

DIAGNÓSTICO DE ENFERMEDADES BUCALES EN IMÁGENES BASADO EN MACHINE LEARNING

DIAGNOSIS OF ORAL DISEASES IN IMAGES BASED ON MACHINE LEARNING

Bryan Steven Guncay Saldaña ^{1*}

¹ Carrera en Tecnologías de la Información, Facultad de Ingeniería Civil, Universidad Técnica de Machala. Ecuador. ORCID: <https://orcid.org/0009-0005-7153-9147>. Correo: bguncay1@utmachala.edu.ec

Bryan Steven Guevara Pando ²

² Carrera en Tecnologías de la Información, Facultad de Ingeniería Civil, Universidad Técnica de Machala. Ecuador. ORCID: <https://orcid.org/0009-0003-3102-6211>. Correo: bguevara2@utmachala.edu.ec

Bertha Eugenia Mazón Olivo ³

³ Carrera en Tecnologías de la Información, Facultad de Ingeniería Civil, Universidad Técnica de Machala. Ecuador. ORCID: <https://orcid.org/0000-0002-2749-8561>. Correo: bmazon@utmachala.edu.ec

Wilmer Braulio Rivas Asanza ⁴

⁴ Carrera en Tecnologías de la Información, Facultad de Ingeniería Civil, Universidad Técnica de Machala. Ecuador. ORCID: <https://orcid.org/0000-0002-2239-3664>. Correo: wrvivas@utmachala.edu.ec

* Autor para correspondencia: bguncay1@utmachala.edu.ec

Resumen

La detección automatizada de enfermedades bucales es un reto crucial en odontología. Este estudio analiza el impacto de técnicas avanzadas de aumento de datos y el buen ajuste de hiperparámetros en el rendimiento del modelo YOLOv8 para la identificación precisa de enfermedades orales. Aplicando la metodología CRISP-DM, se procesaron 3814 imágenes etiquetadas, utilizando transformaciones como rotaciones, ajustes de brillo, contraste, y ruido gaussiano, con el fin de aumentar la variabilidad y robustez del entrenamiento. Además, se optimizaron hiperparámetros clave como la tasa de aprendizaje y el tamaño de lote. Los resultados evidenciaron que la combinación de aumento de datos y ajustes óptimos de hiperparámetros mejoró significativamente el desempeño del modelo, alcanzando una precisión del 89.10%, sensibilidad del 86.07%, F1-Score de 87.56% y un mAP@50 del 90.70%. De igual forma, se logró un mAP@50-95 del 56.40%,

superando las configuraciones estándar utilizadas en estudios previos. La convergencia estable en las curvas de entrenamiento valida la efectividad de las estrategias aplicadas y la capacidad del modelo para generalizar correctamente. Estos hallazgos destacan la relevancia de aplicar técnicas avanzadas de preprocesamiento y optimización en el diagnóstico automatizado de enfermedades bucales, mejorando la precisión y facilitando una detección temprana en la práctica clínica.

Palabras clave: Diagnóstico; enfermedades bucales; aprendizaje automático; imágenes médicas; redes neuronales convolucionales

Abstract

Automated detection of oral diseases is a crucial challenge in dentistry. This study analyzes the impact of advanced data augmentation techniques and proper hyperparameter tuning on the performance of the YOLOv8 model for the accurate identification of oral diseases. Following the CRISP-DM methodology, 3,814 labeled images were processed using transformations such as rotations, brightness and contrast adjustments, and Gaussian noise to enhance variability and robustness during training. Additionally, key hyperparameters such as learning rate and batch size were optimized. The results showed that combining data augmentation with optimal hyperparameter adjustments significantly improved the model's performance, achieving an accuracy of 89.10%, sensitivity of 86.07%, F1-Score of 87.56%, and an mAP@50 of 90.70%. Similarly, an mAP@50-95 of 56.40% was obtained, surpassing the standard configurations used in previous studies. The stable convergence of the training curves validates the effectiveness of the applied strategies and the model's ability to generalize properly. These findings highlight the importance of implementing advanced preprocessing and optimization techniques in automated oral disease diagnosis, enhancing precision and enabling early detection in clinical practice.

Keywords: diagnosis; oral diseases; machine learning; medical images; convolutional neural networks

Fecha de recibido: 08/01/2025

Fecha de aceptado: 11/03/2025

Fecha de publicado: 01/04/2025

Introducción

El diagnóstico oportuno y preciso de las enfermedades bucales es un componente esencial para garantizar la salud bucodental y prevenir complicaciones más graves. Tradicionalmente, los especialistas han utilizado radiografías y evaluaciones clínicas visuales para identificar estas enfermedades. Sin embargo, este enfoque depende en gran medida de la experiencia del clínico, lo que puede llevar a variaciones en los resultados diagnósticos y a errores potenciales debido a la subjetividad inherente al proceso (Martins et al., 2023; Patil et al., 2022).

Las técnicas de aprendizaje profundo han demostrado excelentes capacidades en el reconocimiento de objetos en imágenes fotográficas, y esto podría facilitar la detección temprana de enfermedades bucales (Ding et al., 2021). Al automatizar y simplificar los procedimientos relacionados con la salud, la inteligencia artificial ha transformado los diagnósticos médicos y mejorado la detección temprana de enfermedades orales que afectan a una parte significativa de la población mundial.

En el campo de la odontología, los algoritmos de aprendizaje automatizado, en particular las redes neuronales convolucionales (CNN), han demostrado ser eficaces en la interpretación de imágenes médicas y radiográficas para la detección de enfermedades bucales como caries, lesiones periodontales, y patologías relacionadas con el tejido óseo, proporcionando un diagnóstico más preciso y oportuno (Galdames, 2024; Santos, 2024). Entre los modelos de aprendizaje profundo, YOLOv8 ha demostrado un rendimiento sobresaliente en la detección de características dentales y lesiones específicas, alcanzando una alta precisión en la segmentación de imágenes clínicas y radiográficas (Ozsunkar et al., 2024; Santos, 2024). Incorporar el aprendizaje profundo en la imagenología dental mejora la precisión diagnóstica y permite una atención personalizada. Con más automatización y precisión, este desarrollo podría revolucionar la prestación de atención dental al aumentar la eficacia de la identificación de enfermedades y la planificación del tratamiento (Santos, 2024).

Estudios recientes han logrado una precisión diagnóstica del 85,89%, mediante el uso de modelos sofisticados como VGG19 y DarkNet53 en radiografías panorámicas para examinar conexiones anatómicas intrincadas, como la distancia entre los molares superiores y el suelo del seno maxilar (Kadan et al., 2024). Sin embargo, la familia de arquitecturas YOLO (*You Only Look Once*) se ha vuelto más popular debido a su alta precisión, eficiencia computacional y facilidad de implementación en entornos clínicos. A pesar de la existencia de otros modelos de una etapa, como SSD (Single Shot Detector), D-SSD (*Deconvolutional Single Shot Detector*) y RetinaNet, pruebas comparativas han demostrado que YOLOv8 funciona más rápido y con mayor precisión que YOLOv5 y *Faster R-CNN*, lo que lo convierte en la mejor opción para situaciones donde la velocidad de diagnóstico es crucial (Hussain, 2023).

Además de la elección del modelo, la configuración de sus hiperparámetros juega un papel crucial en el rendimiento de los sistemas de detección de objetos. Ajustes en el umbral de confianza, la tasa de aprendizaje y la selección de optimizadores pueden influir significativamente en la precisión del modelo. (Rozada Raneros et al., s. f., 2021), observó que una reducción en el umbral de confianza puede aumentar la cantidad de detecciones, pero a expensas de una mayor tasa de falsos positivos. Asimismo, la tasa de aprendizaje afecta la estabilidad del entrenamiento, donde un valor demasiado bajo puede prolongar el proceso sin garantizar mejoras sustanciales, mientras que un valor alto podría impedir la convergencia adecuada.

Estas tecnologías de inteligencia artificial están transformando el campo de la odontología, redefiniendo la forma en que se analizan e interpretan las imágenes médicas, facilitando la identificación de estructuras anatómicas y patologías con mayor precisión. Sin embargo, a pesar de estos avances, la implementación de estas tecnologías en la práctica clínica sigue enfrentando desafíos importantes, tales como la necesidad de bases de datos más diversas, énfasis en métricas individuales, el cumplimiento de normativas éticas y la validación de los modelos en entornos clínicos reales (Busch et al., 2024). Estos factores limitan la aplicabilidad de muchos de los modelos de estos trabajos (Galdames, 2024; Martins et al., 2023).

Por ejemplo, estudios que utilizan modelos como Faster R-CNN han demostrado una gran precisión en la segmentación de imágenes dentales; sin embargo, sus largas duraciones de inferencia dificultan la implementación en tiempo real (Ozsunkar et al., 2024). Por otro lado, las investigaciones de YOLOv8 han priorizado la velocidad, pero con compromisos en la sensibilidad para detectar lesiones menores (Hussain, 2023).

Este estudio investigó cómo los cambios en hiperparámetros y técnicas de aumento de datos impactan el rendimiento de modelos de aprendizaje profundo, como YOLOv8, en el diagnóstico de enfermedades bucales mediante imágenes. Su objetivo fue analizar y comparar enfoques para evaluar su efecto en la precisión diagnóstica y aplicabilidad en odontología, planteando la pregunta clave: ¿Cómo influyen las configuraciones de los hiperparámetros y las técnicas de aumento de datos al rendimiento de modelos basados en deep learning, como YOLOv8, en la detección de enfermedades bucales en imágenes?

Para estructurar esta investigación, adoptaremos la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), reconocida por su capacidad para guiar proyectos basados en datos y garantizar la reproducibilidad y eficiencia. Esta metodología será utilizada como marco general para llevar a cabo las fases de comprensión del problema, preparación de datos, modelado, evaluación y despliegue.

Esta metodología permite una integración sistemática y replicable de herramientas de IA en la práctica odontológica, facilitando el análisis de resultados y su comparación con investigaciones previas, que utilizaron el aprendizaje profundo para el diagnóstico oral (Schröer et al., 2021; Wirth & Hipp, 2018).

Los resultados de este trabajo buscan no solo incrementar la precisión diagnóstica en imágenes bucales, sino también establecer un estándar crítico que permita integrar modelos de inteligencia artificial en la práctica clínica odontológica, superando las limitaciones actuales relacionadas con la diversidad de datos y validación en escenarios reales.

Materiales y métodos

La presente investigación se fundamenta en el desarrollo de un modelo de inteligencia artificial basado en redes neuronales convolucionales para la detección de enfermedades bucales en imágenes médicas. Se implementó la metodología CRISP-DM, ampliamente utilizada en la minería de datos y aprendizaje profundo, asegurando un proceso estructurado y replicable en cada fase de la investigación (Schröer et al., 2021).

You Only Look Once (YOLOv8)

Tradicionalmente, la detección de objetos usaba dos redes separadas para detección y clasificación. Sin embargo, esta arquitectura dual suele resultar en procesos lentos, lo que dificulta su aplicación en la detección de objetos en tiempo real. En contraste, YOLO (*You Only Look Once*) redefine la detección de objetos como un problema de regresión único, procesando directamente los píxeles de la imagen para predecir cada punto de coordenada en los cuadros delimitadores, además también las probabilidades de cada clase de manera simultánea con una sola red neuronal convolucional (Scholnik-Elias et al., 2023).

YOLOv8, desarrollada por Ultralytics, se fundamenta en los principios de YOLO, según lo documentado por

(Jocher & Ultralytics, 2023), brindando mayor precisión y velocidad que sus predecesores, destacándose en los modelos ligeros YOLOv8 Nano (YOLOv8n) y YOLOv8 Small (YOLOv8s). Su rendimiento superior se muestra en la Figura 1, superando a otros modelos en detección rápida.

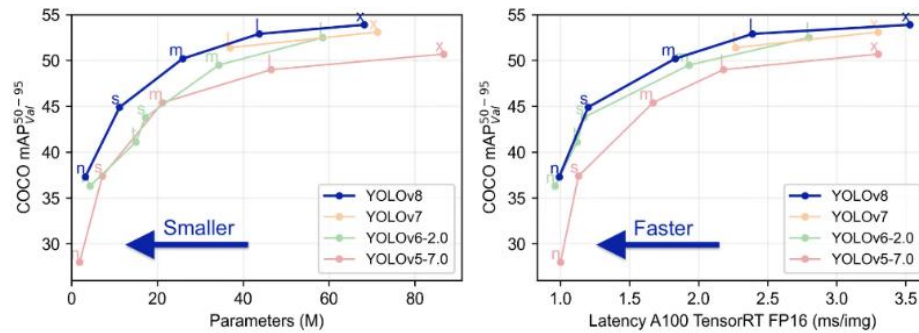


Figura 1. Gráficas comparativas de precisión y velocidad de modelos de YOLO.

Fuente: Tomado de (Jocher & Ultralytics, 2023).

La arquitectura de YOLOv8, representada en la Figura 2, utiliza una Red Piramidal de Características (FPN) para detectar objetos de diferentes tamaños y escalas en una imagen. Se divide en el Backbone, que extrae características con capas convolucionales (Conv y C2f), y la Head, que predice bounding boxes y clases mediante pérdidas Bbox Loss y Cls Loss, empleando operaciones como Concat, MaxPool2d y Upsample (Scholnik-Elias et al., 2023). Este diseño permite usar un modelo preentrenado en el dataset COCO para detección general de objetos.

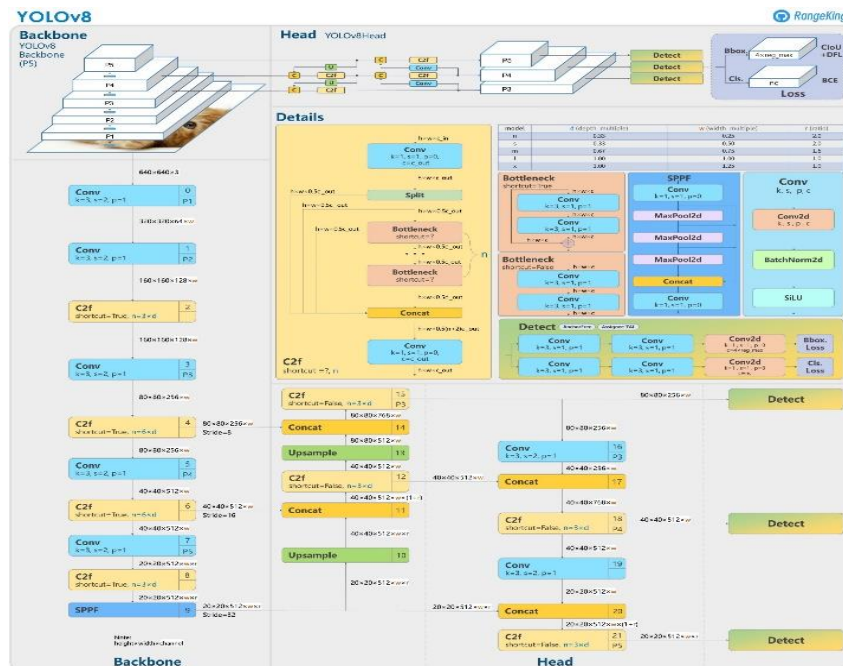


Figura 2. Arquitectura de red YOLOv8

Fuente: Tomado de (RangeKing, 2023)

Entorno computacional

El modelo se desarrolló, entrenó y evaluó en Google Colab Pro, usando una GPU NVIDIA A100 (40 GB VRAM) para procesar grandes volúmenes de datos eficientemente. El entorno incluyó 83.5 GB de RAM, 235.7 GB de almacenamiento y Python 3.9 como lenguaje principal.

Metodología CRISP-DM

El presente estudio tiene un enfoque cuantitativo y experimental, centrado en la evaluación de un modelo de aprendizaje profundo para el diagnóstico de enfermedades bucales a partir de imágenes médicas. La metodología utilizada fue CRISP-DM, debido a su flexibilidad y adaptabilidad en proyectos de minería de datos y machine learning.

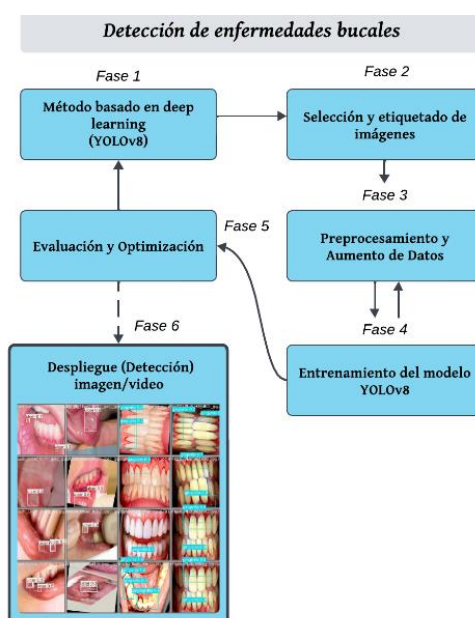


Figura 3. Fases de la metodología CRISP-DM

Fuente: Elaboración propia

Fase 1: Comprensión del negocio

En el campo de la odontología, el diagnóstico de patologías bucales depende en gran medida de la calidad de las imágenes y de las condiciones en las que se toman. Por ejemplo, la iluminación, el ángulo de captura y la resolución de las imágenes pueden diferir significativamente en entornos clínicos, lo que plantea un desafío importante para la precisión del diagnóstico. Las técnicas convencionales, basadas en la evaluación visual por parte de especialistas, no siempre son consistentes con estas variaciones, lo que puede resultar en errores de diagnóstico. Este estudio se enfocó en utilizar y evaluar un modelo de diagnóstico automatizado basado en técnicas de aprendizaje profundo, específicamente la versión YOLOv8m, para detectar enfermedades bucales a partir de imágenes médicas.

Fase 2: Selección y etiquetado de imágenes

En esta etapa se recopiló, seleccionó y etiquetó un conjunto de imágenes para entrenar el modelo YOLOv8, asegurando que fueran representativas de las condiciones reales de la práctica clínica odontológica.

- **Recopilación de datos**

Se obtuvieron 3814 imágenes en formatos .jpg y .png de fuentes confiables como NinjaDataset, Kaggle y Roboflow Universe, con resoluciones entre 640×480 y 1920×1080 píxeles. Las imágenes estaban etiquetadas en clases de enfermedades bucales como caries, gingivitis, úlceras y cálculos dentales.

- **Características del conjunto de datos**

Perspectivas variadas: Capturas desde ángulos frontal, lateral y superior.

Condiciones variables: Imágenes con diferentes niveles de iluminación y calidad, simulando entornos clínicos reales.

- **Etiquetado y preparación**

Se utilizó la herramienta Roboflow Annotate para organizar y verificar las etiquetas de las imágenes, realizando correcciones manuales cuando fue necesario. El conjunto de datos fue exportado en un formato compatible con YOLOv8, incluyendo las coordenadas de bounding boxes, y quedó listo para su integración en Google Colab para el entrenamiento del modelo.

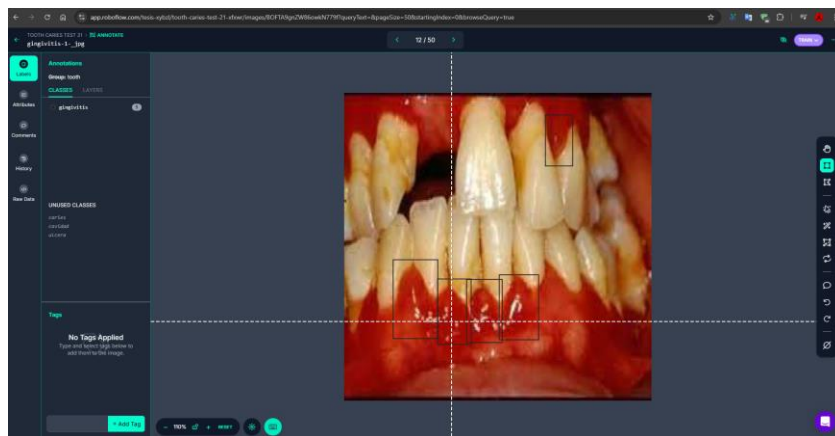


Figura 4. Roboflow annotate (2024, 16 de diciembre)

Fuente: Elaboración propia

Fase 3: Preprocesamiento y aumentación de datos

En esta etapa, se optimizó el conjunto de datos utilizando la plataforma Roboflow para mejorar la calidad de las imágenes y aumentar su diversidad, permitiendo que el modelo YOLOv8 generalice mejor en diversos escenarios clínicos. Roboflow facilitó la aplicación de técnicas de preprocesamiento y aumento de datos de manera eficiente, seguidas de una organización estructurada del dataset para su uso en el entrenamiento.

- **Preprocesamiento de imágenes**

Se utilizaron las siguientes funcionalidades de Roboflow para estandarizar y preparar las imágenes:

- **Auto-Orientación:** Se activó la función "Auto-Orient" para asegurar que todas las imágenes se orientaran correctamente, eliminando inconsistencias causadas por metadatos de orientación incorrectos. Esta función garantiza que el modelo reciba todas las imágenes en la orientación adecuada, mejorando la coherencia del conjunto de datos.
- **Redimensionamiento Uniforme:** Todas las imágenes fueron redimensionadas a una resolución de 640x640 píxeles utilizando la función de redimensionamiento de Roboflow. Esta resolución se eligió para cumplir con los requisitos del modelo YOLOv8 y asegurar que todos los detalles relevantes de las lesiones fueran visibles.

Para enriquecer el conjunto de datos y aumentar su diversidad, se aplicaron técnicas avanzadas de aumentación, generando múltiples variaciones por cada imagen original y clases desbalanceadas:

- **Volteo horizontal y vertical:** Simulación de orientaciones comunes en la captura de imágenes.
- **Rotaciones en distintos ángulos:** Adaptación a diferentes ángulos de captura, incluyendo giros parciales y completos.
- **Recorte y acercamiento:** Simulación de capturas más cercanas para reflejar diferentes enfoques en las imágenes.
- **Rotaciones aleatorias:** Representación de desajustes ligeros en ángulos típicos de captura en entornos clínicos.
- **Ajustes de brillo y contraste:** Simulación de diversas condiciones de iluminación para hacer el modelo más robusto frente a variaciones en calidad visual.



Figura 5. Muestra de preprocesamiento de imágenes
Fuente: Elaboración propia

• Organización del conjunto de datos

El conjunto de datos se dividió de manera estratificada para garantizar una representación equilibrada de las clases en cada subconjunto:

- 80% para entrenamiento: Base principal para que el modelo aprenda patrones.
- 15% para validación: Verificación del rendimiento durante el entrenamiento.
- 5% para prueba: Evaluación final del modelo en datos no vistos previamente.

Esta estrategia garantiza que el modelo se pruebe en escenarios independientes, reduciendo el riesgo de sesgo por reutilización de datos. El dataset está compuesto por 3814 imágenes, distribuidas entre los conjuntos de entrenamiento, validación y prueba. Cabe destacar que una misma imagen puede estar etiquetada con más de una clase, lo que influye en la distribución de las categorías. La Tabla 1 presenta el desglose detallado del conjunto de datos original.

Tabla 1. Distribución del conjunto de datos original

Clase	Entrenamiento	Validación	Prueba
Caries	1253	258	71
Gingivitis	777	143	140
Úlcera	809	140	56
Cálculos	925	173	42
Total	3051	572	191

• Estructura de datos en YOLO

Tras aplicar el preprocesamiento y aumento de datos, se estructuraron en carpetas según el formato requerido por YOLO. Se definieron directorios principales (train/, valid/, test/), que a su vez incluyen subcarpetas para almacenar imágenes (images/) y etiquetas (labels/). Las imágenes y sus etiquetas en el formato requerido (coordenadas de bounding boxes en archivos de texto) se almacenaron en estas subcarpetas para entrenamiento, validación y pruebas. La organización típica de los datos sigue el siguiente formato:

```

▼ test
  ▸ images
  ▸ labels
▼ train
  ▸ images
  ▸ labels
▼ valid
  ▸ images
  ▸ labels
  README.dataset.txt
  README.roboflow.txt
  data.yaml
    
```

Figura 6: Estructura de datos en Yolo

Fuente: Captura de pantalla tomada desde Google Colab.

El archivo `data.yaml`, crucial para la detección, define las rutas de los datos, el número de clases y las clases utilizadas.

```
train: ../train/images
val: ../valid/images
test: ../test/images

nc: 4
names: ['calculos', 'caries', 'gingivitis', 'ulcera']
```

Figura 7. Configuración del archivo data.yaml para el conjunto de datos en YoloV8.

Fuente: Elaboración propia

- **Exportación del conjunto de datos**

El dataset preprocesado y aumentado se exportó en el formato requerido por YOLOv8:

- Archivos de texto (txt): Se generaron archivos de texto con las coordenadas de las bounding boxes para cada imagen, con su estructura correspondiente como se muestra en la Figura 6.

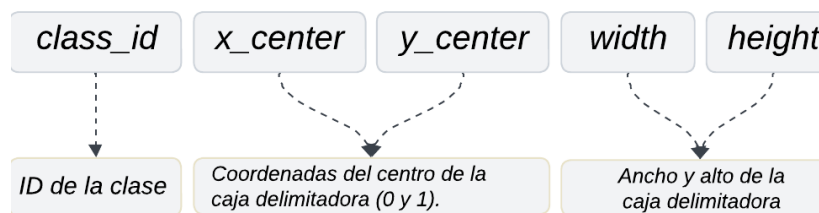


Figura 8. Estructura de etiqueta .txt Darknet de YOLO

Fuente: (*Ultralytics & Roboflow, 2023*)

El conjunto de datos una vez exportado, preprocesado y organizado, ya está listo para ser cargado en Google Colab, donde se realizó el entrenamiento correspondiente. En la siguiente fase, se detallará el proceso de entrenamiento del modelo YOLOv8, incluyendo la configuración de los hiperparámetros y las técnicas de aumento de datos aplicadas.

Fase 4: Entrenamiento del modelo

En esta fase, se llevó a cabo el entrenamiento y la optimización del modelo YOLOv8, aplicando un proceso iterativo para mejorar la detección de enfermedades bucales con buena precisión. Se realizaron cuatro casos de estudio distintos para evaluar el impacto de los hiperparámetros y las técnicas de aumento de datos, utilizando el conjunto de datos optimizado y preparado previamente, diseñado para integrarse sin problemas en entornos de nube como Google Colab.

- **Estrategia de entrenamiento**

- **Iteraciones continuas:** Se realizaron múltiples ciclos de entrenamiento y evaluación, ajustando configuraciones como los hiperparámetros y las etiquetas del conjunto de datos para resolver problemas de sobreajuste y mejorar la precisión.

- **Monitoreo de métricas:** Durante el entrenamiento, se monitorearon métricas clave como mAP, precisión y recall para evaluar el rendimiento del modelo en tiempo real.
- **Optimización de hiperparámetros**

Se realizaron ajustes cuidadosos en los parámetros clave para equilibrar la generalización del modelo y la reducción del sobreajuste:

- Tasa de aprendizaje: Ajustada dinámicamente para lograr una convergencia más estable.
- Tamaño de lote: Configurados para maximizar el uso de los recursos computacionales disponibles.

Tabla 2. Hiperparámetros esenciales para entrenamiento del modelo.

Hiperparámetros esenciales		
Tasa de aprendizaje (learning_rate)	Controla el tamaño del paso en la actualización de los pesos del modelo.	0.001 - 0.0001
Tamaño de lote (Batch size)	Número de imágenes procesadas antes de actualizar los pesos.	8 - 32 (dependiendo de la GPU disponible)
Número de épocas (epochs)	Número de veces que el modelo ve el conjunto de datos completo.	200
Umbral de confianza (confidence_threshold)	Un valor bajo detecta más objetos, pero introduce falsos positivos; un valor alto descarta detecciones inciertas.	0.25 - 0.9
Optimizador	Algoritmo que ajusta los pesos del modelo para minimizar la pérdida.	AdamW
momentum	Valores más altos ayudan a estabilizar la convergencia, evitando oscilaciones.	0.8 - 0.95
Aumento de datos (augmentation)	Aplicación de transformaciones como rotaciones, recortes y cambios de brillo.	True
Weight decay	Término de regularización L2, que penaliza pesos grandes para evitar el sobreajuste.	0.0005
Scheduler (Planificador de tasa de aprendizaje)	Estrategia para ajustar la tasa de aprendizaje durante el entrenamiento.	Reducción adaptativa de la tasa de aprendizaje tras 10 épocas sin mejora en la pérdida de validación.

- **Librerías y entorno de entrenamiento**
 - Framework de aprendizaje profundo: PyTorch (versión 2.5.1+cu121).
 - Librería de detección de objetos: Ultralytics YOLOv8 (versión 8.3.80).
 - Entorno: Google Colab Pro con GPU Nvidia A100.
 - Manipulación de datos: Se utilizó Pandas (versión 2.2.1) y NumPy (versión 1.26.4).
 - Visualización: Se utilizó Matplotlib (versión 3.8.3), Seaborn (versión 0.13.2) y Plotly (versión 5.20.0).
 - Ciencia de datos y métricas: Se utilizó Scikit-learn (versión 1.4.1.post1), para las métricas.

Fase 5: Evaluación

En esta fase, se evaluó el rendimiento del modelo YOLOv8 entrenado para la detección de enfermedades bucales. El objetivo fue determinar el rendimiento del modelo utilizando métricas estadísticas y técnicas de análisis que permitieran identificar fortalezas y áreas de mejora. Además, se realizaron ajustes para optimizar el modelo y garantizar que fuera capaz de generalizar en diferentes escenarios clínicos.

- **Evaluación del modelo con métricas**

Para evaluar el desempeño del modelo, se utilizó un conjunto de datos de prueba que no había sido visto durante el entrenamiento. Este conjunto de prueba permitió verificar la capacidad del modelo para detectar enfermedades bucales en condiciones reales. Las métricas utilizadas para la evaluación incluyeron:

- **Precisión**

La precisión mide la proporción de detecciones correctas entre todas las detecciones realizadas:

$$P = \frac{TP}{TP + FP}$$

TP (True Positives) son los objetos correctamente detectados, y los FP (False Positives) son los objetos detectados incorrectamente. Un valor alto de precisión indica que el modelo tiene pocos falsos positivos. Según (Rozada Raneros et al., 2021) esta métrica es fundamental para evitar sobre detecciones en imágenes con múltiples elementos no relevantes.

- **Sensibilidad (Recall)**

La sensibilidad mide qué proporción de los objetos reales fueron detectados por el modelo. Un alto valor de sensibilidad indica que el modelo detecta la mayoría de los objetos presentes en la imagen. El autor (Redmon et al., 2023), explican que un modelo con alta sensibilidad es útil en aplicaciones críticas como la detección médica.

$$R = \frac{TP}{TP + FN}$$

- **Puntaje F1 (F1-Score)**

Según (Redmon & Farhadi, 2023) el F1-Score es clave para evaluar modelos en los que tanto la precisión como la sensibilidad es importante. El F1-Score es la media armónica entre precisión y sensibilidad, proporcionando un balance entre ambas métricas:

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}$$

- **Precisión Media Promediada (mAP)**

El mAP mide la precisión del modelo en diferentes umbrales de confianza. Se calcula como el promedio de la precisión media (AP) para todas las clases:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i$$

En su estudio, (Redmon et al., 2023) demostraron que YOLO logra una mayor velocidad de detección en comparación con otros métodos, manteniendo un mAP competitivo.

○ Matriz de Confusión

Según el autor (Raneros Rozada, 2021), enfatiza que la matriz de confusión es esencial para visualizar el comportamiento del modelo y ajustar distintos umbrales. La matriz de confusión muestra cómo el modelo clasifica correctamente o incorrectamente las detecciones.

Fase 6: Despliegue

El modelo entrenado fue exportado, generando como resultado el archivo final de entrenamiento best.pt, el cual contiene los pesos optimizados y configuraciones finales que permiten al modelo alcanzar su mejor desempeño. Este archivo fue integrado en un software prototipo desarrollado con Python y Flask para la gestión del backend, mientras que la interfaz de usuario fue diseñada utilizando HTML, CSS y Tailwind CSS.

Este sistema permite la carga y procesamiento de imágenes o videos en un entorno de prueba que simula un escenario de diagnóstico odontológico automatizado. La plataforma cuenta con una interfaz intuitiva en la que los usuarios pueden subir imágenes desde su computadora o mediante un código QR, permitiendo así la carga rápida desde dispositivos móviles.

Además, cuando las imágenes son procesadas, el usuario tiene la opción de ajustar parámetros clave, como el umbral de confianza y la Intersección sobre Unión (IoU), lo que proporciona flexibilidad en la detección de enfermedades y permite adaptar el modelo a diversas condiciones clínicas.

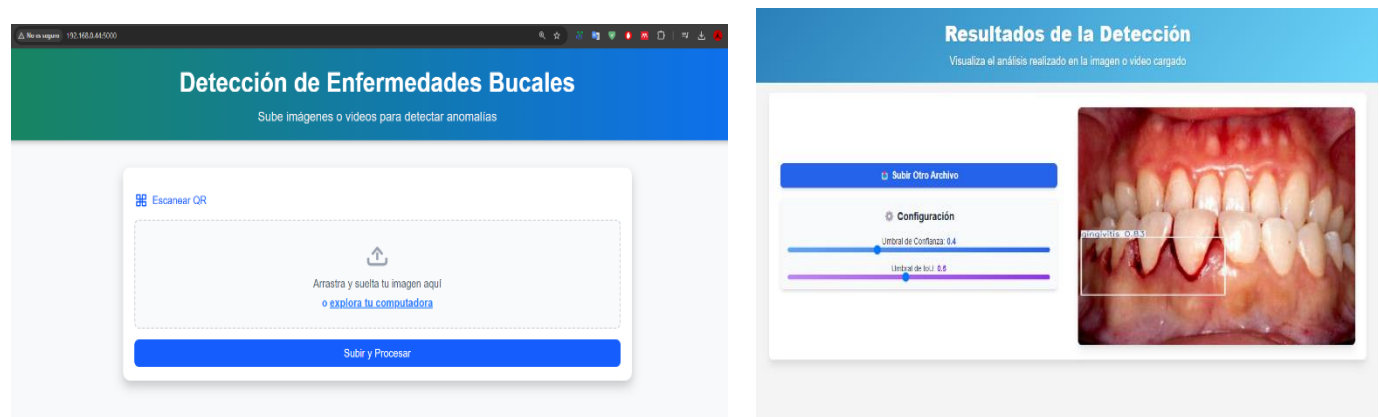


Figura 9. Interfaz del prototipo de diagnóstico odontológico automatizado.

Fuente: Elaboración propia

Resultados y discusión

Los experimentos realizados permitieron evaluar el impacto de diferentes configuraciones en el desempeño del modelo YOLOv8 para la detección de enfermedades bucales. Se analizaron cuatro escenarios experimentales, comparando la precisión, la capacidad de generalización y la estabilidad del entrenamiento.

Tabla 3. Configuración de los experimentos elaborados.

Caso de Estudio	Tamaño del dataset	Técnicas aplicadas	Tamaño de imagen (px)	Optimizador	Hiperparámetros
Caso 1: Dataset Original con Hiperparámetros estándar de YOLOv8	3814	Ninguna	640	AdamW	lr=0.00125, batch=16, epochs=200, momentum=0.9
Caso 2: Dataset Original con Hiperparámetros Optimizados	3814	Ninguna	640	AdamW	lr=0.0001, lrf=0.01, momentum=0.937, weight_decay=0.0005, save_period=15, workers=4, batch=32, epochs=200, imgsz=640, cos_lr=True
Caso 3: Aumento de Datos con Hiperparámetros estándar de YOLOv8	4230	Rotaciones, volteos, ajustes de brillo	640	AdamW	lr=0.001, imgsz=640, batch=16, epochs=200
Caso 4: Aumento de Datos Refinado con Hiperparámetros Avanzados	4230 con transformación generadas aleatorias en tiempo real	Rotaciones sutiles, ajustes de escala, recortes aleatorios, ajustes de brillo y contraste leves, desenfoque gaussiano, transformaciones elásticas, ruido gaussiano	640	AdamW	lr=0.0001, lrf=0.01, momentum=0.937, weight_decay=0.0005, workers=4, cos_lr=True, batch=32, epochs=200, imgsz=640

Con estas configuraciones, se buscó analizar el efecto de la optimización de hiperparámetros y el aumento de datos en el rendimiento del modelo YOLOv8 para la detección de enfermedades bucales.

• Comparación de métricas de rendimiento

Para evaluar la efectividad de cada configuración, se midieron métricas clave como la precisión, recall, F1-score, mAP@50 y mAP@50-95. El mAP@50 mide la precisión media considerando un umbral de IoU de 0.5, mientras que el mAP@50-95 evalúa el desempeño en múltiples umbrales, lo que proporciona una medida más completa de la capacidad de generalización del modelo. Los resultados se presentan en la siguiente tabla:

Tabla 4. Comparación de métricas de rendimiento de los modelos en cada escenario.

Escenario	Precisión (%)	Recall (%)	F1-Score (%)	mAP@50 (%)	mAP@50-95 (%)
Caso 1 (Base)	89.29	82.08	85.53	88.51	51.48
Caso 2 (Hiperparámetros optimizados)	88.03	84.96	86.47	88.83	55.81
Caso 3 (Aumento de datos)	88.50	81.80	85.02	86.90	51.10
Caso 4 (Aumento refinado + Hiperparámetros avanzados)	89.10	86.07	87.56	90.70	56.40

El caso 4 mostró la mejor combinación de métricas, con el mayor recall (86.07%) y F1-score (87.56%), lo que indica un mejor equilibrio entre precisión y sensibilidad. Además, alcanzó el mayor mAP@50-95 (56.40%), con un incremento de 4.92 puntos porcentuales respecto al modelo base, demostrando una mejor capacidad de generalización. Esto evidencia que la combinación de aumento refinado de datos y optimización avanzada de hiperparámetros es la estrategia más efectiva para mejorar el rendimiento del modelo.

Por otro lado, el Caso 1 (Base), que utiliza hiperparámetros estándar, obtuvo una precisión de 89.29%, pero con un recall más bajo (82.08%), lo que indica una menor capacidad para detectar correctamente todos los casos positivos.

El caso 2, que aplica solo optimización de hiperparámetros, logró un recall mejorado de 84.96%, aunque con una ligera disminución en precisión (88.03%). Sin embargo, su mAP@50-95 de 55.81% demuestra que los ajustes en hiperparámetros contribuyen a una mejor generalización del modelo.

En contraste, el caso 3, basado en aumento de datos sin optimización avanzada, no mostró mejoras significativas. Su recall (81.80%) es el más bajo entre los escenarios evaluados y su mAP@50-95 de 51.10% indica que la estrategia no aportó mejoras sustanciales. Esto sugiere que el aumento de datos, sin un refinamiento adecuado o ajustes adicionales, no necesariamente mejora el rendimiento y puede afectar negativamente la sensibilidad del modelo.

• Análisis de las gráficas de entrenamiento y validación del caso 4

El análisis de las gráficas de entrenamiento y validación, nos permiten visualizar el comportamiento del modelo durante el proceso de aprendizaje y validar la efectividad de las estrategias aplicadas. Estas curvas no solo muestran la convergencia del modelo y la ausencia de sobreajuste, sino que también proporcionan evidencia cuantitativa y visual de cómo las optimizaciones mejoran las métricas clave, como la precisión, el recall y el mAP. A través de este análisis, demostramos que una estrategia bien estructurada de optimización y preprocesamiento contribuye significativamente a un rendimiento mejorado, a continuación, en la Figura 9 se muestra los resultados obtenidos:

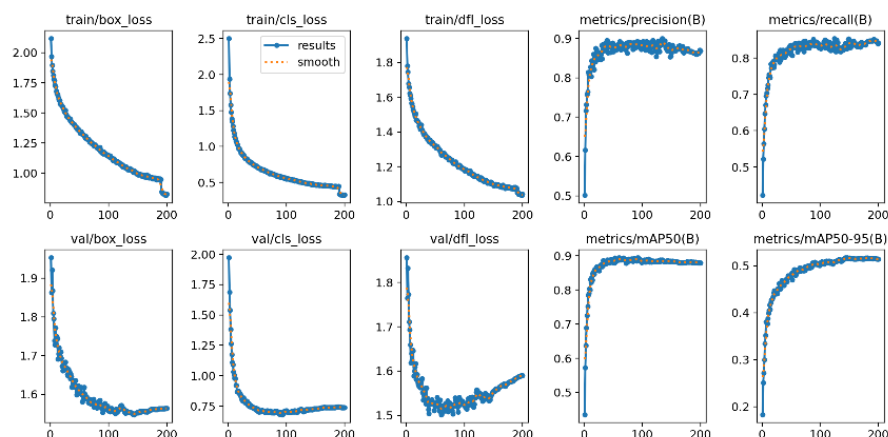


Figura 10. Curvas de pérdida y métricas de rendimiento durante el entrenamiento y validación del caso 4

Fuente: Elaboración propia

Analizando la Figura 10, se observan los valores de *train/box_loss*, *train/cls_loss* y *train/df_l_loss* donde muestran una disminución constante a lo largo de las épocas, indicando una adecuada convergencia del modelo. En validación, los valores de *val/box_loss*, *val/cls_loss* y *val/df_l_loss* reflejan un patrón similar, aunque con una ligera estabilización en épocas posteriores, lo que sugiere que el modelo no presenta sobreajuste significativo. En cuanto a las métricas de desempeño, la *precision(B)*, *recall(B)*, *mAP50(B)* y *mAP50-95(B)* alcanzan su estabilización después de aproximadamente 50 épocas. Se observa que el *mAP50* en validación alcanza valores cercanos a 0,9 en el mejor escenario, lo que indica una buena capacidad del modelo para detectar correctamente las enfermedades bucales. Las métricas de validación indican que el modelo logra un equilibrio entre *precision* y *recall*, alcanzando valores altos en *mAP50* y *mAP50-95*, lo que refuerza su capacidad de generalización.

• Análisis de las curvas de evaluación del caso 4

Las curvas de evaluación, como la curva de Precisión-Recall y la curva F1-Confianza, no solo permiten evaluar el equilibrio entre *precision* y *recall*, sino que también proporcionan información detallada sobre cómo el rendimiento varía en función de las clases y los umbrales de confianza. A continuación, en la Figura 10 se puede observar estas curvas de forma clara para cada una de las clases propuestas.

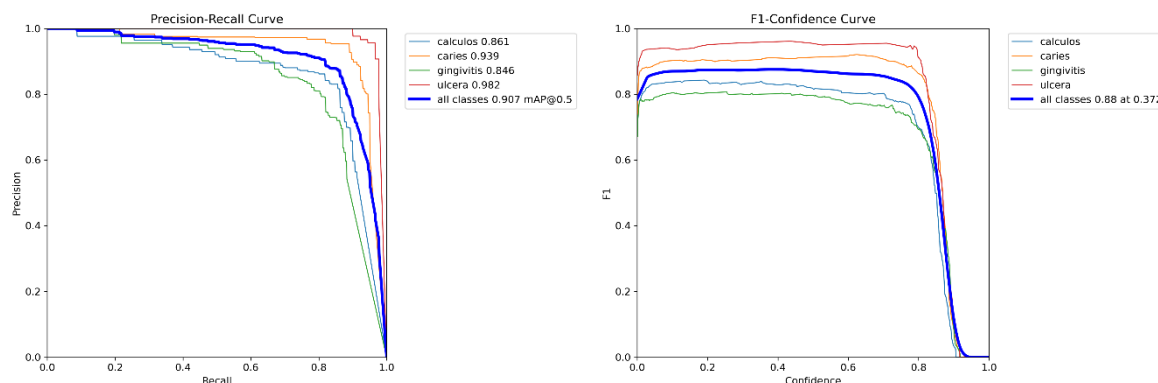


Figura 11. Curvas de Precisión-Recall y F1-Confianza por clase para el Modelo 4 de YOLOv8

Fuente: Elaboración propia

La curva de Precisión-Recall indica el equilibrio entre ambas métricas para cada clase. En el caso 4, el área bajo la curva muestra que la detección de úlcera es la más precisa con un valor de 0.982, seguida de caries (0.939). Las enfermedades como gingivitis (0.846) y cálculo (0.861) presentan un desempeño menor, indicando que estas clases son más difíciles de detectar. Si analizamos también la curva F1-Confianza refleja la relación entre el umbral de confianza y la métrica F1. Se alcanzó un valor óptimo de $F1=0.88$ con un umbral de confianza de 0.372, lo que sugiere que a este nivel el modelo logra un balance adecuado entre precisión y sensibilidad.

• Análisis de detección: Matriz de confusión y errores

El análisis de la matriz de confusión permite identificar qué clases se confunden entre sí con mayor frecuencia. Se observó que la mayor tasa de falsos positivos ocurre entre gingivitis y cálculo, lo que sugiere que estas clases presentan características visuales similares que afectan la detección.

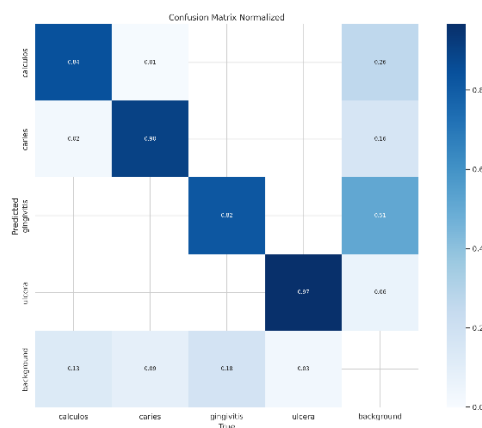


Figura 12. Matriz de Confusión Normalizada del Caso 4: Análisis de Errores en la Detección de enfermedades bucales con YOLOv8

• Predicciones del mejor modelo

Este análisis visualizó las predicciones del modelo YOLOv8 optimizado en el Caso 4, identificado como el mejor desempeño en el diagnóstico de enfermedades bucales, utilizando el archivo best4.pt. Las imágenes mostraron comparaciones entre imágenes reales de cavidades bucales y sus detecciones, destacando la capacidad del modelo para identificar enfermedades como caries, gingivitis y úlceras con cuadros delimitadores y puntuaciones de confianza. Además, se presentó la interfaz del prototipo desarrollado, permitiendo visualizar resultados y ajustar parámetros como umbrales de confianza e IoU. Estos resultados ilustran la precisión y prueba del modelo con imágenes médicas, consolidando su potencial para automatizar diagnósticos odontológicos con gran fiabilidad.

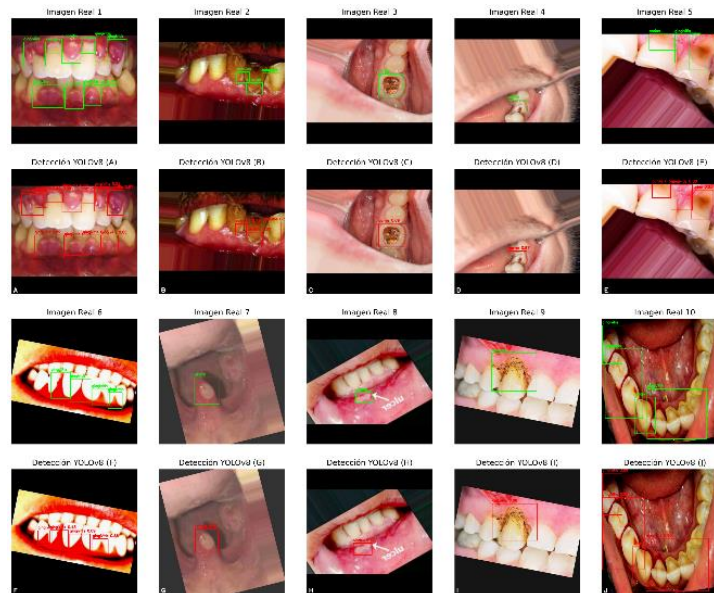


Figura 13. Resultados de detecciones del modelo YOLOv8 Caso 4 en imágenes reales de Enfermedades Bucleas
Fuente: Elaboración propia



Figura 14. Interfaz del Prototipo YOLOv8 para Detección Automatizada en Imágenes Odontológicas
Fuente: Elaboración propia

Conclusiones

El presente estudio permitió evaluar el desempeño del modelo YOLOv8m en el diagnóstico de enfermedades bucales a partir de imágenes médicas, destacando la configuración que integró un aumento de datos refinado y una optimización avanzada de hiperparámetros como la más efectiva. Esta combinación logró los mejores resultados en todas las métricas analizadas, evidenciando una mejora significativa en la precisión, sensibilidad y capacidad de generalización del modelo frente a configuraciones base. Entonces, respondiendo a la pregunta planteada, se concluye que las técnicas avanzadas de aumento de datos y la optimización de hiperparámetros influyen directamente en el rendimiento del modelo, permitiendo detectar enfermedades bucales con mayor precisión y eficacia. Estos hallazgos destacan el potencial de la inteligencia artificial como una herramienta valiosa en el diagnóstico clínico, abriendo la posibilidad de automatizar el análisis de imágenes médicas, reducir tiempos y aumentar la exactitud en la detección de enfermedades bucales.

Se recomienda validar el modelo optimizado en conjuntos de datos más amplios y diversos, con imágenes provenientes de distintas fuentes y condiciones clínicas más reales. Futuros trabajos podrían enfocarse en explorar enfoques híbridos o integrar modelos de aprendizaje profundo adicionales para seguir mejorando la precisión y la interpretabilidad en el diagnóstico automatizado en odontología.

Referencias

- Busch, F., Hoffmann, L., Truhn, D., Ortiz-Prado, E., Makowski, M. R., Bressem, K. K., Adams, L. C., Zhang, L., Zatoński, T., Xu, L., van Wijngaarden, P., van Dijk, E. H. C., Tuncel, M., Truong, M. H., Toapanta-Yanchapaxi, L. N., Thulesius, H. O., Tanioka, S., Takeda, K., Tabakova, N. G., ... Abdala, N. (2024). Global cross-sectional student survey on AI in medical, dental, and veterinary education and practice at 192 faculties. *BMC Medical Education*, 24(1). <https://doi.org/10.1186/s12909-024-06035-4>
- Ding, B., Zhang, Z., Liang, Y., Wang, W., Hao, S., Meng, Z., Guan, L., Hu, Y., Guo, B., Zhao, R., & Lv, Y. (2021). Detection of dental caries in oral photographs taken by mobile phones based on the YOLOv3 algorithm. *Annals of Translational Medicine*, 9(21), 1622-1622. <https://doi.org/10.21037/atm-21-4805>
- Galdames, I. S. (2024). *Aplicaciones de la inteligencia artificial en el diagnóstico dentomaxilofacial*. 2-3. <http://www.revestomatologia.sld.cu/index.php/est/article/view/4934>
- Hussain, M. (2023). YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. En *Machines* (Vol. 11, Número 7). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/machines11070677>
- Jocher, G., & Ultralytics. (2023, diciembre 12). *Ultralytics YOLOv8*. <https://docs.ultralytics.com/es/models/yolov8/#overview>
- Kadan, E. A., Kiliç, R., Miloğlu, Özbek, Y., & Oral, E. A. (2024). Panoramic Radiography in the Evaluation of the Relationship of Maxillary Molar Teeth and Maxillary Sinuses on the Deep Learning Models Improved with the Findings Obtained by Cone Beam Computed Tomography. *Nigerian Journal of Clinical Practice*, 5, 1-9. <https://doi.org/10.4103/njcp.njcp.220.24>

- Martins, M. V., Baptista, L., Luís, H., Assunção, V., Araújo, M. R., & Realinho, V. (2023). Machine Learning in X-ray Diagnosis for Oral Health: A Review of Recent Progress. En *Computation* (Vol. 11, Número 6). MDPI. <https://doi.org/10.3390/computation11060115>
- Ozsunkar, P. S., Özen, D. Ç., Abdelkarim, A. Z., Duman, S., Uğurlu, M., DemİR, M. R., Kuleli, B., ÇelİK, Ö., Imamoglu, B. S., Bayrakdar, I. S., & Duman, S. B. (2024). Detecting white spot lesions on post-orthodontic oral photographs using deep learning based on the YOLOv5x algorithm: a pilot study. *BMC Oral Health*, 24(1). <https://doi.org/10.1186/s12903-024-04262-1>
- Patil, S., Albogami, S., Hosmani, J., Mujoo, S., Kamil, M. A., Mansour, M. A., Abdul, H. N., Bhandi, S., & Ahmed, S. S. S. J. (2022). Artificial Intelligence in the Diagnosis of Oral Diseases: Applications and Pitfalls. En *Diagnostics* (Vol. 12, Número 5). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/diagnostics12051029>
- RangeKing. (2023, enero). *Brief summary of YOLOv8 model structure*. <https://github.com/ultralytics/ultralytics/issues/189>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (s. f.). *You Only Look Once: Unified, Real-Time Object Detection*. Recuperado 23 de febrero de 2025, de <http://pjreddie.com/yolo/>
- Santos, D. F. (2024). *Advancing Automated Dental Diagnostics: YOLOv8 Segmentation and Deep Learning Insights*. <https://doi.org/https://doi.org/10.36227/techrxiv.171995323.35014378/v1>
- Scholnik-Elias, A., Martínez-Díaz, S., Luna-Taylor, J. E., & Castro-Liera, I. (2023). Detección de armas tipo pistola mediante el uso de redes convolucionales con una arquitectura tipo YOLO y estereoscopia. *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, 11(Especial2), 196-204. <https://doi.org/10.29057/icbi.v11iespecial2.10727>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526-534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Raneros Rozada, S. (2021b). *Estudio de la arquitectura YOLO para la detección de objetos mediante deep learning* [UNIVERSIDAD DE VALLADOLID]. <http://uvadoc.uva.es/handle/10324/45359>
- Ultralytics, & Roboflow. (2023). *YOLOv8 PyTorch TXT Annotation Format*. <https://roboflow.com/formats/yolov8-pytorch-txt>
- Wirth, R., & Hipp, J. (s. f.). *CRISP-DM: Towards a Standard Process Model for Data Mining*.