjolsness & DeCoste, 2001), esto permite analizar datos y aprender de ellos, con el objetivo de mejorar la precisión de las predicciones y la capacidad de generalización de la información.

Considerando todo lo anterior, en el presente proyecto se busca implementar un modelo computacional predictivo de cosecha mediante el uso de herramientas tecnológicas avanzadas, orientado a la producción sostenible del cultivo de cacao, una herramienta valiosa cacao para tomar decisiones informadas sobre la gestión de los cultivos y mejorar la productividad y la rentabilidad de los productores de cacao. Este estudio promueve la adopción de tecnologías accesibles y de bajo costo, lo que posibilita su implementación en diversas escalas de producción, beneficiando tanto a pequeños productores como a grandes empresas del sector cacaotero. Es así cómo se genera un impacto técnico, económico y social de las comunidades agrícolas, promoviendo prácticas responsables frente a los retos del cambio climático y las demandas del mercado global.

1.1. Campo temático

La implementación de un modelo computacional predictivo de cosecha de cacao a través de la integración de imágenes es una temática de gran relevancia en la actualidad. En primer lugar,

se debe destacar que el cultivo de cacao es un pilar importante en la economía de muchos países y, por tanto, los avances en la optimización de su producción son de gran interés para el desarrollo económico y social de las comunidades vinculadas a esta actividad.

Además, la implementación de un modelo computacional predictivo de cosecha de cacao también tiene implicaciones ambientales positivas, ya que puede contribuir a una gestión más sostenible y responsable de los recursos naturales, como el agua y los suelos. Así mismo, desde un punto de vista académico, el uso de tecnologías avanzadas en el estudio de la cosecha de cacao también representa un avance significativo en la investigación interdisciplinaria, que es cada vez más necesaria para abordar los desafíos complejos que enfrenta la sociedad en la actualidad. La combinación de diferentes disciplinas, como la biología molecular, la estadística y la informática, puede proporcionar una perspectiva más amplia y profunda de la problemática que se aborda, lo que se traduce en soluciones más innovadoras y efectivas.

El estudio forma parte del Centro de Investigaciones en Medio Ambiente y Desarrollo (CIMAD) de la Universidad de Manizales. Entre los objetivos del grupo se destacan: ampliar el conocimiento sobre el desarrollo económico sostenible, comunicar los resultados a la comunidad académica, fomentar la formación de investigadores de maestría y doctorado, y participar en redes nacionales e internacionales sobre medio ambiente y desarrollo sostenible. De igual forma, la presente tesis está vinculada al grupo de investigación BIOS, perteneciente al Centro de Bioinformática y Biología Computacional de Colombia – BIOS, el cual busca promover el uso de la Infraestructura de Computación de Alto Rendimiento para la optimización de investigaciones en diferentes campos, incluyendo la bioinformática y la biología computacional.

Cabe mencionar que, la presente tesis se realizó en el marco del programa BIOFÁBRICAS: una oportunidad de desarrollo bioeconómico para Caldas a través de la biotecnología - No. 80740-503-2021, financiado por Minciencias bajo la convocatoria 903: Convocatoria para el apoyo a programas y proyectos de I+D+i que contribuyan a resolver los desafíos establecidos en la misión “Bioeconomía para una Colombia potencia viva y diversa hacia una sociedad impulsada por el conocimiento”.

2. Planteamiento del problema

Uno de los cultivos de mayor relevancia en el país, es el cacao (*Theobroma cacao* L.), un cultivo tropical producido en América desde hace miles de años, que representa un sector importante para la economía campesina e agroindustrial con gran impacto en la tradición, cultura y progreso de las regiones. De acuerdo con la Organización Internacional del Cacao (ICCO), en el año 2022 se produjeron 5,8 millones de toneladas de cacao seco en el mundo; lo que demuestra su importancia no solo a nivel local sino también a nivel internacional con gran relevancia para múltiples países, tanto por la venta del producto primario como por la generación de productos derivados de alta calidad, resultado de la sofisticación y mejoramiento de la cadena, con altos precios. Son diferentes factores los que intervienen en la cadena productiva, como la madurez del fruto de cacao en el momento de su recolección (Bueno et al., 2020), que permite alcanzar mejores precios cuando se hace en los tiempos óptimos gracias a la calidad del producto agrícola. Tradicionalmente, considerando las características principales del fruto, éste es ovoide y puede apreciarse su estado inmaduro con los colores (verde y rojo) y textura (lisa o rugosa), dependiendo de su genotipo (Arrazate CH et al., 2018). Dicha información es aprovechada por los productores para identificar el tiempo óptimo de cosecha del fruto.

En el año 2016, se implementó una iniciativa denominada Cacao for peace, a cargo de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID) y el Departamento de Agricultura de Estados Unidos (USDA) cuyo objetivo principal era mejorar la cadena de valor del cacao en Colombia, fortaleciendo los sectores público y privado para aprovechar el rápido aumento de la demanda mundial de este producto, presentando una oportunidad en el contexto del posconflicto e impactando a más de 38.000 familias de agricultores colombianos (Abbott et al., 2017).

El cultivo de cacao presenta varias etapas para su comercialización: en primer lugar, se encuentra la siembra, en la cual se realiza una selección de suelo y de semillas aptas para a productividad, después la cosecha, que se realiza de manera empírica aproximadamente 180 días después de la floración, y finalmente se tiene la fase de la postcosecha, en la cual se incluyen los procesos de fermentación, secado, limpieza, almacenamiento selección y clasificación. Los

inconvenientes en la producción de cacao incluyen esquemas no óptimos de siembra, variabilidad en los precios y gran dependencia de insumos (Espinosa, 2016), los cuales generan impactos negativos en las condiciones agronómicas (Hernández-Núñez et al., 2020).

A pesar de que el cacao colombiano, presenta un alto potencial para competir de manera masiva a nivel mundial por su sabor fino, la carencia de expertos y de tecnologías emergentes enfocadas en su oportuna producción, hace que el proceso resulte poco óptimo y, por ende, reduce el rendimiento de los cultivos (Romero et al., 2022). Actualmente, la mayoría de los procesos en este campo continúan siendo de manera empírica, es decir, en la fase de cosecha, los cacaocultores utilizan parámetros morfológicos como el color, la textura y el tamaño para definir si el fruto ya se puede recoger, lo cual no es eficiente, pues se presentan grandes variedades que tienen un cambio en su tonalidad dependiendo de su estado de madurez, además, los aspectos como la luz y la textura, generan que esta técnica sea subjetiva y condicionada a la hora y el ángulo en el que observan el fruto. Así mismo, los agricultores utilizan tradicionalmente los sonidos que se forman al golpear los frutos de cacao para diferenciar sus niveles de madurez, para esto, utilizan las uñas, los nudillos, el extremo redondeado de un cuchillo, o el propio cuchillo (Bueno et al., 2020). Estas técnicas, en múltiples ocasiones, implica la obtención de frutos inmaduros o que sobrepasan la madurez, lo cual afecta la calidad del fruto, que es medida, teniendo en cuenta estándares mínimos de aceptación (Heredia-Gómez et al., 2020).

Según Yen & Hong (2016), la cosecha inoportuna del cacao (falta de fermentación de calidad) genera una disminución de antioxidantes, pues se minimizan las características organolépticas, que influyen directamente en el sabor, color y olor. De igual forma, existe un aumento en la presencia de oxalatos solubles, los cuales pueden tener repercusiones negativas en el organismo, debido a que contribuyen en la generación de cálculos renales. Además, si se espera demasiado tiempo y el grano comienza a germinar, existe una pérdida total del producto.

Adicionalmente, el rendimiento de cultivos también está afectado por el cambio climático, pues dicha variabilidad propicia un ambiente ideal para la propagación de fitopatógenos y plagas. De hecho, se ha registrado una pérdida de hasta el 50% de la producción atribuida a fitopatógenos por mal proceso en la producción a la hora de realizar el control fitosanitario (McMahon et al.,

2015). Actualmente, existen diversos modelos basados en el aprendizaje de máquina para realizar la predicción de la fecha de cultivo, sin embargo, estos no presentan una validación e impacto significativo en el trabajo de campo con los cacaocultores. Por lo anterior, es indispensable implementar herramientas de fácil uso, portables y de fácil conectividad para efectuar una cosecha oportuna, en el tiempo que requiere para su madurez y considerando factores como exposición a la luz solar, humedad, altura, entre otros factores en los que se encuentra expuesto el cultivo.

Considerando todo lo anterior, se ha propuesto la siguiente pregunta de investigación, ¿Cómo se puede desarrollar un modelo computacional predictivo que permita mejorar la cosecha de cacao a partir de imágenes del cultivo?

3. Justificación

La bioeconomía es utilizada para crear una ruta de crecimiento, en la cual se consideran los resultados investigativos sobre las oportunidades de la biodiversidad. Esta consiste en un modelo económico, en donde la producción de bienes y servicios está enfocada en el uso sostenible de los recursos biológicos. Es así como la bioeconomía busca una respuesta a los retos globales teniendo en cuenta la biodiversidad, no sólo pensando en su conservación, sino también para su aprovechamiento productivo sostenible, enfocado en un desarrollo económico sustentado en conocimiento generado desde la diáspora de Ciencia, Tecnología e Innovación.

La economía colombiana se basa principalmente en la agricultura. Específicamente, Colombia es el décimo productor mundial de cacao, con una producción registrada de 51.827 toneladas en 2022 (FAO, 2022), donde su producción contribuye de manera significativa a la sostenibilidad económica de aproximadamente 65.341 agricultores (Ministerio de Agricultura y Desarrollo Rural, 2021a). La producción de grano fermentado y seco consta de las siguientes etapas: cultivo y cosecha, desgranado, fermentación, secado y almacenamiento; una vez que los granos han sido fermentados y secos, pueden procesarse para producir una variedad de productos como manteca, licor y cacao en polvo, así como chocolate (ICCO, 2016). Los inconvenientes en la producción de cacao incluyen estrategias no óptimas en la siembra y cosecha, además de la fluctuación en los precios (Espinosa, 2016), teniendo una repercusión negativa en la cadena productiva del fruto.

En particular, para realizar el proceso de cosecha, los cacaocultores, se basan en su mayoría, en las características morfológicas de la mazorca del cacao, lo cual no resulta eficiente y objetivo al momento de realizar la extracción del fruto. De ahí, la importancia de generar estrategias basadas en las tecnologías emergentes, que permitan automatizar este proceso, enfocado en la agricultura de precisión (Pierce & Nowak, 1999). Lo anterior, considerando que se pueden realizar estudios de bioprospección, que consiste en la identificación de potenciales beneficios de los metabolitos detectados en las especies del país. Además, se realizan análisis sobre su potencial aplicación en diversas industrias, como la cosmética, nutracéutica y farmacéutica (Fujimura et al., 2011). Cabe destacar que este estudio, al estar enfocado en una cosecha oportuna, también se obtienen los

derivados de este con mejores propiedades, lo cual genera una ventaja para el desarrollo de bioproductos, es así como se puede garantizar un mejor aprovechamiento de su biomasa.

La Corporación Colombiana de Investigación Agropecuaria - AGROSAVIA (CORPOICA, 2000), reporta que cada tonelada de semilla seca equivale a aproximadamente 10 toneladas de cáscara del cacao; otros autores reportan que únicamente es aprovechado entre el 10% y 21% de la mazorca para grano fermentado y seco y que el porcentaje restante (entre 90% y 79%) corresponde a cáscara de cacao, la cual queda en el cultivo sin ningún aprovechamiento (Encalada & Jácome, 2018)(Torres Cabezas, 2019) (Murillo & Méndez, 2019). En el país son escasas las iniciativas enfocadas a biopreparados de cáscaras de cacao (i.e. mezclas de cáscaras con otros residuos de las fincas), o no han sido reportadas en la literatura, siendo necesario darle valor agregado a un producto derivado en diversos mercados, apreciando el potencial de aprovechamiento biotecnológico de una biomasa residual.

Hay muchos ejemplos de valorización de los subproductos de los cultivos, en concreto para el cacao, se presenta la obtención de energía, fertilizantes, alimentos funcionales, cosmética y en la industria farmacéutica (Meza-Sepúlveda et al., 2021). La cosecha de calidad y la identificación de vías de valorización de los residuos de cáscara de cacao puede aumentar los ingresos de los agricultores, debido a la mejora del precio de venta del fruto con mayor calidad, al igual que proporcionar una solución medioambiental para el tratamiento de estos residuos, mediante una economía circular.

El entendimiento y uso responsable de la biodiversidad es uno de los retos más importantes que enfrenta actualmente Colombia. Es así, como se han desarrollado múltiples estrategias a partir de políticas públicas, con el objetivo de aumentar la competitividad mediante el uso de los recursos naturales. A nivel internacional, se presenta el Objetivo de Desarrollo Sostenible (ODS) 2, el ODS 8 y el ODS 12 los cuales, corresponden a hambre cero, a trabajo decente y crecimiento económico y a producción y consumos responsables respectivamente, al tener un país megadiverso como Colombia, se puede enfocar la variedad de recursos biológicos en el marco de estos objetivos.

Entre las políticas a nivel nacional, se presenta el CONPES 3934 de 2018, que busca “impulsar a 2030 el aumento de la productividad y la competitividad económica del país, al tiempo que se asegura el uso sostenible del capital natural y la inclusión social, de manera compatible con el clima” (DNP, 2018), pues en el presente trabajo se enfoca en identificar estrategias que permitan el rendimiento de la producción y el rendimiento. Así mismo, el país aspira a posicionarse en el biocomercio para el año 2025, lo cual resulta ser la línea base para la implementación de herramientas para el uso de la biotecnología en la agricultura, en el ambiente y en la salud (Rodríguez Leon, 2017). De igual forma, la presente propuesta se encuentra en el marco del CONPES 4069, considerando que este busca incrementar la contribución de la Ciencia, Tecnología e Innovación al desarrollo social, económico, ambiental, y sostenible, del país con un enfoque diferencial, territorial, y participativo, para contribuir a lograr los cambios culturales que promuevan una sociedad del conocimiento (CONPES, 2021).

De la misma manera, dentro de los enfoques que propone la Misión de Sabios de 2019, se identifican: Biotecnología, medio ambiente, bioeconomía e industria 4.0, las cuales son áreas y temáticas claves y relevantes para establecer la ruta para el avance de la Ciencia, la Tecnología y la Innovación como base del desarrollo y la competitividad del país (Gobierno de Colombia, 2019). Finalmente, el presente estudio considera la Política Nacional para la Transformación Digital e Inteligencia Artificial, establecida en el CONPES 3975 la cual define que la IA es un “campo de la informática dedicado a resolver problemas cognitivos comúnmente asociados con la inteligencia humana o seres inteligentes, entendidos como aquellos que pueden adaptarse a situaciones cambiantes. Su base es el desarrollo de sistemas informáticos, disponibilidad de datos y los algoritmos” y a su vez, busca crear condiciones habilitantes para disminuir las barreras que existen en el país al momento de adoptar e incorporar estas tecnologías en el sector productivo.

4. Objetivos

4.1. Objetivo general

Implementar un modelo computacional predictivo de cosecha mediante el uso de herramientas tecnológicas avanzadas, orientado a la producción sostenible del cultivo de cacao.

4.2. Objetivos específicos

1. Diseñar un sistema de seguimiento de los árboles de cacao desde la flor hasta el fruto, garantizando la trazabilidad de la maduración.
2. Desarrollar un modelo de predicción de cosecha del cultivo de cacao sustentado en tecnologías avanzadas y emergentes basado en datos de imágenes.
3. Validar el modelo computacional mediante datos de imágenes del cultivo de cacao.

5. Referente teórico y Antecedentes

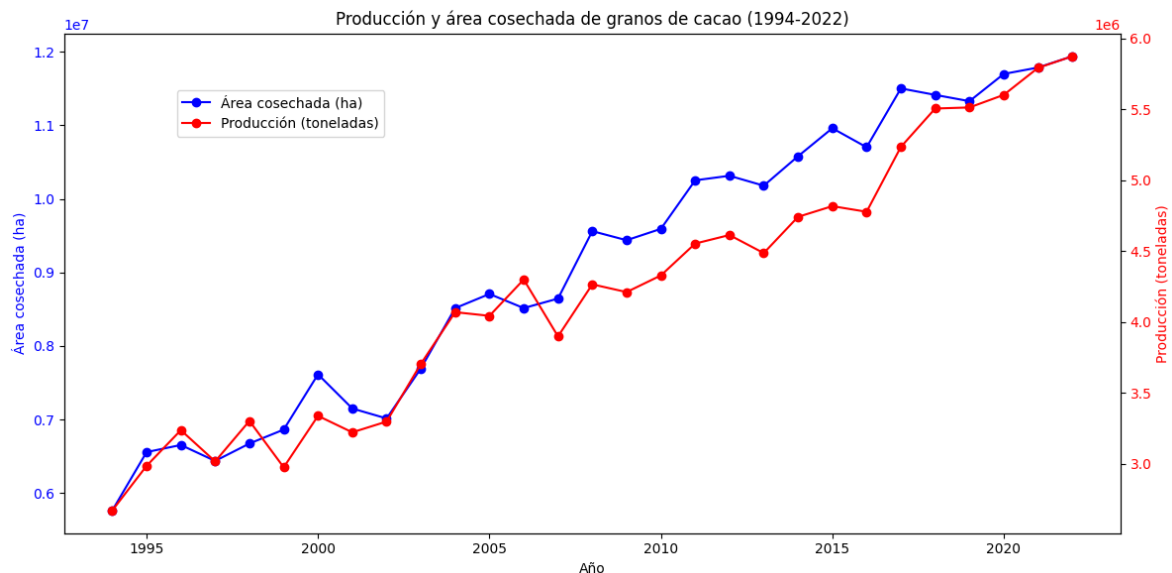
5.1. Generalidades

Theobroma cacao L. es un árbol económicamente importante a nivel mundial. Tiene una altura de 4 a 8 m, de la familia Malvaceae y es originario de la región tropical de América. Los granos de cacao presentan gran cantidad de polifenoles, los cuales corresponden aproximadamente al 10% del peso seco del grano entero, y uno de sus derivados, el chocolate negro, se considera uno de los mayores contribuyentes de antioxidantes a la dieta estadounidense después de las frutas y las verduras (Rusconi & Conti, 2010).

Es importante destacar que, la producción de cacao representa una actividad económica de gran relevancia a nivel global. En la figura 1 se presentan los datos de la producción mundial de los granos de cacao, así como el área cosechada en el periodo comprendido entre 1994 y 2022, en la cual se identifica una tendencia en aumento, registrando en el año 2022 5.874.581,94 toneladas, abarcando una superficie de cultivo de 11.939.436 hectáreas (FAO, 2022).

Figura 1

Producción y área cosechada de granos de cacao en el mundo (1994-2022).

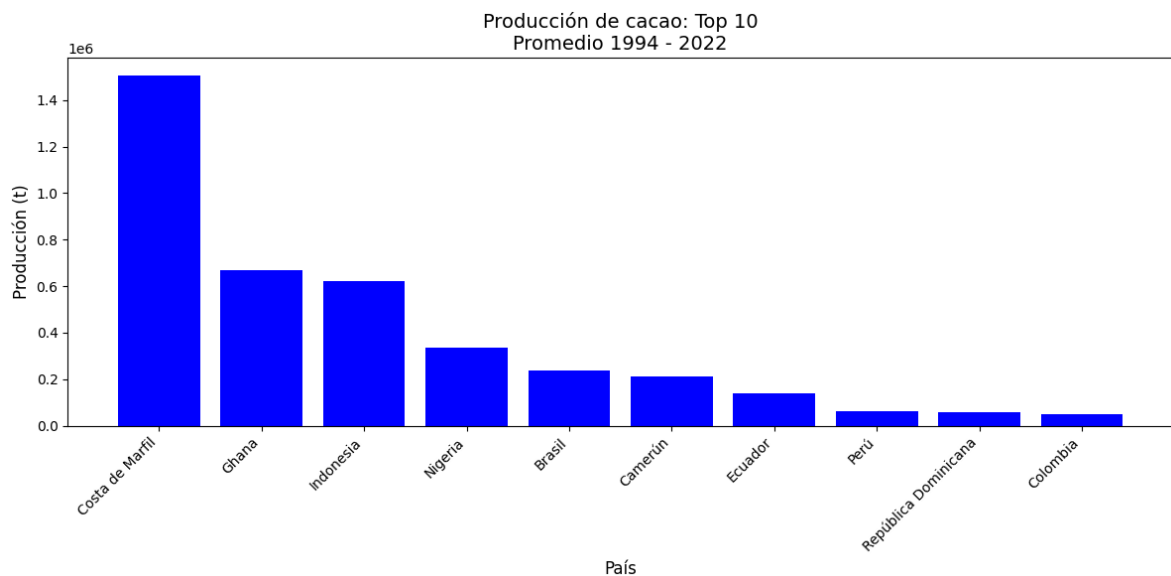


Nota. Adaptado de (FAO, 2022)

Específicamente, los principales países productores incluyen Costa de Marfil, Ghana, Indonesia, Nigeria y Brasil (FAO, 2022). En la figura 2, se aprecian los datos del promedio de producción entre los años 1994 y 2022 de los 10 países con mayor promedio a nivel mundial. En este contexto, Colombia se posiciona como el décimo mayor productor de cacao a nivel mundial y el quinto en Latinoamérica, con una producción registrada de 51.827 toneladas en 2022 (FAO, 2022).

Figura 2

10 países con mayor producción de cacao (1994-2022).



Nota. Adaptado de (FAO,2022)

5.2.Cadena Productiva de Cacao

Una cadena productiva, también conocida como cadena de suministro, es el conjunto de actividades que se llevan a cabo desde la producción de materias primas hasta la entrega del producto final al consumidor. En otras palabras, es el proceso completo que involucra la transformación de materias primas en productos terminados. Esta incluye diferentes etapas como la extracción de materias primas, el transporte, la transformación, el almacenamiento, la distribución y la comercialización de los productos. Cada una de estas etapas es fundamental para garantizar la calidad del producto final y la satisfacción del cliente (Cayeros et al., 2016).

Particularmente, para el cultivo de cacao, las etapas que conforman la cadena productiva son: producción de semillas de cacao, siembra y cultivo, cosecha y poscosecha, procesamiento y comercialización. En la figura 3, se puede identificar una ilustración de este proceso (Cristancho et al., 2021).

Figura 3
Cadena productiva del Cacao.



Nota. Recuperado de: (Cristancho et al., 2021)

En Colombia, según Fedecacao, existen tres tipos de cacao: criollo, forastero y el híbrido o trinitario (Superintendencia de industria y comercio, 2011). El cacao criollo es considerado la variedad más fina y delicada. Se cultiva principalmente en América Latina, por otro lado, el cacao forastero es la variedad más común y la que se utiliza con mayor frecuencia en la elaboración de chocolate, finalmente, el cacao trinitario (o híbrido) es una variedad que surge del cruce entre el criollo y el forastero, esta variedad de cacao tiene un sabor equilibrado entre el criollo y el forastero (Superintendencia de industria y comercio, 2011). Dependiendo de la variedad podrá ser aprovechado para aplicaciones diferentes.

Actualmente, Caldas es uno de los principales departamentos productores de cacao a nivel nacional en Colombia. Según cifras del DANE, en Caldas se registró un 3,55% de las exportaciones de cacao en grano (Vásquez-Barajas et al., 2018). Este departamento cuenta con una larga tradición en la producción de cacao, y ha venido desarrollando iniciativas y proyectos para mejorar la calidad y la productividad de su producción cacaotera. Además, el cacao de Caldas es reconocido por su sabor y calidad, lo que le ha permitido posicionarse en el mercado nacional e internacional de chocolates finos y otros productos derivados del cacao (Osorio Arias, 2022).

Cabe destacar que en este departamento existen múltiples cadenas productivas priorizadas Plan Departamental de Extensión Agropecuaria Departamento de Caldas 2020-2023; entre estas se presenta el café, caña panelera, hortofrutícola, plátano, aguacate, cítricos, cacao, caucho, guadua, forestal, agroganadera, piscicultura y especies menores (Gobernación de Caldas, 2020). Adicionalmente, se identifica la importancia de realizar estudios enfocados en las anteriores cadenas productivas, debido a que se les ha dado una alta importancia y se han establecido como áreas principales de enfoque para el desarrollo de la ciencia, la tecnología y la innovación en el sector agropecuario del país.

5.2.1. Bioeconomía

Esta se define como la producción, utilización y conservación de recursos biológicos, incluyendo conocimientos de ciencia, tecnología e innovación, para proporcionar información, productos, procesos y servicios en todos los sectores económicos, con el propósito de avanzar hacia una economía sostenible (Gobierno de Colombia, 2020).

5.2.2. Agricultura sostenible

Es un enfoque de producción agrícola que busca maximizar la productividad de la tierra y los recursos naturales. La agricultura sostenible implica prácticas de cultivo y manejo de ganado que son respetuosas con el medio ambiente, socialmente justas y económicamente viables. Algunas prácticas comunes en la agricultura sostenible incluyen el uso de técnicas de cultivo que conservan

el suelo, la reducción del uso de productos químicos sintéticos y pesticidas, la gestión del agua y la energía, la conservación de la biodiversidad y la promoción de la salud y el bienestar de los animales de granja (Condiza, 1998).

El objetivo de la agricultura sostenible es proporcionar una fuente de alimentos segura y nutritiva, así como mejorar la calidad de vida de las comunidades rurales, minimizando el impacto ambiental y respetando los derechos y necesidades de los trabajadores agrícolas y las comunidades locales (Condiza, 1998).

5.3. Agricultura 4.0

En la agricultura, se ha empezado a incorporar tecnologías emergentes para la generación de agricultura de precisión, estas se definen como las herramientas que se encuentran en proceso de consolidación, novedosas y de reciente aparición (Bosch, 2013).

5.3.1. *Inteligencia Artificial*

La inteligencia artificial (IA) es una rama de la informática que se enfoca en el desarrollo de sistemas o programas informáticos capaces de realizar tareas o actividades, tratando de emular las decisiones del ser humano (Rouhiainen, 2018). La IA se basa en algoritmos y modelos matemáticos que imitan el comportamiento y la toma de decisiones de los seres humanos. Dentro de esta técnica se presentan enfoques como el aprendizaje automático, el procesamiento del lenguaje natural, la visión artificial, la robótica y la automatización de procesos, entre otros (Incio Flores et al., 2021).

Según Expósito Gallardo & Ávila Ávila (2008), la inteligencia artificial tiene una amplia variedad de aplicaciones en áreas como la medicina, la industria, la logística, la seguridad, el entretenimiento, la educación y muchos otros campos. Específicamente, existen múltiples investigaciones centradas en la agricultura, particularmente, las investigaciones actuales se han enfocado en el uso de tecnologías emergentes, como la IA para realizar predicciones sobre el tiempo de cosecha o para la detección temprana de patologías del cultivo.

5.3.1.1. Aprendizaje Automático (ML)

El ML es utilizado para resolver varios problemas basado en la experiencia previa, con esto se pueden abordar enfoques de clasificación y predicción para generar conclusiones estadísticas sobre los datos (Mjolsness & DeCoste, 2001). Su función principal es calibrar los parámetros para optimizar el rendimiento sobre los datos de entrenamiento y los datos de entrada posteriores (Zou et al., 2019). En particular, el uso de modelos de ML suele requerir experiencia en la codificación para el procesamiento de datos, el entrenamiento, la validación y la prueba del conjunto de datos. Además, existen técnicas como tuneo de hiperparámetros, que permiten establecer los mejores valores que requiere el modelo para generar un ajuste correcto a los datos (Li et al., 2020).

Actualmente en el estudio realizado por Romero et al. (2022), se presenta un modelo computacional, basado técnicas de ML, específicamente en regresión logística con datos de variables agroclimáticas para determinar la fecha exacta de cosecha. Para esto recopilaron datos de variables ambientales, de suelo y de los cultivos de cacao mediante técnicas de muestreo sistemático. Posteriormente, se utilizó un enfoque de sistemas dinámicos para construir el modelo de cultivo, esto se basa en una serie de ecuaciones matemáticas que describen las interacciones entre las variables ambientales, de suelo y de los cultivos de cacao, consiguiendo un ajuste con un RRMSE del 7.2% para predicciones de rendimiento. Esos resultados muestran que las condiciones climáticas necesarias para la maduración de los frutos varían significativamente entre regiones, identificando que la práctica tradicional de cosechar a los 180 días después de la floración no siempre es adecuada, de ahí la importancia de evaluar nuevos modelos que permitan identificar los tiempos óptimos de cosecha y que sean la base para la toma de decisiones de los productores.

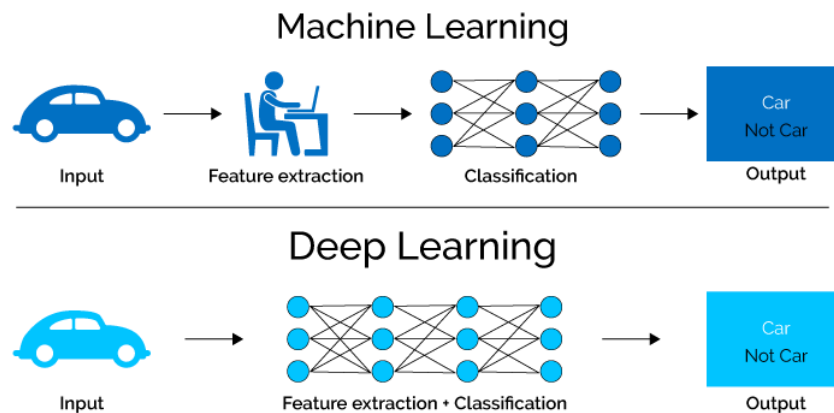
5.3.1.2. Aprendizaje Profundo (DL)

Es una rama de la IA que se enfoca en entrenar redes neuronales artificiales de múltiples capas para aprender representaciones jerárquicas de los datos. Posee múltiples aplicaciones como el reconocimiento de objetos, el procesamiento del lenguaje natural, la detección de fraudes, entre otras (Lecun et al., 2015). Para poder utilizar los algoritmos de aprendizaje profundo no es necesario contar con las características de los datos, pues él es capaz de extraer las características

principales que le permite realizar la clasificación de los datos. Esta es la principal diferencia con modelos de aprendizaje automático tradicionales, pues estos requieren una base de datos con características diseñadas manualmente (Robinson et al., 2020). Esto le permite al DL generar modelos más precisos y útiles para una amplia gama de tareas. En la figura 4, se puede apreciar la ilustración de un algoritmo basado en técnicas de DL y su comparación al solo utilizar las técnicas de ML convencional.

Figura 4

Esquema de la metodología de DL y ML.



Nota. Recuperado de: (Semiconductor, 2018)

Existen muchos parámetros como el número de iteraciones, la tasa de aprendizaje y la regularización que representan un gran reto para encontrar los valores óptimos que generen modelos complejos que se ajusten adecuadamente al comportamiento de los datos (Qi et al., 2021).

Basado en esta metodología existen varios modelos orientados en la cosecha del cultivo de cacao. Concretamente, en el proyecto liderado por Heredia-Gómez et al. (2020), se presenta un modelo automatizado, mediante el desarrollo de una herramienta portable, basada en una red neuronal convolucional (CNN), una de las arquitecturas utilizadas en el aprendizaje de máquina, para indicar si una mazorca de cacao se encuentra en el momento oportuno para ser cosechada. Como resultado se realizó un sistema embebido con una precisión de 34.83% mAP, demostrando el reto para realizar este tipo de predicciones. Este porcentaje sirve como una referencia para

futuros estudios, en los que se avance en aumento de información original, balanceo de datos y mejora en el entrenamiento de la red.

De igual forma, en investigaciones desarrolladas por Bueno et al. (2020) se presenta una arquitectura, basada en CNN, en la cual se extraen características reconocibles para el proceso de entrenamiento y obtuvieron una precisión de 97,13% para la clasificación binaria según su madurez. Sin embargo, se aprecia que los datos iniciales, eran frecuencias de sonido, no tienen en cuenta las condiciones climáticas en las que se encuentran los cultivos, además, la captura de dicha información puede resultar más compleja para el agricultor.

Dentro de este enfoque, se han identificado diversos modelos que han sido ampliamente utilizados en datos de imágenes y cuyo funcionamiento es posible extrapolar al cultivo de cacao. El modelo de Mask R-CNN, desarrollado por He et al. (2020), tiene como objetivo la segmentación y detección de objetos. Este puede ser extrapolado a los datos que se tienen en el momento, con el propósito de ver el rendimiento preliminar del modelo computacional.

De igual manera, se identifica el modelo *You Only Look Once* (YOLO) (Diwan et al., 2023), el cual ha sido ampliamente empleado en la identificación de objetos, realizando un proceso de detección en tiempo real mediante la conversión en problemas de regresión. A lo largo del tiempo se ha ido mejorando dicho modelo, generando nuevas versiones que aumenten la precisión y confiabilidad (Terven et al., 2023). Específicamente se ha usado en campos, como la agricultura, en donde se presentan varias investigaciones relevantes, como la realizada por (Lippi et al., 2021), en la cual se genera un sistema de detección de plagas (insectos) para avellanos basado en datos de imágenes, con el cual se obtuvo un rendimiento en términos de promedio de precisión de 94.5%. De igual forma, en la investigación propuesta por (Yue et al., 2023), identificaron cómo YOLO puede ser extrapolado para la detección de enfermedades en el cultivo del tomate, presentando un F1-score de 88.7%.

Adicionalmente, se presentan los modelos basados en *Vision Transformer*, los cuales aplican la arquitectura de *Transformers* en tareas de visión por computadora, como clasificación de imágenes, detección de objetos o segmentación semántica (Mia et al., 2023). Al igual que el

modelo de YOLO, también ha sido usado en el sector agrícola. En la investigación realizada por (Bandi et al., 2023), se utiliza *vision transformer* para detectar la enfermedad en las hojas de la planta, mediante una clasificación realizada sobre la base de la gravedad de la infección de la hoja, con el cual se obtuvo un F1-score de 0,758. Así mismo, en el estudio desarrollado por (Maillet et al., 2023), realiza una aproximación para la detección de patógenos fúngicos en cultivos de uva, a partir del uso de *Vision Transformer* (ViT) para procesar imágenes y un modelo de fusión multimodal que integra datos climáticos y visuales en un único modelo. Dicho sistema logra un porcentaje de precisión del 97% en la detección de patologías, identificando el potencial uso de estas herramientas en campo.

El diseño e implementación de un modelo computacional implica varios retos, en primer lugar, se requiere de espacio para almacenar y procesar los datos a utilizar para el entrenamiento, validación y prueba del modelo. Así mismo, es necesario resaltar el uso de una Unidad de Procesamiento Gráfico (GPU) para aquellos modelos que requieran técnicas más robustas de aprendizaje de máquina. Adicionalmente, si se requiere mejorar la eficiencia de entrenamiento y obtención de los datos, se debe usar una capacidad de procesamiento y cómputo de alto desempeño y rendimiento (HPC).

6. Materiales y métodos

6.1. Especificaciones de Hardware

Para los experimentos de Deep Learning se utilizó la plataforma de Google Collaboratory (Bisong, 2019), esta posee una GPU NVIDIA A100-SXM4-40GB con 40,514 MiB de memoria proporciona el poder computacional necesario para manejar grandes volúmenes de datos, así mismo se utilizó Python 3.10.12, un lenguaje versátil y ampliamente soportado en el ecosistema de ciencia de datos, y PyTorch 2.5.1+cu121, una librería de aprendizaje profundo optimizada para GPU. En la figura 5, se aprecia con mayor detalle estas prestaciones del Hardware.

Figura 5

Prestaciones del Hardware usadas para el entrenamiento del modelo.

NVIDIA-SMI 535.104.05				Driver Version: 535.104.05		CUDA Version: 12.2	
GPU	Name	Persistence-M	Bus-Id	Disp.A	Volatile	Uncorr. ECC	
Fan	Temp	Perf	Pwr:Usage/Cap	Memory-Usage	GPU-Util	Compute M.	MIG M.
0	NVIDIA A100-SXM4-40GB	Off	00000000:00:04.0	Off	0	0	0
N/A	31C	P0	43W / 400W	2MiB / 40960MiB	0%	Default	Disabled

Processes:							
GPU	GI	CI	PID	Type	Process name	GPU Memory	
ID	ID	ID				Usage	
No running processes found							

6.2. Diseño Metodológico

La presente investigación cuenta con un enfoque mixto, debido a que se realizó una revisión documental por medio de una selección de fuentes de información, y un análisis cuantitativo, optimizando el porcentaje de predicción de un modelo computacional para la cosecha de cacao; de igual forma se trabajó con imágenes, es decir, computacionalmente se representan con matrices numéricas, además cuando se presenten los resultados, se realizará una interpretación cualitativa del estado en el que se encuentra en cultivo (Floración, desarrollo, maduración o aborto); el alcance del estudio es descriptivo y predictivo, pues se realizará el modelo computacional y se generará una sugerencia a partir de la información utilizada para entrenar el modelo. El tipo de la investigación es experimental, pues antes del desarrollo del modelo se recolectarán las imágenes del cultivo a partir de una aproximación longitudinal del dato.

6.3. Diseño Experimental

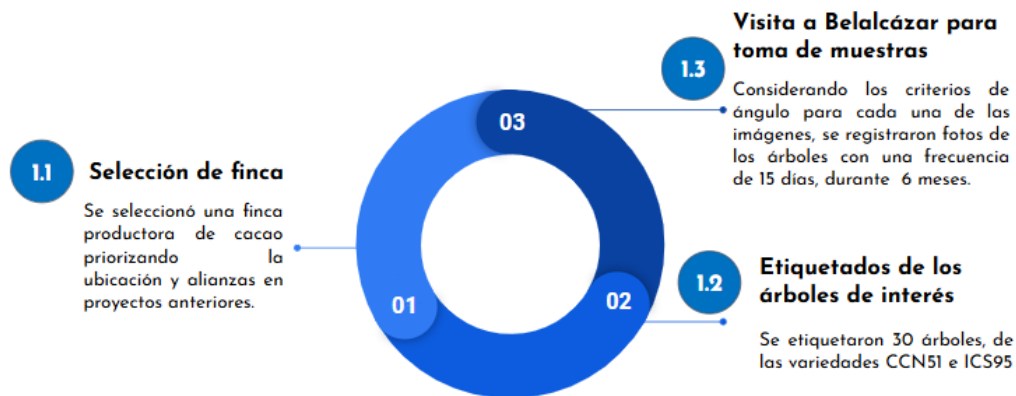
6.3.1. Objetivo Específico 1

OE1 (Objetivo Específico 1): Diseñar un sistema de seguimiento de los árboles de cacao desde la flor hasta el fruto, garantizando la trazabilidad de la maduración

Para dar cumplimiento a este objetivo, en primera instancia se realizó la selección de la finca, mediante un enfoque estratégico, considerando los criterios de: ubicación geográfica, consulta con expertos de la región, análisis de accesibilidad y logística y priorización de alianzas previas, considerando el proyecto Kocolatl, desarrollado en el periodo 2020-2021. Posteriormente, se realizó el etiquetado de 30 árboles en la finca, correspondientes a las variedades **CCN51** e **ICS95**, considerando la productividad y el perfil organoléptico del grano, para este etiquetado, se utilizó un número consecutivo dentro de su grupo varietal, teniendo en cuenta dos parcelas de la finca. Finalmente, se realizó el seguimiento a cada uno de los árboles, tomando un registro fotográfico con una periodicidad de 15 días, durante 6 meses, lo cual garantizó el monitoreo de los frutos desde la etapa de floración a la etapa de cosecha, para esto se utilizó un celular Xiaomi Redmi Note 8 y se garantizó que cada una de las imágenes se tomaran en el mismo ángulo de registro. En la figura 6, se presenta un diagrama que resume las actividades realizadas.

Figura 6

Metodología para el cumplimiento del objetivo específico 1.



6.3.2. Objetivo Específico 2

OE2: Desarrollar un modelo de predicción de cosecha del cultivo de cacao sustentado en tecnologías avanzadas y emergentes basado en imágenes.

Después de tener la estructuración del repositorio, se procede con el etiquetado y organización manual del conjunto de datos de imágenes. Las etiquetas de interés corresponden a imágenes de flor, progreso (desde ~1 mes hasta ~5 meses de crecimiento), madurez y aborto. Para ello, se utilizó la plataforma CVAT (Krenzer et al., 2022) (<https://www.cvat.ai/>).

Finalmente, se procedió con la realización de experimentos computacionales, con el objetivo de apoyar la expansión y validación de modelo productivo de cacao. Para esto, se instaló la librería de Ultralytics 8.3.49, la cual es una versión robusta de la biblioteca YOLO, que permite la detección de los objetos. Y se realizaron ajustes en el modelo, cambiando los parámetros del optimizador, épocas, dropout y la tasa de aprendizaje, en la tabla 1, se presenta la descripción de cada uno de los parámetros y el rango de valores que se cambiaron.

Tabla 1

Parámetros que se ajustan en el modelo computacional

Parámetro	Descripción	Valores
optimizer	Optimizador del modelo.	‘Adam’
	Afecta a la velocidad de convergencia y a la estabilidad	‘AdamW’
epochs	Número total de épocas de entrenamiento. Cada	60
	época corresponde a un paso completo por el conjunto de datos.	100
dropout	Tasa de abandono para la	0.0
	regularización en tareas de	0.1
	clasificación, con este se evita el sobreajuste del modelo.	0.2

lr0	Tasa de aprendizaje inicial.	0,01
	Este valor influye en la rapidez de actualización de las ponderaciones del modelo.	0,001

En la figura 7, se identifica el flujo de trabajo establecido para el segundo objetivo.

Figura 7

Metodología para el cumplimiento del objetivo específico 2.



6.3.3. Objetivo Específico 3

OE3: Validar el modelo computacional mediante datos de imágenes del cultivo de cacao.

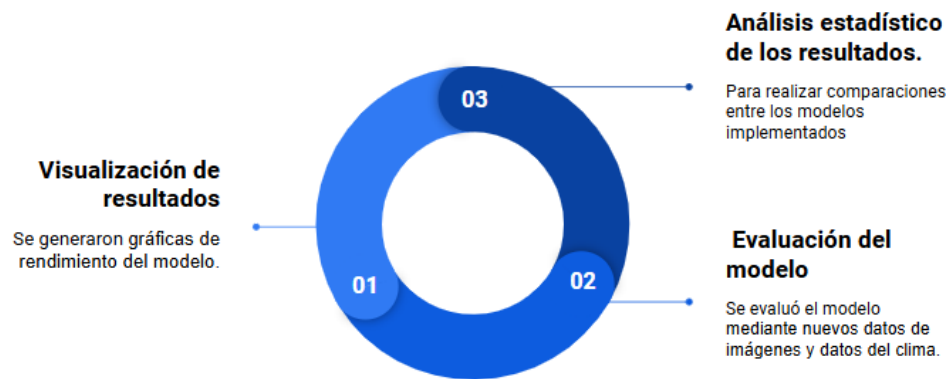
En primer lugar, se generaron las gráficas de rendimiento del modelo, que incluye la matriz de confusión y métricas relacionadas con el porcentaje de precisión del modelo, gráficas de pérdida y curva precisión-confianza. Posteriormente, se realizó una evaluación del modelo con nuevos datos,

con el objetivo de validar los resultados obtenidos con el modelo, generando de igual forma un análisis estadístico con la comparación de los resultados obtenidos.

En la figura 8, se presenta la gráfica con las actividades generales que se desarrollaron.

Figura 8

Metodología para el cumplimiento del objetivo específico 3.



Con el objetivo de generar una aplicabilidad del modelo desarrollado, se implementó una herramienta que permite apoyar la toma de decisiones de los productores. Particularmente el proceso de análisis y diseño de la estrategia de software para la aplicación destinada a los cacaocultores está basado en la identificación de requisitos, la definición de una arquitectura escalable y el uso de tecnologías adecuadas para su implementación. Así mismo, se considera un enfoque modular, teniendo en cuenta una arquitectura basada en microservicios, con un backend en Python y una base de datos optimizada para gestionar grandes volúmenes de información. La metodología de desarrollo ágil permite iteraciones rápidas y adaptabilidad a mejoras, mientras que la integración del modelo de inteligencia artificial asegura un análisis eficiente de los datos.

7. Materiales y métodos

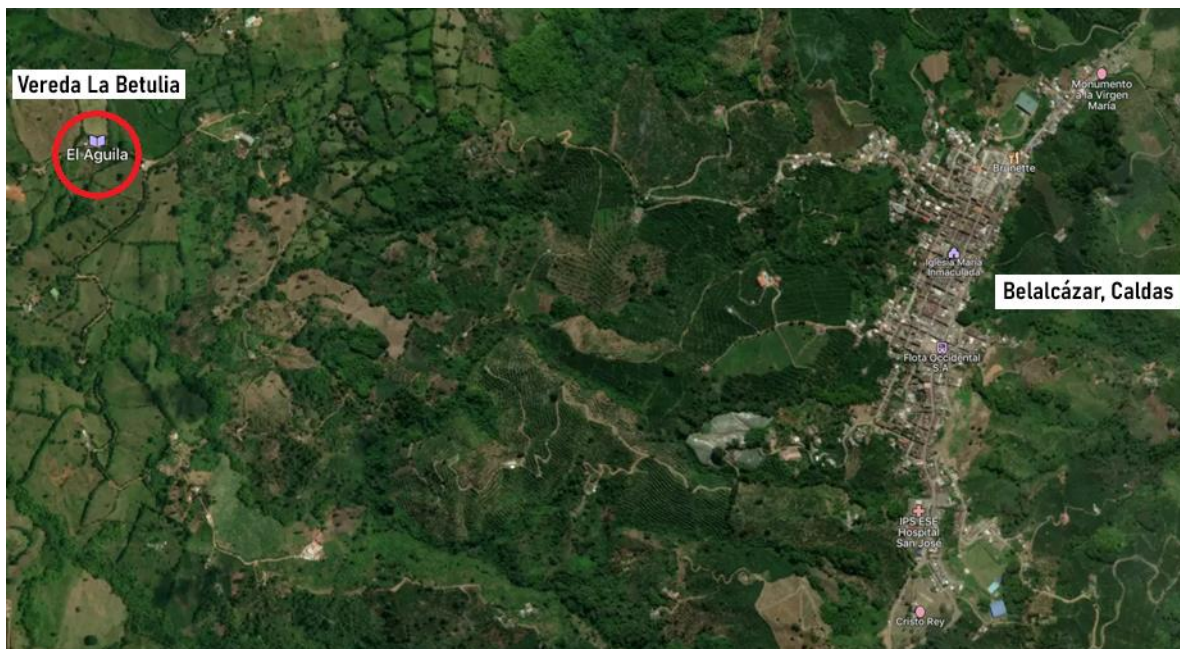
7.1. Resultados y discusión

Selección de la finca.

Se seleccionó una finca productora de cacao priorizando la ubicación y alianzas en proyectos anteriores. Particularmente, la asociación Asoprob el apoyó en la identificación de la finca productora de cacao fino en aroma y sabor. En la figura 9, se aprecia la ubicación de la finca, la cual se localiza en la Vereda La Betulia, Caldas.

Figura 9

Ubicación de la finca La Esmeralda.



Finca: La Esmeralda, localizada en:

Latitud: 4° 59' 52" norte

Longitud: 75° 49' 55" oeste

La zona presenta una temperatura que oscila entre 18°C y 27°C, por ende, es un territorio en donde además del cultivo de cacao, se presenta plátano y café. Particularmente, para el cultivo

de cacao, cuenta con 1.700 árboles, en donde se encuentran variedades como ICS95 y CCN51. En la figura 10, se puede apreciar una imagen de la finca La Esmeralda, en donde se realizó el proceso de recolección de información.

Figura 10
Finca La Esmeralda.



Etiquetado de árboles.

Una vez se contó con el permiso de los productores de la finca, se etiquetaron los árboles de interés, en la figura 11, se presenta una imagen como ejemplo de las etiquetas utilizadas.

Figura 11
Etiquetado de 30 árboles en la finca La Esmeralda.



Visita a Belalcázar para toma de muestras

Se realizaron las visitas a Belalcázar para la toma del registro fotográfico, en esta etapa es crucial mantener el ángulo de captura de la imagen, con el objetivo de establecer el seguimiento a los árboles seleccionados. En la figura 12, se aprecia el proceso llevado a cabo en dicho registro, en donde se realiza un ajuste de la altura y el ángulo de la cámara, y se estableció un protocolo de repetibilidad para cada visita, considerando las imágenes tomadas con anterioridad, con el objetivo de que las imágenes capturadas a lo largo del tiempo pudieran compararse de manera precisa.

Figura 12

Registro fotográfico de los árboles seleccionados.



Con el objetivo de mantener una estructuración de la información, se utilizó la plataforma Kobotoolbox, la cual permite ingresar la fecha, la finca, el identificador del árbol y el estado de la mayoría de los frutos que se registran. Es importante destacar que, además de las imágenes registradas en el periodo del 5 de enero del 2024 y el 19 de julio de 2024, se cuenta con imágenes adicionales que se registraron previamente en la finca. En la figura 13, se puede identificar (13A) la plataforma utilizada y (13B) la organización de esta información.

Figura 13

A. Desarrollo de aplicación en Kobotoolbox para la adquisición de datos. B. Estructura de la información recopilada en la aplicación

A



B

Validation	start	end	Fecha	Finca	ID Arbol	Estado	Foto	_version_
1 - 30	720 results							
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	31	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	19	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	48	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	44	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	2	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	43	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	29	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	38	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	3	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	1	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	24	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	11	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	23	Desarrollo de fruto	vjRMojW8aM6b
☐	—	Feb 23, 202...	Feb 23, 202...	2024-02-19	La Esmeralda	30	Desarrollo de fruto	vjRMojW8aM6b

El repositorio generado incluye la base de datos creada para este proyecto, la cual contiene 684 imágenes que documentan las diferentes etapas del desarrollo del fruto, proporcionando un registro visual exhaustivo que sirve de apoyo a los procesos de formación y validación del modelo. En la figura 14, se presenta un ejemplo de estructura de información, que permite identificar la trazabilidad del fruto a lo largo del tiempo, particularmente, es posible establecer la trazabilidad desde la etapa de floración, hasta la última etapa, correspondiente a la cosecha del fruto.

Figura 14

Ejemplo de seguimiento a árbol de interés.



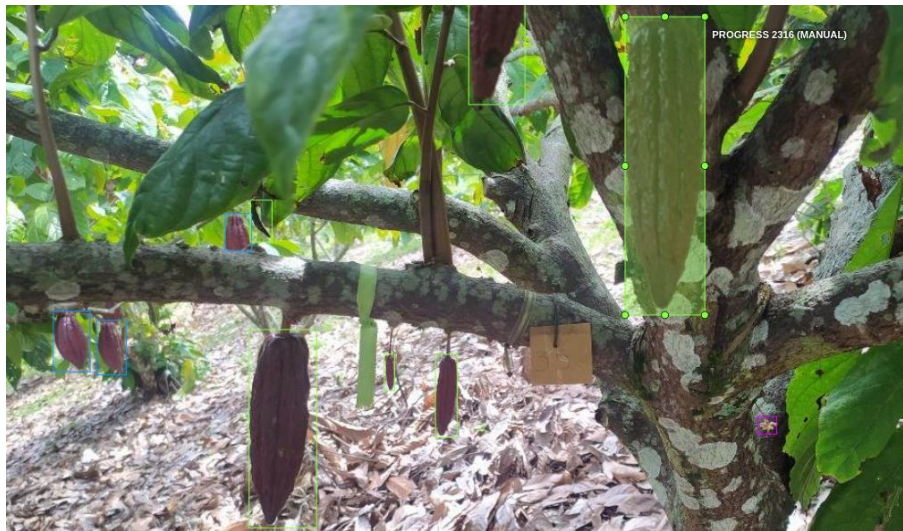
El desarrollo de una base de datos longitudinal con imágenes del cultivo de cacao demostró ser un recurso relevante para el desarrollo de modelos predictivos enfocados en el análisis del crecimiento y desarrollo de los frutos, de ahí la importancia de establecer una sistematización en el registro fotográfico para garantizar que los datos recopilados sean consistentes, ordenados y útiles para los análisis posteriores. Este diseño apoya la identificación de patrones de desarrollo en el cultivo, y constituye un papel fundamental para la implementación de estrategias de agricultura de precisión, mejorando la toma de decisiones en el manejo del cultivo de cacao.

Procesamiento y etiquetado de los datos.

Antes de realizar los experimentos computacionales, se delimitaron manualmente las imágenes con los objetos que se identificarán. En la figura 15, se puede identificar el resultado del proceso llevado a cabo con la plataforma CVAT (Krenzer et al., 2022).

Figura 15

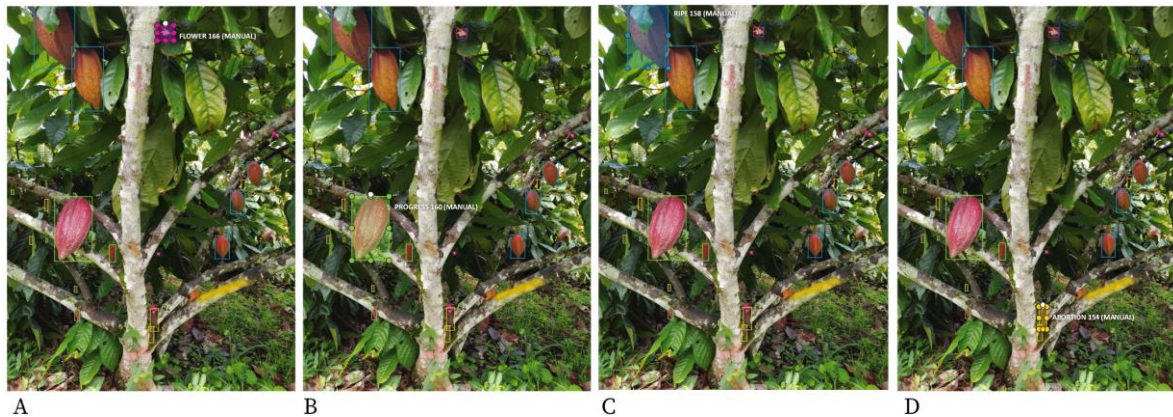
Uso de herramienta CVAT para el etiquetado manual de los objetos.



Particularmente, en el proceso se anotaron los objetos de: flor, progreso (correspondiente a estado intermedio de desarrollo del fruto), cosecha (que corresponde a la maduración del fruto) y aborto, cuando el fruto no logran alcanzar la madurez. En la figura 16, se puede identificar dicha anotación para cada clase.

Figura 16

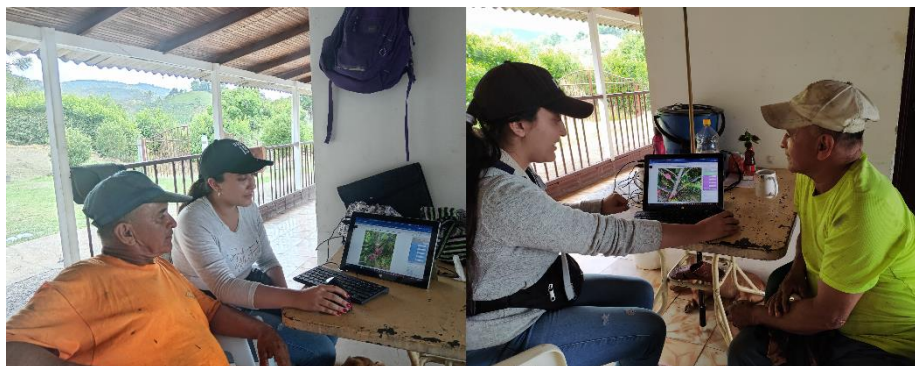
Etiquetas de los objetos identificados en el árbol de cacao. Etiquetas de A.Flor, B.Progreso, C.Maduración, D.Aborto.



Cabe destacar que las etiquetas anteriores, se realizaron con el apoyo de los agricultores, en la figura 17 se aprecia el proceso llevado a cabo con su asesoramiento.

Figura 17

Apoyo de productores en el etiquetado de los datos.



La realización de un proceso correcto de etiquetado, en conjunto con los productores expertos, es fundamental para garantizar la precisión y uso de los modelos predictivos. Tener en el estudio la asesoría de especialistas permite delimitar correctamente los frutos, reduciendo la confusión en las categorías y minimizando los errores en el proceso de anotación. Este enfoque permite la mejora del desempeño del modelo al proporcionar información confiable y representativa, y a su vez, establece una base sólida para investigaciones futuras.

Como resultado de este proceso se presenta un archivo en formato .txt, que contiene las anotaciones de cada una de las imágenes. En la figura 18, se presenta un ejemplo de esta anotación, específicamente, la primera columna corresponde al índice de la clase (0:ripe, 1:progress, 2:flower y 3:abortion). Posteriormente, se encuentran las coordenadas normalizadas con las posiciones de cada objeto en la imagen, en la segunda y tercera columna se encuentra la posición del centro del cuadro delimitador (x_center, y_center), la cuarta columna corresponde al ancho del cuadro delimitador (width) y la última columna a la altura del cuadro delimitador (height).

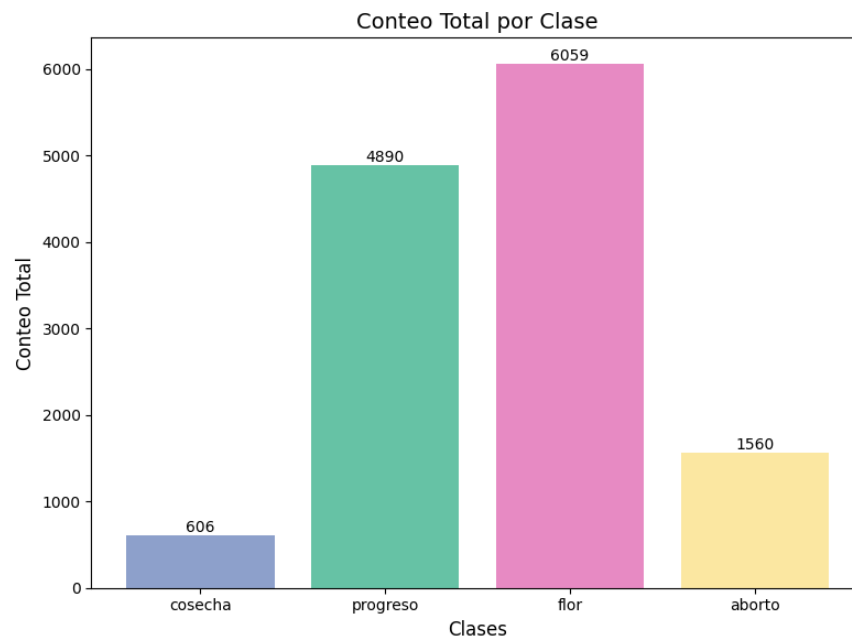
Cabe destacar, que los valores de las coordenadas están normalizadas respecto al tamaño de la imagen, es por esto que únicamente, se aprecian valores en un rango entre 0 y 1.

Figura 18

Ejemplo de la anotación requerida para el entrenamiento del modelo.

```
3 0.561725 0.335530 0.007100 0.019807
1 0.794700 0.119825 0.274700 0.183250
2 0.182154 0.650517 0.014712 0.012100
3 0.229141 0.392085 0.028888 0.127097
2 0.293625 0.576958 0.010050 0.009883
3 0.552460 0.339348 0.009730 0.024770
2 0.623406 0.397703 0.003838 0.006060
3 0.372300 0.610983 0.013750 0.036500
2 0.256024 0.325062 0.007593 0.015143
0 0.917767 0.357833 0.122465 0.331067
1 0.113500 0.595050 0.015700 0.035233
1 0.282129 0.581400 0.008177 0.022467
1 0.390810 0.602128 0.024530 0.060343
1 0.437846 0.534917 0.018508 0.022233
1 0.081032 0.423095 0.007985 0.021543
2 0.182402 0.700123 0.003905 0.006953
2 0.293758 0.556133 0.002450 0.003067
```

Después de realizar la delimitación, se realizó el conteo para cada una de las clases analizadas, la cual se puede apreciar en la figura 19.

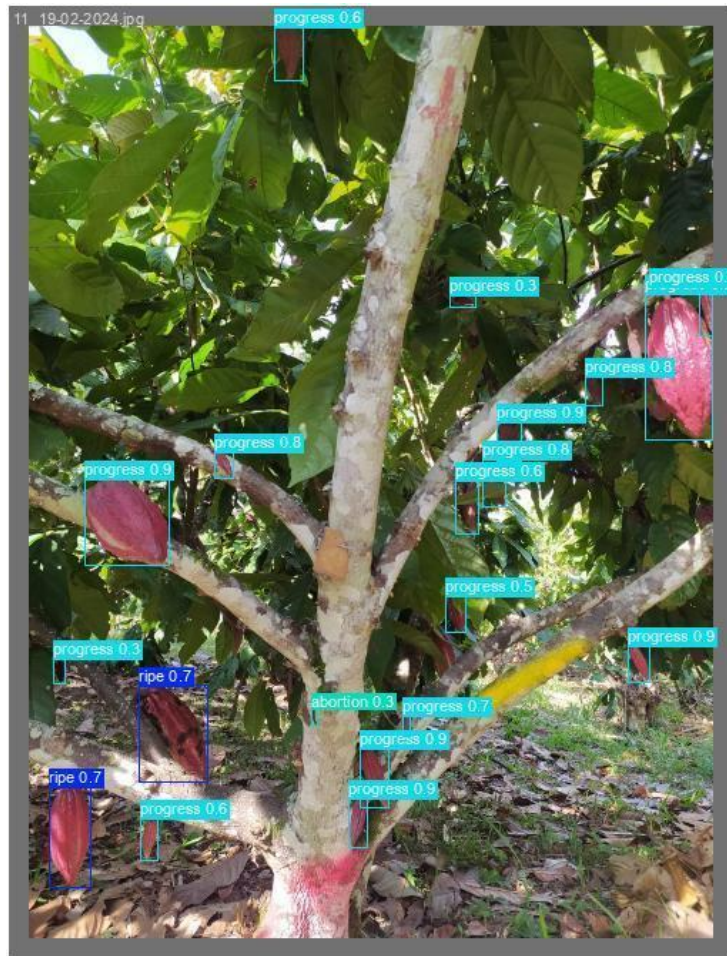
Figura 19*Conteo de las clases de interés*

Con lo anterior, es posible afirmar que se presenta un problema de desbalance de clases, pues la mayoría de la información proviene de las flores en el árbol. Con estos datos, se procede a entrenar el modelo YOLO versión 8.

La figura 20 muestra los resultados de rendimiento del modelo en términos de segmentación y clasificación de las imágenes considerando las cuatro etapas de desarrollo. Aquí se visualiza cómo el modelo identifica y clasifica cada etapa. Para cada uno de los frutos que identifica, se tiene la clase a la que pertenece (*ripe*, *progress*, *flower* o *abortion*) y a la vez se cuenta con la confiabilidad que el modelo genera para cada una de las muestras.

Figura 20

Resultados del modelo en términos de segmentación y clasificación de las imágenes.



Asimismo, en la figura 21, se puede observar el rendimiento y la validación del modelo implementado. En esta figura se destacan los indicadores clave que confirman la precisión y eficacia del modelo, apoyando su aplicabilidad para el seguimiento y análisis del cacao

Figura 21

Validación del modelo de clasificación.



Los resultados completos de los experimentos realizados pueden apreciarse en el Anexo 1. En particular, se obtuvo que la combinación con la que se obtenía un mejor rendimiento corresponde a:

Epochs: 60

Dropout: 0.1

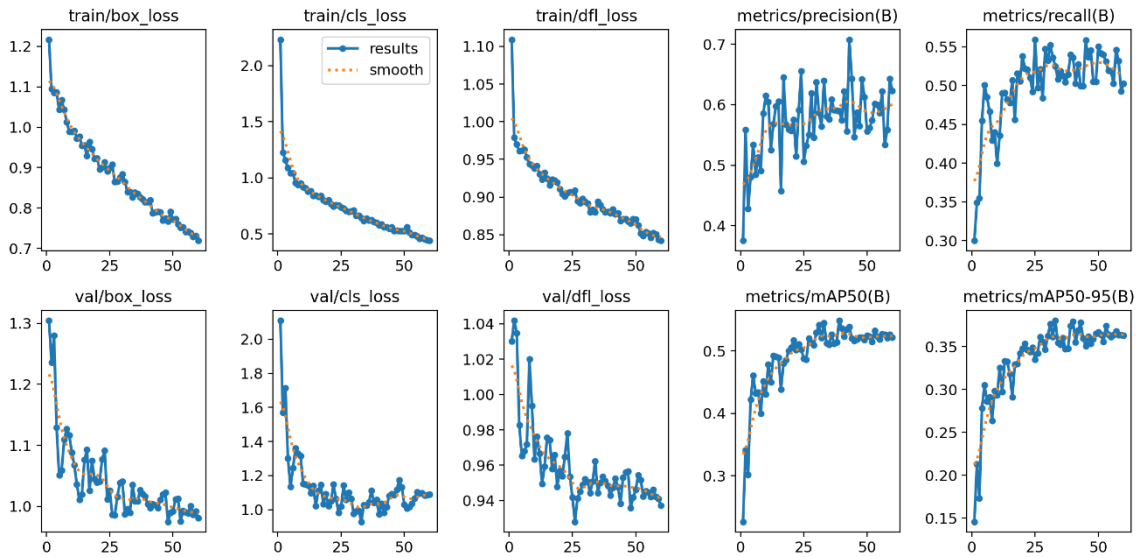
Optimizer: Adam

Lr0: 0.001

Dicho modelo tardó 15,96 minutos en entrenar, con el cual se obtiene una métrica de precisión de 63,7% y una sensibilidad de 52,4%. Los gráficos presentados en la figura 22 ilustran las métricas evaluadas durante el desarrollo del modelo, incluidas las pérdidas observadas en las etapas de entrenamiento y validación, así como los porcentajes de precisión y sensibilidad a lo largo de las épocas en el entrenamiento.

Figura 22

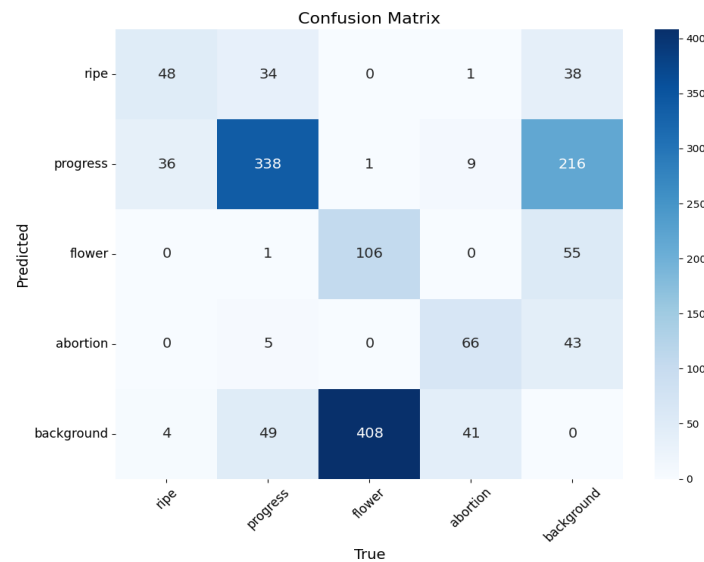
Rendimiento del modelo de clasificación de la cosecha de cacao.



En la figura 22, se aprecian los resultados de pérdida, tanto para los datos de entrenamiento, como para los datos de la validación, a su vez se presenta información sobre la precisión, sensibilidad y media de la precisión para los datos de validación. En estas se identifica que a medida que aumenta la cantidad de épocas, existe una disminución en la función pérdida del modelo, lo cual permite medir qué tan lejos están las predicciones del modelo de los valores reales. Para el caso de la precisión, sensibilidad y la media de precisión se identifica que a medida que aumentan las épocas, incrementa el valor de su precisión.

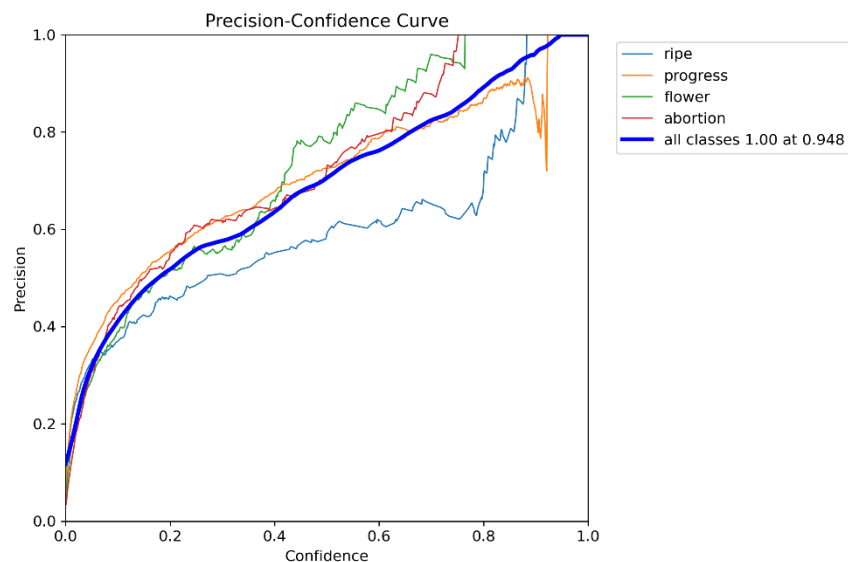
Así mismo, en la figura 23, se puede apreciar la matriz de confusión del modelo de clasificación llevado a cabo. En dicha imagen se identifica que existe una gran cantidad de muestras que el modelo no puede predecir como flores, y las clasifica como contexto (*background*), este comportamiento se debe refinar en posteriores estudios, pues para objetos pequeños, el modelo aún no los puede diferenciar con certeza.

Figura 23
Matriz de confusión del modelo con mejor rendimiento.



En la figura 24, se aprecia la curva precisión-confianza, cuyo comportamiento es similar al esperado, en la cual se presenta de manera inicial valores muy altos de confianza y una precisión que tiende a 100%, pero al momento de reducir los valores de confianza, también se disminuye el valor de la precisión, pues está incrementando la probabilidad de que algunas de las clases identificadas sean erróneas.

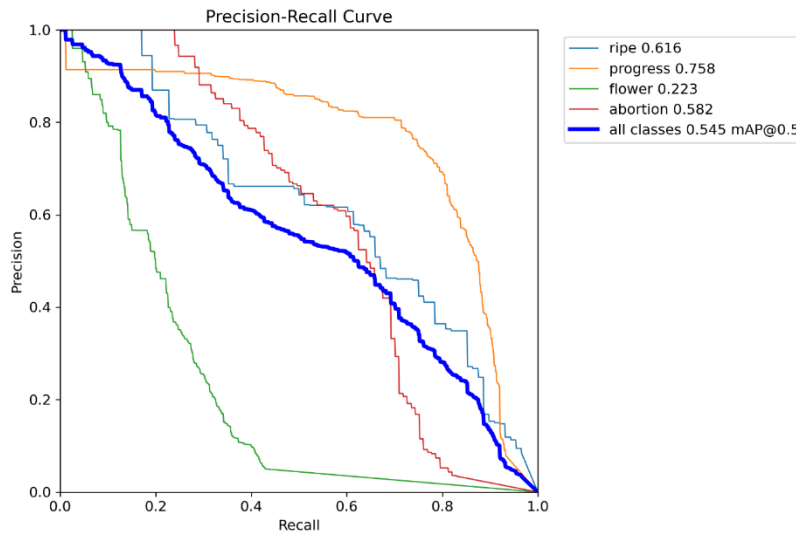
Figura 24
Curva precisión-confianza para el modelo con el mejor rendimiento.



De igual manera, para el modelo con el mejor rendimiento se obtuvo la curva *precision-recall* presente en la figura 25, en donde se detecta que para altos valores de *recall*, el cual es la capacidad que tiene el modelo para detectar correctamente la mayoría de los casos positivos, la precisión suele disminuir, pues el modelo podría incluir un mayor número de falsas positivas al maximizar la detección de verdaderos positivos.

Figura 25

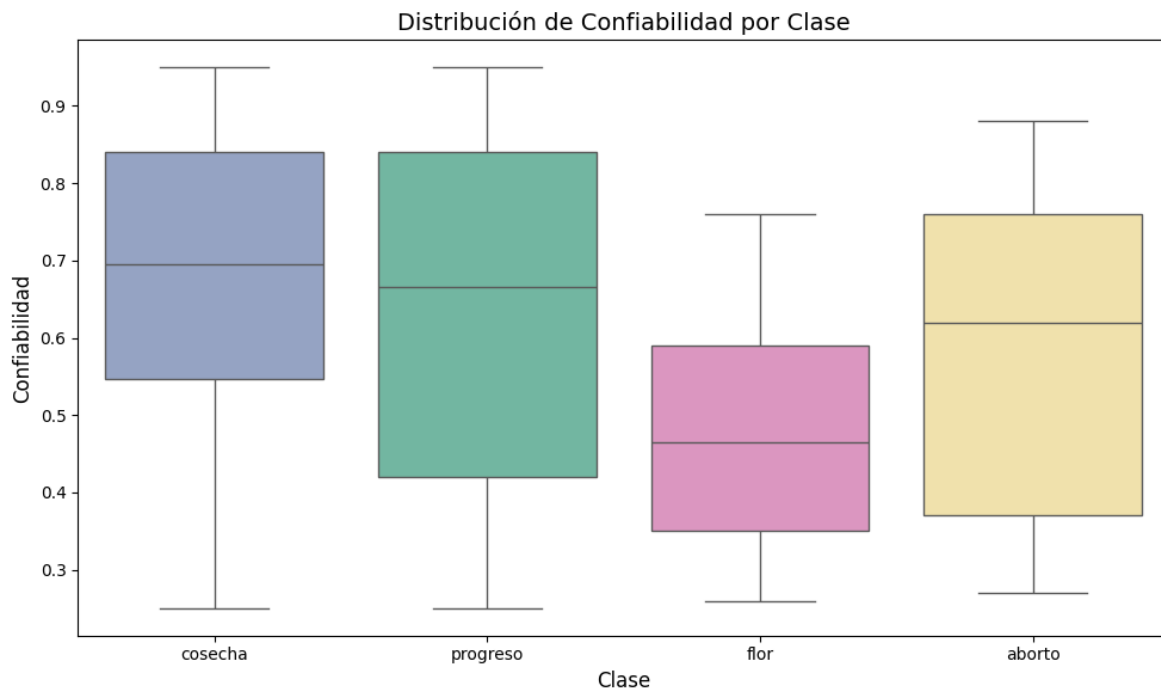
Curva precision-recall con el modelo que presenta mayor rendimiento.



De igual forma, se establecen las métricas para cada una de las clases evaluadas, particularmente, se identifica que con las muestras de las flores se tiene un menor nivel de confiabilidad que con las demás clases, esto va ligado a la matriz de confusión obtenida anteriormente, pues el modelo tiene la tendencia de confundir la flor con otros objetos de la imagen, lo anterior se identifica en la figura 26. Cabe destacar que, la confiabilidad se refiere a la probabilidad asociada a la clase predicha por el modelo, generando un valor de veracidad para la predicción individual de cada muestra.

Figura 26

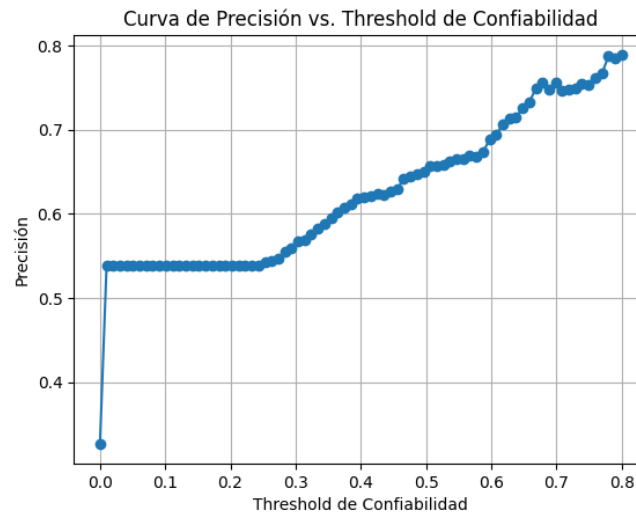
Confiabilidad en datos de prueba para cada una de las clases



Cabe destacar que, a partir de los resultados anteriores, se realizó un análisis considerando la confiabilidad de los datos para cada una de las muestras, para esto se generó la gráfica de la precisión respecto a la confiabilidad, con estos valores, es posible realizar un filtro en el modelo, para considerar valores superiores de confiabilidad de las clases y, de esta manera, impactar la precisión del modelo. Esto es útil, pues al generarse una confusión entre la clase flor y el fondo (*background*), se puede establecer un valor mínimo de confiabilidad para que las muestras que está detectando como flor y que se presentan de baja probabilidad, se eliminen del modelo. La gráfica realizada se puede apreciar en la figura 27.

Figura 27

Gráfica de Precisión vs Confiabilidad del modelo.



Así mismo, se realizaron experimentos considerando únicamente 3 clases, progreso, aborto y cosecha. Con estas clases, el modelo también puede identificar aquellos frutos que se pueden cosechar. Los resultados completos de los experimentos realizados pueden apreciarse en el Anexo 2. En particular, se obtuvo que la combinación con la que se obtenía un mejor rendimiento corresponde a:

Epochs: 60

Dropout: 0.0

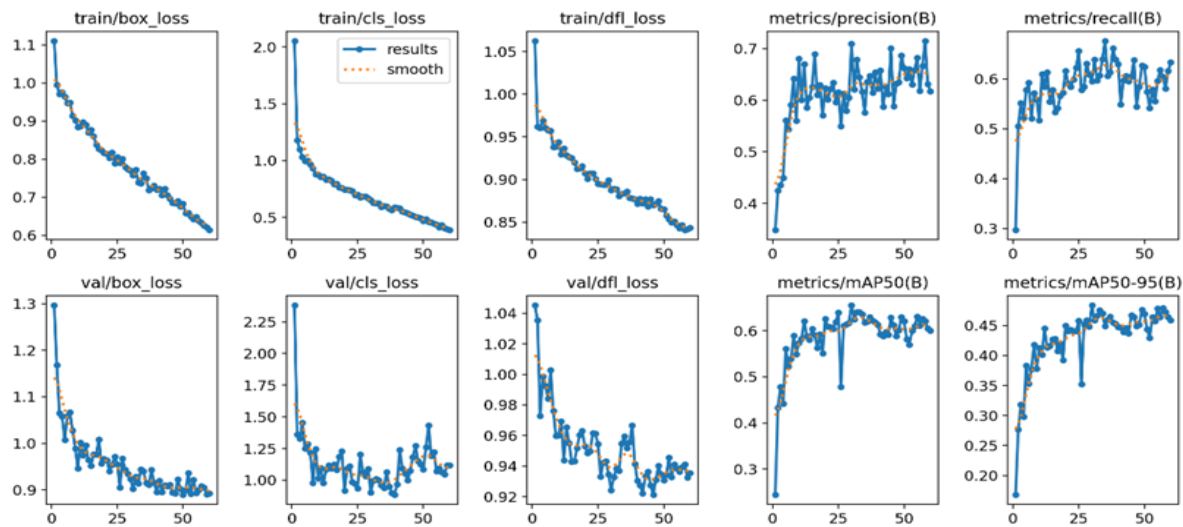
Optimizer: Adam

Lr0: 0.001

Dicho modelo tardó 16,38 minutos en entrenar, con el cual se obtiene una métrica de precisión de 71% y una sensibilidad de 59,6%. Los gráficos presentados en la figura 28 ilustran las métricas evaluadas durante el desarrollo del modelo, incluidas las pérdidas observadas en las etapas de entrenamiento y validación, así como los porcentajes de precisión y sensibilidad a lo largo de las épocas en el entrenamiento.

Figura 28

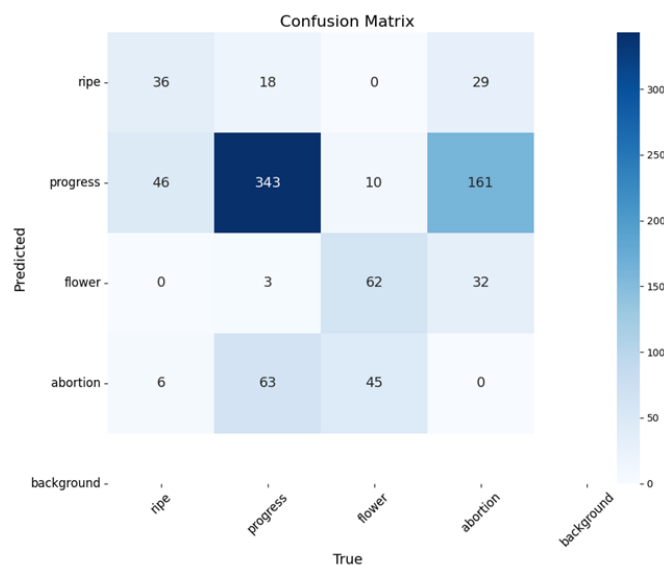
Rendimiento del modelo de clasificación de la cosecha de cacao (3 clases)



Así mismo, en la figura 29, se puede apreciar la matriz de confusión del modelo de clasificación llevado a cabo. En dicha imagen se identifica que existe una gran cantidad de muestras que el modelo no puede predecir como flores, y las clasifica como contexto (*background*), este comportamiento se debe refinar en posteriores estudios, pues para objetos pequeños, el modelo aún no los puede diferenciar con certeza.

Figura 29

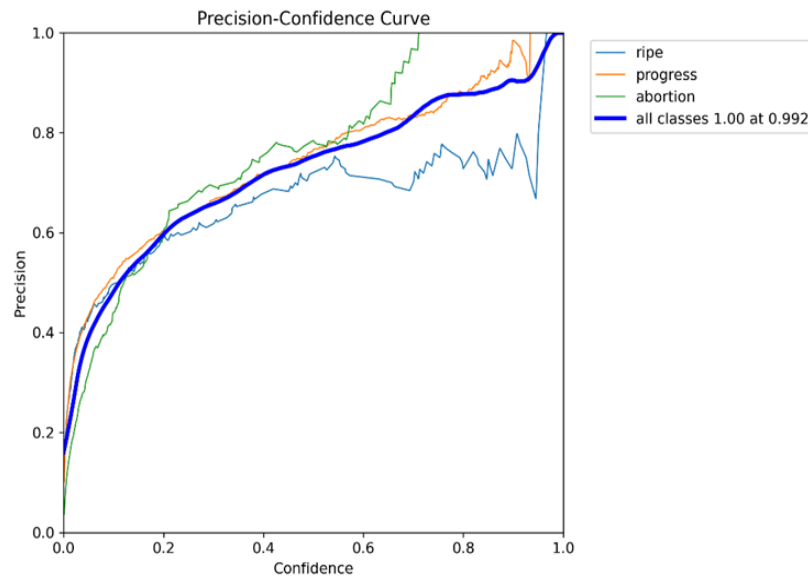
Matriz de confusión del modelo con mejor rendimiento (3 clases).



En la figura 30, se aprecia la curva precisión-confianza, cuyo comportamiento es similar al esperado, en la cual se presenta de manera inicial valores muy altos de confianza y una precisión que tiende a 100%, pero al momento de reducir los valores de confianza, también se disminuye el valor de la precisión, pues está incrementando la probabilidad de que algunas de las clases identificadas sean erróneas.

Figura 30

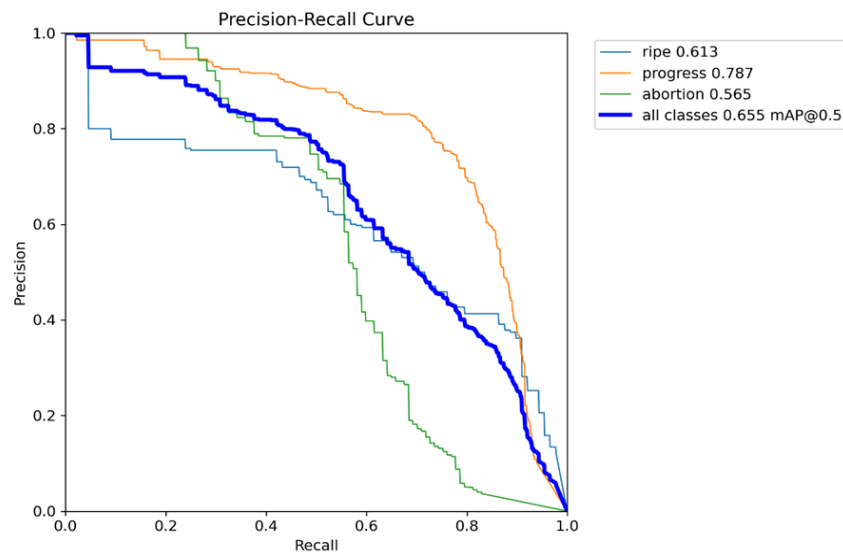
Curva precisión-confianza para el modelo con el mejor rendimiento (3 clases).



De igual manera, para el modelo con el mejor rendimiento se obtuvo la curva *precision-recall* presente en la figura 31, en donde se detecta que para altos valores de *recall*, el cual es la capacidad que tiene el modelo para detectar correctamente la mayoría de los casos positivos, la precisión suele disminuir, pues el modelo podría incluir un mayor número de falsas positivas al maximizar la detección de verdaderos positivos.

Figura 31

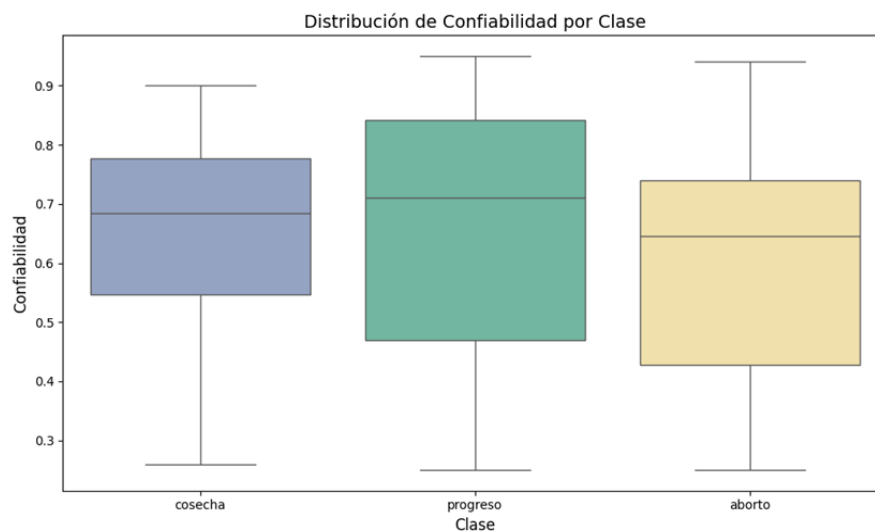
Curva precision-recall con el modelo que presenta mayor rendimiento (3 clases)



A partir de lo anterior, se puede construir la distribución de confianza para cada una de las clases analizadas. Este resultado se aprecia en la figura 32. En dicha gráfica se identifica la precisión media para cada una de las clases, en el caso de la cosecha se obtiene un valor de 68%, para progreso de 72% y para aborto del 65%. En este resultado se aprecia la necesidad de obtener más muestras para el caso de la clase “cosecha”.

Figura 32

Confianza en datos de prueba para cada una de las clases (3 clases)



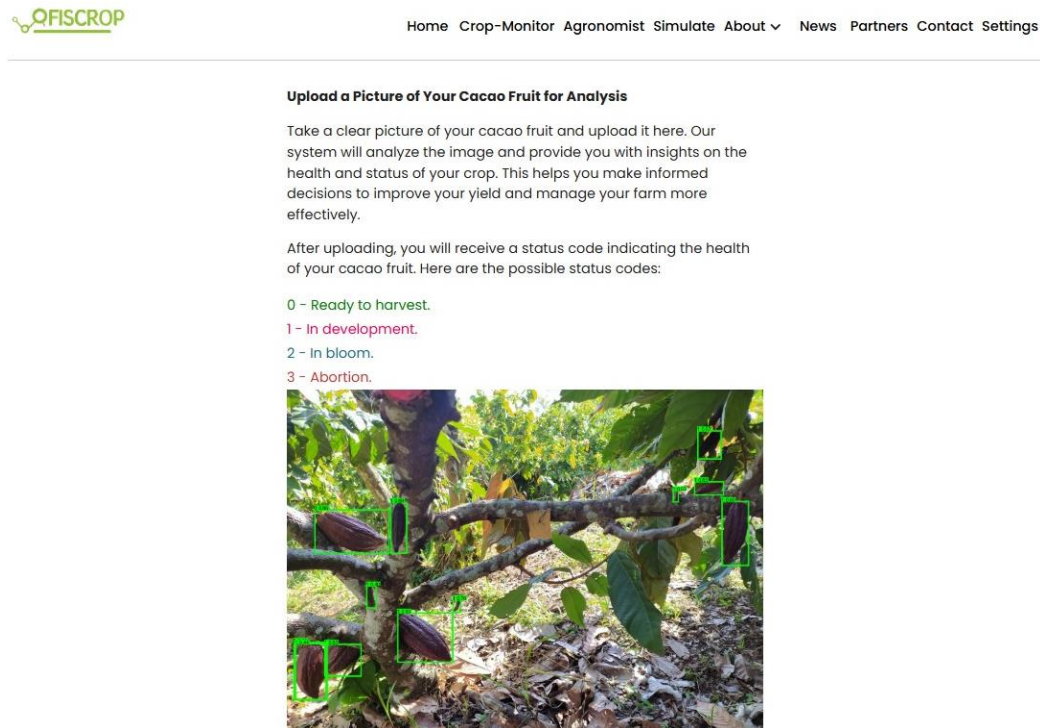
En los análisis desarrollados para la generación del modelo se resalta la importancia de ajustar las estrategias de clasificación para incrementar la precisión del modelo computacional predictivo. En primer lugar, con la clasificación de cuatro clases, el modelo presentó una precisión media de aproximadamente 63%. Con este resultado se evidencia una limitación en la capacidad del modelo para distinguir entre flores y el fondo de la imagen (*background*), lo que ocasiona confusión en la asignación de etiquetas y afectó el desempeño general. Para abordar esta problemática, se implementaron estrategias alternativas, incluyendo la reducción del número de clases a tres: progreso, cosecha y aborto. Este enfoque permitió concentrar los esfuerzos del modelo en las categorías más relevantes para el objetivo general del proyecto, que era identificar los frutos óptimos de cosecha. Como resultado, se logró un aumento significativo en la precisión del modelo, alcanzando un 71%.

Así mismo, se destaca la importancia de realizar experimentos ajustando los parámetros del modelo, con el fin de identificar las condiciones óptimas que maximicen su rendimiento. En este análisis, se encontró que la función de optimización Adam ofrece el mejor desempeño, utilizando una tasa de aprendizaje de 0,001, lo cual se pudo comprobar tanto en el modelo desarrollado para cuatro clases como en el de tres clases. La función Adam es conocida por su capacidad de ajustar dinámicamente las tasas de aprendizaje de cada parámetro, su eficiencia computacional y su resistencia a datos que presenten ruido o gradientes variables. Adicionalmente, presenta un diseño de carácter adaptativo, lo cual asegura una convergencia rápida y precisa.

Por último, la figura 33 corresponde a la información que se integró en una aplicación diseñada para uso de los cacaocultores. Esta herramienta proporciona a los agricultores un acceso fácil y directo a los datos y análisis generados, apoyando la toma de decisiones informadas en la gestión y optimización de sus cultivos de cacao.

Figura 33

Aplicación móvil para supervisar los frutos de cacao en los árboles.



La producción de cacao se enfrenta a numerosos retos, incluyendo la variabilidad climática, numerosas patologías y la complejidad de determinar el momento óptimo de cosecha (Heredia-Gómez et al., 2020). Con el presente modelo predictivo se integran técnicas avanzadas de análisis de imágenes, proporcionando una solución a estos desafíos, que se enfoca en una necesidad crítica dentro de la industria del cacao, considerando mejoras en las prácticas de cosecha para optimizar tanto el rendimiento como la calidad, al tiempo que se promueve la sostenibilidad del cultivo (Lu et al., 2018).

Específicamente, el uso de técnicas avanzadas de procesamiento de imágenes para evaluar las fases de maduración de los frutos del cacao representa una innovación significativa. Los métodos tradicionales para determinar la madurez de los frutos suelen basarse en evaluaciones visuales subjetivas, que pueden dar lugar a incoherencias y errores (Rojas et al., 2020). Mediante el uso de técnicas computacionales, el modelo ofrece un medio objetivo de evaluar la madurez del fruto. Este avance es fundamental para garantizar que los frutos se cosechen en su punto óptimo,

mejorando así el perfil de sabor y aroma y la comerciabilidad de los granos de cacao (Lamos-Díaz et al., 2020).

Los resultados obtenidos en este estudio muestran una mejora significativa en la precisión y aplicabilidad del modelo desarrollado por Heredia-Gómez et al. (2020), en donde se presenta un modelo automatizado mediante CNN, en dicho estudio se reportó un valor de 34.83% mAP, demostrando el reto para realizar este tipo de predicciones. En el presente trabajo se obtuvo un valor de 65,5% mAP50, con una precisión del 71%, con lo cual se evidencia que esta nueva estrategia es significativa al momento de establecer el tiempo óptimo de la cosecha del cultivo de cacao.

En la investigación realizada por Bueno et al. (2020), se propone una arquitectura basada en CNN, que se enfoca en la extracción de características a partir de frecuencias, logrando una precisión del 97,13% para la clasificación binaria según el estado de madurez del fruto. A diferencia de esto, el presente trabajo utiliza datos de imágenes directamente capturadas en campo, lo que permite que el modelo sea empleado por los agricultores sin necesidad de equipos adicionales. Además, la clasificación que se realiza en el presente proyecto abarca tres clases (multiclase), ofreciendo una caracterización más completa del desarrollo de los frutos.

El cacao se cultiva en diversas condiciones climáticas, cada uno con retos y requisitos únicos. Al implementar un modelo de predicción que puede extrapolarse a diferentes regiones y variedades de cacao, se está proponiendo una herramienta con potencial de escalabilidad (Durán-Soria et al., 2021), lo cual aporta a los agricultores locales a optimizar sus prácticas en función de las condiciones específicas a las que se enfrentan (Meza-Sepúlveda et al., 2021).

7.2. Productos

- **Generación de nuevo conocimiento:**

- Participación en evento internacional de divulgación científica. Modalidad virtual: ***3rd Women in Bioinformatics & Data Science LA Conference.***

- Título de la ponencia: **Calibration and implementation of a predictive computational harvest model for cacao quality improvement**

- Participación en evento internacional de divulgación científica. Modalidad presencial: **ISCB-Latin America SoIBio CCBCOL Conference on Bioinformatics 2024.**

- Título del póster: **Computational Predictive Model for Sustainable Optimization of Cocoa Harvesting by using Advanced Technological Tools**

8. Conclusiones

En el presente proyecto se desarrolló un modelo predictivo para la cosecha del cultivo de cacao con un porcentaje de precisión del 71%. Este avance representa una herramienta valiosa para los productores, ya que facilita la toma de decisiones estratégicas al optimizar la planificación de la cosecha. Como resultado, se espera una reducción de pérdidas asociadas a un manejo ineficiente y un impacto positivo en los ingresos económicos de los productores. Además, esta estrategia contribuye a la sostenibilidad del cultivo pues promueve prácticas eficientes, garantizando un uso adecuado de los recursos y fortaleciendo la resiliencia del sector cacaotero considerando los desafíos actuales e impactando la bioeconomía de la región.

El enfoque multiclase empleado para clasificar las etapas de desarrollo del cacao permite realizar un análisis detallado de datos, en comparación con modelos binarios, esto permite un entendimiento más integral del cultivo y refuerza la capacidad de los productores para intervenir de manera oportuna en cada fase de desarrollo. A su vez, se evidencia el rendimiento de modelos como YOLO, para el proceso de detección y clasificación de imágenes en el sector agrícola, el cual permite extrapolar la información del modelo desarrollado en diferentes cultivos.

Este trabajo presenta un aporte significativo en la gestión del cultivo del cacao debido al uso de tecnologías emergentes en el sector agrícola, integrando información de imágenes para optimizar las prácticas de cosecha. A partir de esta línea base, pueden desencadenarse futuras investigaciones orientadas hacia el escalamiento de la aplicación, identificando nuevas funciones y mejorando el modelo según la retroalimentación que brindan los agricultores. Además, se puede centrar en la integración de esta tecnología con otras prácticas agrícolas y otros estudios como los relacionados con ciencias ómicas, para generar una correlación con los resultados del modelo que permitan mejorar la gestión de los cultivos en diversos sectores.

9. Impactos y perspectivas

La implementación de tecnologías avanzadas y la analítica de datos en la producción de cacao representan un cambio significativo hacia una bioeconomía más sostenible y eficiente, debido a la mejora en la rentabilidad de la producción de cacao que se traduce en un uso racional de los recursos. Este enfoque permite a los agricultores emplear prácticas agrícolas más precisas, optimizando el uso del agua y los fertilizantes, y reduciendo desperdicios. Adicionalmente, la incorporación de herramientas basadas en inteligencia artificial proporciona una ventaja competitiva para la industria, maximizando los rendimientos de manera sostenible.

De igual forma, el impulso a la investigación y desarrollo fomenta la innovación en la industria del cacao, generando nuevas oportunidades y perspectivas para abordar desafíos críticos como el cambio climático y las enfermedades del cultivo. La metodología aplicada en el presente estudio aporta establecer las bases para una agricultura más resiliente, que tenga en cuenta la calidad y la sostenibilidad de la producción. Adicionalmente, los avances hacia una caracterización más específica de los estadios de los frutos permiten una mejor planificación de cosecha y refuerzan la capacidad que tienen los productores para adaptarse a condiciones cambiantes, lo cual es fundamental para promover una economía circular y mejorar el impacto ambiental del cultivo de cacao.

Con el objetivo de garantizar la permanencia y escalabilidad de los avances actuales, se proponen diversas líneas de trabajo futuro que continúen impactando la sostenibilidad y la bioeconomía del sector cacaotero. Principalmente, se requiere la incorporación de información temporal sobre los estadios de los frutos en función de los meses, la cual se puede derivar del repositorio actual, en donde se generen nuevas etiquetas temporales para una caracterización más precisa del desarrollo del cultivo, facilitando la identificación de patrones claves como los índices de abortos o pérdidas en el cultivo por enfermedades. Este análisis longitudinal es bastante relevante para empresas como Luker, pues se optimiza la planificación agrícola y se reducen las pérdidas relacionadas con manejo inadecuado o condiciones desfavorables en el cultivo.

A su vez, es prioritario incrementar las muestras para la clase de "cosecha", lo cual permite el fortalecimiento del modelo desarrollado. Esta estrategia mejorará la precisión de la herramienta y también facilitará su implementación a gran escala, promoviendo su adopción en diversos contextos agrícolas. Otro aspecto relevante es el uso de esta metodología para la identificación temprana de enfermedades, lo que permitirá un manejo sanitario más eficiente y tendría un impacto en la reducción de pérdidas económicas y mejoramiento de la resiliencia del cultivo frente a amenazas fitosanitarias.

Adicionalmente, el auge de la integración multimodal de datos ofrece una oportunidad de mejora del modelo actual, en donde se incorpore información climática como temperatura, humedad, precipitaciones y radiación solar a los modelos actuales, lo cual genera predicciones más precisas y contextualizadas, mejorando la toma de decisiones de los agricultores.

Finalmente, la implementación de herramientas basadas en el análisis de video del cultivo podría generar una facilidad a los productores para la obtención de datos de manera más eficiente al transformar videos en una fuente automatizada de información sobre el desarrollo de los frutos, permitiendo el registro de mayor área del cultivo. Además, se podría integrar información de otras variedades de cacao y extrapolar el modelo desarrollado a los nuevos datos, lo cual genera una escalabilidad y permitirá explorar características específicas de cada variedad.

10. Recomendaciones

La optimización del modelo predictivo es una etapa clave para incrementar los valores de precisión y aplicabilidad en la cadena productiva del cacao. Es así como la incorporación de nuevas variables y de información adicional, de la mano con la calibración de parámetros, permitirá obtener predicciones más acertadas del tiempo óptimo de la cosecha. Aunado a esto, se debe centrar esfuerzos en la integración de prácticas de MLOps, que permitan la automatización del entrenamiento, el despliegue del modelo, considerando las actualizaciones constantes y un monitoreo en tiempo real para mantener la efectividad del modelo.

Cabe resaltar que, es fundamental capacitar a los productores en el uso y aplicabilidad del modelo, con el objetivo de promover su uso en diferentes regiones productoras de cacao. A su vez, la consolidación con sistemas agrícolas inteligentes podrá generar una gestión más sostenible del cultivo, realizando una revisión del impacto económico y ambiental del modelo a largo plazo, a partir de una comparación de la producción de las fincas antes y después de la aplicabilidad del modelo. El presente desarrollo funciona como base para la tecnificación de otros cultivos, lo que facilitaría su validación en distintos contextos agrícolas; para generar una escalabilidad de estas iniciativas, se recomienda la implementación de normativas que impulsen el uso de tecnología en el sector agrícola y la creación de programas de financiamiento que impacten a pequeños y medianos productores.

Para finalizar, el presente trabajo respalda futuras investigaciones que se enfoquen en algoritmos avanzados de inteligencia artificial, como lo son redes neuronales profundas, modelos de ensamble, e integración de datos multimodales para mejorar la efectividad del sistema. De igual forma, se sugiere ampliar las bases de datos utilizadas, incorporando información que se recolecte *in situ* de sensores IoT e imágenes satelitales, lo que tendrá un impacto directo en los resultados del análisis predictivo. Del mismo modo, la validación del modelo en distintas regiones productoras mediante estudios comparativos va a contribuir a determinar su adaptabilidad, y a su vez, la identificación de limitaciones actuales en las unidades productivas permitirá buscar estrategias para la mejora continua del sistema.

11. Referencias

- Abbott, P. C., Benjamin, T. J., Burniske, G. R., Croft, M. M., Fenton, M., Kelly, C. R., Lundy, M., Rodriguez Camayo, F., & Wilcox M.D. (2017). An Analysis on the Supply Chain of Lansones. *Análisis de La Cadena Productiva Del Cacao En Colombia*, November, 208.
- ANDI. (2019). Chocolatería, una industria que crece sosteniblemente en el país. <https://www.andi.com.co/Home/Noticia/15568-chocolateria-una-industria-que-crece-so#:~:text=Las industrias que transforman el,aportado %24782.000 millones al PIB>.
- Arrazate CH, A. o, Bolaños M, M., & López, M. (2018). The Cocoa (*Theobroma cacao* L.) native in Mexico. *Biodiversity International Journal*, 2(6), 535–536. <https://doi.org/10.15406/bij.2018.02.00109>
- Bandi, R., Swamy, S., & Arvind, C. S. (2023). Leaf disease severity classification with explainable artificial intelligence using transformer networks. *International Journal of Advanced Technology and Engineering Exploration*, 10(100), 278–302. <https://doi.org/10.19101/IJATEE.2022.10100136>
- Bisong, E. (2019). Google Colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform* (pp. 59–64). Apress. https://doi.org/10.1007/978-1-4842-4470-8_7
- Bosch, M. (2013). Reflexiones acerca de las Tecnologías Emergentes en la Agricultura Introducción. 2009, 1–22.
- Bueno, G. E., Valenzuela, K. A., & Arboleda, E. R. (2020). Maturity classification of cacao through spectrogram and convolutional neural network. *Jurnal Teknologi Dan Sistem Komputer*, 8(3), 228–233. <https://doi.org/10.14710/jtsiskom.2020.13733>
- CAF. (2019). Proyecto de investigación regional sobre producción de cacao sostenible adaptado al clima y cambio. *Observatorio Del Cacao Fino y de Aroma Para América Latina*, 1–15. http://bit.ly/CIAT_CGIAR_
- Cayeros, S. E., Robles, F. J., & Soto, E. (2016). Cadenas Productivas y Cadenas de Valor. *Revista Educateconciencia*, 10(11), 6–12. <file:///C:/Users/alumno.le1/Downloads/129-630-1-PB.pdf>
- CNCH. (2019). BPA - Buenas Prácticas Agrícolas en el cultivo, beneficio y comercialización de Cacao (*Theobroma cacao* L.).

- Condiza, C. A. (1998). Programa Nacional De Transferencia De Tecnología Agropecuaria Pronatta Agricultura Sostenible Sena Fundación Integración Campesina Servicio Nacional De Aprendizaje “Pinte C” “S E N a.” Agricultura Sostenible, 114. [http://bibliotecadigital.agronet.gov.co/bitstream/11348/4741/1/20061127145142_Definicion de agricultura sostenible%5B1%5D.pdf](http://bibliotecadigital.agronet.gov.co/bitstream/11348/4741/1/20061127145142_Definicion%20de%20agricultura%20sostenible%5B1%5D.pdf)
- CONPES. (2021). Documento CONPES 4069: Política Nacional de Ciencia, Tecnología e Innovación 2022-2031. 108. <https://colaboracion.dnp.gov.co/CDT/Conpes/Economicos/3582.pdf>.
- CORPOICA. (2000). Corporación Colombiana de investigación Agropecuaria.
- Cristancho, E., López, L., Mojica, A., Pedraza, E., & Valqui, A. (2021). Informe ejecutivo: Estudio sobre las necesidades y brechas de calidad en la cadena productiva de cacao y sus derivados y Plan de Acción Santander y su zona de influencia, Colombia. Proyecto ColombiaMide, 1–50. <https://colombiamide.inm.gov.co/wp-content/uploads/2021/08/INFORME-EJECUTIVO-CACAO.pdf>
- Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimedia Tools and Applications*, 82(6), 9243–9275. <https://doi.org/10.1007/s11042-022-13644-y>
- DNP. (2018). POLÍTICA DE CRECIMIENTO VERDE, CONPES 3934.
- Durán-Soria, S., Pott, D. M., Will, F., Mesa-Marín, J., Lewandowski, M., Celejewska, K., Masny, A., Żurawicz, E., Jennings, N., Sønsteby, A., Krüger, E., & Osorio, S. (2021). Exploring genotype-by-environment interactions of chemical composition of raspberry by using a metabolomics approach. *Metabolites*, 11(8), 1–18. <https://doi.org/10.3390/metabo11080490>
- Encalada, J., & Jácome, T. (2018). Determinación de Parámetros Cinéticos en la Devolatilización de Biomasa Residual de Cacao Ecuatoriano. In Universidad Central de Ecuador.
- Espinosa, J. A. (2016). Características estructurales y funcionales de un faro agroecológico a partir de las experiencias de productores cacaoteros de las regiones de los departamentos de Nariño, Meta, Caquetá y Tolima. 236.
- Expósito Gallardo, M. del C., & Ávila Ávila, R. (2008). Aplicaciones de la inteligencia artificial en la Medicina: perspectivas y problemas. *Acimed*, 17(5), 0–0. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352008000500005&lng=es&nrm=iso&tlng=es

- FAO. (2022). Información estadística de cultivos disponible en la plataforma (FAOSTAT). Universidad de Cordoba. <https://www.fao.org/faostat/en/#data/QCL>
- Fujimura, Y., Kurihara, K., Ida, M., Kosaka, R., Miura, D., Wariishi, H., Maeda-Yamamoto, M., Nesumi, A., Saito, T., Kanda, T., Yamada, K., & Tachibana, H. (2011). Metabolomics-driven nutraceutical evaluation of diverse green tea cultivars. *PLoS ONE*, 6(8). <https://doi.org/10.1371/journal.pone.0023426>
- Gobernación de Caldas. (2020). PLAN DEPARTAMENTAL DE EXTENSIÓN AGROPECUARIA DEPARTAMENTO DE CALDAS 2020-2023. Agencia de Desarrollo Rural ADR, 28. <http://190.0.34.154:7000/planes/descargar?id=5>
- Gobierno de Colombia. (2019). Misión internacional de sabios para el avance de la Ciencia, la Tecnología y la Innovación. Pacto por la Ciencia, la Tecnología y la Innovación: Un sistema para construir el conocimiento del futuro.
- Gobierno de Colombia. (2020). Bioeconomía para un crecimiento sostenible (p. 48).
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 386–397. <https://doi.org/10.1109/TPAMI.2018.2844175>
- Heredia-Gómez, J. F., Rueda-Gómez, J. P., Talero-Sarmiento, L. H., Ramírez-Acuña, J. S., & Coronado-Silva, R. A. (2020). Cocoa pods ripeness estimation, using convolutional neural networks in an embedded system. *Revista Colombiana de Computacion*, 21(2), 42–55. <https://doi.org/10.29375/25392115.4030>
- Hernández-Núñez, H. E., Gutiérrez-Montes, I., Sánchez-Acosta, J. R., Rodríguez-Suárez, L., Gutiérrez-García, G. A., Suárez-Salazar, J. C., & Casanoves, F. (2020). Agronomic conditions of cacao cultivation: its relationship with the capitals endowment of Colombian rural households. *Agroforestry Systems*, 94(6), 2367–2380. <https://doi.org/10.1007/s10457-020-00556-9>
- ICCO. (2016). International Cocoa Organization.
- Incio Flores, F. A., Capuñay Sanchez, D. L., Estela Urbina, R. O., Valles Coral, M. Á., Vergara Medrano, E. E., & Elera Gonzales, D. G. (2021). Inteligencia artificial en educación: una revisión de la literatura en revistas científicas internacionales. *Apuntes Universitarios*, 12(1), 353–372. <https://doi.org/10.17162/au.v12i1.974>

- Krenzer, A., Makowski, K., Hekalo, A., Fitting, D., Troya, J., Zoller, W. G., Hann, A., & Puppe, F. (2022). Fast machine learning annotation in the medical domain: a semi-automated video annotation tool for gastroenterologists. *BioMedical Engineering Online*, 21(1), 1–23. <https://doi.org/10.1186/s12938-022-01001-x>
- Lamos-Díaz, H., Puentes-Garzón, D. E., & Zarate-Caicedo, D. A. (2020). Comparison Between Machine Learning Models for Yield Forecast in Cocoa Crops in Santander, Colombia. *Revista Facultad de Ingenieria*, 29(54). <https://doi.org/10.19053/01211129.v29.n54.2020.10853>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Li, L., Jamieson, K., Rostamizadeh, A., Gonina, E., Moritz, J. B., Recht, B., & Talwalkar, A. (2020). A system for massively parallel hyperparameter tuning. *Proceedings*.
- Lippi, M., Bonucci, N., Carpio, R. F., Contarini, M., Speranza, S., & Gasparri, A. (2021). A YOLO-based pest detection system for precision agriculture. 2021 29th Mediterranean Conference on Control and Automation, MED 2021, 342–347. <https://doi.org/10.1109/MED51440.2021.9480344>
- Lu, F., Rodriguez-Garcia, J., Van Damme, I., Westwood, N. J., Shaw, L., Robinson, J. S., Warren, G., Chatzifragkou, A., McQueen Mason, S., Gomez, L., Faas, L., Balcombe, K., Srinivasan, C., Picchioni, F., Hadley, P., & Charalampopoulos, D. (2018). Valorisation strategies for cocoa pod husk and its fractions. *Current Opinion in Green and Sustainable Chemistry*, 14, 80–88. <https://doi.org/10.1016/j.cogsc.2018.07.007>
- Maillet, W., Ouhami, M., & Hafiane, A. (2023). Fusion of Satellite Images and Weather Data with Transformer Networks for Downy Mildew Disease Detection. *IEEE Access*, 11, 5406–5416. <https://doi.org/10.1109/ACCESS.2023.3237082>
- McMahon, P., Purung, H. bin, Lambert, S., Mulia, S., Nurlaila, Susilo, A. W., Sulistyowati, E., Sukanto, S., Israel, M., Saftar, A., Amir, A., Purwantara, A., Iswanto, A., Guest, D., & Keane, P. (2015). Testing local cocoa selections in three provinces in Sulawesi: (i) Productivity and resistance to cocoa pod borer and *Phytophthora* pod rot (black pod). *Crop Protection*, 70, 28–39. <https://doi.org/10.1016/j.cropro.2015.01.001>
- Meza-Sepúlveda, D. C., Castro, A. M., Zamora, A., Arboleda, J. W., Gallego, A. M., & Camargo-Rodríguez, A. V. (2021). Bio-based value chains potential in the management of cacao pod

- waste in colombia, a case study. *Agronomy*, 11(4), 1–17. <https://doi.org/10.3390/agronomy11040693>
- Mia, M. S., Arnob, A. B. H., Naim, A., Voban, A. A. B., & Islam, M. S. (2023). ViTs are Everywhere: A Comprehensive Study Showcasing Vision Transformers in Different Domain. 2023 International Conference on Cognitive Computing and Complex Data, ICCD 2023, 101–117. <https://doi.org/10.1109/ICCD59681.2023.10420683>
- Ministerio de Agricultura y Desarrollo Rural. (2021a). CADENA DE CACAO - Dirección de Cadenas Agrícolas y Forestales.
- Ministerio de Agricultura y Desarrollo Rural. (2021b). Cadena de valor de Cacao. Direccion de Cadenas Agrícolas y Forestales, 27. [https://sioc.minagricultura.gov.co/Cacao/Documentos/2021-03-31 Cifras Sectoriales.pdf](https://sioc.minagricultura.gov.co/Cacao/Documentos/2021-03-31%20Cifras%20Sectoriales.pdf)
- Mjolsness, E., & DeCoste, D. (2001). Machine learning for science: State of the art and future prospects. *Science*, 293(5537), 2051–2055. <https://doi.org/10.1126/science.293.5537.2051>
- Murillo, W., & Méndez, J. J. (2019). Aprovechamiento de la Biomasa Lignocelulósica de la Cadena del Cacao, Un Camino a la Bioeconomía Circular. Universidad Del Tolima.
- Ojeda Beltran, A. (2022). Plataformas Tecnológicas en la Agricultura 4.0: una Mirada al Desarrollo en Colombia. *Computer and Electronic Sciences: Theory and Applications*, 3(1), 9–18. <https://doi.org/10.17981/cesta.03.01.2022.02>
- Osorio Arias, E. L. (2022). Caracterización de los Productores de Cacao (*Theobroma Cacao*) del Municipio de Marquetalia, Caldas [Universidad de Caldas]. In Universidad de Caldas. <https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>
- Pierce, F. J., & Nowak, P. (1999). Aspects of Precision Agriculture. *Advances in Agronomy*, 67(C), 1–85. [https://doi.org/10.1016/S0065-2113\(08\)60513-1](https://doi.org/10.1016/S0065-2113(08)60513-1)
- PROAMAZONIA. (2003). Caracterización De Las Zonas Productoras de cacao en Perú y su competitividad. 221. http://agroaldia.minag.gob.pe/biblioteca/download/pdf/manuales-%0Ahttp://infocafes.com/portal/wp-content/uploads/2016/03/cacao_completo.pdf
- Proexport Colombia. (2012). Cacao colombiano fino y de aroma. 15. [http://www.inviertaencolombia.com.co/images/Perfil Cacao 2012.pdf](http://www.inviertaencolombia.com.co/images/Perfil%20Cacao%202012.pdf)
- Qi, W., Xu, C., & Xu, X. (2021). AutoGluon: A revolutionary framework for landslide hazard analysis. *Natural Hazards Research*, 1(3), 103–108. <https://doi.org/10.1016/j.nhres.2021.07.002>

- Quintana Fuentes, L. F., García Jerez, A., & Moreno Martínez, E. (2018). Perfil sensorial de cuatro modelos de siembra de cacao en Colombia. *Entramado*, 14(2), 256–268. <https://doi.org/10.18041/1900-3803/entramado.2.4756>
- Robinson, M. C., Glen, R. C., & Lee, A. A. (2020). Validating the validation: reanalyzing a large-scale comparison of deep learning and machine learning models for bioactivity prediction. *Journal of Computer-Aided Molecular Design*, 34(7), 717–730. <https://doi.org/10.1007/s10822-019-00274-0>
- Rodriguez Leon, L. (2017). Estudio de la situación en Colombia acerca de la producción de Fitoterapéuticos a través de procesos biotecnológicos. 21. <https://repository.javeriana.edu.co/bitstream/handle/10554/33799/RodriguezLeonLeonardo2017.pdf?sequence=5&isAllowed=y>
- Rojas, K. E., García, M. C., Cerón, I. X., Ortiz, R. E., & Tarazona, M. P. (2020). Identification of potential maturity indicators for harvesting cacao. *Heliyon*, 6(2), 0–7. <https://doi.org/10.1016/j.heliyon.2020.e03416>
- Romero, A., Camargo, A., Ramirez, O. D., Arenas, P., & Gallego, A. M. (2022). A Crop Modelling Strategy to Improve Cacao Quality and Productivity. 1–16.
- Rouhiainen, L. (2018). Inteligencia artificial 101 cosas que debes saber hoy sobre nuestro futuro. Alienta Editorial, 22. https://planetadelibrosar0.cdnstatics.com/libros_contenido_extra/40/39307_Inteligencia_artificial.pdf
- Rusconi, M., & Conti, A. (2010). Theobroma cacao L., the Food of the Gods: A scientific approach beyond myths and claims. *Pharmacological Research*, 61(1), 5–13. <https://doi.org/10.1016/j.phrs.2009.08.008>
- Semiconductor. (2018). Deep Learning Spreads. <https://semiengineering.com/deep-learning-spreads/>
- Superintendencia de industria y comercio. (2011). Cadena productiva del cacao: Diagnostico de libre competencia. Superintendencia Industria y Comercio. Superintendencia De Industria Y Comercio, 1–6. https://www.sic.gov.co/recursos_user/documentos/promocion_competencia/Estudios_Economicos/Cacao.pdf

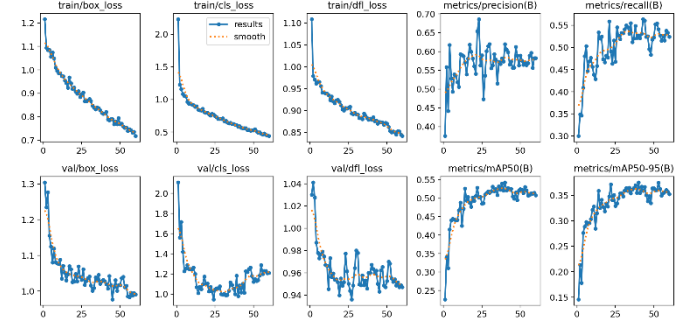
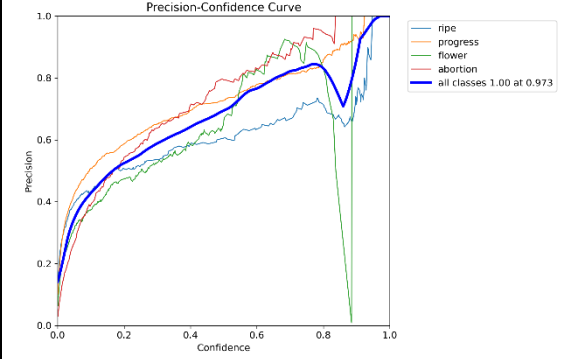
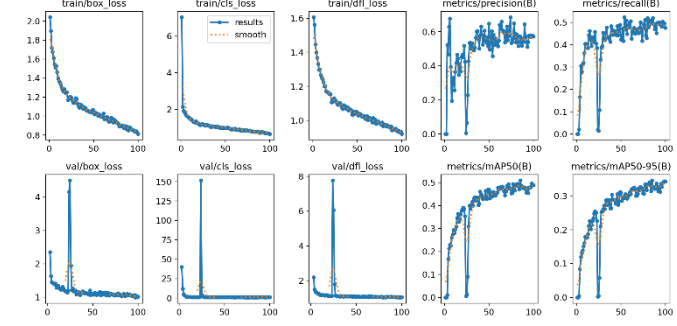
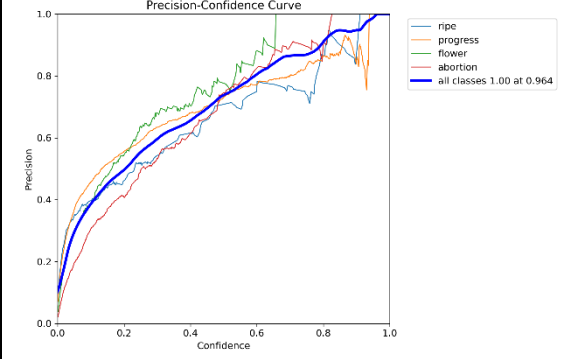
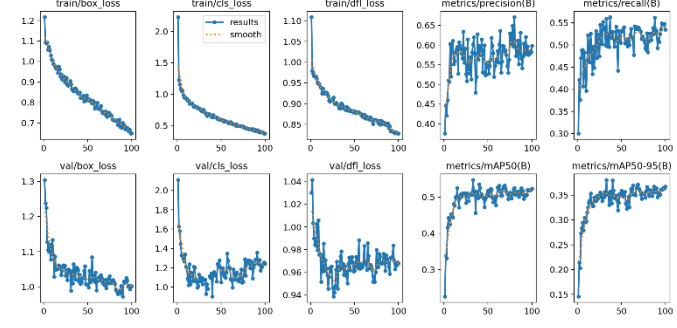
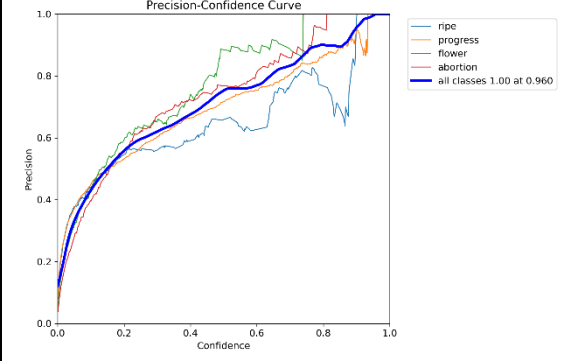
- Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5(4), 1680–1716. <https://doi.org/10.3390/make5040083>
- Torres Cabezas, M. A. (2019). Obtención de celulosa a partir de la cáscara de cacao ecuatoriano (*Theobroma cacao* L.) mediante hidrólisis térmica para la elaboración de pulpa de papel. 1–19.
- USAID. (2021). SEED - Cacao for Peace. <https://www.usaid.gov/colombia/fact-sheets/red-cacao-peace>
- Vásquez-Barajas, E. F., García-Torres, N. E., Bastos-Osorio, L. M., & Lázaro-Pacheco, J. M. (2018). Análisis económico del sector cacaotero en Norte de Santander, Colombia y a nivel internacional. *Revista de Investigación, Desarrollo e Innovación*, 8(2), 237–250. <https://doi.org/10.19053/20278306.v8.n2.2018.7963>
- Yen, K., & Hong, N. (2016). Effects of Maturity Stages and Fermentation of Cocoa Beans on Total Phenolic Contents and Antioxidant Capacities in Raw Cocoa. 14(4), 743–752.
- Yue, X., Qi, K., Na, X., Zhang, Y., Liu, Y., & Liu, C. (2023). Improved YOLOv8-Seg Network for Instance Segmentation of Healthy and Diseased Tomato Plants in the Growth Stage. *Agriculture (Switzerland)*, 13(8). <https://doi.org/10.3390/agriculture13081643>
- Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2019). A primer on deep learning in genomics. *Nature Genetics*, 51(1), 12–18. <https://doi.org/10.1038/s41588-018-0295-5>

12. Anexos

Anexo 1. Experimentos en YOLO para 4 clases

Epochs	Dropout	Optimizer	lr0	Tiempo [horas]	Precisi on	Recall	mAP5 0	Métricas de pérdida y precisión	Curva Precisión-Confianza
60	0.0	Adam	0.01	0.27	0.6	0.484	0.499		
60	0.0	Adam	0.001	0.267	0.588	0.545	0.545		

60	0.1	Adam	0.01	0.268	0.588	0.489	0.497	
60	0.1	Adam	0.001	0.266	0.637	0.524	0.545	
60	0.2	Adam	0.01	0.267	0.57	0.47	0.49	

60	0.2	Adam	0.001	0.265	0.62	0.527	0.543		
100	0.0	Adam	0.01	0.423	0.573	0.505	0.509		
100	0.0	Adam	0.001	0.42	0.604	0.56	0.548		

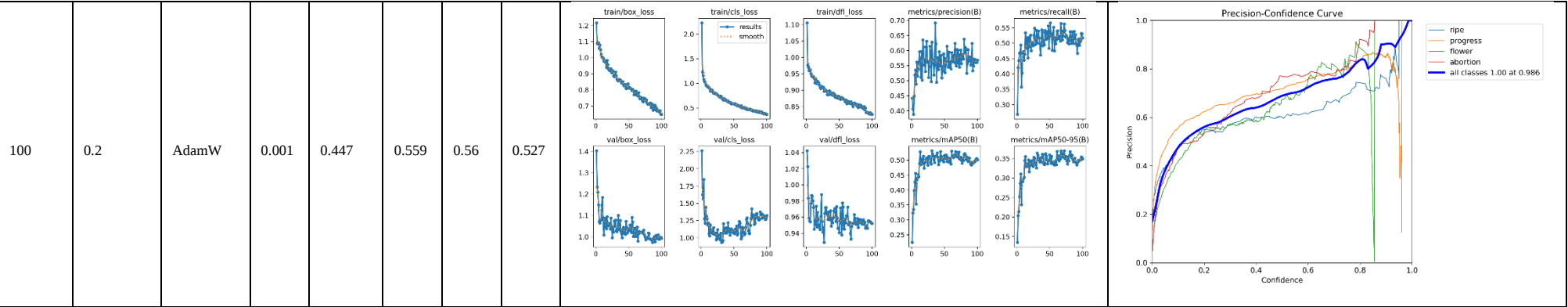
100	0.1	Adam	0.01	0.419	0.603	0.477	0.505	
100	0.1	Adam	0.001	0.416	0.6	0.574	0.541	
100	0.2	Adam	0.01	0.424	0.574	0.513	0.508	

100	0.2	Adam	0.001	0.418	0.598	0.518	0.535	
60	0.0	AdamW	0.01	0.255	0.529	0.516	0.496	
60	0.0	AdamW	0.001	0.278	0.623	0.519	0.548	

60	0.1	AdamW	0.01	0.272	0.55	0.509	0.488	
60	0.1	AdamW	0.001	0.272	0.583	0.556	0.552	
60	0.2	AdamW	0.01	0.276	0.627	0.458	0.491	

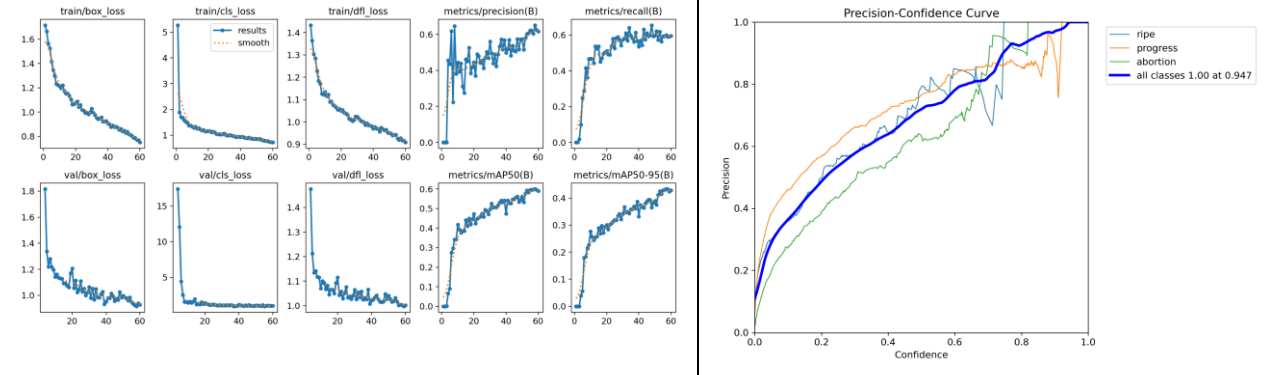
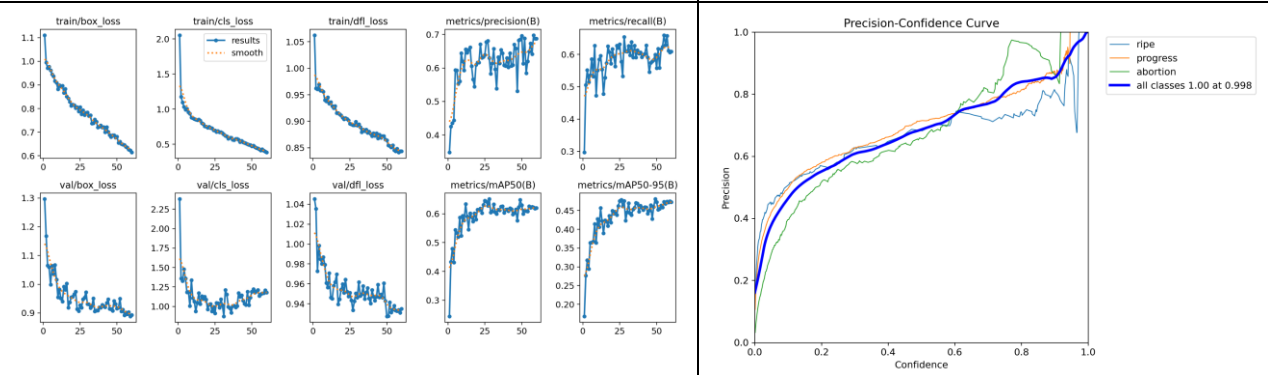
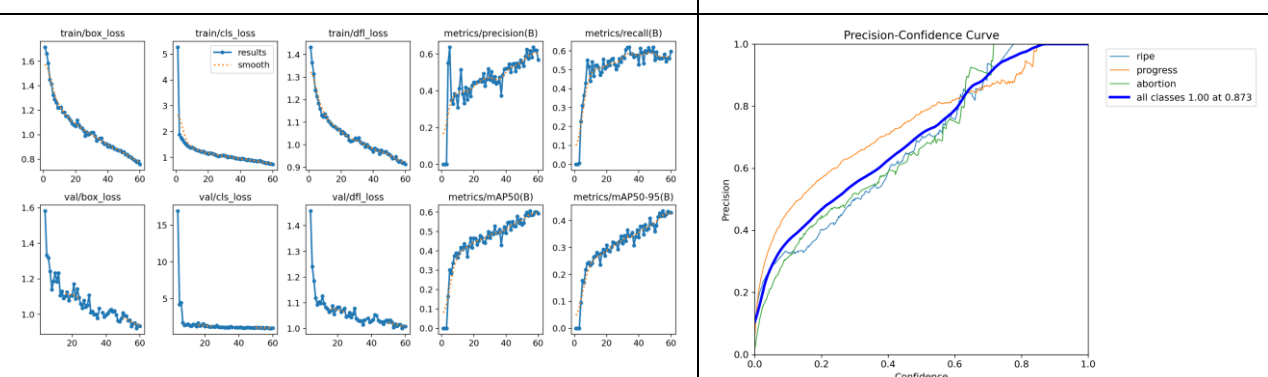
60	0.2	AdamW	0.001	0.268	0.629	0.497	0.518		
100	0.0	AdamW	0.01	0.429	0.544	0.511	0.505		
100	0.0	AdamW	0.001	0.44	0.579	0.515	0.528		

100	0.1	AdamW	0.01	0.446	0.569	0.478	0.489		
100	0.1	AdamW	0.001	0.446	0.583	0.533	0.526		
100	0.2	AdamW	0.01	0.427	0.576	0.521	0.513		



Anexo 2. Experimentos en YOLO para 3 clases

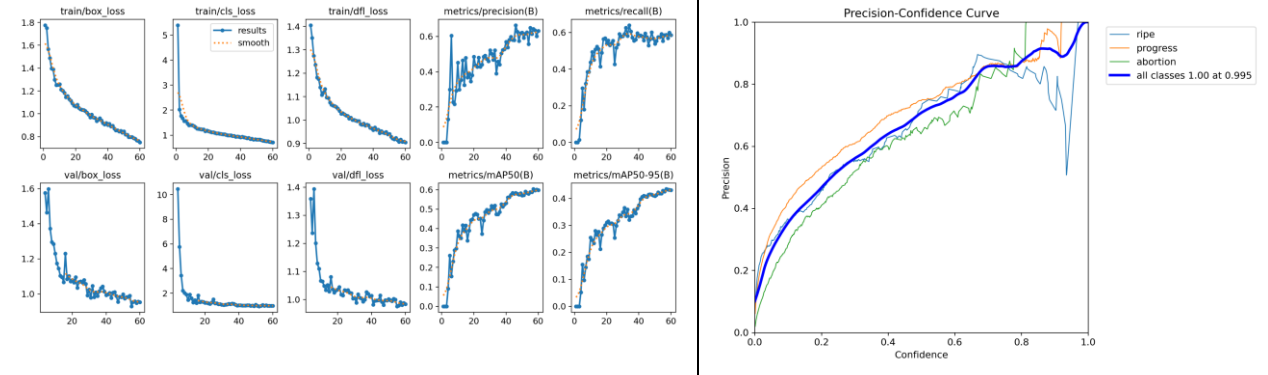
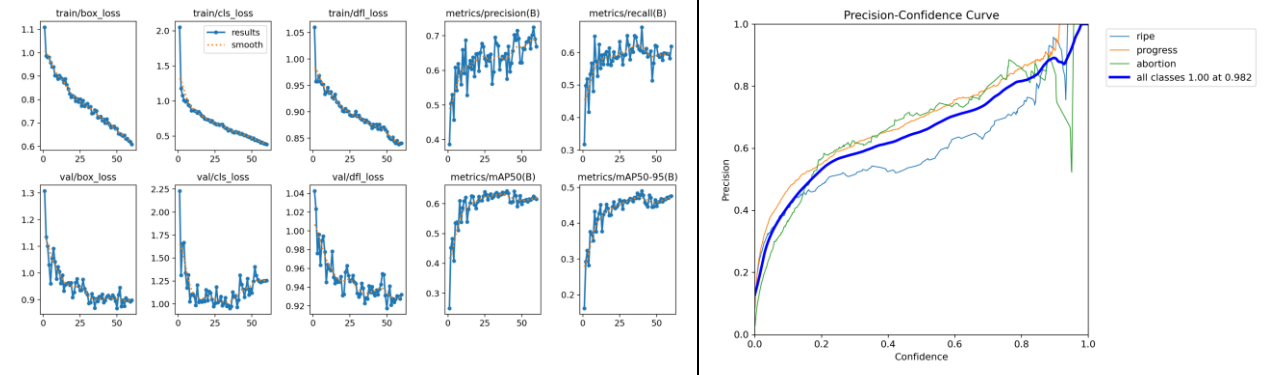
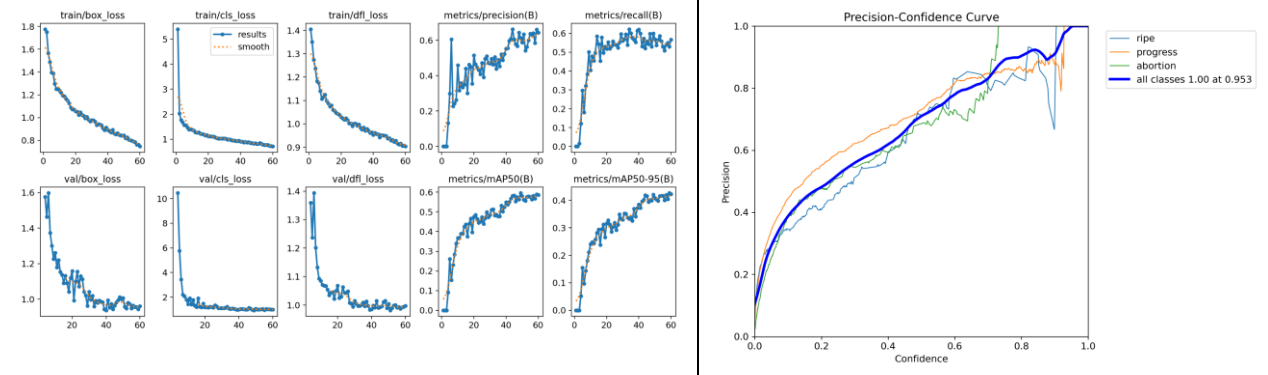
Epochs	Dropout	Optimizer	lr0	Tiempo [horas]	Precisi on	Recall	mAP50	Métricas de pérdida y precisión	Curva Precisión-Confianza
60	0.0	Adam	0.01	0.257	0.618	0.546	0.586		
60	0.0	Adam	0.001	0.273	0.71	0.596	0.655		

60	0.1	Adam	0.01	0.282	0.605	0.6	0.598	
60	0.1	Adam	0.001	0.278	0.685	0.6	0.648	
60	0.2	Adam	0.01	0.283	0.619	0.563	0.606	

60	0.2	Adam	0.001	0.274	0.707	0.611	0.634		
100	0.0	Adam	0.01	0.431	0.705	0.563	0.63		
100	0.0	Adam	0.001	0.425	0.704	0.551	0.622		

100	0.1	Adam	0.01	0.428	0.622	0.625	0.628		
100	0.1	Adam	0.001	0.421	0.619	0.627	0.639		
100	0.2	Adam	0.01	0.426	0.578	0.641	0.63		

100	0.2	Adam	0.001	0.43	0.669	0.61	0.642		
60	0.0	AdamW	0.01	0.281	0.652	0.591	0.598		
60	0.0	AdamW	0.001	0.28	0.656	0.649	0.671		

60	0.1	AdamW	0.01	0.282	0.642	0.585	0.604	
60	0.1	AdamW	0.001	0.28	0.575	0.679	0.642	
60	0.2	AdamW	0.01	0.283	0.66	0.53	0.588	

60	0.2	AdamW	0.001	0.281	0.675	0.654	0.651		
100	0.0	AdamW	0.01	0.433	0.621	0.592	0.609		
100	0.0	AdamW	0.001	0.423	0.654	0.644	0.65		

100	0.1	AdamW	0.01	0.427	0.612	0.582	0.618		
100	0.1	AdamW	0.001	0.423	0.697	0.593	0.662		
100	0.2	AdamW	0.01	0.43	0.573	0.621	0.607		

