Public-Image Corpus Design and an Explainable, Robust Deep Learning Framework for Detection and Classification of Maca (*Lepidium meyenii*) Hypocotyl

Aradiel Castañeda Hilario, Doctor[1]; Mas Azahuanche Guillermo Antonio, Doctor[1]; ;Castillo Paredes, Omar Tupac Amaru, Msc[1] ; Reinoso Palacios, Artemio Rubén, Msc[1]; Zevallos vera , Erika Juana; Meza Ramos, Mihuller Rushbeer[1]; [Universidad Nacional del Callao, Perú--@unac.edu.pe: haradielc, gamasa, otacastillop, arreinosop, ejzevallosv, mrmezar]

Abstract*—Open image repositories contain photographs of maca, but they do not provide a ready-to-use benchmark for hypocotyl color phenotyping. This study develops a standalone, data-centric framework for constructing a secondary RGB corpus and for evaluating explainable and robust deep-learning models for detection and classification of maca (Lepidium meyenii) hypocotyl color phenotypes. The source audit identifies Wikimedia Commons, iNaturalist and GBIF as primary sources; EOL as a complementary aggregator; POWO and digital herbaria as taxonomic-validation resources; and the image collection used by PIS-Net as a conditional source requiring authorization. Because records may be replicated across aggregators, the framework distinguishes recovered records, canonical images, duplicate clusters and independent observations using canonical identifiers, cryptographic and perceptual hashes, authorship and provenance metadata. The revised experimental design adds an explicit feasibility analysis before defining the final classification task, compares a conventional image-level random split with a data-centric grouped workflow, distinguishes chromatic normalization from probabilistic calibration, formalizes confidence-based abstention, and quantifies Grad-CAM localization. At the present stage, the verified public-source audit supports the feasibility of constructing an auditable secondary corpus, but final effective sample size and model-performance estimates require completion of screening, annotation and computational experiments. Accordingly, no unexecuted metric is reported as validated evidence, and the scope remains bounded by the scarcity, heterogeneity and retrospective nature of public-image labels.

***Keywords*—maca, public image corpus, deep learning, explainable artificial intelligence, robustness.**

Diseño de un corpus de imágenes públicas y un framework de aprendizaje profundo explicable y robusto para la detección y clasificación de fenotipos de color del hipocótilo de maca (*Lepidium meyenii*).

Resumen*—Los repositorios abiertos contienen fotografías de maca, pero no ofrecen una base de datos directamente utilizable para el fenotipado cromático del hipocótilo. Este estudio desarrolla un marco autónomo y centrado en datos para construir un corpus secundario de imágenes RGB y evaluar modelos explicables y robustos destinados a detectar y clasificar fenotipos de color de maca (Lepidium meyenii). La auditoría de fuentes identifica a Wikimedia Commons, iNaturalist y GBIF como fuentes primarias; a EOL como agregador complementario; a POWO y los herbarios digitales como recursos de validación taxonómica; y al conjunto utilizado por PIS-Net como fuente condicionada a autorización. Debido a la posible replicación de registros entre agregadores, el diseño distingue registros recuperados, imágenes canónicas, clústeres de duplicados y observaciones independientes mediante identificadores canónicos, hashes criptográficos y perceptuales, autoría y metadatos de procedencia. La versión revisada incorpora un análisis explícito de viabilidad antes de definir la tarea final de clasificación, compara una partición aleatoria convencional por imagen con un workflow data-centric agrupado, distingue normalización cromática de calibración probabilística, formaliza la abstención basada en confianza y cuantifica la localización de Grad-CAM. En la etapa actualmente verificable, la auditoría confirma la factibilidad de construir un corpus secundario auditable; sin embargo, el tamaño efectivo final y las métricas de desempeño requieren completar el cribado, la anotación y los experimentos computacionales. Por ello, no se presentan métricas no ejecutadas como evidencia validada y el alcance se limita por la escasez, heterogeneidad y naturaleza retrospectiva de las etiquetas públicas.

***Palabras clave*—maca, corpus público de imágenes, aprendizaje profundo, inteligencia artificial explicable, robustez.**

I. INTRODUCCIÓN

En primer lugar, la maca es una especie andina de la familia Brassicaceae, cultivada principalmente por su eje hipocótilo-raíz engrosado. La literatura describe fenotipos amarillos, rojos, negros, violetas, crema y combinaciones intermedias, y ha demostrado que el color, el lugar de cultivo y las condiciones ambientales pueden relacionarse con variaciones en metabolitos secundarios [1]–[3]. Por ello, el color del hipocótilo constituye un atributo agronómico y comercial que requiere procedimientos reproducibles de documentación.

Asimismo, la identificación cromática manual puede verse afectada por suciedad superficial, deshidratación, madurez, sombras, reflejos, balance de blancos y diferencias entre dispositivos. El problema se vuelve más complejo cuando una imagen contiene varios hipocótilos, productos procesados, cortes, etiquetas o fondos que podrían ser apren

didos como correlaciones espurias por una red neuronal.

Por otra parte, el aprendizaje profundo ha mostrado capacidad para representar patrones visuales complejos mediante transferencia de aprendizaje. Sin embargo, su desempeño depende de la calidad de los datos, de la independencia entre entrenamiento y prueba, de la calibración de la confianza y de la evaluación frente a cambios de dominio. En un problema con pocos datos y procedencias heterogéneas, una partición aleatoria por archivo puede colocar copias o vistas del mismo registro en subconjuntos distintos y producir estimaciones artificialmente optimistas.

Actualmente no se ha identificado un benchmark público, equilibrado y etiquetado específicamente para detectar y clasificar fenotipos de color del hipocótilo de maca en imágenes RGB. En cambio, existen fotografías dispersas en repositorios de biodiversidad, colecciones multimedia y plataformas abiertas. Esta situación transforma el problema: antes de entrenar un modelo, es necesario diseñar un corpus secundario trazable, deduplicado y jurídicamente reutilizable.

En consecuencia, este artículo desarrolla una investigación metodológica y computacional centrada en datos para integrar fuentes públicas, establecer criterios reproducibles de inclusión, deduplicación y anotación, determinar empíricamente la viabilidad de la clasificación fenotípica y comparar un workflow convencional con un workflow data-centric. El diseño incorpora además calibración probabilística, abstención, explicabilidad cuantitativa y evaluación de robustez. En coherencia con el estado actual de ejecución, se diferencian explícitamente los hallazgos ya verificables de la auditoría de fuentes y los resultados computacionales que requieren completar el corpus definitivo.

II. PROBLEMA, PREGUNTAS Y CONTRIBUCIONES

A. Problema científico

El problema consiste en determinar si las fotografías públicas disponibles permiten construir un corpus suficientemente trazable e independiente para entrenar y evaluar un sistema que localice hipocótilos y clasifique su fenotipo cromático bajo variaciones de fuente, fondo y captura. La dificultad no es únicamente algorítmica: incluye escasez de imágenes pertinentes, ausencia de etiquetas fenotípicas estandarizadas, duplicación entre agregadores, incertidumbre cromática y restricciones de licencia.

B. Preguntas de investigación

PI1. ¿Qué repositorios contienen imágenes reutilizables de maca y qué función metodológica puede cumplir cada uno?
PI2. ¿Qué proporción del material recuperado muestra hipocótilos íntegros y permite una etiqueta cromática razonablemente confiable?
PI3. ¿Cómo deben deduplicarse y agruparse los registros para evitar que una misma observación aparezca en entrenamiento y prueba?
PI4. ¿En qué medida el recorte o segmentación del hipocótilo reduce la dependencia del fondo y mejora la coherencia de Grad-CAM?
PI5. ¿Cuánto se degrada el desempeño al evaluar una fuente, proveedor o condición visual no utilizada durante el entrenamiento?
PI6. ¿La cantidad, distribución e independencia de las observaciones disponibles después de la depuración permiten sostener

empíricamente una clasificación multiclase de los fenotipos cromáticos del hipocótilo?
PI7. ¿En qué medida el workflow data-centric modifica las estimaciones de desempeño, calibración y generalización respecto de un workflow convencional basado en partición aleatoria por imagen?

C. Contribuciones

Las contribuciones del estudio son: (1) una taxonomía funcional de fuentes públicas y condicionadas; (2) un protocolo reproducible de cosecha, licenciamiento, depuración, deduplicación y anotación retrospectiva; (3) una definición explícita de imagen canónica, clúster de duplicados y observación independiente; (4) una estrategia de partición agrupada compatible con registros públicos heterogéneos; (5) una caracterización cuantitativa del corpus después de su depuración; (6) un análisis de viabilidad que condiciona la formulación final de la tarea al soporte empírico disponible; (7) una comparación controlada entre un workflow convencional y el workflow data-centric; y (8) un marco de calibración, abstención, explicabilidad y robustez sensible al dominio de origen.

III. FUNDAMENTO BIOLÓGICO Y DESAFÍOS VISUALES

A. Fenotipos cromáticos

Los fenotipos de maca no se reducen a tres categorías rígidas. Aunque amarillo, rojo y negro son los grupos más difundidos, se han descrito tonalidades violetas, crema, grises y combinaciones bicolores [1]–[3]. Además, Clément et al. encontraron que el color y el sitio de cultivo deben considerarse conjuntamente al interpretar la composición de metabolitos [2], [3]. En términos computacionales, esto implica que una etiqueta visual puede presentar fronteras difusas y no siempre coincide con una separación cromática simple.

C. Factores de confusión

Una fotografía de maca puede representar la planta completa, hojas, hipocótilos frescos, hipocótilos secos, rodajas, harina, cápsulas o material de exhibición. El color visible también puede cambiar por exposición, compresión, temperatura de color, balance de blancos, suciedad, envejecimiento y posprocesamiento. Por ello, el corpus no debe mezclar indiscriminadamente órganos y estados de procesamiento. La clasificación fenotípica se limitará al subconjunto donde la superficie externa del hipocótilo sea visible y no existan transformaciones que invaliden la interpretación.

C. Dos tareas diferenciadas

La investigación separará dos tareas. La primera será la detección de hipocótilos en imágenes con uno o varios objetos. La segunda será la clasificación cromática de cada recorte detectado. Las imágenes de plantas sin hipocótilo visible podrán apoyar la identificación general de la especie o funcionar como material auxiliar, pero no ingresarán al clasificador de fenotipos. De manera análoga, las imágenes de polvo, cápsulas y productos comerciales se conservarán como categorías de exclusión o ejemplos negativos.

IV. FUENTES PÚBLICAS DE IMÁGENES Y TRABAJOS RELACIONADOS

A. Fuentes que integrarán el estudio

La Fig. 1 presenta la jerarquía de las fuentes. Wikimedia Commons, iNaturalist y GBIF proporcionarán el material primario. EOL se utilizará de manera complementaria después de comprobar que sus archivos no procedan ya de Wikimedia u otros proveedores. POWO y los herbarios digitales servirán principalmente para validación taxonómica. Finalmente, el conjunto empleado en PIS-Net solo se incorporará si los autores suministran las imágenes y una licencia de reutilización compatible.

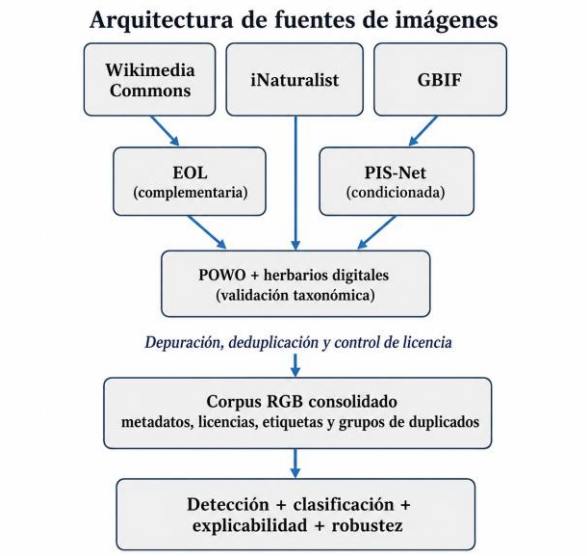


Fig. 1. Arquitectura de fuentes y función metodológica dentro del corpus consolidado.

TABLA I
FUENTES DE IMÁGENES QUE SERÁN UTILIZADAS

Fuente	Contenido, función y rol
Wikimedia Commons	Fuente primaria. 28 archivos con plantas, hipocótilos, rodajas, productos y escenas de cultivo [4]; la licencia se verificará por archivo.
iNaturalist	Fuente primaria. 28 observaciones con fotografías y metadatos [5]; todas las fotografías de una observación se mantendrán agrupadas.
GBIF	Fuente primaria. Registros de ocurrencia y extensión multimedia; puede agregar iNaturalist y otros proveedores [7]–[9], por lo que requiere deduplicación prioritaria.
EOL	Fuente complementaria. Colección multimedia con archivos de Wikimedia e instituciones botánicas [10]; solo se incorporarán imágenes independientes.
POWO y herbarios	Fuente de validación. Imágenes taxonómicas y especímenes curatoriales [11] para comprobar identidad y morfología, no como referencia principal de color.
PIS-Net	Fuente condicionada. Estudio que utilizó imágenes de maca para segmentación [12]; se incorporará únicamente si los autores proporcionan los datos y autorización de reutilización.

B. Dependencia entre repositorios

Las bases no pueden asumirse como independientes. GBIF no aloja necesariamente la fotografía original, sino que agrega enlaces y extensiones multimedia de proveedores [8], [9]. EOL también muestra archivos procedentes de Wikimedia Commons y de instituciones botánicas [10]. Por tanto, sumar el número de registros de cada portal sobreestimaría el tamaño real del corpus. La unidad contabilizada deberá ser la imagen canónica o el clúster de imágenes visualmente equivalentes, no el número de páginas donde aparece.

C. Antecedente computacional directo

Chen et al. propusieron PIS-Net para segmentación de instancias débilmente supervisada mediante puntos en escenarios agrícolas [12]. El artículo reportó 13 puntos por instancia, AP de 38,5 y AP50 de 68,3, equivalente aproximadamente al 90 % del desempeño completamente supervisado informado por los autores. Este antecedente demuestra que existen imágenes de maca utilizadas en una tarea de segmentación; no obstante, el objetivo fue el reconocimiento de malezas en arrozales y no la clasificación de colores del hipocótilo. Además, no se localizó un enlace de descarga directa del conjunto, por lo que su incorporación se considera condicionada.

D. Brecha de investigación

Los trabajos de maca se han concentrado en composición química, metabolitos, espectroscopía y autenticación. En contraste, no se identificó una base pública diseñada para combinar detección de múltiples hipocótilos, clasificación cromática, explicabilidad visual y pruebas de robustez. La brecha no se resolverá simplemente acumulando imágenes de internet: requiere control de procedencia, licencia, independencia estadística y validez de la etiqueta.

V. METODOLOGÍA

A. Tipo y diseño del estudio

Se plantea un estudio tecnológico aplicado, retrospectivo y centrado en datos públicos. La primera fase consiste en una auditoría y construcción de corpus; la segunda, en el entrenamiento y evaluación de modelos. El diseño no presupone que todas las imágenes recuperadas serán utilizables. Cada registro conservará su estado de inclusión, motivo de exclusión, fuente, licencia y nivel de confianza de la etiqueta.

B. Estrategia de búsqueda y cosecha

Las búsquedas utilizarán *Lepidium meyenii*, *Lepidium peruvianum*, "maca", "maca hypocotyl", "maca root", "hipocótilo de maca", "maca amarilla", "maca roja", "maca negra", "yellow maca", "red maca" y "black maca". Los nombres científicos se utilizarán como eje principal para reducir falsos positivos. Cuando la plataforma disponga de API o descarga estructurada, se priorizará ese mecanismo sobre la extracción manual. En GBIF se solicitará el formato Darwin Core Archive, conservando `occurrence.txt` y `multimedia.txt` [8].

C. Criterios de inclusión y exclusión

TABLA II

CRITERIOS DE SELECCIÓN DEL MATERIAL VISUAL

Criterio	Regla de decisión
Identidad	Incluir registros identificados como *L. meyenii* o sinónimo aceptado y con evidencia visual coherente; excluir especies dudosas o identificaciones contradictorias.
Órgano	Incluir hipocótilos visibles para color y plantas útiles para tareas auxiliares; excluir polvo, cápsulas y alimentos preparados del clasificador cromático.
Calidad	Incluir resolución suficiente, superficie visible y color no saturado; excluir desenfoque severo, compresión extrema, oclusión dominante o edición cromática evidente.
Licencia	Redistribuir CC0, CC BY o CC BY-SA; documentar otras licencias para uso interno y excluir derechos reservados sin permiso.
Independencia	Conservar el registro canónico y su clúster de duplicados; excluir copias sin trazabilidad o procedencia verificable.

D. Gestión de licencias y atribución

Cada imagen mantendrá autor, URL original, fecha de acceso, licencia y proveedor. Para un corpus redistribuible se priorizarán CC0, CC BY y CC BY-SA. Las fotografías con cláusula no comercial podrán analizarse internamente solo si el uso previsto y la política editorial lo permiten, pero no se incluirán automáticamente en un paquete público. Las imágenes con derechos reservados se excluirán salvo autorización escrita. iNaturalist advierte que las licencias de la observación y de cada fotografía pueden ser diferentes [6].

E. Deduplicación, agrupamiento y definición de la unidad independiente

La deduplicación operará en cuatro niveles: coincidencia de URL canónica; hash criptográfico de archivo; hash perceptual para copias redimensionadas, recortadas o recomprimidas; y revisión de coincidencias por autor, fecha, descripción y observación. Todas las variantes de una imagen se asignarán a un mismo `duplicate_cluster_id`. En iNaturalist, todas las fotografías de una misma observación compartirán un `group_id`. Ningún grupo podrá aparecer simultáneamente en entrenamiento, validación y prueba.

Para evitar que el volumen bruto de archivos sobreestime el tamaño efectivo del corpus, se distinguirán cuatro niveles de conteo: registro recuperado, archivo de imagen, imagen canónica y observación independiente. Las copias exactas o perceptualmente equivalentes se asignarán al mismo `duplicate_cluster_id`, mientras que las fotografías correspondientes a una misma observación biológica compartirán un `group_id`. En consecuencia, el tamaño efectivo utilizado para la caracterización estadística y la partición experimental se expresará principalmente mediante el número de observaciones independientes, complementado por el número de imágenes canónicas y el soporte por fenotipo

F. Esquema de anotación

La anotación se realizará en tres niveles. Primero, se clasificará el tipo de contenido: hipocótilo íntegro, hipocótilo cortado, material seco, planta completa, parte aérea, producto procesado o no utilizable. Segundo, cuando sea válido, se asignará el fenotipo amarillo, rojo, negro, violeta, blanco/crema, mixto o indeterminado. Tercero, se generarán cajas delimitadoras y, en una submuestra, máscaras de segmentación. Dos evaluadores independientes revisarán las etiquetas y la concordancia se estimará con kappa de Cohen [25]. Los casos sin consenso permanecerán como indeterminados y no serán forzados a una clase.

G. Metadatos mínimos

El archivo maestro contendrá image_id, source_database, source_record_id, original_url, canonical_media_url, scientific_name, image_type, phenotype_label, label_confidence, author, license, country, locality, capture_date, duplicate_cluster_id, group_id, quality_status, exclusion_reason y split. Los metadatos no se utilizarán como variables predictoras cuando puedan revelar la clase; su función será documentar procedencia, estratificar el análisis y detectar sesgos.

H. Control y normalización cromática

Debido al carácter retrospectivo y multifuente del corpus, no se asumirá equivalencia colorimétrica absoluta entre fotografías que carezcan de una referencia cromática conocida. Las imágenes serán examinadas para identificar saturación severa, dominantes cromáticas extremas, edición evidente y condiciones de iluminación que comprometan la interpretación del fenotipo. Los casos no interpretables serán excluidos de la clasificación cromática o permanecerán como indeterminados. Cuando resulte metodológicamente viable, se evaluará una estrategia reproducible de normalización cromática conservando las imágenes originales como condición de referencia, de modo que su efecto pueda cuantificarse mediante una comparación controlada. La normalización cromática se distinguirá explícitamente de la calibración probabilística del clasificador.

I. Análisis de viabilidad de la tarea de clasificación

Antes del entrenamiento se evaluará la viabilidad de la clasificación fenotípica a partir del número efectivo de observaciones independientes por clase, su distribución entre proveedores, la calidad de las imágenes, el nivel de confianza de las etiquetas y la concordancia entre evaluadores. Las categorías con soporte insuficiente, elevada incertidumbre o concordancia inadecuada podrán fusionarse con una categoría fenotípicamente justificable, excluirse del análisis confirmatorio o mantenerse como indeterminadas. Sobre esta evidencia se definirá la formulación experimental final como clasificación multiclase completa, multiclase reducida, binaria o estudio de viabilidad. Los aumentos de datos, recortes derivados y duplicados no se contabilizarán como nuevas observaciones independientes.

J. Partición experimental

La partición inicial será 70/15/15 para entrenamiento, validación y prueba, agrupada por observación y clúster de duplicados. Si el corpus final es pequeño, se utilizará validación cruzada agrupada y repetida. Además, se realizará una evaluación externa por fuente o proveedor: el modelo se entrenará sin un dominio y se evaluará en ese dominio retenido. Debido a que GBIF y EOL pueden contener imágenes provenientes de iNaturalist o Wikimedia, la definición de dominio se basará en el proveedor original, no únicamente en el portal de acceso.

K. Workflow convencional y workflow data-centric propuesto

Para cuantificar el efecto de la metodología propuesta se establecerá una línea base convencional basada en partición aleatoria por imagen y entrenamiento mediante transferencia de aprendizaje. Esta condición se comparará con el workflow data-centric, que incorporará deduplicación, agrupamiento por observación y partición agrupada antes del entrenamiento. Cuando corresponda, ambas condiciones utilizarán la misma arquitectura, configuración de entrenamiento, semillas y métricas, con el propósito de atribuir las diferencias observadas principalmente al procedimiento de construcción y partición del corpus y no al cambio de modelo. De forma complementaria, se evaluará una configuración ampliada que incorpore calibración probabilística y abstención. Un desempeño puntual superior del workflow convencional no se interpretará automáticamente como evidencia de mayor generalización si existe dependencia entre imágenes o fuga entre particiones.



Fig. 2. Flujo reproducible para construir el corpus y evaluar el sistema.

L. Arquitecturas y entrenamiento

La detección se evaluará con modelos YOLO de diferente capacidad, incluyendo una variante compacta y YOLOv5X como referencia de alta capacidad. Los recortes detectados alimentarán clasificadores Inception V3, ResNet-50, EfficientNet-B0 y MobileNetV3, inicializados con pesos de ImageNet [13]–[18]. En una primera etapa se congelará el extractor; posteriormente se aplicará ajuste fino con tasa de aprendizaje reducida. La selección se realizará exclusivamente con validación y F1 macro, complementado por calibración y latencia.

TABLA III

CONFIGURACIÓN EXPERIMENTAL Y CONTROLES

Componente	Especificación y control
Partición	70/15/15 o validación cruzada agrupada; ningún grupo ni duplicado podrá aparecer en subconjuntos distintos.
Aumento	Rotación, escala, recorte, brillo, contraste y desenfoque moderado, aplicados solo al entrenamiento.
Desbalance	Pesos de clase, muestreo estratificado y F1 macro; reportar soporte y resultados por clase.
Optimización	AdamW o SGD, parada temprana y scheduler; registrar semillas, versiones, hardware y tiempo.
Calibración	Temperature scaling en validación [22]; seleccionar umbrales sin utilizar el conjunto de prueba.
Rechazo	Abstención por confianza y detección fuera de distribución; revisión humana en casos ambiguos.

M. Calibración probabilística y criterio de abstención

Las probabilidades del clasificador se calibrarán mediante temperature scaling ajustado exclusivamente sobre el conjunto de validación. La calidad de la calibración se evaluará mediante Expected Calibration Error (ECE), Brier score y diagramas de confiabilidad. El sistema incorporará además una regla explícita de abstención: para una observación x, se aceptará la clase de máxima probabilidad únicamente cuando p_max(x) = max_k p(y=k|x) sea mayor o igual que un umbral τ; en caso contrario, la salida será "indeterminado/revisión humana". El valor de τ se seleccionará exclusivamente en validación y permanecerá fijo durante la evaluación final. Se reportarán cobertura, tasa de abstención, error selectivo y desempeño de las predicciones aceptadas, evitando seleccionar el punto operativo a partir del conjunto de prueba.

N. Explicabilidad y métrica de localización

Se generarán mapas Grad-CAM para aciertos, errores y casos de alta confianza incorrecta [19]. Además de la revisión visual, la localización de la explicación se cuantificará mediante la proporción de energía dentro del objeto: $E_{obj} = \Sigma[H(x,y)\cdot M(x,y)] / \Sigma H(x,y)$, donde H(x,y) representa la intensidad no negativa del mapa Grad-CAM y M(x,y) la máscara binaria —o, cuando no exista máscara, la caja delimitadora— del hipocótilo. E_{obj} tomará valores entre 0 y 1, donde valores mayores indican una mayor concentración de la activación dentro del objeto. Esta medida se complementará con una prueba contrafactual contextual: primero se ocultará el hipocótilo conservando el fondo y, posteriormente, se neutralizará el fondo conservando el objeto. La persistencia de la predicción después de ocultar el hipocótilo se interpretará como evidencia de posible dependencia contextual, no como prueba causal definitiva.

O. Robustez y generalización

Las pruebas de robustez priorizarán perturbaciones directamente relacionadas con el problema: iluminación y temperatura de color, contraste/saturación, desenfoque o compresión, variaciones de fondo y, cuando el tamaño efectivo lo permita, ruido, rotación, escala y oclusión [20]. La robustez relativa se expresará mediante la variación de F1 o mAP respecto del conjunto limpio, reportando también la condición de peor desempeño. La generalización de dominio se evaluará mediante fuente o proveedor retenido y, cuando exista metadato suficiente, país o localidad retenida [21]. La selección final de perturbaciones se limitará a aquellas que puedan evaluarse con soporte suficiente y sin presentar análisis no ejecutados como evidencia.

P. Métricas y análisis estadístico

TABLA IV
MÉTRICAS DE EVALUACIÓN

Dimensión	Medidas
Clasificación	Exactitud balanceada, precisión, sensibilidad, especificidad, F1 macro y por clase, AUC uno-contra-resto.
Detección	AP por clase, mAP@0,5, mAP@0,5:0,95, IoU, sensibilidad y latencia.
Calibración	Expected Calibration Error, Brier score y diagramas de confiabilidad.
Explicabilidad	Energía dentro del objeto, revisión experta y comportamiento contrafactual.
Robustez	Degradación relativa por perturbación y peor dominio.
Incertidumbre	Intervalos bootstrap del 95 %, variación entre semillas y soporte por clase.

La comparación entre clasificadores utilizará las mismas imágenes y particiones. Se reportarán varias semillas y se estimarán intervalos bootstrap. Cuando corresponda, se emplearán pruebas pareadas sobre errores de clasificación. No se interpretará como mejora sustantiva una diferencia pequeña sin intervalo de confianza, sin estabilidad entre semillas o acompañada de peor calibración.

Q. Auditoría de licencias y procedencia

La aptitud jurídica de cada archivo se evaluará de manera independiente de su disponibilidad técnica. Para cada imagen se conservarán la licencia declarada, el titular de los derechos, la plataforma de acceso, el proveedor original y la fecha de consulta. Cuando una plataforma actúe como agregador, la licencia válida será la indicada por el proveedor primario y no la del portal intermediario. Los registros con derechos reservados, licencia ausente o términos incompatibles con la redistribución se mantendrán fuera del paquete público, aunque sus metadatos puedan emplearse para documentar la búsqueda.

Asimismo, se generará una matriz de atribución que vincule cada image_id con autor, URL canónica, licencia y texto de reconocimiento requerido. Este control reduce el riesgo de incorporar imágenes técnicamente accesibles pero jurídicamente no reutilizables. La versión liberada del corpus contendrá únicamente archivos cuya licencia permita la redistribución; para el resto se publicarán identificadores, hashes y scripts de recuperación, siempre que las condiciones del proveedor lo autoricen.

R. Validez de las etiquetas y concordancia

La etiqueta cromática se considerará una inferencia retrospectiva y no una determinación química o genética. Dos evaluadores revisarán de forma independiente la superficie externa del hipocótilo, su estado de conservación y la consistencia del color en distintas regiones del objeto. Además de la clase nominal, se registrará el nivel de confianza —alta, media o baja— y el motivo de incertidumbre, por ejemplo, iluminación dominante, suciedad, deshidratación, corte del órgano o posprocesamiento. Los desacuerdos se resolverán por consenso y los casos persistentes permanecerán en la clase indeterminada.

La concordancia se informará mediante kappa de Cohen y porcentaje de acuerdo [25]. No se utilizará un único valor global para ocultar clases problemáticas: también se presentará una matriz de desacuerdos y el soporte por fenotipo. Si una categoría exhibe baja concordancia o un número insuficiente de observaciones independientes, se fusionará con una clase más amplia o se excluirá del análisis confirmatorio. Esta regla evita que el modelo aprenda etiquetas cuya incertidumbre supera la variación que se pretende reconocer.

S. Análisis estadístico preespecificado

El análisis se definirá antes de abrir el conjunto de prueba. Para clasificación se priorizarán F1 macro, exactitud balanceada y métricas por clase; para detección se utilizarán AP, mAP e IoU. Los intervalos de confianza se estimarán mediante remuestreo bootstrap agrupado por observación, de modo que todas las imágenes pertenecientes a un mismo grupo se remuestreen conjuntamente. La comparación de clasificadores se realizará sobre las mismas particiones y semillas, evitando comparar ejecuciones obtenidas con conjuntos diferentes.

Las decisiones no dependerán exclusivamente del mayor valor puntual. Se considerarán simultáneamente la estabilidad entre semillas, la calibración, el peor dominio, la latencia y la coherencia de las explicaciones. Cuando dos modelos presenten intervalos ampliamente superpuestos, se preferirá la alternativa más simple o eficiente, siempre que mantenga un desempeño equilibrado. Todas las pruebas exploratorias se distinguirán de los análisis confirmatorios en el reporte final.

VI. ARQUITECTURA DEL CORPUS Y DEL SISTEMA

A. Organización de archivos

El corpus se organizará por fuente original y no por clase durante la fase de cosecha. Después de la auditoría se generarán directorios de imágenes aceptadas, excluidas, anotaciones, metadatos y licencias. Esta separación evita perder el contexto de procedencia y permite reconstruir cada decisión. Los recortes y aumentos se considerarán derivados; nunca sustituirán el archivo original.

B. Módulos funcionales

El sistema incluirá módulos de cosecha, control de integridad, deduplicación, revisión taxonómica, anotación, entrenamiento, evaluación, explicabilidad y reporte. Cada predicción conservará identificador de imagen, modelo, versión, umbral, clase, probabilidad calibrada, caja, mapa Grad-CAM y condición de rechazo. La interfaz no devolverá únicamente una etiqueta textual.

C. Ciencia abierta

La documentación seguirá principios de fichas de datos y tarjetas de modelo [23], [24]. El repositorio publicará scripts de descarga cuando las condiciones de uso lo permitan, hashes, metadatos, particiones, diccionario de datos, configuración y métricas. Las imágenes solo se redistribuirán si su licencia lo autoriza. Cuando no sea posible, el registro conservará la URL y la información necesaria para reproducir la descarga desde el proveedor.

VII. RESULTADOS

A. Construcción, depuración y caracterización del corpus

En primer lugar, la auditoría de fuentes confirmó que la disponibilidad bruta de fotografías no equivale al tamaño efectivo del corpus. En la fecha de corte documentada, Wikimedia Commons registraba 28 archivos en la categoría Lepidium meyenii [4], mientras que iNaturalist registraba 28 observaciones [5]. La inspección de estas fuentes mostró una composición heterogénea que incluye plantas completas, hipocótilos, rodajas secas, escenas de cultivo, exhibiciones y productos procesados. Por consiguiente, estos conteos se consideran registros de descubrimiento y no observaciones fenotípicas independientes.

Asimismo, la revisión de procedencia confirmó que los portales consultados no pueden sumarse como fuentes independientes. GBIF puede exponer extensiones multimedia provenientes de iNaturalist u otros publicadores [7]–[9], mientras que EOL incorpora material procedente de Wikimedia Commons y de instituciones botánicas [10]. En consecuencia, el tamaño efectivo deberá establecerse después de aplicar URL canónica, hash criptográfico, hash perceptual, duplicate_cluster_id y group_id. Con la evidencia actualmente disponible en el manuscrito no es metodológicamente válido informar todavía un número final de imágenes canónicas, clústeres de duplicados, hipocótilos utilizables u observaciones independientes; esos valores deberán provenir del archivo maestro generado por la depuración y no de una extrapolación de los conteos de los portales.

Por otra parte, las fuentes mostraron utilidades diferentes. Wikimedia Commons aporta archivos de alta resolución y licencias explícitas [4]; iNaturalist aporta unidades de observación y metadatos asociados [5], [6]; GBIF facilita estructura Darwin Core y acceso a múltiples proveedores [7]–[9]; EOL funciona principalmente como mecanismo complementario de descubrimiento [10]; y POWO y los herbarios digitales son más apropiados para validación taxonómica que para establecer el color original de hipocótilos frescos [11]. PIS-Net se mantuvo como fuente condicionada porque el estudio utiliza imágenes de maca para segmentación, pero no se identificó en la revisión documental un conjunto directamente reutilizable con licencia verificada [12].

B. Viabilidad de la tarea de clasificación

En segundo lugar, la auditoría confirma que la viabilidad de una clasificación cromática multiclase no puede determinarse a partir de los 28 archivos de Wikimedia Commons y las 28 observaciones de iNaturalist por sí solos. La unidad relevante es la observación independiente después de deduplicación, control de calidad y revisión de etiquetas. En particular, la presencia de imágenes sin hipocótilo visible, material seco o procesado y fotografías con condiciones cromáticas no controladas reduce el conjunto potencialmente clasificable. Por ello, la decisión entre clasificación multiclase completa, multiclase reducida, tarea binaria o estudio de viabilidad quedará condicionada al soporte independiente por fenotipo y a la concordancia entre evaluadores, estimada mediante kappa de Cohen [25].

Este resultado evita convertir aumentos de datos, recortes derivados o múltiples fotografías de una misma observación en evidencia de mayor diversidad biológica. Hasta completar el cribado y la doble anotación, no se atribuye viabilidad confirmatoria a ninguna formulación multiclase. La formulación definitiva deberá quedar registrada antes de abrir el conjunto de prueba.

C. Desempeño cuantitativo de los modelos

En tercer lugar, el manuscrito revisado preespecifica F1 macro como métrica primaria de clasificación, complementada por exactitud balanceada, precisión, sensibilidad, especificidad, F1 por clase y AUC uno-contra-resto. Para detección se consideran AP, mAP e IoU, y la incertidumbre se expresará mediante intervalos bootstrap del 95 % agrupados por observación. Sin embargo, el material documental disponible para esta revisión no contiene ejecuciones verificables de los clasificadores o detectores ni archivos de predicciones a partir de los cuales calcular estas métricas. En consecuencia, no se incorporan valores simulados ni estimaciones hipotéticas como resultados. Este apartado deberá completarse únicamente con salidas reproducibles del experimento definido en la Sección V.

D. Comparación entre workflow convencional y data-centric

En cuarto lugar, la comparación metodológica queda definida sobre una diferencia controlada: una línea base con partición aleatoria por imagen frente al workflow data-centric con deduplicación, agrupamiento por observación y partición agrupada. Esta comparación responde al riesgo de que copias o vistas relacionadas aparezcan en entrenamiento y prueba y produzcan estimaciones artificialmente optimistas. Para que la diferencia sea atribuible al workflow, ambas condiciones deberán emplear, cuando corresponda, la misma arquitectura, configuración de entrenamiento, semillas y métricas. A la fecha del documento, no existen resultados experimentales verificables que permitan cuantificar la diferencia entre ambas condiciones; por ello, no se afirma superioridad de ninguna de ellas.

E. Calibración, abstención y explicabilidad

En quinto lugar, se establecieron tres mecanismos complementarios para evaluar la confiabilidad de las predicciones. La variabilidad cromática se tratará mediante control de calidad y, cuando sea viable, normalización cromática, sin equipararla a calibración colorimétrica absoluta en fotografías que carecen de una referencia física. La calibración probabilística se evaluará mediante temperature scaling, Expected Calibration Error y Brier score [22]. Finalmente, el criterio de abstención utilizará un umbral τ seleccionado exclusivamente en validación y reportará cobertura, tasa de abstención y error selectivo.

La explicabilidad se evaluará mediante Grad-CAM [19] y la proporción de energía de activación localizada dentro del hipocótilo, E_obj. Esta medida se complementará con ocultamiento del objeto y neutralización del fondo para identificar posible dependencia contextual. Debido a que todavía no se dispone de mapas, máscaras y predicciones experimentales consolidadas en los materiales suministrados, este apartado define los resultados que deben reportarse, pero

no atribuye valores de E_obj ni efectos contrafactuales no observados.

F. Robustez, generalización y análisis de errores

Finalmente, la heterogeneidad entre repositorios constituye una fuente observable de variación de dominio y justifica una evaluación por proveedor retenido. Las pruebas priorizarán cambios de iluminación y temperatura de color, desenfoque o compresión, variaciones de fondo y dominio de origen, siguiendo principios de evaluación de robustez frente a cambios de distribución [20], [21]. El reporte final deberá incluir la degradación respecto del conjunto limpio y la condición de peor desempeño. Dado que el manuscrito no contiene todavía ejecuciones verificables de estas pruebas, no se presentan magnitudes de degradación ni conclusiones sobre generalización externa.

VIII. DISCUSIÓN

A. Del volumen bruto al tamaño efectivo del corpus

En primer lugar, los resultados de la auditoría respaldan que el principal desafío no es acumular enlaces o archivos, sino establecer cuántas observaciones independientes y cromáticamente interpretables permanecen después de controlar procedencia, duplicación, calidad y licencia. Los 28 archivos identificados en Wikimedia Commons [4] y las 28 observaciones registradas en iNaturalist [5] describen disponibilidad inicial, pero no constituyen por sí mismos un tamaño muestral efectivo. Esta distinción es especialmente relevante cuando GBIF y EOL actúan como agregadores de material disponible en otras plataformas [8]–[10].

Desde una perspectiva data-centric, esta reducción potencial del tamaño aparente no debe interpretarse como pérdida de información, sino como una condición necesaria para evitar pseudorreplicación y fuga entre particiones. La documentación mediante identificadores canónicos, hashes y grupos de observación también se alinea con prácticas de documentación explícita de conjuntos de datos y modelos [23], [24].

B. Viabilidad y límites de la clasificación fenotípica

En segundo lugar, la evidencia disponible aconseja separar la existencia de fotografías de maca de la existencia de ejemplos válidos para clasificación cromática. Una imagen taxonómicamente correcta puede carecer de hipocótilo visible o mostrar material seco, cortado o procesado, mientras que iluminación, balance de blancos, suciedad y posprocesamiento pueden alterar el color aparente. Por ello, la viabilidad de una tarea multiclase debe sustentarse en el número efectivo por fenotipo, la calidad de la etiqueta y la concordancia entre evaluadores, y no en el número total de archivos recuperados.

Además, incluso una clasificación visual reproducible debe interpretarse dentro de su constructo. El color inferido retrospectivamente a partir de una fotografía no demuestra composición química, procedencia geográfica o identidad varietal. La literatura sobre maca muestra asociaciones entre color, condiciones de cultivo y metabolitos [1]–[3], pero esas asociaciones no autorizan a inferir dichas propiedades a partir de una etiqueta visual aislada. En consecuencia, el alcance del sistema debe mantenerse en la documentación y preclasificación del fenotipo cromático observable.

C. Efecto esperado de controlar dependencia y fuga de información

En tercer lugar, la comparación entre partición aleatoria y partición agrupada es central para interpretar cualquier métrica posterior. Si varias fotografías de una misma observación o copias perceptualmente equivalentes aparecen en subconjuntos diferentes, el modelo puede beneficiarse de similitudes que no representan generalización hacia nuevas observaciones. Por ello, una eventual disminución del desempeño al aplicar grouped split no deberá interpretarse automáticamente como deterioro del método; podría reflejar una estimación más conservadora y más próxima al escenario de generalización que se pretende estudiar.

Por la misma razón, la comparación con un workflow convencional debe centrarse no solo en F1 macro, sino también en estabilidad entre semillas, calibración, peor clase, peor dominio y presencia de fuga. La superioridad metodológica no puede deducirse de una única diferencia puntual sin intervalos de confianza y sin mantener constantes los demás componentes experimentales.

D. Calibración, abstención y utilidad predictiva

En cuarto lugar, la confiabilidad del sistema requiere distinguir dos problemas diferentes. El primero es cromático: las imágenes públicas no poseen necesariamente una referencia física que permita reconstruir el color real. El segundo es probabilístico: una red neuronal puede producir probabilidades excesivamente confiadas incluso cuando su clasificación es incorrecta. El uso de temperature scaling y métricas como ECE y Brier score aborda este segundo problema [22], pero no corrige por sí mismo la incertidumbre de adquisición de la imagen.

La abstención constituye, por tanto, un mecanismo operativo para evitar que toda entrada reciba obligatoriamente una etiqueta. La selección de τ exclusivamente en validación y el análisis conjunto de cobertura y error selectivo permitirán evaluar el compromiso entre responder con mayor frecuencia y restringir las predicciones a casos de mayor confianza. Su utilidad deberá juzgarse con los resultados reales y no mediante un umbral fijado retrospectivamente sobre el conjunto de prueba.

E. Explicabilidad y dependencia contextual

En quinto lugar, Grad-CAM debe utilizarse como instrumento de auditoría y no únicamente como recurso ilustrativo [19]. En este problema, una activación concentrada sobre etiquetas comerciales, fondos, recipientes o elementos de exhibición podría revelar correlaciones contextuales que comprometan la interpretación del clasificador. La métrica E_obj permite convertir parte de esta inspección en una medida reproducible de localización, aunque una alta concentración espacial tampoco constituye por sí sola evidencia causal.

Por ello, el análisis contrafactual mediante ocultamiento del hipocótilo y neutralización del fondo resulta complementario. La persistencia de una predicción después de retirar visualmente el objeto sería consistente con dependencia contextual y debería analizarse especialmente en errores de alta confianza. Las conclusiones sobre este punto permanecerán

abiertas hasta disponer de los mapas y pruebas contrafactuales generados por los modelos entrenados.

F. Robustez, generalización y alcance externo

Finalmente, la diversidad de plataformas puede aprovecharse como una prueba de cambio de dominio, pero no sustituye una validación agronómica independiente. La evaluación leave-one-source/provider-out permitirá estimar cuánto depende el sistema de un proveedor digital particular, mientras que las perturbaciones controladas permitirán cuantificar sensibilidad a cambios visuales [20], [21]. Aun con resultados favorables, la generalización hacia centros de acopio, parcelas o dispositivos móviles en condiciones productivas requerirá un conjunto primario independiente.

En conjunto, la principal contribución potencial del enfoque radica en hacer explícita la relación entre calidad del dato, independencia estadística y credibilidad de la evaluación. Esta interpretación es coherente con la orientación del manuscrito hacia documentación reproducible del corpus y del modelo [23], [24]. No obstante, mientras no se completen la depuración, el etiquetado y las ejecuciones experimentales, las conclusiones cuantitativas deben permanecer restringidas a la auditoría de fuentes y al diseño metodológico verificable.

IX. AMENAZAS A LA VALIDEZ, LIMITACIONES Y REPRODUCIBILIDAD

A. Validez interna

Las principales amenazas internas son la fuga de información entre particiones, la presencia de duplicados residuales, los desacuerdos de anotación y la selección retrospectiva de umbrales. El diseño metodológico reduce estos riesgos mediante agrupamiento por observación y clúster de duplicados, bloqueo del conjunto de prueba, selección de umbrales exclusivamente en validación y trazabilidad de las transformaciones. No obstante, mientras el corpus definitivo y los experimentos no hayan sido completados, estos controles deben entenderse como procedimientos preespecificados y no como evidencia de que todo riesgo residual haya sido eliminado.

B. Validez de constructo

El color observado en una fotografía pública no equivale necesariamente al fenotipo agronómico medido bajo condiciones controladas, ni permite inferir composición química, identidad genética o procedencia geográfica. La ausencia de cartas de color impide asumir calibración colorimétrica absoluta. Por ello, la variable objetivo se restringe al fenotipo cromático visual del hipocótilo bajo la evidencia disponible, y la normalización cromática se evalúa como procedimiento computacional separado de la calibración probabilística del clasificador [1]–[3], [22].

C. Validez externa

La validez externa está condicionada por la distribución de las imágenes públicas, los proveedores, los fotógrafos y las condiciones de captura. Un desempeño favorable en repositorios abiertos no garantiza funcionamiento en parcelas, centros de acopio o dispositivos móviles. La evaluación por fuente o proveedor retenido permitirá estimar generalización entre dominios digitales [21], pero una validación agronómica o comercial requerirá posteriormente un conjunto primario, controlado e independiente de hipocótilos procedentes de contextos productivos relevantes.

D. Validez estadística

La inferencia dependerá del número de observaciones independientes y no del número bruto de archivos. Se reportarán soporte por clase, F1-macro, exactitud balanceada, intervalos bootstrap del 95 % agrupados por observación y variación entre semillas. Las diferencias pequeñas no se interpretarán como mejoras sustantivas cuando presenten intervalos ampliamente superpuestos, inestabilidad entre semillas, peor calibración o degradación marcada en el dominio de peor desempeño. Las clases con soporte insuficiente o concordancia inadecuada no serán forzadas a un análisis multiclase confirmatorio.

E. Reproducibilidad y ciencia abierta

La fecha de auditoría, consultas, versiones de API, entornos de software, semillas, particiones y reglas de inclusión se conservarán de forma versionada. La ficha del corpus documentará composición, licencias, exclusiones y procedencia, siguiendo principios de documentación de conjuntos de datos [23]. La tarjeta del modelo registrará propósito, clases, dominios no cubiertos, métricas, umbral de abstención, hardware y limitaciones [24]. Las imágenes solo se redistribuirán cuando su licencia lo autorice; en los demás casos se conservarán identificadores, URL, hashes y scripts de recuperación cuando los términos del proveedor lo permitan. **X. CONCLUSIONES**

En primer lugar, la revisión de fuentes confirma que no existe un benchmark público listo para entrenar y validar directamente un clasificador de fenotipos cromáticos del hipocótilo de maca. Wikimedia Commons, iNaturalist y GBIF constituyen fuentes potencialmente efectivas; EOL amplía el descubrimiento; POWO y los herbarios digitales apoyan la validación taxonómica; y PIS-Net permanece como fuente condicionada al acceso y a la licencia [4]–[12]. El aporte verificable en esta etapa es, por tanto, la definición de una estrategia para transformar registros dispersos en un corpus secundario trazable, deduplicado y auditable.

En segundo lugar, la revisión metodológica muestra que el tamaño bruto de archivos no debe utilizarse como sustituto del tamaño efectivo de muestra. La unidad relevante para la evaluación será la observación independiente después de controlar imágenes canónicas y clústeres de duplicados. La formulación final de la clasificación —multiclase completa, multiclase reducida, binaria o estudio de viabilidad— quedará condicionada al soporte por fenotipo, la concordancia de las etiquetas y la diversidad entre proveedores, evitando inflar el tamaño muestral mediante aumentos, recortes o copias derivadas.

En tercer lugar, el diseño experimental queda preparado para cuantificar el efecto del enfoque data-centric mediante una

comparación controlada con un workflow convencional. La partición aleatoria por imagen se contrastará con deduplicación, agrupamiento por observación y partición agrupada, utilizando configuraciones comparables. Asimismo, la confianza del sistema se evaluará mediante calibración probabilística, abstención seleccionada en validación y explicabilidad cuantitativa basada en Grad-CAM y localización sobre el objeto [19], [22].

Finalmente, las conclusiones predictivas definitivas requieren completar el cribado, la anotación y los experimentos computacionales. Por esta razón, esta versión no atribuye valores de F1, kappa, ECE, Brier score, umbrales de abstención, \(E_{\mathrm{obj}}\) ni intervalos de confianza que todavía no hayan sido obtenidos. El alcance científico permanece limitado al fenotipo cromático visual inferible desde imágenes públicas y no se extiende a composición química, autenticidad geográfica o identidad varietal. Una validación aplicada requerirá posteriormente datos primarios y controlados del contexto productivo peruano.

DECLARACIÓN DE DISPONIBILIDAD DE DATOS

Las imágenes cuya licencia permita redistribución se publicarán junto con sus metadatos y atribuciones. Para los registros no redistribuibles se compartirán, cuando los términos del proveedor lo permitan, URL de procedencia, identificadores, hashes, decisiones de inclusión y scripts de cosecha. Las particiones, clústeres de duplicados y versiones del archivo maestro se congelarán antes de la evaluación final, con el fin de preservar la trazabilidad y evitar modificaciones posteriores al análisis.

AGRADECIMIENTOS

Los agradecimientos institucionales, la Universidad Nacional del Callao, la financiación y las contribuciones individuales se incorporarán después de la revisión por pares, de acuerdo con el procedimiento de evaluación doble ciego.

REFERENCIAS

[1] G. F. Gonzales, L. Villaorduna, M. Gasco, J. Rubio, and C. Gonzales, "Maca (Lepidium meyenii Walp), a review of its biological properties," Revista Peruana de Medicina Experimental y Salud Pública, vol. 31, no. 1, pp. 100–110, 2014, doi: 10.17843/rpmesp.2014.311.15. https://pubmed.ncbi.nlm.nih.gov/24718534/

[2] C. Clément et al., "Influence of colour type and previous cultivation on secondary metabolites in hypocotyls and leaves of maca (Lepidium meyenii Walpers)," Journal of the Science of Food and Agriculture, vol. 90, no. 5, pp. 861–869, 2010, doi: 10.1002/jsfa.3896. https://doi.org/10.1002/jsfa.3896

[3] C. Clément et al., "Secondary metabolites in maca as affected by hypocotyl color, cultivation history, and site," Agronomy Journal, vol. 102, no. 2, pp. 431–439, 2010, doi: 10.2134/agronj2009.0315. https://doi.org/10.2134/agronj2009.0315

[4] Wikimedia Commons, "Category: Lepidium meyenii," accessed Jul. 25, 2026. https://commons.wikimedia.org/wiki/Category:Lepidium_meyenii

[5] iNaturalist, "Lepidium meyenii," accessed Jul. 25, 2026. https://www.inaturalist.org/taxa/372676-Lepidium-meyenii

[6] iNaturalist, "Can I use the photos and sounds that are posted on iNaturalist?," accessed Jul. 25, 2026. https://help.inaturalist.org/en/support/solutions/articles/151000169918

[7] Global Biodiversity Information Facility, "Lepidium meyenii Walp.," accessed Jul. 25, 2026. https://www.gbif.org/taxon/6PBH6

[8] Global Biodiversity Information Facility, "Occurrence download formats," accessed Jul. 25, 2026. https://techdocs.gbif.org/en/data-use/download-formats

[9] Global Biodiversity Information Facility, "Sharing images, sounds and videos on GBIF," accessed Jul. 25, 2026. https://techdocs.gbif.org/en/data-publishing/multimedia-publishing

[10] Encyclopedia of Life, "Lepidium meyenii Walp.: Media," accessed Jul. 25, 2026. https://eol.org/pages/483599/media

[11] Royal Botanic Gardens, Kew, "Lepidium meyenii Walp.," Plants of the World Online, accessed Jul. 25, 2026. https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:286286-1/images

[12] H. Chen et al., "PIS-Net: Efficient weakly supervised instance segmentation network based on annotated points for rice field weed identification," Smart Agricultural Technology, vol. 9, art. no. 100557, 2024, doi: 10.1016/j.atech.2024.100557. https://doi.org/10.1016/j.atech.2024.100557

[13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[14] C. Szegedy et al., "Rethinking the Inception architecture for computer vision," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[16] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th Int. Conf. Machine Learning, vol. 97, 2019, pp. 6105–6114. https://proceedings.mlr.press/v97/tan19a.html

[17] A. Howard et al., "Searching for MobileNetV3," in Proc. IEEE/CVF Int. Conf. Computer Vision, 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[19] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Computer Vision, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.

[20] D. Hendrycks and T. Dietterich, "Benchmarking neural network robustness to common corruptions and perturbations," in Proc. Int. Conf. Learning Representations, 2019. https://openreview.net/forum?id=HJz6tiCqYm

[21] I. Gulrajani and D. Lopez-Paz, "In search of lost domain generalization," in Proc. Int. Conf. Learning Representations, 2021. https://openreview.net/forum?id=lQdXeXDoWtI

[22] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proc. 34th Int. Conf. Machine Learning, vol. 70, 2017, pp. 1321–1330. https://proceedings.mlr.press/v70/guo17a.html

[23] T. Gebru et al., "Datasheets for datasets," Communications of the ACM, vol. 64, no. 12, pp. 86–92, 2021, doi: 10.1145/3458723.

[24] M. Mitchell et al., "Model cards for model reporting," in Proc. Conf. Fairness, Accountability, and Transparency, 2019, pp. 220–229, doi: 10.1145/3287560.3287596.

[25] J. Cohen, "A coefficient of agreement for nominal scales," Educational and Psychological Measurement, vol. 20, no. 1, pp. 37–46, 1960, doi: 10.1177/001316446002000104.