

<https://dx.doi.org/10.17488/RMIB.46.1.2>

E-LOCATION ID: 1478

Evaluación de Tubos de Recolección de Muestras de Sangre Utilizando Deep Learning

Evaluation of Blood Sample Collection Tubes Using Deep Learning

Ignacio Franco-Alucano¹ , Julian Aguilar-Duque¹ , Yolanda Baez-Lopez¹ , Jorge Limon-Romero¹ ,
María Marcela Solís-Quinteros² , Diego Tlapa¹  

¹Universidad Autónoma de Baja California, Facultad de Ingeniería, Arquitectura y Diseño, Baja California - México

²Universidad Autónoma de Baja California, Facultad de Contaduría y Administración, Baja California - México

RESUMEN

La flebotomía es un procedimiento para obtener muestras de sangre principalmente para análisis clínicos en laboratorios. La cantidad de sangre, identificación de tubos y el uso del tubo adecuado son características que el profesional de la salud inspecciona visualmente. Al ser una actividad manual, la posibilidad de error está presente pudiendo tener efectos tanto en la calidad, como en el flujo de trabajo y eficiencia. A pesar del avance de las tecnologías de la industria 4.0, incluida la inteligencia artificial (IA), hay poca evidencia de aplicaciones en laboratorios clínicos. Este estudio tiene como objetivo evaluar la idoneidad de utilizar el aprendizaje profundo o *deep learning* (DL) en la inspección de tubos con muestras de sangre. Particularmente se prueban tres arquitecturas YOLOv5, YOLOv7 y YOLOv8 en la detección de seis clases incluyendo color de tapa y presencia de etiqueta. El mayor desempeño de precisión se presentó con el modelo YOLOv8 obteniendo una precisión de 0.927 en la detección, lo que evidencia una alta capacidad para inspeccionar características importantes en el servicio de flebotomía, siendo DL una alternativa viable para asistir a los profesionales de la salud en actividades de inspección. Trabajo futuro incluye ampliar el número de imágenes de manera balanceada.

PALABRAS CLAVE: deep learning, detección de objetos, laboratorio clínico, muestra de sangre, red neuronal convolucional, YOLO

ABSTRACT

Phlebotomy is a procedure to obtain blood samples, mainly for laboratory clinical analysis. The amount of blood, tube identification, and the use of the appropriate tube are characteristics that the health professional visually inspects. Being a manual activity, the possibility of error is latent and can affect quality, workflow, and efficiency. Despite the advancement of industry 4.0 technologies, including artificial intelligence (AI), there is little evidence of applications in clinical laboratories. This study aims to evaluate the suitability of using deep learning (DL) in inspecting tubes with blood samples. Specifically, three architectures, YOLOv5, YOLOv7, and YOLOv8, are tested to detect six classes, including cap color and the presence of labels. The highest precision performance was presented by the YOLOv8 model, obtaining a precision of 0.927 in detection, which shows a high capacity to inspect important characteristics in the phlebotomy service. Therefore, being DL is a suitable alternative to assist health professionals in inspection activities. Future work includes expanding the number of images in a balanced manner.

KEYWORDS: deep learning, object detection, clinical laboratory, blood samples, convolutional neural networks, YOLO

Autor de correspondencia

DESTINATARIO: **Diego Tlapa**

INSTITUCIÓN: **UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA**

DOMICILIO: **CARRETERA TRANSPENINSULAR
ENSENADA-TIJUANA NÚMERO 3917, COLONIA
PLAYITAS. ENSENADA, B.C., C.P. 22860, MÉXICO**
CORREO ELECTRÓNICO: **diegotlapa@uabc.edu.mx**

Recibido:

15 Octubre 2024

Aceptado:

09 Diciembre 2024

INTRODUCCIÓN

En el campo de la atención médica, los servicios de análisis clínicos son fundamentales para el diagnóstico y tratamiento de pacientes. Al respecto, la flebotomía es el procedimiento médico que consiste en extraer sangre de un paciente para realizar pruebas de laboratorio, transfusiones de sangre o donaciones. La evaluación de una muestra de sangre de un paciente es un paso crítico en la atención médica^[1]. Los laboratorios clínicos se rigen por diversos lineamientos sanitarios y estándares, de los cuales, la norma ISO 15189:2022 establece los requisitos de un sistema de gestión de la calidad en laboratorios clínicos, incluyendo lineamientos para procesos pre-examen, durante el examen y post-examen^[2]. Algunos lineamientos incluyen el volumen de las muestras, tipo de tubo y sustancia preservativa a utilizar para asegurar que no existe faltante ni excedente de muestras colectadas^[2], ya que cualquier desviación es crítica y debe ser corregida^[3]. Además, el estándar requiere lineamientos para la identificación correcta de muestras para garantizar la seguridad del paciente, optimizar los procesos de laboratorio y entregar resultados de pruebas para el posterior diagnóstico y seguimiento de enfermedades.

Dentro de las tres fases del periodo analítico (preanalítica, analítica y postanalítica), es en las fases preanalítica y postanalítica donde son más propensas a errores que la fase analítica^{[4][5][6]}. En la fase preanalítica destacan errores como solicitudes incorrectas, mala identificación del paciente o muestras, recolección de muestras inadecuada (volumen insuficiente), uso de contenedores incorrectos y fallos en el manejo, almacenamiento o transporte de la muestra^[4]. En cuanto la etapa postanalítica, los errores incluyen validación incorrecta de datos, retrasos en la entrega de resultados críticos, errores en la entrada de datos y tiempos de respuesta excesivos^{[7][8]}. Estos errores potencialmente llevan al rechazo de muestras por parte del laboratorio. En la fase preanalítica, enfermería desempeña un papel crítico, al ser la responsable de la obtención de las muestras de sangre, mientras que en la fase postanalítica es el personal administrativo quien es el responsable de la recopilación y envío de resultados.

En las unidades médicas, el área de flebotomía es uno de los lugares con mayor cantidad de pacientes en un hospital^[9], lo que implica grandes flujos de pacientes que pueden experimentar demoras por existencia de cuellos de botella. Al ser el servicio de obtención de muestras un proceso manual, éste conlleva la inspección visual de numerosos recipientes de muestras de sangre y es susceptible a errores humanos, lo que podría dar lugar a diagnósticos erróneos, retrasos en el tratamiento o atención deficiente al paciente. En este contexto, existe una creciente necesidad de sistemas apoyados en tecnologías que puedan ayudar a inspeccionar recipientes con muestras de sangre en laboratorios clínicos, incluyendo la reciente tendencia a la integración de la inteligencia artificial (IA) en el área de hematología^[10] y química sanguínea mediante la automatización del análisis, la interpretación de datos, la segmentación precisa de imágenes para aislar células sanguíneas y el análisis de datos clínicos a nivel molecular y del paciente^[11]. Sin embargo, persisten desafíos como la generalización de modelos, el desempeño en conjuntos de datos externos, la adopción de arquitecturas más actuales^[11], la falta de interpretación, el volumen de datos, así como la confidencialidad de la información^[12]. Existen además oportunidades de mejora en la inspección de muestras de sangre para la identificación de errores cuando la muestra de sangre recolectada de un paciente, se etiqueta incorrectamente con información de otro paciente, pudiendo ser difícil para el laboratorio reconocer este tipo de error^[13].

En este sentido, mejorar la calidad y la eficiencia en los servicios de salud implica un reto recurrente al que se enfrenta el sector salud^[14], por lo que en años recientes se ha incrementado la digitalización de los servicios de salud, conocida como *Healthcare 4.0*^[15], que involucra la implementación de variadas tecnologías y con fines distintos

incluyendo *deep learning* (DL)^[16], *machine learning* (ML)^[17], *big data*^[18], automatización^[19], simulación^[20], entre otras. Particularmente, herramientas de IA desempeñan un papel crucial en el campo médico, transformando la manera en que se diagnostican enfermedades, se realizan tratamientos y se gestionan los datos clínicos. Estas tecnologías han demostrado un potencial significativo para mejorar la precisión, la eficiencia y la atención médica en general. Algunas de las actividades clave para el sector médico donde se emplea IA son el análisis de grandes volúmenes de datos y el reconocimiento de patrones u objetos en imágenes médicas, como rayos X^[21], tomografías computarizadas^[22], resonancias magnéticas^[23] y ultrasonidos^[24]. DL es una subrama de la IA, la cual ha presenciado avances notables en los últimos años, ofreciendo oportunidades para el análisis de imágenes y las tareas de detección de objetos. Los modelos de DL se componen por múltiples capas que permiten aprender una secuencia en sus distintos niveles^[25]. Al respecto, las redes neuronales convolucionales (CNN) procesan imágenes y otras formas de datos con estructura espacial, donde la clasificación se basa en regiones creando un cuadro delimitador alrededor de la región de interés, como la densidad, el color, etc. Los métodos estándares involucran algoritmos de una etapa^[26] y dos etapas^[27], siendo YOLO (*You Only Look Once*)^[28] una familia de algoritmos de una sola etapa que ha mostrado ventajas incluyendo menor tiempo de entrenamiento^[29] así como su eficiencia y precisión en la localización y clasificación simultánea de múltiples objetos en tiempo real en imágenes^[28].

Este estudio prueba la factibilidad de los algoritmos YOLO para la detección de características relevantes de tubos con muestras de sangre en un laboratorio clínico, centrándose en tres tareas principales: (1) verificar la presencia de identificación en cada recipiente, (2) verificar la presencia de sangre contenida en el recipiente y (3) detectar el adecuado recipiente para la prueba a realizar en función del color de la tapa. Al abordar estas tareas, se busca contribuir a mejorar la eficiencia y confiabilidad del análisis de muestras de sangre, lo que en última instancia mejora la atención al paciente y los flujos de trabajo de laboratorio.

MATERIALES Y MÉTODOS

Con la finalidad de una evaluación integral de los modelos YOLO, se consideró las arquitecturas YOLOv5, YOLOv7 y YOLOv8 debido a características específicas que aportan cada uno en términos de precisión, velocidad y optimización respecto a sus predecesores. YOLOv5 es reconocido por su balance entre eficiencia y facilidad de implementación, facilitando así la inspección en tiempo real. YOLOv7 incluye mejoras en la velocidad y precisión en diferentes tipos de *hardware*. YOLOv8 incluye avances de arquitectura de red y procesamiento, resultando en mayor adaptabilidad y desempeño en escenarios complejos. Estas diferencias permitieron evaluar cada versión en la identificación precisa y rápida de seis clases de objetos, incluyendo: (i) etiqueta presente en tubos (*label*), (ii) tubo con llenado insuficiente (*not_ok*), (iii) tubo con llenado ok (*good*), (iv) tubo con tapa azul (*blue_cap*) que contienen citrato de sodio y se utilizan para pruebas de coagulación, (v) tubo con tapa morada (*purple_cap*) utilizadas en pruebas de hematología y (vi) tubo con tapa amarilla (*yellow_cap*) con anticoagulante para pruebas de química sanguínea. Se utilizó un conjunto de datos compuesto por 3000 imágenes, desglosadas en 2104 imágenes (70 %) para entrenamiento, 604 imágenes para validación (20 %) y 292 imágenes (10 %) para prueba y se realizó las anotaciones buscando un balance en las clases como se describe en la Tabla 1.

TABLA 1. Anotaciones por clase.

Descripción	Clases	Anotaciones
Detección de etiqueta presente en	label	2,378
Tubo con llenado insuficiente	not_ok	1,918
Tubo con llenado ok	good	2,580
Tubo con tapa azul	blue_cap	2,047
Tubo con tapa morada	purple_cap	1,998
Tubo con tapa amarilla	yellow_cap	1,725

El entrenamiento de los datos empleó el acelerador Tesla T4 incluido en la versión gratuita de Google Colab. Este acelerador, es una unidad de procesamiento gráfico (GPU) diseñada para tareas de alto rendimiento y cuenta con 40 núcleos, una frecuencia de 1.59 GHz y una capacidad de procesamiento de FP16 de 65 TFLOPS, lo cual, lo vuelve una herramienta versátil, potente y accesible. Por su parte, el etiquetado de las imágenes se realizó mediante la segmentación, empleando recuadros delimitadores “*bounding boxes*” en la herramienta *Roboflow*, la cual es una plataforma en línea para la clasificación de imágenes y la gestión de datos^[30]. Las imágenes utilizadas presentan una resolución de 640x640 píxeles en formato JPG y se cuentan con acceso libre al conjunto de datos en la siguiente liga: <https://doi.org/10.34740/kaggle/dsv/9616793>. Los resultados se evaluaron como verdadero positivo (TP, *true positive*), es decir, el número de objetos detectados correctamente; verdadero negativo (TN, *true negative*), es decir, muestras que se rechazan correctamente de la clase; falso positivo (FP, *false positive*), es decir, el número de objetos detectados erróneamente; y falso negativo (FN, *false negative*), es decir, el número de objetos omitidos. Para evaluar las diferentes arquitecturas se utilizaron índices de desempeño como precisión (*precision*), sensibilidad (*recall*) y precisión media promedio (*mAP*). Precisión, precisión por clase (PRC)^[31] o valor predictivo positivo (VPP)^[32], mide la proporción de objetos correctamente detectados en relación con los objetos detectados^[33]. La ecuación 1 muestra la fórmula utilizada en este estudio.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Recall, indica la proporción de objetos detectados correctamente en relación con todos los objetos presentes en el conjunto de datos^[34], por lo tanto, *recall* es la precisión de las instancias predichas positivamente que describen cuántas se etiquetaron correctamente^[35], lo que muestra el nivel de un modelo para predecir la clase positiva cuando la clase real es positiva^[36]. En este estudio calculamos *recall* con la ecuación (2).

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Precisión media promedio (*mAP*) mide en qué porcentaje el algoritmo predice el objeto correctamente de todas las clases individuales^[37], siendo una métrica clave para evaluar los algoritmos de detección de objetos^[38] y para la comparación entre modelos^[39]. La ecuación 3 muestra la descripción matemática de *mAP*, donde AP_k representa la precisión promedio de las clases y n representa el número de clases^[40].

$$mAP = \frac{1}{n} \sum_{k=1}^{k=n} AP_k \quad (3)$$

En conjunto, estos métricos proporcionan una evaluación global del rendimiento del modelo y son ampliamente utilizados en la evaluación de algoritmos de detección de objetos ya que permiten comparar el desempeño del modelo en diferentes aspectos de la detección^{[41][42]}. El estudio también se enfoca en analizar el comportamiento de los modelos en distintas épocas de entrenamiento. Se seleccionaron las épocas 20, 50 y 100 como puntos de referencia para examinar la evolución de los modelos durante el proceso de entrenamiento y evaluar su capacidad de aprendizaje a lo largo del tiempo. Para complementar el resultado, se analizaron las matrices de confusión normalizadas y las gráficas de pérdidas para evaluar y visualizar el rendimiento de un modelo.

Equipo y Herramientas

El análisis se llevó a cabo en un equipo con un procesador AMD Ryzen 5 4650G con Radeon R7 Renoir, funcionando a una velocidad de 3.70 GHz, con 32 GB de RAM y sistema operativo Windows 11 de 64 bits. Además, se empleó una tarjeta gráfica NVIDIA GeForce GTX 1660 SUPER VENTUS XS OC con 6 GB de memoria dedicada para el procesamiento computacional. Para la captura de imágenes se utilizaron dos dispositivos: una cámara marca Salandens, modelo B0872YBHBV, con una resolución de 1080p, y la cámara de un teléfono móvil Huawei Nova Y9 SE, con una resolución de 108 MP. El entrenamiento de los modelos se realizó utilizando Google Colab en su versión gratuita, la cual ofrece acceso a una GPU T4 con 12.7 GB de RAM, 15 GB de memoria dedicada para la GPU, y un espacio en disco de 112.6 GB.

RESULTADOS Y DISCUSIÓN

El desempeño de los diferentes modelos de YOLO se muestra en la Tabla 2, siendo YOLOv8 la arquitectura que presentó la mayor precisión, 0.927, con 50 épocas mientras que YOLOv5, 0.924, con 100 épocas y YOLOv7, 0.917 con 100 épocas. Respecto a *recall*, YOLOv8 presentó los mayores valores, destacando 0.914 con 100 épocas. Al evaluar la capacidad del modelo para detectar imágenes de distintas clases, YOLOv8 presentó un valor de *mAP* de 0.941 y 0.940 con 100 y 50 épocas respectivamente. En términos de tiempo de entrenamiento, el modelo más rápido fue YOLOv5 con 20 épocas (15 min, 4 s), en contraparte, YOLOv7 con 100 épocas presentó la mayor duración (3 h, 27 min, 44 s).

TABLA 2. Resultados del rendimiento de los modelos de YOLOv5, YOLOv7 y YOLOv8.

Modelo	Épocas	Precisión	Sensibilidad	<i>mAP</i>	Tiempo
YOLOv5	20	0.909	0.897	0.919	15 min 4 s
	50	0.923	0.900	0.934	37 min 45 s
	100	0.924	0.900	0.935	1 h 14 min 10 s
YOLOv7	20	0.682	0.665	0.675	42 min 36 s
	50	0.903	0.852	0.856	1 h 44 min 24 s
	100	0.917	0.878	0.906	3 h 27 min 44 s
YOLOv8	20	0.922	0.907	0.933	34 min 19 s
	50	0.927	0.909	0.940	1 h 26 min 13 s
	100	0.917	0.914	0.941	2 h 59 min 56 s

En términos generales, YOLOv8 presenta un mayor desempeño, logrando una mayor precisión y *mAP* en todas las épocas analizadas. Su tiempo de entrenamiento, aunque mayor que YOLOv5, es más corto que YOLOv7, lo cual lo convierte en una opción eficiente en términos de rendimiento y costo computacional. YOLOv5, aunque no alcanza los niveles de YOLOv8, ofrece un desempeño consistente con una buena precisión y *mAP*, y tiempos de entrena-

miento significativamente menores, haciéndolo atractivo en escenarios donde los recursos de tiempo y procesamiento son limitados. YOLOv7 comienza con un rendimiento inferior incluyendo precisión y *mAP* bajos con 20 épocas, pero mejora significativamente con más épocas de entrenamiento, aunque su tiempo de entrenamiento es considerablemente mayor que los otros dos modelos, lo cual podría ser un factor limitante dependiendo de la aplicación. La Figura 1 muestra ejemplos de detección de clases empleando YOLOv8.

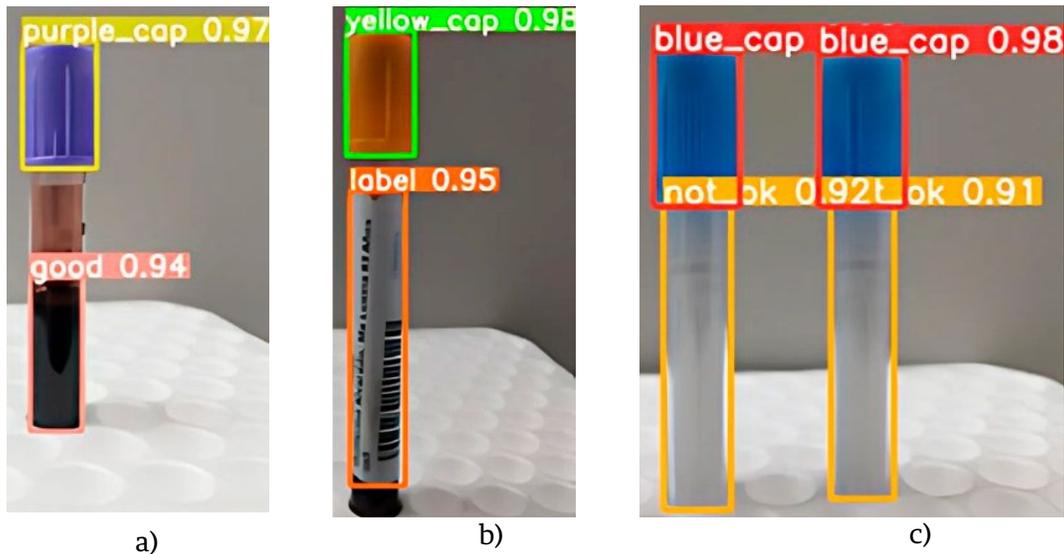


FIGURA 1. (A) Ejemplo de detección de las clases "purple_cap" y "good". (B) Demostración de la detección de las clases "yellow_cap" y "label". (C) Representación de las clases "blue_cap" y "not_ok" empleando YOLOv8.

El desempeño de las tres arquitecturas por clase al utilizar 50 épocas se muestra en las Tablas 3-5. El resto de resultados con 20 y 100 épocas se agregó como material suplementario al artículo. Particularmente, la Tabla 3 muestra el desempeño de YOLOv5 con 50 épocas, destacando una precisión de hasta 0.980 para tapa morada "purple_cap" y una mínima de 0.876 para identificar tubos con tapa amarilla "yellow_cap". *Recall* presenta valores de 0.978 hasta 0.85 para la detección de tubos con tapa morada "purple_cap" y tapa amarilla "yellow_cap" respectivamente. La clase con mayor *mAP* (0.99) es para tapa morada "purple_cap", mientras que el menor valor fue 0.893 para la clase tapa amarilla "yellow_cap".

TABLA 3. Resultados del rendimiento del modelo de YOLOv5 con 3000 imágenes entrenado con 50 épocas

Clase	Precisión	Recall	<i>mAP</i>
blue_cap	0.924	0.92	0.951
good	0.934	0.915	0.938
label	0.93	0.851	0.927
not_ok	0.896	0.886	0.903
purple_cap	0.98	0.978	0.99
yellow_cap	0.876	0.85	0.893

Por su parte, la Tabla 4 muestra el desempeño de YOLOv7 con 50 épocas, logrando una precisión de hasta 0.966 para tapa morada "purple_cap" y una mínima de 0.821 para identificar tubos con tapa amarilla "yellow_cap". *Recall* presentó valores de 0.963 hasta 0.708 para la detección de tubos con tapa morada "purple_cap" y tubos con etiqueta "label" respectivamente. La clase con el mayor *mAP* de 0.978 fue tapa morada "purple_cap", mientras que

el menor valor fue 0.782 para la clase que reporta los tubos con tapa amarilla “*yellow_cap*”.

TABLA 4. Resultados del rendimiento del modelo de YOLOv7 con 3000 imágenes entrenado con 50 épocas

Clase	Precisión	Recall	mAP
blue_cap	0.906	0.914	0.877
good	0.901	0.860	0.87
label	0.944	0.708	0.826
not_ok	0.88	0.817	0.805
purple_cap	0.966	0.963	0.978
yellow_cap	0.821	0.847	0.782

La Tabla 5 muestra los resultados de cada clase utilizando YOLOv8 con 50 épocas, donde se observa una precisión máxima de 0.984 y mínima de 0.883 para identificar los tubos de sangre con tapa morada (“*purple_cap*”) y los tubos vacíos o que no contienen sangre (“*not_ok*”) respectivamente. Respecto a *recall*, el valor mayor es de 0.976 para la detección de tubos con tapa morada (“*purple_cap*”), mientras que 0.840 fue el valor menor para tubos con tapa amarilla (“*yellow_cap*”). La clase con mayor *mAP* de hasta 0.993 es la de tapa morada (“*purple_cap*”), mientras que el menor valor fue 0.891 para la clase que reporta los tubos que no contienen sangre (“*not_ok*”).

TABLA 5. Resultados del rendimiento del modelo de YOLOv8 con 3000 imágenes entrenado con 50 épocas.

Clase	Precisión	Recall	mAP
blue_cap	0.923	0.914	0.955
good	0.933	0.921	0.942
label	0.944	0.901	0.952
not_ok	0.883	0.901	0.891
purple_cap	0.984	0.976	0.993
yellow_cap	0.895	0.840	0.908

La Figura 2 muestra la matriz de confusión normalizada para YOLOv8 con 50 épocas donde cada fila representa los porcentajes de instancias reales de cada clase, mientras que las columnas muestran las predicciones del modelo. La diagonal normalizada muestra valores altos en la tasa de acierto, incluyendo “*purple_cap*” con 0.98, mientras “*yellow_cap*” muestra un valor de 0.91, siendo el menor.

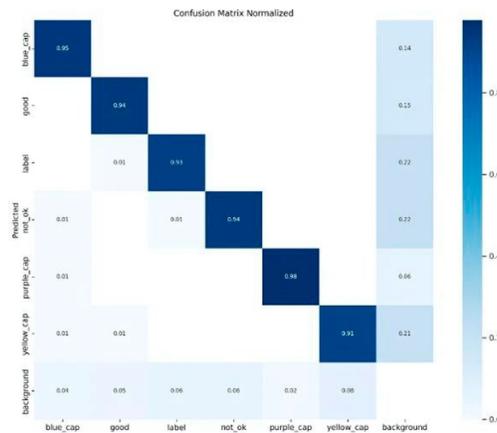


FIGURA 2. Matriz de confusión normalizada de YOLOv8 con 50 épocas.

Por su parte, la Figura 3 muestra la gráfica de pérdida para YOLOv8 con 50 épocas, destacando cómo la función de pérdida del modelo disminuye a lo largo del tiempo durante el entrenamiento. Una disminución continua en la pérdida indica que el modelo está aprendiendo correctamente a medida que avanza el entrenamiento. La estabilización de la curva o una disminución más lenta después de muchas épocas sugiere que el modelo ha alcanzado su punto de convergencia y que entrenar por más tiempo no mejoraría significativamente su desempeño. Si bien es cierto que extender el tiempo de entrenamiento puede resultar en mejoras marginales en precisión, es necesario considerar el costo computacional y la aplicación práctica para sistemas que trabajan en tiempo real. En conjunto, estas figuras reflejan un buen desempeño de YOLOv8, con una matriz de confusión que revela una alta precisión en la clasificación de los defectos, y una gráfica de pérdida que sugiere que el modelo se ha aprendido eficientemente a lo largo del proceso de entrenamiento.

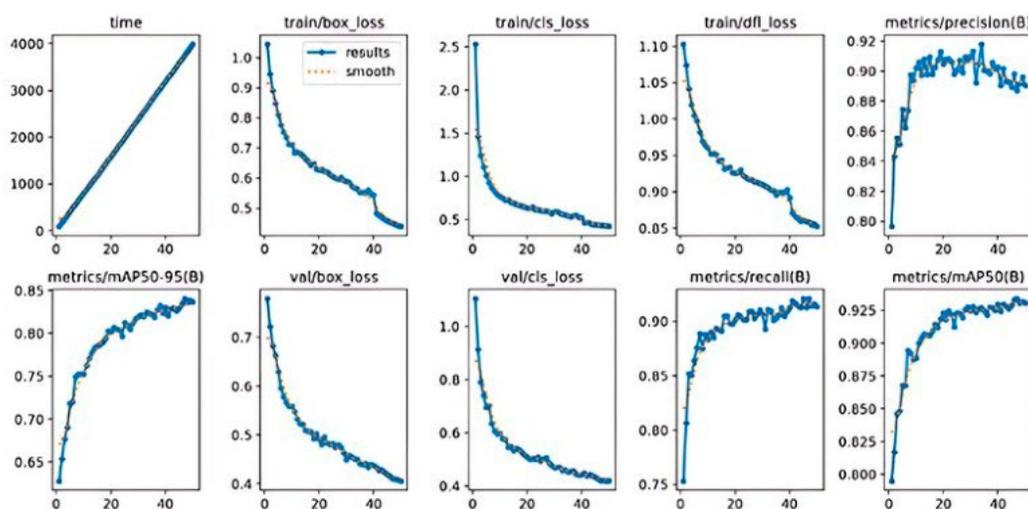


FIGURA 3. Gráficas de pérdida para YOLOv8 con 50 épocas.

Adicionalmente, el desempeño de YOLOv8 con 100 épocas se resume en la Tabla 6, destacando una precisión de hasta 0.970 para tapa morada (“purple_cap”) y una mínima de 0.892 para identificar tubos vacíos o que no contienen sangre (“not_ok”). *Recall* presenta valores de 0.976 hasta 0.877 para la detección de tubos con tapa morada (“purple_cap”) y tapa amarilla (“yellow_cap”) respectivamente. La clase con mayor *mAP* (0.993) es para tapa morada (“purple_cap”), mientras que el menor valor fue 0.898 para la clase que reporta los tubos que no contienen sangre (“not_ok”). En resumen, YOLOv8 presenta desempeño similar tanto en 50 y 100 épocas.

TABLA 6. Resultados del rendimiento del modelo de YOLOv8 con 3000 imágenes entrenado con 100 épocas.

Clase	Precisión	Recall	mAP
blue_cap	0.901	0.895	0.953
good	0.928	0.928	0.943
label	0.928	0.915	0.947
not_ok	0.892	0.892	0.898
purple_cap	0.970	0.976	0.993
yellow_cap	0.881	0.877	0.916

Implicaciones para laboratorios clínicos

Los laboratorios clínicos tienen un rol significativo al proveer de información para la toma de decisiones de los profesionales de salud. No obstante, existe un grado de falibilidad en pruebas o dispositivo de laboratorio, por lo que la ocurrencia de errores es factible en las diferentes fases de una prueba de laboratorio (preanalítica, analítica y postanalítica)^[43]. Este estudio aborda la evaluación de diferentes modelos de DL para identificar objetos en imágenes de tubos con muestras de sangre de un laboratorio clínico como un medio para asistir al profesional de salud en la inspección de riesgos en la etapa preanalítica. Respecto a precisión, este métrico fluctúa entre 0 y 1, donde valores cercanos a 1 indican una alta tasa de detección de clases de manera correcta^[33], siendo YOLOv8 la arquitectura que presentó el mayor valor, 0.927, con 50 épocas. Respecto a *recall*, es decir, la capacidad del modelo de identificar todas las clases en la imagen^{[44][45][46]}, el modelo YOLOv8 presentó los mayores valores, destacando 0.914 con 100 épocas. Al evaluar la capacidad del modelo para detectar imágenes de distintas clases^[47], YOLOv8 presentó un valor de *mAP* de 0.941 y 0.940 con 100 y 50 épocas respectivamente, demostrando un equilibrio entre la precisión y *recall* en todas las clases durante la validación^[45]. Estos resultados indican que el modelo YOLOv8 presentó un mejor desempeño general, seguido por YOLOv5 y YOLOv7.

Estos resultados presentan a los algoritmos de DL como una herramienta factible para asistir al técnico flebotomista en la inspección de características importantes que pueden ser omitidas por la naturaleza repetitiva del proceso y dependiente del factor humano, incluyendo omitir etiqueta de identificación, extraer cantidad no adecuada de sangre, utilizar tubo incorrecto (con o sin preservativo o anticoagulante), entre otros. Particularmente, la identificación errónea del paciente^[6] y el consecuente problema para comunicar resultados, afectan a la prestación de servicios de diagnóstico, siendo reconocidos como objetivos para la mejora de la calidad en este sector^[4]. Estos errores pueden provocar la retoma de muestra, retraso del análisis, alteración o contaminación de la sangre, teniendo un impacto en el diagnóstico y salud de los pacientes, así como en la calidad y eficiencia del proceso. Aunque se tiende a responsabilizar al personal de atención médica por los errores, en realidad la mayoría de estos problemas surgen debido a la falta de procesos seguros y bien estructurados^[5]. A pesar de su importancia, existe poca información sobre el nivel de precisión al realizar actividades de inspección en el sector salud, particularmente en laboratorios clínicos. En sectores industriales, la tasa de errores al realizar inspección visual varía por múltiples factores^[48], con tasas entre 20 % y 30 % de manera general^[8], mientras que en el sector de construcción va de 19 % a 48 %^[49] y de 17.8 % a 29.8 % en sector metalmecánica^[50].

En la evaluación de la calidad en la atención médica existen varias herramientas, incluyendo los indicadores de calidad (QIs), los cuales proporcionan medidas objetivas basadas en evidencia para evaluar diferentes aspectos críticos del cuidado de manera consistente^[51]. Al respecto, la clasificación de errores en laboratorio según su gravedad, es crucial para identificar áreas prioritarias de mejora en la calidad^[4], por lo que, el entendimiento de los QIs permite a los laboratorios clínicos identificar las acciones correctivas y mejoras más adecuadas para la resolución de problemas^[52]. Los indicadores han sido agrupados en seis etapas del proceso de pruebas de laboratorio: i) orden de la prueba, ii) identificación del paciente y recolección de la muestra, iii) identificación, preparación y transporte de la muestra, iv) análisis, v) informe de resultados, así como vi) interpretación de resultados y acciones posteriores^[53]. A pesar de que la implementación sistemática los QIs en laboratorios puede ser efectiva para reducir los errores, mejorar la seguridad del paciente y cumplir con los requisitos de la norma ISO 15189, en la práctica existen dificultades para mantener la recolección de datos de manera estandarizada y sistemática, así como para fomentar un continuo interés, compromiso y dedicación en todo el personal^[52]. Al respecto, los resultados de este estudio contribuyen a

brindar información sobre la factibilidad detectar desviaciones y evaluar indicadores correspondientes a las primeras etapas (i, ii e iii), con posibilidad de expansión al resto. En este sentido se ha sugerido que los laboratorios deben crear un mapa de proceso que describa todos los pasos del proceso de prueba desde la orden del médico hasta la entrega del resultado^[6].

Las herramientas de IA como DL han apoyado en el diagnóstico médico permitiendo reducir el tiempo para iniciar un tratamiento^[54], mediante la detección y clasificación del cáncer^[55], diagnósticos mediante electrocardiogramas^[56], detección de glaucoma^[57] o la clasificación de tumores cerebrales^[58], lo que permite reducir el tiempo para iniciar un tratamiento^[54]. Particularmente, diferentes versiones de YOLO han sido evaluados en contextos variados como el farmacéutico para la detección de defectos en blíster alcanzando una precisión de 0.974^[59], para detectar equipo médico de protección personal logrando un *mAP* de 97.2 %^[60], o para detectar máscaras y tubos de ventilación en pacientes logrando una exactitud (*accuracy*) de 93 %^[61]. Sin embargo, la evidencia es escasa sobre estudios que aborden DL para la evaluación y análisis de actividades de inspección visual en sector salud y particularmente en laboratorios clínicos. El presente estudio contribuye en la discusión sobre el uso de DL como apoyo al técnico flebotomista en la etapa preanalítica. En este sentido, estudios previos han identificado que el uso de IA puede facilitar la colaboración entre especialistas de la salud^[62], permitiendo una mejor la utilización de servicios con el fin de optimizar los recursos^[63]. Al respecto, es crucial involucrar al equipo del sector de la salud y centrarse en fortalecer las capacidades locales y adaptar la tecnología disponible a las necesidades específicas de los países latinoamericanos^[64]. En la región, la IA se encuentra en fase inicial y muchas áreas carecen de la infraestructura necesaria^[65]. Sin embargo, tecnologías como la IA tienen el potencial de mejorar el acceso a servicios, reducir consultas innecesarias, brindar comodidad a los usuarios, reducir la carga de trabajo, disminuir costos^[66] o incluso disminuir las readmisiones no planificadas^[67], lo cual, tiene un efecto directo a la calidad de atención al paciente^[68].

Inspección visual y aprendizaje profundo (DL)

Con respecto al desempeño de actividades de inspección visual en laboratorios clínicos, existe poca información sobre el grado de precisión o exactitud de los profesionales de salud. En comparación con actividades de inspección en sectores industriales, comúnmente la inspección visual humana es superada por algoritmos de DL^{[69][70]}, aunque existen excepciones mostrando resultados mixtos o comparables^{[71][72]} o incluso un rendimiento inferior de los modelos en comparación con los inspectores humanos^[73]. El presente estudio destaca una brecha entre la inspección totalmente automatizada y la necesidad continua de participación humana, principalmente en actividades críticas como las asociadas a la salud. En algunos escenarios, los algoritmos comienzan la inspección y los inspectores humanos intervienen en el caso de elementos inciertos o que caen por debajo de un umbral establecido^{[74][75]}. De esta manera, los modelos de DL tienen el potencial de ayudar de manera consistente a los procesos de inspección al minimizar la participación humana y aliviar la fatiga física y mental. La utilización de algoritmos para detectar objetos empleando redes neuronales convolucionales, como YOLO, ofrece la posibilidad de realizar esta tarea de manera automatizada y eficiente. Adicionalmente, las herramientas de DL presentan la capacidad de identificar objetos que pudieran no ser evidentes a simple vista u omitidos debido a condiciones humanas de error. Considerando que hasta dos tercios de decisiones de los profesionales de salud son basadas en resultados de pruebas de laboratorio^[76], DL representa un soporte para los profesionales de la salud y en última instancia, un apoyo para mejorar de la calidad de la atención médica.

Retos Prácticos

Algunos retos identificados incluyen aspectos técnicos para la implementación de sistemas de inspección en laboratorio clínico basados en IA, incluyendo la dependencia al volumen y calidad de conjuntos de datos para su rendimiento^[77]. Cuando no se dispone de grandes conjuntos de datos, es necesario realizar un entrenamiento previo^[78] o transferencia de aprendizaje^[79] como alternativas para trabajar con conjunto de datos pequeños. Otro aspecto y tal vez más relevante es el factor humano, donde la transición de inspección tradicionalmente realizada por un profesional de la salud, a una inspección asistida por tecnología puede representar un reto, principalmente por desconocimiento, falta de infraestructura, resistencia al cambio, entre otros. Adicionalmente, un reto futuro en contextos de laboratorios clínicos es el uso de más arquitecturas de DL, incluyendo modelos de dos etapas como redes neuronales convolucionales basadas en regiones (R-CNN), *Fast R-CNN*^[80], *Faster R-CNN*^[81] y *Mask R-CNN*^[82], que, aunque requieren conjuntos de datos más grandes, pueden aportar una comparativa integral.

Limitaciones

El estudio no está exento de limitaciones, incluyendo el tamaño del conjunto de datos, que puede considerarse pequeño, por lo que un trabajo futuro incluye su ampliación. El estudio consideró una variabilidad limitada en las condiciones del entorno de las imágenes incluyendo la iluminación y los tipos de tubos, por lo que se proyecta ampliar estas características. Adicionalmente, la incorporación de más y diferentes fondos de imagen (*background*) podrá ayudar a mejorar aún más el desempeño de las arquitecturas evaluadas para detectar objetos con diferentes fondos. Un aspecto importante es el desequilibrio de clases, el cual en este estudio no es un problema crítico, ya que el rango de anotaciones va desde 1,725 a 2,580, como se mostró en Tabla 1, sin embargo, un trabajo futuro será balancear las anotaciones de clase, aumentar el número de imágenes con respecto a las clases a identificar, así como implementar data augmentation para el incremento de imágenes. Finalmente, las clases referentes a volumen de sangre, solo consideraron presencia o ausencia de sangre en el tubo, por lo que el siguiente paso será entrenar al modelo en diferentes niveles o cantidades presentes en el tubo.

CONCLUSIONES

El presente estudio abordó el uso de tecnología asociada a inteligencia artificial como lo es *deep learning*, en la detección de características relevantes al realizar inspecciones en el área de flebotomía de un laboratorio clínico público. Los resultados indican que los modelos YOLO y particularmente YOLOv8 mostraron un buen desempeño en identificar adecuadamente las clases analizadas. Si bien no se cuenta con información sobre el nivel de precisión del personal de salud al realizar actividades de inspección en laboratorios clínicos, estos valores rondan entre 85 y 90 % en otros sectores, por lo que los resultados de este estudio son alentadores y sugieren que los modelos de DL son capaces de detectar características relevantes en la fase preanalítica con una precisión superior. Además, estos resultados abren posibilidades de desarrollo para sistemas de inspección que asistan a los profesionales de salud y complementen la función de inspección que realizan. El estudio brinda información valiosa para futuras investigaciones, incluyendo la inspección de más variables críticas en laboratorios clínicos, la realización de análisis más detallados, así como optimizaciones específicas para cada arquitectura, teniendo en cuenta las necesidades y requisitos específicos del laboratorio clínico. Dada la escasa literatura sobre aplicaciones específicas de DL en el llenado e identificación de muestras de sangre en laboratorios clínicos, los hallazgos respaldan el potencial de estas herramientas para asistir a los profesionales de salud en actividades de inspección visual y con ello contribuir a la mejora de los servicios de salud.

DECLARACIÓN ÉTICA

El presente trabajo no involucró la participación de seres humanos ni animales. Por lo tanto, no se requirió el consentimiento informado de personas. El protocolo de investigación fue debidamente registrado ante el Instituto de Servicios de Salud Pública del Estado de Baja California.

AGRADECIMIENTOS

Agradecemos al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) y a la Universidad Autónoma de Baja California por el apoyo brindado para la realización del presente estudio.

CONTRIBUCIÓN DE LOS AUTORES

I. F.-A. conceptualización, curación de datos, análisis formal y escritura de manuscrito original; J. A.-D. conceptualización, análisis formal, escritura de manuscrito original y revisión final; Y. B.-L. investigación, metodología, administración de proyecto y validación; J. L.-R. curación de datos, software, validación y revisión final; M. M. S.-Q. investigación, visualización y administración de proyecto; D. T. conceptualización, análisis formal, supervisión, escritura de manuscrito original y revisión final.

REFERENCIAS

- [1] R. Singh, A. Sharma, N. Sharma, R. Gupta, "Impact of Adam, Adadelta, SGD on CNN for White Blood Cell Classification," en 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 1702-1709, doi: <https://doi.org/10.1109/ICSSIT55814.2023.10061068>
- [2] International Standard Organization, "ISO-15189 Medical laboratories-Requirements for quality and competence," 2022. [Online]. Available: <https://www.iso.org/standard/76677.html>
- [3] G. Lima-Oliveira et al., "Laboratory Diagnostics and Quality of Blood Collection," J. Med. Biochem., vol. 34, no. 3, pp. 288-294, 2015, doi: <https://doi.org/10.2478/jomb-2014-0043>
- [4] M. Plebani, "The detection and prevention of errors in laboratory medicine," Ann. Clin. Biochem., vol. 47, no. 2, pp. 101-110, 2010, doi: <https://doi.org/10.1258/acb.2009.009222>
- [5] A. Aita, A. Padoan, G. Antonelli, L. Sciacovelli, M. Plebani, "Patient safety and risk management in medical laboratories: theory and practical application," J. Lab. Precis. Med., vol. 2, no. 2, pp. 75-75, 2017, doi: <https://doi.org/10.21037/jlpm.2017.08.14>
- [6] J. H. Nichols, "Laboratory quality control based on risk management," Ann. Saudi Med., vol. 31, no. 3, pp. 223-228, 2011, doi: <https://doi.org/10.4103/0256-4947.81526>
- [7] M. Plebani, "Exploring the iceberg of errors in laboratory medicine," Clin. Chim. Acta, vol. 404, no. 1, pp. 16-23, 2009, doi: <https://doi.org/10.1016/j.cca.2009.03.022>
- [8] J. Hickner et al., "Testing process errors and their harms and consequences reported from family medicine practices: A study of the American Academy of Family Physicians National Research Network," Qual Saf. Health Care, vol. 17, no. 3, pp. 194-200, 2008, doi: <https://doi.org/10.1136/qshc.2006.021915>
- [9] H. Javadifard et al., "Predicting Patient Waiting Time in Phlebotomy Units Using a Deep Learning Method," en Proceedings - 2019 Innovations in Intelligent Systems and Applications Conference, ASYU 2019, Izmir, Turkey, 2019, pp. 1-4, doi: <https://doi.org/10.1109/ASYU48272.2019.8946380>
- [10] B. Fridath, A. Gildas, A. Dooguy, "White Blood Cells Recognition and Classification using Convolutional Neural Network," en 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAIC), Salem, India, 2023, pp. 145-150, doi: <https://doi.org/10.1109/ICAIC56838.2023.10140293>
- [11] J. Wang, "Deep Learning in Hematology: From Molecules to Patients," Clin. Hematol. Int., vol. 6, no. 4, pp. 19-42, 2024, doi: <https://doi.org/10.46989/001c.124131>
- [12] R. Chandra, A. K. Geico, "Deep Learning Diagnostics: A Revolutionary Approach to Healthcare Insurance," NeuroQuantology, vol. 19, no. 12, pp. 745-754, 2021, doi: <https://doi.org/10.48047/nq.2021.19.12.NQ21278>

- [13] H. Hou, R. Zhang, J. Li, "Artificial intelligence in the clinical laboratory," *Clin. Chim. Acta*, vol. 559, 2024, art. no. 119724, doi: <https://doi.org/10.1016/j.cca.2024.119724>
- [14] D. Tlapa et al., "Effects of Lean Interventions Supported by Digital Technologies on Healthcare Services: A Systematic Review," *Int. J. Environ. Res. Public Health*, vol. 19, no. 15, 2022, art. no. 9018, doi: <https://doi.org/10.3390/ijerph19159018>
- [15] G. L. Tortorella et al., "Effects of contingencies on healthcare 4.0 technologies adoption and barriers in emerging economies," *Technol. Forecast Soc. Change*, vol. 156, 2020, art. no. 120048, doi: <https://doi.org/10.1016/j.techfore.2020.120048>
- [16] M. Ibrahim et al., "Augmentation in Healthcare: Augmented Biosignal Using Deep Learning and Tensor Representation," *J. Healthc Eng.*, vol. 2021, 2021, art. no. 6624764, doi: <https://doi.org/10.1155/2021/6624764>
- [17] E. Lee et al., "Transforming hospital emergency department workflow and patient care," *Interfaces*, vol. 45, no. 1, pp. 58-82, 2015, doi: <https://doi.org/10.1287/inte.2014.0788>
- [18] G. Arcidiacono, A. Pieroni, "The revolution Lean Six Sigma 4.0," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 1, pp. 141-149, 2018, doi: <https://doi.org/10.18517/ijaseit.8.1.4593>
- [19] J. De Mast et al., "Process improvement in healthcare: Overall resource efficiency," *Qual. Reliab. Eng. Int.*, vol. 27, no. 8, pp. 1095-1106, 2011, doi: <https://doi.org/10.1002/qre.1198>
- [20] D. Marshall et al., "Selecting a dynamic simulation modeling method for health care delivery research - Part 2: Report of the ISPOR dynamic simulation modeling emerging good practices task force," *Value Health*, vol. 18, no. 2, pp. 147-160, 2015, doi: <https://doi.org/10.1016/j.jval.2015.01.006>
- [21] A. M. Ismael, A. Şengür, "Deep learning approaches for COVID-19 detection based on chest X-ray images," *Expert. Syst. Appl.*, vol. 164, 2021, art. no. 114054, doi: <https://doi.org/10.1016/j.eswa.2020.114054>
- [22] J. I. Arreola Minjarez et al., "Detection of COVID-19 Lung Lesions in Computed Tomography Images Using Deep Learning," *Rev. Mex. Ing. Biomed.*, vol. 43, no. 1, pp. 7-18, 2022, doi: <https://doi.org/10.17488/RMIB.43.1.1>
- [23] R. Indraswari et al., "Brain Tumor Detection on Magnetic Resonance Imaging (MRI) Images Using Convolutional Neural Network (CNN)," en 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Jakarta, Indonesia, 2022, pp. 367-373, doi: <https://doi.org/10.23919/EECSI56542.2022.9946622>
- [24] S. Guizani, N. Guizani, S. Gharsallaoui, "A Hybrid CNN-SVM Prediction Approach for Breast Cancer Ultrasound Imaging," en 2023 International Wireless Communications and Mobile Computing (IWCMC), Marrakesh, Morocco, 2023, pp. 1574-1578, doi: <https://doi.org/10.1109/iwcmc58020.2023.10182874>
- [25] V. H. Benitez Baltazar et al., "Autonomic Face Mask Detection with Deep Learning: an IoT Application," *Rev. Mex. Ing. Biomed.*, vol. 42, no. 2, pp. 160-170, 2021, doi: <https://doi.org/10.17488/RMIB.42.2.13>
- [26] J. Fan, T. Huo, X. Li, "A review of one-stage detection algorithms in autonomous driving," en 2020 4th CAA International Conference on Vehicular Control and Intelligence (CVCI), Hangzhou, China, 2020, pp. 210-214, doi: <https://doi.org/10.1109/CVICI51460.2020.9338663>
- [27] Q. Tang et al., "A Two-stage Raman Imaging Denoising Algorithm Based on Deep Learning," en 2022 Asia Communications and Photonics Conference (ACP), Shenzhen, China, 2022, pp. 2096-2099, doi: <https://doi.org/10.1109/ACP55869.2022.10088904>
- [28] J. Redmon et al., "You only look once: Unified, real-time object detection," en 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, doi: <https://doi.org/10.1109/CVPR.2016.91>
- [29] S.-H. Chen et al., "A surface defect detection system for golden diamond pineapple based on CycleGAN and YOLOv4," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 10, pp. 8041-8053, 2022, doi: <https://doi.org/10.1016/j.jksuci.2022.07.018>
- [30] J. M. Kimeu et al., "Deep learning-based mobile application for the enhancement of pneumonia medical imaging analysis: A case-study of West-Meru Hospital," *Inform. Med. Unlocked*, vol. 50, 2024, art. no. 101582, doi: <https://doi.org/10.1016/j.imu.2024.101582>
- [31] S. Shahinfar, P. Meek, G. Falzon, "How many images do I need? Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring," *Ecol. Inform.*, vol. 57, 2020, art. no. 101085, doi: <https://doi.org/10.1016/j.ecoinf.2020.101085>
- [32] A. Cruz et al., "Detection of grapevine yellows symptoms in *Vitis vinifera* L. with artificial intelligence," *Comput. Electron. Agric.*, vol. 157, pp. 63-76, 2019, doi: <https://doi.org/10.1016/j.compag.2018.12.028>
- [33] D. E. Brown et al., "Neural network methods for diagnosing patient conditions from cardiopulmonary exercise testing data," *BioData Min.*, vol. 15, no. 1, 2022, art. no. 16, doi: <https://doi.org/10.1186/s13040-022-00299-6>
- [34] H. Wang, Q. Li, D. Zhang, D. Li, Q. Chen, "Key Components of Deep Metric Learning," en 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 2022, pp. 648-651, doi: <https://doi.org/10.1109/ICCECE54139.2022.9712771>
- [35] V. Gonzalez-Huitron et al., "Disease detection in tomato leaves via CNN with lightweight architectures implemented in Raspberry Pi 4," *Comput. Electron. Agric.*, vol. 181, 2021, art. no. 105951, doi: <https://doi.org/10.1016/j.compag.2020.105951>
- [36] P. T. Q. Anh, D. Q. Thuyet, Y. Kobayashi, "Image classification of root-trimmed garlic using multi-label and multi-class classification with deep convolutional neural network," *Postharvest Biol. Technol.*, vol. 190, 2022, art. no. 111956, doi: <https://doi.org/10.1016/j.postharvbio.2022.111956>
- [37] M. Glučina et al., "Automated Detection and Classification of Returnable Packaging Based on YOLOV4 Algorithm," *Appl. Sci.*, vol. 12, no. 21, 2022, art. no. 11131, doi: <https://doi.org/10.1109/ACP55869.2022.10088904>

- [38] Y. Ke et al., "Empowering Intelligent Home Safety: Indoor Family Fall Detection with YOLOv5," en 2023 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), Abu Dhabi, United Arab Emirates, 2023, pp. 0942-0949, doi: <https://doi.org/10.1109/DASC/PiCom/CBDCom/Cy59711.2023.10361490>
- [39] G. P. Bhat, N. G. Cholli, "Effective object detection using Tensorflow facilitated YOLOv3 model," en 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 2021, pp. 1-8, doi: <https://doi.org/10.1109/CSITSS54238.2021.9683109>
- [40] M. L. Francies, M. M. Ata, M. A. Mohamed, "A robust multiclass 3D object recognition based on modern YOLO deep learning algorithms," *Concurr. Comput. Pract. Exp.*, vol. 34, no. 1, 2022, art. no. e6517, doi: <https://doi.org/10.1002/cpe.6517>
- [41] R. Ammu, N. Sinha, "Small Segment Emphasized Performance Evaluation Metric for Medical Images," en 2020 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 2020, pp. 1-5, doi: <https://doi.org/10.1109/SPCOM50965.2020.9179617>
- [42] S. Raj, Y. Gupta, R. Malhotra, "License Plate Recognition System using Yolov5 and CNN," en 2022 8th International Conference on Advanced Computing and Communication Systems, ICACCS, Coimbatore, India, 2022, pp. 372-377, doi: <https://doi.org/10.1109/ICACCS54159.2022.9784966>
- [43] S. W. Njoroge, J. H. Nichols, "Risk management in the clinical laboratory," *Ann. Lab. Med.*, vol. 34, no. 4, pp. 274-278, doi: <https://doi.org/10.3343/alm.2014.34.4.274>
- [44] R. K. Bania, "Ensemble of deep transfer learning models for real-time automatic detection of face mask," *Multimed. Tools Appl.*, vol. 82, no. 16, pp. 25131-25153, 2023, doi: <https://doi.org/10.1007/s11042-023-14408-y>
- [45] I. C. Juanatas, R. A. Juanatas, "Convolution Neural Network Approach for Facial Mask Detection," en 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE), Nara, Japan, 2023, pp. 1152-1155, doi: <https://doi.org/10.1109/GCCE59613.2023.10315441>
- [46] L. Kamal, R. J. R. Raj, "Harnessing deep learning for blood quality assurance through complete blood cell count detection," *e-Prime - Adv. Electr. Eng. Electron. Energy*, vol. 7, 2024, art. no. 100450, doi: <https://doi.org/10.1016/j.prime.2024.100450>
- [47] P. Henderson, V. Ferrari, "End-to-end training of object class detectors for mean average precision," en 13th Asian Conference on Computer Vision, Taipei, Taiwan, 2017, pp. 198-213, doi: https://doi.org/10.1007/978-3-319-54193-8_13
- [48] J. E. See, "Visual Inspection: A Review of the Literature," Sandia National Laboratories, Albuquerque, New Mexico, Rep. no. SAND2012-8590, 2012. [Online]. Available: doi: <https://www.osti.gov/servlets/purl/1055636>
- [49] B. A. Graybeal et al., "Visual Inspection of Highway Bridges," *J. Nondestruct. Eval.*, vol. 21, pp. 67-83, 2002, doi: <https://doi.org/10.1023/A:1022508121821>
- [50] M. M. Stallard (Voelker), C. A. MacKenzie, F. E. Peters, "A probabilistic model to estimate visual inspection error for metalcastings given different training and judgment types, environmental and human factors, and percent of defects," *J. Manuf. Syst.*, vol. 48, pp. 97-106, 2018, doi: <https://doi.org/10.1016/j.jmsy.2018.07.002>
- [51] M. Plebani et al., "Quality indicators in laboratory medicine: A fundamental tool for quality and patient safety," *Clin. Biochem.*, vol. 46, no. 13-14, pp. 1170-1174, 2013, doi: <https://doi.org/10.1016/j.clinbiochem.2012.11.028>
- [52] L. Sciacovelli et al., "Quality Indicators in Laboratory Medicine: The status of the progress of IFCC Working Group 'laboratory Errors and Patient Safety' project," *Clin. Chem. Lab. Med.*, vol. 55, no. 3, pp. 348-357, 2017, doi: <https://doi.org/10.1515/cclm-2016-0929>
- [53] S. Shahangian, S. R. Snyder, "Laboratory medicine quality indicators: A review of the literature," *Am. J. Clin. Pathol.*, vol. 131, no. 3, pp. 418-431, 2009, doi: <https://doi.org/10.1309/ajcpj8ji4zldque>
- [54] P. E. A. Kuriki, F. C. Kitamura, "Artificial Intelligence in Radiology: A Private Practice Perspective From a Large Health System in Latin America," *Semin. Roentgenol.*, vol. 58, no. 2, pp. 203-207, 2023, doi: <https://doi.org/10.1053/j.ro.2023.01.006>
- [55] R. Pandian, V. Vedanarayanan, D. N. S. Ravi Kumar, R. Rajakumar, "Detection and classification of lung cancer using CNN and Google net," *Meas.: Sens.*, vol. 24, 2022, art. no. 100588, doi: <https://doi.org/10.1016/j.measen.2022.100588>
- [56] I. Shahin, A. B. Nassif, M. B. Alsabek, "COVID-19 Electrocardiograms Classification using CNN Models," en 14th International Conference on Developments in eSystems Engineering, DeSE 2021, Sharjah, Emiratos Árabes Unidos, 2021, pp. 448-452, doi: <https://doi.org/10.1109/DESE54285.2021.9719358>
- [57] H. J. Sandoval-Cuellar et al., "Image-based Glaucoma Classification Using Fundus Images and Deep Learning," *Rev. Mex. Ing. Biomed.*, vol. 42, no. 3, pp. 28-41, 2021, doi: <https://doi.org/10.17488/RMIB.42.3.2>
- [58] A. U. Haq et al., "MCNN: a multi-level CNN model for the classification of brain tumors in IoT-healthcare system," *J. Ambient. Intell. Humaniz. Comput.*, vol. 14, no. 5, pp. 4695-4706, 2022, doi: <https://doi.org/10.1007/s12652-022-04373-z>
- [59] R. Patgiri, V. Ajantha, S. Bhuvanewari, V. Subramaniaswamy, "Intelligent Defect Detection System in Pharmaceutical Blisters Using YOLOv7," en 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), Vellore, India, 2024, pp. 1-7, doi: <https://doi.org/10.1109/ic-ETITE58242.2024.10493735>
- [60] B. Wu, C. Pang, X. Zeng, X. Hu, "ME-YOLO: Improved YOLOv5 for Detecting Medical Personal Protective Equipment," *Appl. Sci.*, vol. 12, no. 23, 2022, art. no. 11978, doi: <https://doi.org/10.3390/app122311978>
- [61] Q. T. Do, J. Chaudri, "Creating Computer Vision Models for Respiratory Status Detection," en 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Reino Unido, 2022, pp. 1350-1353, doi: <https://doi.org/10.1109/EMBC48229.2022.9871978>

- [62] D. Aguilar Bucheli et al., "Artificial Intelligence in Medical Education: Latin American context," *Metro Ciencia*, vol. 31, no. 2, pp. 21-34, 2023, doi: <https://doi.org/10.47464/MetroCiencia/vol31/2/2023/21-34>
- [63] F. M. Delpino et al., "Emergency department use and Artificial Intelligence in Pelotas: design and baseline results," *Rev. Bras. Epidemiol.*, vol. 26, 2023, art. no. e230021, doi: <https://doi.org/10.1590/1980-549720230021>
- [64] R. Garcia Alonso, U. Thoene, D. Davila Benavides, "Digital Health and Artificial Intelligence: Advancing Healthcare Provision in Latin America," *IT Prof.*, vol. 24, no. 2, pp. 62-68, 2022, doi: <https://doi.org/10.1109/MITP.2022.3143530>
- [65] F. C. Kitamura et al., "Forging Connections in Latin America to Advance AI in Radiology," *Radiol. Artif. Intell.*, vol. 4, no. 5, 2022, art. no. e220125, doi: <https://doi.org/10.1148/ryai.220125>
- [66] M. Álvarez Vega, L. M. Quirós Mora, M. V. Cortés Badilla, "Inteligencia artificial y aprendizaje automático en medicina," *Rev. Med. Sinerg.*, vol. 5, no. 8, 2020, art. no. e557, doi: <https://doi.org/10.1109/CVCI51460.2020.9338663>
- [67] C. Fierro, J. Pérez, J. Mora, "Predicting Unplanned Readmissions with Highly Unstructured Data," 2020, arXiv:2003.11622, doi: <https://doi.org/10.48550/arXiv.2003.11622>
- [68] M. Jamei et al., "Predicting all-cause risk of 30-day hospital readmission using artificial neural networks," *PLoS One*, vol. 12, no. 7, 2017, art. no. e0181173, doi: <https://doi.org/10.1371/journal.pone.0181173>
- [69] F. Chang et al., "A mobile vision inspection system for tiny defect detection on smooth car-body surfaces based on deep ensemble learning," *Meas. Sci. Technol.*, vol. 30, no. 12, 2019, art. no. 125905, doi: <https://doi.org/10.1088/1361-6501/ab1467>
- [70] A. Rachman, R. M. C. Ratnayake, "Machine learning approach for risk-based inspection screening assessment," *Reliab. Eng. Syst. Saf.*, vol. 185, pp. 518-532, 2019, doi: <https://doi.org/10.1016/j.res.2019.02.008>
- [71] J. Aust, D. Pons, "Comparative Analysis of Human Operators and Advanced Technologies in the Visual Inspection of Aero Engine Blades," *Appl. Sci.*, vol. 12, no. 4, 2022, art. no. 2250, doi: <https://doi.org/10.3390/app12042250>
- [72] O. Wosner, G. Farjon, A. Bar-Hillel, "Object detection in agricultural contexts: A multiple resolution benchmark and comparison to human," *Comput. Electron. Agric.*, vol. 189, 2021, art. no. 106404, doi: <https://doi.org/10.1016/j.compag.2021.106404>
- [73] T. Y. Kim et al., "A Deep Learning Technique for Optical Inspection of Color Contact Lenses," *Appl. Sci.*, vol. 13, no. 10, 2023, art. no. 5966, doi: <https://doi.org/10.3390/app13105966>
- [74] K. Y. Chan et al., "Ball bonding inspections using a conjoint framework with machine learning and human judgement," *Appl. Soft. Comput.*, vol. 102, 2021, art. no. 107115, doi: <https://doi.org/10.1016/j.asoc.2021.107115>
- [75] I. Rio-Torto et al., "Hybrid Quality Inspection for the Automotive Industry: Replacing the Paper-Based Conformity List through Semi-Supervised Object Detection and Simulated Data," *Appl. Sci.*, vol. 12, no. 11, 2022, art. no. 5687, doi: <https://doi.org/10.3390/app12115687>
- [76] M. S. Shaikh et al., "Performance evaluation of a coagulation laboratory using Sigma metrics," *Int. J. Health Care Qual. Assur.*, vol. 31, no. 6, pp. 600-608, 2018, doi: <https://doi.org/10.1108/ijhcqa-07-2017-0134>
- [77] M. Cheng et al., "MicroCrack-Net: A Deep Neural Network with Outline Profile-Guided Feature Augmentation and Attention-Based Multiscale Fusion for MicroCrack Detection of Tantalum Capacitors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 6, pp. 5141-5152, 2022, doi: <https://doi.org/10.1109/TAES.2022.3181117>
- [78] S. Feng, H. Zhou, H. Dong, "Using deep neural network with small dataset to predict material defects," *Mater. Des.*, vol. 162, pp. 300-310, 2019, doi: <https://doi.org/10.1016/j.matdes.2018.11.060>
- [79] D. Wang, D. He, "Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning," *Biosyst. Eng.*, vol. 210, pp. 271-281, 2021, doi: <https://doi.org/10.1016/j.biosystemseng.2021.08.015>
- [80] R. Girshick, "Fast R-CNN," en 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1440-1448, doi: <https://doi.org/10.1109/ICCV.2015.169>
- [81] S. Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149, 2017, doi: <https://doi.org/10.1109/TPAMI.2016.2577031>
- [82] K. He et al., "Mask R-CNN," en 2017 IEEE International Conference on Computer Vision (ICCV), Venecia, Italia, 2017, pp. 2980-2988, doi: <https://doi.org/10.1109/ICCV.2017.322>

MATERIAL SUPLEMENTARIO

TABLA 7. Resultados del rendimiento del modelo de YOLOv5 con 3000 imágenes entrenado con 20 épocas.

Clase	Precisión	Recall	mAP
blue_cap	0.913	0.916	0.941
good	0.93	0.909	0.923
label	0.937	0.841	0.906
not_ok	0.875	0.876	0.872
purple_cap	0.969	0.979	0.987
yellow_cap	0.831	0.863	0.886

TABLA 8. Resultados del rendimiento del modelo de YOLOv5 con 3000 imágenes entrenado con 100 épocas.

Clase	Precisión	Recall	mAP
blue_cap	0.918	0.911	0.949
good	0.934	0.922	0.94
label	0.94	0.867	0.932
not_ok	0.911	0.867	0.897
purple_cap	0.975	0.981	0.993
yellow_cap	0.867	0.855	0.898

TABLA 9. Resultados del rendimiento del modelo de YOLOv7 con 3000 imágenes entrenado con 20 épocas.

Clase	Precisión	Recall	mAP
blue_cap	0.567	0.899	0.829
good	0.722	0.666	0.694
label	0.618	0.355	0.406
not_ok	0.639	0.558	0.562
purple_cap	0.919	0.707	0.838
yellow_cap	0.627	0.808	0.72

TABLA 10. Resultados del rendimiento del modelo de YOLOv7 con 3000 imágenes entrenado con 100 épocas.

Clase	Precisión	Recall	mAP
blue_cap	0.917	0.914	0.928
good	0.919	0.899	0.902
label	0.946	0.784	0.876
not_ok	0.894	0.857	0.86
purple_cap	0.981	0.963	0.986
yellow_cap	0.845	0.853	0.886

TABLA 11. Resultados del rendimiento del modelo de YOLOv8 con 3000 imágenes entrenado con 20 épocas.

Clase	Precisión	Recall	<i>mAP</i>
blue_cap	0.928	0.914	0.947
good	0.933	0.923	0.935
label	0.929	0.879	0.936
not_ok	0.893	0.892	0.89
purple_cap	0.99	0.981	0.99
yellow_cap	0.857	0.856	0.9

TABLA 12. Resultados del rendimiento del modelo de YOLOv8 con 3000 imágenes entrenado con 100 épocas.

Clase	Precisión	Recall	<i>mAP</i>
blue_cap	0.901	0.895	0.953
good	0.928	0.928	0.943
label	0.928	0.915	0.947
not_ok	0.892	0.892	0.898
purple_cap	0.97	0.976	0.993
yellow_cap	0.881	0.877	0.916