

MAESTRÍA EN CIENCIA DE DATOS Y MÁQUINAS DE APRENDIZAJE CON MENCIÓN EN INTELIGENCIA ARTIFICIAL

**Trabajo previo a la obtención de título de
Magister en Ciencia de Datos y Máquinas
de Aprendizaje con mención en
Inteligencia Artificial**

AUTOR/ES:

Bonilla Zarate, María Camila
Carranza Villarroel, Carlos Daniel
Lincango Simbaña, Betsy Belén
Saguay Sagui, Bryan Alexanders

TUTOR/ES:

Alejandro Cortés López MSc.
Karla Stefania Mora Cajas

Desarrollo de un Modelo de Clasificación para la Detección
Temprana de la Retinopatía Diabética mediante Aprendizaje
Automático y Procesamiento de Imágenes Retinales

Certificación de autoría

Nosotros, Bonilla Zarate María Camila, Carranza Villarroel Carlos Daniel, Lincango Simbaña, Betsy Belén y Saguay Saguay Bryan Alexanders declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.

MARIA CAMILA
BONILLA ZARATE

Digitally signed by MARIA CAMILA
BONILLA ZARATE
Date: 2026.01.25 23:04:55 -05'00'

Firma
Bonilla Zarate María Camila

CARLOS DANIEL
CARRANZA
VILLARROEL

Firmado digitalmente por
CARLOS DANIEL CARRANZA
VILLARROEL
Fecha: 2026.01.25 23:07:24 -05'00'

Firma
Carranza Villarroel Carlos Daniel

BETSY BELEN
LINCANGO SIMBANA

Firmado digitalmente por BETSY
BELEN LINCANGO SIMBANA
Fecha: 2026.01.25 23:01:50 -05'00'

Firma
Lincango Simbaña, Betsy Belén

BRYAN ALEXANDERS
SAGUAY SAGUAY

Firmado digitalmente por BRYAN
ALEXANDERS SAGUAY SAGUAY
Fecha: 2026.01.25 22:44:19
-05'00'

Firma
Saguay Saguay Bryan Alexanders

Autorización de Derechos de Propiedad Intelectual

Nosotros, Bonilla Zarate María Camila, Carranza Villarroel Carlos Daniel, Lincango Simbaña, Betsy Belén y Saguay Saguay Bryan Alexanders en calidad de autores del trabajo de investigación titulado Desarrollo de un Modelo Predictivo para la Detección Temprana de Retinopatía Diabética mediante Aprendizaje Automático y Procesamiento de Imágenes Retinales, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

14. Enero. Quito, (enero 2026)

MARIA CAMILA
BONILLA ZARATE

Digitally signed by MARIA
CAMILA BONILLA ZARATE
Date: 2026.01.25 23:05:27 -05'00'

Firma
Bonilla Zarate María Camila

CARLOS DANIEL
CARRANZA
VILLARROEL

Firmado digitalmente por
CARLOS DANIEL CARRANZA
VILLARROEL
Fecha: 2026.01.25 23:07:56 -05'00'

Firma
Carranza Villarroel Carlos Daniel

BETSY BELEN
LINCANGO SIMBANA

Firmado digitalmente por BETSY
BELEN LINCANGO SIMBANA
Fecha: 2026.01.25 23:02:19
-05'00'

Firma
Lincango Simbaña, Betsy Belén

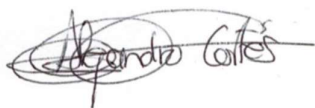
BRYAN
ALEXANDERS
SAGUAY SAGUAY

Firmado digitalmente por BRYAN
ALEXANDERS SAGUAY SAGUAY
Fecha: 2026.01.25 22:49:09
-05'00'

Firma
Saguay Saguay Bryan Alexanders

APROBACIÓN DE DIRECCIÓN Y COORDINACIÓN DEL PROGRAMA

Nosotros, **Alejandro Cortés López** Director EIG y **Karla Estefanía Mora Cajas** Coordinadora UIDE, declaramos que: **Bonilla Zarate María Camila, Carranza Villarroel Carlos Daniel, Lincango Simbaña Betsy Belén y Saguay Saguiay Bryan Alexanders**, son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortés López

Director de la

Maestría en Ciencia de Datos y Maquinas
de Aprendizaje con Mención en Inteligencia
Artificial



Karla Estefanía Mora Cajas

Coordinadora de la

Maestría en Ciencia de Datos y Maquinas de
Aprendizaje con Mención en Inteligencia
Artificial

DEDICATORIA

Dedicamos este proyecto a todos quienes son afectados por enfermedades, y a la ciencia para que sus avances estén orientados a mejorar la calidad de vida.

AGRADECIMIENTOS

Agradecemos a Dios, a la institución, a nuestras familias y amigos por el apoyo y las enseñanzas recibidas a lo largo de esta maestría.

RESUMEN

La retinopatía diabética (RD) representa una de las principales complicaciones microvasculares de la diabetes y una causa crítica de pérdida visual prevenible. Ante la limitada escalabilidad del tamizaje manual, el presente estudio desarrolla y evalúa un modelo de clasificación multiclase de imágenes de fondo de ojo para identificar el grado de severidad de la RD. Utilizando el conjunto de datos público Diabetic Retinopathy de Mendeley Data, se empleó la metodología CRISP-DM para estructurar el ciclo de vida del proyecto. La fase de modelado integró la arquitectura VGG16 mediante Transfer Learning y un posterior proceso de Fine-Tuning sobre el último bloque convolucional. Se aplicaron técnicas de preprocesamiento y aumento de datos para mitigar el impacto del desbalance de clases y mejorar la generalización. Los resultados alcanzaron una exactitud global del 79% en validación, con un rendimiento sobresaliente en la detección de pacientes sanos (No_DR) y un desempeño aceptable en estadios moderados; no obstante, se identificó una menor sensibilidad en las etapas severas y proliferativas debido a la similitud visual entre estadios adyacentes. Como factor diferenciador, se incorporó Grad-CAM para proporcionar explicabilidad visual, permitiendo identificar las regiones anatómicas que fundamentan la predicción. Finalmente, el modelo se desplegó mediante un prototipo funcional basado en FastAPI y Streamlit, demostrando la viabilidad técnica de integrar sistemas de inteligencia artificial en flujos de salud digital con fines de apoyo al diagnóstico académico.

Palabras Clave: retinopatía diabética, clasificación de imágenes, aprendizaje profundo, VGG16, transfer learning, fine-tuning, Grad-CAM, FastAPI, Streamlit.

ABSTRACT

Diabetic retinopathy (DR) represents one of the primary microvascular complications of diabetes and a critical cause of preventable vision loss. Given the limited scalability of manual screening, this study develops and evaluates a multiclass classification model of fundus images to identify the degree of DR severity. Utilizing the Diabetic Retinopathy public dataset from Mendeley Data, the CRISP-DM methodology was employed to structure the project life cycle. The modeling phase integrated the VGG16 architecture through Transfer Learning and a subsequent Fine-Tuning process on the last convolutional block. Preprocessing and data augmentation techniques were applied to mitigate the impact of class imbalance and improve generalization. The results achieved an overall accuracy of 79% in validation, with outstanding performance in detecting healthy patients (No_DR) and acceptable performance in moderate stages; however, lower sensitivity was identified in severe and proliferative stages due to visual similarity between adjacent stages. As a differentiating factor, Grad-CAM was incorporated to provide visual explainability, identifying the anatomical regions that underpin the prediction. Finally, the model was deployed through a functional prototype based on FastAPI and Streamlit, demonstrating the technical feasibility of integrating artificial intelligence systems into digital health workflows for academic diagnostic support purposes.

Keywords: diabetic retinopathy, image classification, deep learning, VGG16, transfer learning, fine-tuning, Grad-CAM, FastAPI, Streamlit.

TABLA DE CONTENIDOS (Índice)

CAPITULO 1.....	1
1. Introducción	1
1.1 Definición Del Proyecto	1
1.2 Justificación E Importancia Del Trabajo De Investigación	2
1.3 Alcance	4
1.4 Objetivos	7
1.4.1 Objetivo general.....	7
1.4.2 Objetivos específicos	7
CAPITULO 2.....	9
2. Estado del Arte	9
2.1 Marco Teórico.....	13
2.2 Diabetes Mellitus y sus complicaciones.....	13
2.3 Retinopatía Diabética	15
2.4 Inteligencia Artificial y Machine Learning en medicina.....	17
2.5 Transfer Learning aplicado a imágenes médicas.....	20
2.6 Deep Learning y Redes Neuronales Convolucionales (CNN).....	21
2.7 Fine-Tuning en redes neuronales convolucionales	23
2.8 Explicabilidad en modelos de Deep Learning.....	23
2.9 Métricas de evaluación en modelos de diagnóstico médico.....	24
2.9.1 Exactitud (Accuracy)	25
2.9.2 Sensibilidad o Recall.....	25
2.9.3 Especificidad	26
2.9.4 Precisión.....	26
2.9.5 F1-score	27
2.9.6 Matriz de confusión	27
2.9.7 Curva ROC y Área Bajo la Curva (AUC-ROC)	28
CAPITULO 3.....	29
3. DESARROLLO.....	29
3.1 Metodología CRISP-DM	29
3.1.1 Comprensión del negocio (Business Understanding)	30
3.1.2 Comprensión de los datos (Data Understanding)	30
3.1.3 Preparación de los datos (Data Preparation).....	31
3.1.4 Modelado (Modeling).....	32
3.1.5 Evaluación (Evaluación).....	33

3.1.6 Despliegue (Deployment).....	45
CAPITULO 4.....	57
4.1 Pruebas de Concepto.....	57
4.2 Análisis de Resultados	57
CAPITULO 5.....	70
5. Conclusiones y Recomendaciones	70
5.1 Conclusiones	70
5.2 Recomendaciones	72
Referencias.....	73
Anexos.....	77

LISTA DE TABLAS (Índice de tablas)

Tabla 1 Algoritmos utilizados para la detección de retinopatía diabética	11
---	----

LISTA DE FIGURAS (Índice de figuras)

Figura 1. Diagrama esquemático de la metodología CRISP-DM.....	29
Figura 2. Parámetros de entrenamiento.....	32
Figura 3. Métricas de evaluación y reporte de clasificación utilizando transfer learning	33
Figura 4. Matriz de confusión usando transfer learning.....	36
Figura 5. F1-Scores por Clase de Retinopatía.....	37
Figura 6. Curvas Roc (One-vs-Rest)	38
Figura 7. Curvas Precision, Recall and Clasificación Multiclase	39
Figura 8. Resumen de Métricas globales del modelo	40
Figura 9. AUC-ROC por Clase y Average Precision por Clase	41
Figura 10. Distribución de Probabilidades de Predicción por Clase	42
Figura 11. Reporte de clasificación Fine-Tuning.....	43
Figura 12. Reporte de clasificación Fine-Tuning.....	44
Figura 13. Explicabilidad con Grad-CAM	45
Figura 14. Arquitectura del proyecto.....	46
Figura 15. Arquitectura del Front- End y Back-End.....	48
Figura 16. Application Programming Interface (API)	50
Figura 17. Instrucciones de uso del Sistema de Detección de Retinopatía diabética.....	52
Figura 18. Sistema de Detección de Retinopatía módulo de clasificación	54
Figura 19. Sistema de Detección de Retinopatía: Módulo subir imagen	56
Figura 20. Clasificación Imagen: Retinopatía Leve	59
Figura 21. Clasificación Imagen: Retinopatía Moderada.....	60
Figura 22. Clasificación Imagen: Sin Retinopatía.....	61
Figura 23. Clasificación Imagen: Retinopatía Proliferativa.....	62
Figura 24. Clasificación Imagen: Retinopatía Severa	63
Figura 25. Comparación con el modelo entrenado desde cero.....	64
Figura 26. Matriz de confusión Transfer Learning	65
Figura 27. Resultados Fine Tuning	66
Figura 28. Matriz de confusión Fine Tuning.....	67
Figura 29. Resultados comparación general.....	68
Figura 30. Comparación de Rendimiento: Transfer Learning vs Fine Tuning	69

CAPITULO 1

En el presente capítulo se tiene como propósito describir el contexto en el que surge la necesidad de un clasificador de imágenes de fondo de ojo orientado a la detección temprana de la retinopatía diabética, una de las principales causas de pérdida visual prevenible a nivel mundial. Esta patología, asociada a la diabetes mellitus, requiere diagnósticos oportunos para evitar complicaciones irreversibles, lo que hace indispensable el desarrollo de herramientas tecnológicas que apoyen al personal médico en la identificación temprana de la enfermedad.

1. Introducción

1.1 Definición Del Proyecto

La retinopatía diabética es una de las principales causas de pérdida visual prevenible a nivel mundial y constituye un problema relevante de salud pública asociado al incremento sostenido de la diabetes. La detección temprana de esta enfermedad es fundamental para evitar su progresión hacia estadios severos; sin embargo, los métodos tradicionales de evaluación oftalmológica dependen en gran medida de la interpretación manual de imágenes por parte de especialistas, lo que implica altos costos, demanda de tiempo y una limitada capacidad de escalabilidad, especialmente en contextos con acceso restringido a servicios de salud.

En este contexto, el presente proyecto se orienta al desarrollo de un modelo de clasificación de imágenes retinianas, basado en aprendizaje profundo, para la detección temprana de la retinopatía diabética. El estudio se fundamenta en el análisis de imágenes de fondo de ojo obtenidas de un conjunto de datos público disponible en Mendeley Data, correspondiente al dataset Diabetic Retinopathy (Tuna, 2025), el cual contiene imágenes etiquetadas y documentadas para fines de investigación académica.

El problema se aborda como una tarea de clasificación multiclase, en la que el modelo aprende a identificar la presencia y el grado de severidad de la retinopatía diabética a partir de patrones visuales presentes en las imágenes. Para ello, se emplea una red neuronal convolucional preentrenada VGG16, utilizando técnicas de Transfer Learning y Fine-Tuning, que permiten aprovechar el conocimiento previamente aprendido sobre características visuales generales y adaptarlo al dominio específico de las imágenes médicas.

Como resultado complementario, el proyecto incorpora una técnica de explicabilidad basada en Grad-CAM, que permite visualizar las regiones de las imágenes retinianas que influyen en las decisiones del modelo, fortaleciendo la interpretabilidad y transparencia de los resultados. Adicionalmente, se propone un prototipo conceptual de sistema de apoyo al diagnóstico, entendido como una interfaz demostrativa que permite visualizar las predicciones del modelo con fines académicos, sin sustituir el criterio clínico ni constituir un sistema de diagnóstico médico.

1.2 Justificación E Importancia Del Trabajo De Investigación

La detección temprana de la retinopatía diabética constituye un desafío prioritario para los sistemas de salud, dado el aumento sostenido de la población con diabetes y el impacto directo de esta enfermedad en la pérdida visual prevenible. Tradicionalmente, el diagnóstico se ha basado en la evaluación clínica de imágenes retinianas por parte de especialistas, apoyada en tecnologías como cámaras de retina no midriáticas. Si bien estos métodos han demostrado ser costo-efectivos en programas de tamizaje temprano, su implementación a gran escala continúa estando limitada por la disponibilidad de recursos humanos, infraestructura especializada y tiempos de evaluación prolongados (Sender Palacios et al., 2003).

En este contexto, la inteligencia artificial emerge como una alternativa complementaria capaz de superar varias de las limitaciones de los enfoques tradicionales. Estudios recientes

destacan que los sistemas basados en aprendizaje automático y aprendizaje profundo permiten analizar grandes volúmenes de imágenes retinianas de manera automática, consistente y reproducible, reduciendo la dependencia exclusiva del juicio humano y facilitando la detección temprana en escenarios de alta demanda clínica (De et al., 2025; Martínez López et al., 2024). Estos enfoques resultan especialmente relevantes en programas de tamizaje masivo y en contextos donde el acceso a oftalmólogos es limitado.

Así pues, (Dai et al., 2021) desarrollaron un sistema de Deep Learning mediante detección de imágenes respecto a lesiones retinales. El objetivo era detectar desde etapas tempranas hasta tardías de retinopatía diabética usando 666,383 imágenes de fondo de ojo de 173,346 pacientes, las cuales alcanzaron áreas bajo la curva (AUC) de 0.916 a 0.970 para la clasificación de grados Nature y anotó lesiones retinales incluyendo microaneurismas, manchas algodinosas, exudados duros y hemorragias en 14,901 imágenes.

En particular, las redes neuronales convolucionales (CNN) han demostrado un desempeño destacado en tareas de clasificación de imágenes médicas, al ser capaces de aprender representaciones visuales complejas directamente a partir de los datos. Investigaciones recientes evidencian que las CNN aplicadas a la detección de retinopatía diabética logran resultados competitivos en la identificación de la enfermedad y sus distintos grados de severidad, consolidándose como una de las técnicas más prometedoras dentro del campo de la visión por computador aplicada a la salud (Patiño-Pérez et al., 2025). A diferencia de los métodos clásicos basados en reglas o extracción manual de características, estos modelos permiten capturar patrones sutiles difíciles de identificar mediante enfoques tradicionales.

Desde la perspectiva de la salud pública, la incorporación de modelos de clasificación automatizados puede contribuir a optimizar los procesos de detección temprana, permitiendo

una mejor priorización de pacientes y una asignación más eficiente de los recursos médicos. Al facilitar la identificación temprana de casos sospechosos, estos sistemas pueden complementar las estrategias existentes de tamizaje, reduciendo el riesgo de progresión hacia estadios avanzados y disminuyendo los costos asociados a tratamientos tardíos y pérdida de la visión.

Asimismo, el uso de un dataset público y debidamente documentado, como el conjunto Diabetic Retinopathy disponible en Mendeley Data (Tuna, 2025), garantiza la reproducibilidad y transparencia del estudio, alineándose con las buenas prácticas de investigación científica. En conjunto, este trabajo se justifica por su aporte académico y tecnológico al integrar técnicas actuales de aprendizaje profundo con un problema clínico de alto impacto, sentando bases para futuras aplicaciones en telemedicina y sistemas de apoyo al diagnóstico oftalmológico.

1.3 Alcance

El presente proyecto se circunscribe al diseño, implementación y evaluación experimental de un modelo de clasificación automática de imágenes retinianas de fondo de ojo, basado en técnicas de aprendizaje profundo (Deep Learning). El estudio se enfoca específicamente en el uso de redes neuronales convolucionales preentrenadas, aplicando estrategias de Transfer Learning y Fine-Tuning sobre la arquitectura VGG16, con el propósito de analizar su viabilidad para la detección temprana de la retinopatía diabética.

La investigación se desarrolla exclusivamente a partir de imágenes provenientes del conjunto de datos público Diabetic Retinopathy, disponible en Mendeley Data (Tuna, 2025). En este sentido, el trabajo no contempla la recolección de datos propios, la incorporación de imágenes clínicas adicionales, ni el uso de información complementaria de tipo demográfico, bioquímico o clínico (como edad, sexo, duración de la diabetes o antecedentes médicos). El

análisis se limita estrictamente al contenido visual de las imágenes retinianas proporcionadas por el dataset seleccionado.

El alcance funcional del modelo se restringe a la clasificación multiclase de imágenes retinianas, considerando la identificación de los distintos grados de severidad de la retinopatía diabética definidos en el conjunto de datos, a saber: sin retinopatía (No_DR), leve (Mild), moderada (Moderate), severa (Severe) y retinopatía diabética proliferativa (Proliferate_DR). No se aborda la detección de otras patologías oftalmológicas, como glaucoma, degeneración macular u otras enfermedades de la retina, ni se realizan tareas de segmentación de lesiones, localización de estructuras anatómicas o análisis longitudinal de la progresión de la enfermedad.

Desde el punto de vista metodológico, el proyecto se limita al entrenamiento, validación y análisis del desempeño del modelo en un entorno controlado, utilizando métricas estándar de clasificación multiclase, tales como exactitud, precisión, recall y F1-score, así como herramientas de análisis visual como matrices de confusión y curvas ROC. Adicionalmente, se incorporan técnicas de explicabilidad basadas en Grad-CAM, con el objetivo de analizar las regiones de las imágenes retinianas que influyen en las decisiones del modelo. La inclusión de métodos de Explainable Deep Learning resulta especialmente relevante en el ámbito de las imágenes médicas, donde la interpretabilidad de los modelos es un factor clave para generar confianza y apoyar la toma de decisiones clínicas. En este sentido, Grad-CAM se ha consolidado como una técnica base ampliamente utilizada para resaltar las áreas más relevantes de una imagen en tareas de clasificación, permitiendo comprender y justificar el comportamiento de modelos profundos en aplicaciones de diagnóstico médico (Suara et al., 2024). La evaluación se realiza únicamente sobre los conjuntos de datos definidos

experimentalmente, sin incluir validación clínica externa ni pruebas con datos reales de pacientes en entornos hospitalarios.

Finalmente, el proyecto posee un enfoque estrictamente académico y experimental, orientado a demostrar la factibilidad técnica del uso de modelos de aprendizaje profundo como herramientas de apoyo al análisis de imágenes retinianas. En consecuencia, no se contempla su implementación clínica directa, certificación médica ni uso como sistema de diagnóstico definitivo, sino como un prototipo conceptual que evidencia el potencial de estas tecnologías para futuras aplicaciones en sistemas de apoyo a la decisión clínica, telemedicina o investigación en salud digital.

1.4 Objetivos

1.4.1 Objetivo general

Desarrollar y evaluar un modelo de clasificación de imágenes retinianas, basado en aprendizaje profundo mediante Transfer Learning y Fine-Tuning con la arquitectura VGG16, para la detección temprana de la retinopatía diabética a partir de imágenes de fondo de ojo provenientes de un conjunto de datos público.

1.4.2 Objetivos específicos

- Realizar una revisión del estado del arte sobre la aplicación de aprendizaje profundo y modelos de clasificación de imágenes médicas en la detección de la retinopatía diabética.
- Construir y preparar un conjunto de datos de imágenes retinianas a partir del dataset público Diabetic Retinopathy disponible en Mendeley Data (Tuna, 2025), asegurando su correcta organización para tareas de clasificación multiclase.
- Aplicar técnicas de preprocesamiento y aumento de datos, tales como redimensionamiento, normalización y data augmentation, con el fin de mejorar la capacidad de generalización del modelo.
- Implementar un modelo de clasificación multiclase basado en la arquitectura VGG16, utilizando técnicas de Transfer Learning mediante el congelamiento de capas convolucionales y la incorporación de una cabeza de clasificación adaptada al problema.
- Aplicar Fine-Tuning sobre las capas superiores del modelo VGG16, ajustando hiperparámetros como la tasa de aprendizaje y el número de épocas, con el fin de especializar el modelo en patrones propios de las imágenes retinianas.
- Evaluar el desempeño del modelo mediante métricas estándar de clasificación, tales como exactitud, precisión, recall y F1-score, así como matrices de confusión para el

análisis por clase.

- Incorporar una técnica de explicabilidad basada en Grad-CAM para analizar y visualizar las regiones de las imágenes que influyen en las decisiones del modelo, fortaleciendo la interpretabilidad de los resultados.

CAPITULO 2

REVISIÓN DE LITERATURA

2. Estado del Arte

La identificación temprana de la retinopatía diabética se ha realizado tradicionalmente a través de procedimientos clínicos convencionales y, más recientemente, mediante enfoques computacionales apoyados en inteligencia artificial. Diversos estudios evidencian una transición gradual desde métodos de aprendizaje automático basados en la selección manual de características hacia modelos de aprendizaje profundo, los cuales permiten extraer representaciones directamente a partir de imágenes de la retina (De et al., 2025; Martínez López et al., 2024).

En las primeras investigaciones se emplearon métodos tradicionales como la regresión logística, las máquinas de soporte vectorial (SVM), el algoritmo de k-vecinos más cercanos (KNN) y Naive Bayes para la detección de la retinopatía diabética. Sin embargo, estos métodos suelen necesitar una etapa previa de selección y extracción de características, como textura, color, bordes o vasos sanguíneos. Esta dependencia puede reducir su capacidad para identificar patrones complejos y poco evidentes en imágenes médicas con alta variabilidad (Patiño-Pérez et al., 2025).

En los primeros estudios computacionales orientados a la detección de la retinopatía diabética se utilizaron algoritmos clásicos de aprendizaje automático que dependían de la extracción manual de características. Un ejemplo de ello es el trabajo realizado por Bhatia et al. (2017) en el cual se aplicaron distintos clasificadores de *machine learning*, como árboles de decisión alternantes, AdaBoost, Naive Bayes, Random Forest y máquinas de soporte vectorial (SVM). En este estudio, el diagnóstico se basó en características obtenidas del procesamiento de imágenes de la retina, incluyendo información del disco óptico, la presencia de lesiones

como microaneurismas y exudados, así como atributos generales de la imagen relacionados con su calidad y el preprocesamiento.

Aunque estos métodos permitieron avanzar en la automatización de la detección de la retinopatía diabética, los propios autores indican que su desempeño depende en gran medida de la calidad de la extracción manual de características y del diseño de algoritmos específicos para cada tipo de lesión. Estas restricciones limitan la capacidad de los modelos para generalizar y dificultan su aplicación en conjuntos de datos grandes y diversos. Debido a ello, surgió el interés por desarrollar enfoques basados en aprendizaje profundo, los cuales pueden aprender directamente de las imágenes sin requerir una etapa previa de ingeniería manual de variables Bhatia et al. (2017).

En una etapa intermedia entre los métodos clásicos de *machine learning* y los modelos de aprendizaje profundo de extremo a extremo, Math & Fatima (2020) propusieron un enfoque de aprendizaje adaptativo basado en segmentación para la detección de la retinopatía diabética. Según los autores, aunque los métodos tradicionales presentan limitaciones por depender de características definidas manualmente, el uso de segmentación permite aprender de forma conjunta tanto las características relevantes como los clasificadores directamente a partir de los datos.

En dicha propuesta, los autores adaptaron una red neuronal convolucional preentrenada para estimar la presencia de retinopatía diabética a nivel de segmentos de la imagen, combinando posteriormente la información de todas las regiones para obtener una clasificación final. Este enfoque resulta adecuado para tratar lesiones irregulares propias de la enfermedad. Los resultados obtenidos muestran un área bajo la curva ROC de 0.963 en el conjunto de datos Kaggle, además de altos valores de sensibilidad y especificidad (96.37 %), superando a varios métodos previos. Estos resultados evidencian la ventaja de los modelos basados en

aprendizaje profundo frente a aquellos que dependen de características definidas manualmente (Math & Fatima, 2020).

En comparación con los enfoques basados en aprendizaje automático que requieren una etapa separada de extracción de características, las redes neuronales convolucionales (CNN) han mostrado un mejor desempeño en la clasificación de imágenes retinianas, al integrar de manera automática la extracción de características y la clasificación dentro de un mismo modelo. En particular, el uso de arquitecturas preentrenadas mediante *Transfer Learning*, como VGG16, permite aprovechar el conocimiento aprendido a partir de grandes conjuntos de imágenes generales, como ImageNet, y adaptarlo al contexto médico, incluso cuando la disponibilidad de datos es limitada (Martínez López et al., 2024).

La Tabla 1 presenta una comparación general de los principales enfoques utilizados en la literatura para la detección de retinopatía diabética, considerando criterios relevantes como rapidez de entrenamiento, requerimientos computacionales, capacidad de generalización e interpretabilidad, en relación con los objetivos del presente trabajo.

Tabla 1

Algoritmos utilizados para la detección de retinopatía diabética

Enfoque	Velocidad	Costo	Extracción de características	Generalización	Interpretación	Aplicación en RD
Regresión Logística	Rápido	Bajo	Alta	Limitada	Alta	Utilizada en enfoques iniciales de detección de RD basados en características manuales; limitada para patrones visuales complejos (Bhatia et al., 2017; Tăbăcaru et al., 2023; Wu et al., 2021)
SVM	Medio	Medio	Alta	Media	Media	Empleada en métodos clásicos de clasificación de RD con buen desempeño en datasets pequeños, dependiente del kernel y las características extraídas (Bhatia et al., 2017; Tăbăcaru et al., 2023; Youldash et al., 2024).
KNN	Rápido	Bajo	Alta	Baja	Alta	Utilizado como clasificador base en estudios tempranos de retinopatía diabética por su simplicidad conceptual. No obstante, es sensible al ruido, a la alta dimensionalidad y presenta baja capacidad de generalización en imágenes médicas complejas (Bhatia et al., 2017).

Naive Bayes	Rápido	Bajo	Alta	Baja	Alta	Aplicado en sistemas tempranos de clasificación de RD; limitado por la suposición de independencia entre características (Bhatia et al., 2017; Wu et al., 2021).
CNN desde cero	Lento	Alto	No	Alta (con muchos datos)	Baja	Utilizadas para detección de RD con buen desempeño cuando se dispone de grandes volúmenes de datos, pero con alto costo computacional (Oltu et al., 2023).
CNN con Transfer Learning (VGG16)	Moderado	Medio	No	Alta	Media	Enfoque más utilizado y efectivo en la clasificación de RD, especialmente con datasets limitados, al aprovechar conocimiento preentrenado (Das et al., 2022; Oltu et al., 2023).

Nota: En la tabla 1 se observa seis enfoques y su respectiva aplicación en la retinopatía diabética.

La comparación presentada muestra que los modelos de aprendizaje automático basados en características manuales ofrecen ventajas como una implementación sencilla, mayor facilidad de interpretación y un menor costo computacional. No obstante, su rendimiento se ve restringido cuando se aplican a la clasificación de imágenes médicas complejas, como las utilizadas en la detección de la retinopatía diabética, principalmente por su dependencia de características previamente definidas. Esta situación ha sido señalada de forma consistente en estudios recientes (De et al., 2025).

En este contexto, las redes neuronales convolucionales han demostrado una mayor capacidad para identificar patrones visuales relevantes directamente a partir de las imágenes, lo que ha llevado a su adopción como uno de los enfoques más utilizados en investigaciones recientes. En particular, la aplicación de *Transfer Learning* con arquitecturas como VGG16 ofrece un balance adecuado entre desempeño, costo computacional y estabilidad durante el entrenamiento, permitiendo obtener resultados competitivos incluso cuando se trabaja con conjuntos de datos de tamaño moderado (Martínez López et al., 2024; Patiño-Pérez et al., 2025).

De manera similar, Patiño-Pérez et al. (2025) presentan el modelo DiaGAN-CNN, el cual combina *transfer learning* con una arquitectura avanzada de redes neuronales convolucionales. En su estudio, los autores reportan una precisión del 98 %, superando a varios modelos

previamente propuestos. Estos resultados ponen en evidencia la efectividad del uso parcial de *transfer learning*, mediante la congelación selectiva de capas, una estrategia especialmente útil cuando se dispone de recursos computacionales limitados y se busca aprovechar conocimiento preentrenado. En este sentido, el nivel de precisión alcanzado establece un referente relevante para investigaciones posteriores.

En este contexto, el presente trabajo se alinea con las tendencias identificadas en el estado del arte, al emplear un modelo de clasificación basado en aprendizaje profundo con VGG16, incorporando además técnicas de Fine-Tuning y explicabilidad mediante Grad-CAM, con el objetivo de no solo mejorar el desempeño del modelo, sino también aportar transparencia e interpretabilidad a sus decisiones, aspectos clave en aplicaciones médicas.

2.1 Marco Teórico

La literatura reciente coincide en señalar que la inteligencia artificial desempeña un rol cada vez más relevante en la detección temprana de la retinopatía diabética, particularmente mediante el uso de modelos de clasificación basados en redes neuronales convolucionales. Estudios como los de De et al. (2025) y Martínez López et al. (2024) resaltan el potencial de estas técnicas para mejorar la eficiencia, precisión y escalabilidad de los procesos de tamizaje, mientras que trabajos como el de Patiño-Pérez et al. (2025) evidencian el impacto positivo de las CNN en el análisis automatizado de imágenes retinianas. Estos avances se complementan con investigaciones previas sobre costo-efectividad en detección temprana, que refuerzan la necesidad de integrar nuevas tecnologías en los sistemas de salud (Sender Palacios et al., 2003).

2.2 Diabetes Mellitus y sus complicaciones

La diabetes mellitus es una enfermedad crónica de tipo metabólico que se caracteriza por la presencia de niveles elevados de glucosa en la sangre, como consecuencia de una producción insuficiente de insulina, una respuesta inadecuada del organismo a esta hormona o

la combinación de ambas. Con el paso del tiempo, esta condición puede generar daños en distintos órganos y sistemas, como el corazón, los vasos sanguíneos, los riñones, los nervios y los ojos. La diabetes tipo 2 es la forma más frecuente y ha mostrado un aumento constante en las últimas décadas, especialmente en países de ingresos bajos y medianos, asociado a factores como el envejecimiento de la población, la falta de actividad física y cambios en los patrones de alimentación (Organización Panamericana de la Salud, 2019, párr. 1).

Desde una perspectiva global y regional, la diabetes se ha posicionado como uno de los principales desafíos en salud pública. Organismos como la OMS y la OPS advierten que el diagnóstico tardío y las limitaciones en el acceso a tratamientos adecuados incrementan el riesgo de desarrollar complicaciones y de presentar mortalidad prematura. Por ello, se destaca la importancia de la detección temprana y del seguimiento continuo de las personas con diabetes. Medidas como el control regular de la glucosa, la adopción de hábitos de vida saludables y el acceso oportuno a los servicios de salud resultan claves para prevenir o retrasar las complicaciones de esta enfermedad (Organización Panamericana de la Salud, 2019).

La evolución de la diabetes mellitus puede dar lugar a diversas complicaciones microvasculares, entre las que se incluyen la nefropatía, la neuropatía y los trastornos oculares. En este último caso, la diabetes representa una de las principales causas de ceguera prevenible a nivel mundial, afectando principalmente a personas en edad laboral. Estas complicaciones no solo reducen de manera significativa la calidad de vida de los pacientes, sino que también generan una carga económica considerable para los sistemas de salud.

En la Región de las Américas, el panorama de la diabetes resulta especialmente alarmante. Según estimaciones de la OMS y la OPS, alrededor de 112 millones de adultos viven actualmente con esta enfermedad, cifra que se ha triplicado desde 1990. Además, la diabetes se encuentra entre las principales causas de ceguera, insuficiencia renal,

amputaciones de extremidades inferiores, enfermedades cardiovasculares y muerte prematura. En este contexto, la enfermedad ocupa el tercer lugar como causa de años de vida ajustados por discapacidad (AVAD) en la región (Organización Panamericana de la Salud, 2019).

En el caso de Ecuador, la diabetes mellitus se ha convertido en un problema de salud pública en constante aumento. Datos de la Encuesta Nacional de Salud y Nutrición (ENSANUT) indican que la prevalencia de diabetes en personas entre 10 y 59 años es del 1,7 %. No obstante, esta cifra se incrementa notablemente a partir de los 30 años, alcanzando cerca del 10 % en personas de 50 años, lo que significa que aproximadamente uno de cada diez ecuatorianos en este grupo etario presenta la enfermedad. Esta tendencia refleja un crecimiento de la diabetes en edades productivas, lo cual eleva el riesgo de complicaciones crónicas y genera una mayor demanda sobre el sistema de salud del país (Dávila Flores et al., 2023).

Desde el punto de vista económico, la diabetes mellitus tipo 2 implica una carga significativa para el sistema de salud ecuatoriano. Arias-Pacheco et al. (2024) estiman que el costo promedio de la atención integral por paciente alcanza aproximadamente 9.619 dólares internacionales, mientras que el tratamiento de complicaciones específicas puede ser considerablemente más elevado, como en el caso de la retinopatía diabética, cuyo costo se estima en 5.879 dólares internacionales. Al considerar una prevalencia nacional cercana al 5,5 %, el gasto total asociado al tratamiento de la diabetes tipo 2 y sus complicaciones supera los 11.500 millones de dólares internacionales, lo que pone en evidencia que las estrategias preventivas resultan más costo-efectivas que la atención en etapas avanzadas de la enfermedad.

2.3 Retinopatía Diabética

La retinopatía diabética (RD) es una complicación microvascular asociada a la diabetes mellitus que afecta directamente a los vasos sanguíneos de la retina, siendo considerada una

forma de microangiopatía retiniana. Su aparición está relacionada principalmente con la hiperglucemia sostenida, el tiempo de evolución de la diabetes y otros factores metabólicos vinculados a la enfermedad. Estas condiciones generan un daño progresivo en los capilares de la retina, alterando su permeabilidad y el flujo sanguíneo, lo que da lugar a lesiones características como microaneurismas, hemorragias intrarretinianas y exudados, así como a procesos de neovascularización anormal en etapas más avanzadas (Shukla & Tripathy, 2023, párr. 1).

Desde el punto de vista epidemiológico, la retinopatía diabética representa un problema de salud pública de alta relevancia. Se la reconoce como la tercera causa de ceguera a nivel mundial y como la principal causa de pérdida visual en personas en edad laboral en países en desarrollo. De acuerdo con estimaciones de la Organización Mundial de la Salud, esta enfermedad es responsable de aproximadamente el 5 % de los 37 millones de personas con ceguera en el mundo, lo que refleja su importante impacto social, económico y sanitario (OMS, 2005).

En el ámbito clínico, la retinopatía diabética suele clasificarse en diferentes niveles según la gravedad de la enfermedad: ausencia de retinopatía (No_DR), retinopatía leve (Mild), moderada (Moderate), severa (Severe) y retinopatía diabética proliferativa (Proliferative_DR). Esta clasificación facilita la evaluación de la progresión de la enfermedad, orienta el seguimiento médico y permite definir estrategias de tratamiento adecuadas para cada etapa.

Las complicaciones microvasculares asociadas a la diabetes mellitus afectan principalmente a órganos como la retina y los riñones, existiendo una relación cercana entre la retinopatía diabética y la presencia de microalbuminuria. Esta última es considerada un indicador temprano de daño renal y un marcador de nefropatía diabética, además de estar vinculada a un mayor riesgo de desarrollar enfermedades cardiovasculares, como la hipertensión arterial. La microalbuminuria se define como la excreción continua de albúmina en

la orina en niveles superiores a los normales, por lo que su detección temprana resulta clave para identificar el riesgo de progresión de la enfermedad renal en personas con diabetes (OMS, 2005).

De manera tradicional, el diagnóstico de la retinopatía diabética se ha llevado a cabo mediante la evaluación manual de imágenes del fondo de ojo realizada por oftalmólogos, utilizando procedimientos como la oftalmoscopia y la retinografía digital. Aunque estos métodos son efectivos desde el punto de vista clínico, presentan ciertas limitaciones, como la dependencia del criterio del especialista, la disponibilidad restringida de profesionales capacitados y la dificultad para aplicar estos procedimientos a gran escala en programas de tamizaje. Frente a este escenario, la evidencia clínica y epidemiológica resalta la necesidad de desarrollar herramientas automatizadas de apoyo al diagnóstico, que faciliten una detección temprana, objetiva y consistente de la retinopatía diabética (Vacacela Márquez et al., 2022)..

2.4 Inteligencia Artificial y Machine Learning en medicina

La inteligencia artificial (IA) engloba un conjunto de técnicas computacionales orientadas a que los sistemas puedan imitar ciertas capacidades propias del razonamiento humano, como el aprendizaje, el análisis de información y la toma de decisiones (Rai et al., 2019, p. 16). Dentro de este campo, el *machine learning* (ML) o aprendizaje automático se centra en el diseño de algoritmos capaces de aprender a partir de los datos y de identificar patrones sin necesidad de ser programados de forma específica para cada tarea (Mitchell, 1997).

En el área de la medicina, las técnicas de *machine learning* han sido aplicadas en distintos ámbitos, entre los que destacan el análisis de imágenes médicas, la clasificación de enfermedades, el apoyo al diagnóstico clínico y la evaluación de riesgos. Estas herramientas permiten manejar grandes volúmenes de información de manera eficiente y consistente, complementando el trabajo de los profesionales de la salud y favoreciendo una atención más

oportuna y objetiva. Su aplicación resulta especialmente relevante en enfermedades crónicas de alta prevalencia, donde la detección temprana juega un papel fundamental para reducir la aparición de complicaciones (Goodfellow et al., 2016).

Diversas investigaciones señalan que la retinopatía diabética constituye un campo adecuado para la aplicación de técnicas de inteligencia artificial, debido a que su diagnóstico depende en gran medida del análisis de imágenes del fondo de ojo. De acuerdo con Das et al. (2022), esta enfermedad representa una de las principales causas de deterioro visual y ceguera en personas con diabetes de larga evolución, estimándose que una proporción significativa de pacientes con más de 15 a 20 años de enfermedad desarrolla esta complicación. Los autores indican que, si bien el diagnóstico manual realizado por especialistas es clínicamente efectivo, resulta exigente y difícil de aplicar a gran escala, lo que ha impulsado el desarrollo de soluciones automatizadas basadas en machine learning y deep learning

En una etapa inicial, se desarrollaron diversos modelos de *machine learning* para la detección temprana de la retinopatía diabética, los cuales se apoyaban en la extracción manual de características visuales. Sin embargo, estos enfoques han presentado limitaciones relacionadas con su capacidad de generalización, la fuerte dependencia del diseño de dichas características y su desempeño frente a conjuntos de datos grandes o con alta variabilidad. Como resultado, estas dificultades han motivado la búsqueda de alternativas más robustas y eficientes.

Bajo este escenario, el *deep learning*, como una rama del *machine learning*, ha surgido como una opción relevante para el análisis de imágenes médicas. En particular, los modelos de aprendizaje profundo, como las redes neuronales convolucionales (CNN), permiten aprender representaciones directamente a partir de las imágenes, integrando en un solo proceso la extracción de características y la clasificación. Según Das et al. (2022), este tipo de modelos ha

mostrado mejores resultados en tareas de detección y clasificación de la retinopatía diabética, posicionándose como una de las principales soluciones en este campo.

Por su parte, la visión por computador, entendida como un área de la inteligencia artificial, cumple un rol fundamental en este tipo de aplicaciones al centrarse en el análisis automático de imágenes para obtener información relevante. En el ámbito de la imagen médica, estas técnicas permiten realizar tareas como clasificación, segmentación y detección de anomalías, facilitando la identificación de patrones visuales asociados a distintas patologías.

El procesamiento digital de imágenes médicas representa una etapa clave antes de la aplicación de modelos de clasificación e incluye técnicas como la normalización, la mejora del contraste, la reducción de ruido y el aumento de datos. En el caso de las imágenes de retina, se presentan retos adicionales asociados a variaciones en la iluminación, la presencia del disco óptico, la complejidad de la red vascular y posibles artefactos durante la captura. Estas condiciones hacen necesario el uso de modelos robustos y métodos avanzados de aprendizaje automático que permitan obtener resultados de clasificación más precisos y confiables (Litjens et al., 2017).

En este sentido, Bhimavarapu & Battineni (2022) propusieron un enfoque basado en *deep learning* que integra diversas técnicas de procesamiento de imágenes, junto con métodos de extracción y selección de características. Su estudio, titulado *Deep Learning for the Detection and Classification of Diabetic Retinopathy with an Improved Activation Function*, reporta una precisión de clasificación del 94,20 % y una medida F de 93,51 %. Para ello, utilizaron el conjunto de datos APTOS 2019, compuesto por 3.662 imágenes de fondo de ojo provenientes de zonas rurales de la India, y emplearon características relacionadas con la intensidad en escala de grises y texturas, combinadas mediante un esquema de *ensemble learning* basado en árboles de decisión.

De manera complementaria, Rao et al. (2020) evaluaron el desempeño de distintas redes neuronales convolucionales preentrenadas para la clasificación de la retinopatía diabética a partir de imágenes de fondo de ojo. A través del uso de *Transfer Learning* y técnicas de aumento de datos, compararon arquitecturas como VGG16, VGG19 y ResNet50 en un problema de clasificación multiclase. Los resultados indicaron que los modelos preentrenados alcanzaron una mayor precisión y mejor capacidad de generalización en la identificación de los diferentes niveles de severidad de la enfermedad, lo que respalda la efectividad del aprendizaje profundo en este tipo de aplicaciones.

2.5 Transfer Learning aplicado a imágenes médicas

El *Transfer Learning* se basa en el uso de modelos de aprendizaje profundo que han sido entrenados previamente con grandes conjuntos de datos generales, los cuales pueden adaptarse posteriormente a tareas específicas mediante el ajuste parcial de sus parámetros. Según Shin et al. (2016) en el contexto de las imágenes médicas, esta estrategia ha mostrado buenos resultados, ya que las primeras capas de las redes neuronales convolucionales suelen aprender características visuales generales, como bordes, texturas y patrones espaciales, que pueden reutilizarse en distintos tipos de imágenes.

En este sentido, arquitecturas como VGG16 han sido empleadas con frecuencia en estudios de clasificación de imágenes de la retina, debido a su diseño profundo y su comportamiento estable durante el entrenamiento. Como establecen Jaderberg et al. (2015), la aplicación de *Transfer Learning* permite reducir el tiempo necesario para entrenar los modelos, disminuir el riesgo de sobreajuste y mejorar la capacidad de generalización, aspectos especialmente relevantes en el ámbito médico, donde el acceso a grandes volúmenes de datos etiquetados suele ser limitado.

2.6 Deep Learning y Redes Neuronales Convolucionales (CNN)

El *deep learning* constituye una rama avanzada del aprendizaje automático orientada a modelar relaciones complejas a partir de grandes volúmenes de datos mediante el uso de redes neuronales profundas. Dentro de este enfoque, las redes neuronales convolucionales (CNN) han adquirido un rol protagónico en el análisis de imágenes, debido a su capacidad para aprender de forma automática representaciones jerárquicas que capturan información visual relevante, como bordes, texturas y estructuras espaciales de mayor complejidad (Goodfellow et al., 2016).

Desde el punto de vista estructural, las CNN están conformadas por una secuencia de capas convolucionales, funciones de activación, capas de *pooling* y capas densas, las cuales permiten extraer y combinar características visuales de manera progresiva. En aplicaciones médicas, una estrategia ampliamente adoptada es el *transfer learning*, que consiste en reutilizar modelos previamente entrenados sobre grandes conjuntos de datos generales y adaptarlos a tareas específicas mediante procesos de ajuste fino. Este enfoque resulta particularmente ventajoso en escenarios donde la disponibilidad de datos etiquetados es limitada, ya que reduce el tiempo de entrenamiento, mitiga el sobreajuste y mejora la capacidad de generalización del modelo.

En el contexto específico de la retinopatía diabética, las CNN han demostrado un desempeño notable en la clasificación automática de imágenes de fondo de ojo, permitiendo identificar tanto la presencia de la enfermedad como sus distintos niveles de severidad. Al integrar en un solo modelo la extracción de características y la etapa de clasificación, estos enfoques superan las limitaciones de los métodos tradicionales basados en características diseñadas manualmente, consolidándose como el estado del arte en el análisis automatizado de imágenes retinianas (Gulshan et al., 2016; Saint-Estevan et al., 2022).

Diversos estudios han confirmado la eficacia de los modelos basados en *deep learning* para el análisis automatizado de imágenes retinianas, evidenciando mejoras sustanciales en la detección y clasificación de la retinopatía diabética (Gulshan et al., 2016; Saint-Estevan et al., 2022). En esta línea, Patiño-Pérez et al. (2025) realizaron un análisis comparativo entre tres arquitecturas de redes neuronales convolucionales —ResNet-18, ResNet-50 y una CNN diseñada de forma personalizada— con el objetivo de clasificar imágenes de fondo de ojo en cinco categorías de severidad de retinopatía diabética. Los resultados mostraron que el modelo no preentrenado alcanzó una precisión del 91%, superando de manera consistente a las arquitecturas basadas en *transfer learning*. En contraste, ResNet-18 presentó un desempeño inestable, con una reducción significativa de la precisión desde un 70% en las etapas iniciales hasta un 26% en la fase final del entrenamiento, mientras que ResNet-50 evidenció una mejora progresiva, logrando una precisión final del 83%.

De forma complementaria, Herrero Tudela (2021), desarrolló un sistema automático basado en *deep learning* para la clasificación del grado de severidad de la retinopatía diabética a partir de retinografías. Este enfoque permitió reducir considerablemente el tiempo necesario para el diagnóstico clínico y disminuir la carga operativa de los especialistas oftalmólogos. Entre los principales aportes del estudio se destaca la capacidad del modelo para detectar y clasificar tanto casos derivables como no derivables, así como para estimar automáticamente el nivel de severidad de la enfermedad. Además, los autores resaltan que aproximadamente el 90% de los casos de retinopatía diabética pueden prevenirse mediante una detección temprana y un tratamiento oportuno, subrayando el valor de estas herramientas automatizadas. El modelo propuesto, además, fue concebido con un enfoque generalizable, permitiendo su adaptación a otros tipos de imágenes médicas.

2.7 Fine-Tuning en redes neuronales convolucionales

El *fine-tuning* puede entenderse como una fase posterior al *transfer learning*, en la cual se reajustan de forma selectiva los parámetros de un modelo previamente entrenado, concentrándose principalmente en las capas más profundas de la red. Este procedimiento permite que la red neuronal se especialice en las particularidades del dominio de interés, refinando las representaciones aprendidas inicialmente a partir de grandes conjuntos de datos generales. En el ámbito de las imágenes médicas, el *fine-tuning* resulta especialmente útil para capturar patrones clínicos sutiles que no siempre están presentes en los datos utilizados durante el entrenamiento original (Goodfellow et al., 2016; Tajbakhsh et al., 2016).

Diversos estudios coinciden en que un proceso de *fine-tuning* cuidadosamente controlado —empleando tasas de aprendizaje reducidas y un número moderado de épocas de entrenamiento— permite mejorar el rendimiento del modelo sin afectar negativamente su capacidad de generalización. Este enfoque es particularmente relevante en la clasificación de la retinopatía diabética, donde las diferencias entre los distintos estadios de la enfermedad suelen ser graduales y visualmente poco evidentes, lo que exige modelos capaces de discriminar cambios morfológicos mínimos en las imágenes retinianas (Gulshan et al., 2016; Tajbakhsh et al., 2016).

2.8 Explicabilidad en modelos de Deep Learning

Uno de los principales retos asociados al uso de modelos de *deep learning* en el ámbito clínico es su naturaleza poco transparente, ya que los procesos internos que conducen a una predicción suelen ser difíciles de interpretar. Esta limitación ha impulsado el desarrollo de enfoques de *inteligencia artificial explicable* (*Explainable Artificial Intelligence*, XAI), cuyo objetivo es hacer más comprensibles las decisiones de los modelos, facilitando su validación y aumentando la confianza de los profesionales de la salud en su uso clínico (Guidotti et al., 2018).

Dentro de este conjunto de técnicas, Grad-CAM (*Gradient-weighted Class Activation Mapping*) se ha consolidado como una de las herramientas más empleadas para la interpretación de redes neuronales convolucionales aplicadas a imágenes. Este método permite identificar de forma visual las regiones de la imagen que han tenido mayor influencia en la predicción del modelo. En el caso específico de la retinopatía diabética, Grad-CAM contribuye a resaltar zonas relevantes de la retina asociadas a lesiones microvasculares o alteraciones estructurales, proporcionando una validación cualitativa del razonamiento del modelo y reforzando su coherencia desde el punto de vista clínico (Das et al., 2022; Selvaraju et al., 2017).

2.9 Métricas de evaluación en modelos de diagnóstico médico

La evaluación del rendimiento de los modelos de clasificación en el ámbito médico exige el uso de métricas que permitan analizar su comportamiento desde una perspectiva integral y clínicamente significativa. A diferencia de otros contextos de aplicación, en el diagnóstico de enfermedades no es suficiente considerar únicamente la tasa global de aciertos, ya que una clasificación incorrecta puede derivar en consecuencias directas para la salud del paciente. En este sentido, métricas como la exactitud (*accuracy*), la sensibilidad (*recall*), la especificidad, la precisión y el F1-score resultan fundamentales, ya que permiten examinar distintos aspectos del desempeño del modelo, especialmente en escenarios caracterizados por desbalance entre clases (Powers, 2020; Sokolova & Lapalme, 2009).

De manera complementaria, el uso de herramientas de análisis visual, como la matriz de confusión, junto con métricas basadas en curvas, tales como la curva ROC y el área bajo la curva (AUC-ROC), ofrece una visión más detallada del comportamiento del clasificador frente a diferentes umbrales de decisión. Estas métricas son ampliamente empleadas en el análisis de imágenes médicas y en entornos clínicos, ya que facilitan la evaluación de la confiabilidad del modelo y su viabilidad como sistema de apoyo al diagnóstico, contribuyendo a reducir el riesgo

de errores críticos en la toma de decisiones médicas Altman & Bland, 1994, p. 1552; Fawcett, 2006, p. 863).

2.9.1 Exactitud (Accuracy)

La exactitud o *accuracy* representa la proporción de predicciones correctas realizadas por el modelo con respecto al total de observaciones evaluadas. Se calcula como la suma de los verdaderos positivos y verdaderos negativos dividida para el número total de predicciones. Esta métrica proporciona una visión general del desempeño del modelo y es comúnmente utilizada como primer indicador de calidad en tareas de clasificación (Stehman, 1997).

No obstante, en aplicaciones médicas la exactitud debe interpretarse con cautela, especialmente cuando existe un desbalance entre las clases, como ocurre frecuentemente en conjuntos de datos de retinopatía diabética. En estos casos, un modelo puede alcanzar una alta exactitud simplemente favoreciendo la clase mayoritaria, ocultando un desempeño deficiente en la detección de casos patológicos. Por ello, la exactitud no debe analizarse de forma aislada, sino en conjunto con métricas más específicas orientadas al contexto clínico (Sokolova & Lapalme, 2009).

2.9.2 Sensibilidad o Recall

La sensibilidad, también conocida como *recall*, mide la capacidad del modelo para identificar correctamente los casos positivos. En términos clínicos, esta métrica indica qué proporción de pacientes que presentan la enfermedad es correctamente detectada por el sistema. En el contexto de la retinopatía diabética, una alta sensibilidad es fundamental, ya que reduce el número de falsos negativos, es decir, casos enfermos que no son detectados por el modelo (Altman & Bland, 1994, p. 1552).

Un bajo valor de sensibilidad implica que el modelo puede omitir casos reales de la enfermedad, lo cual resulta especialmente crítico en programas de tamizaje, donde el objetivo principal es identificar tempranamente a los pacientes en riesgo. Por esta razón, la sensibilidad

es considerada una de las métricas más relevantes en sistemas de apoyo al diagnóstico médico (Parikh et al., 2008).

2.9.3 Especificidad

La especificidad mide la capacidad del modelo para reconocer correctamente los casos negativos, es decir, aquellas imágenes o pacientes en los que no se presenta la enfermedad. Esta métrica se expresa como la proporción de verdaderos negativos sobre el total de casos realmente negativos y es en especial importante para limitar la aparición de falsos positivos, los cuales pueden generar alertas innecesarias o derivaciones clínicas sin justificación. (Altman & Bland, 1994, p. 1552).

En el contexto clínico, resulta fundamental mantener un balance adecuado entre sensibilidad y especificidad. Mientras que una sensibilidad elevada reduce la probabilidad de pasar por alto casos con patología, una buena especificidad ayuda a disminuir evaluaciones adicionales que no son necesarias. Por esta razón, ambas métricas deben considerarse de forma conjunta para lograr una interpretación adecuada del desempeño del modelo. (Parikh et al., 2008).

2.9.4 Precisión

La precisión mide la proporción de verdaderos positivos respecto al total de predicciones positivas realizadas por el modelo. Esta métrica indica qué tan confiables son las predicciones positivas, es decir, cuántas de las imágenes clasificadas como retinopatía diabética corresponden realmente a casos verdaderos (Powers, 2020).

En problemas de clasificación multiclase, como la identificación de distintos grados de severidad de la retinopatía diabética, la precisión permite evaluar el desempeño del modelo para cada clase específica. Valores bajos de precisión pueden indicar confusión entre categorías clínicas adyacentes, lo cual es común en enfermedades con progresión gradual, como la retinopatía diabética (Goutte & Gaussier, 2005).

2.9.5 F1-score

El *F1-score* es una métrica que integra la precisión y la sensibilidad en un único indicador, mediante el cálculo de su media armónica. Su principal ventaja radica en que ofrece una evaluación más balanceada del modelo, especialmente en escenarios donde las clases no se encuentran distribuidas de manera uniforme. En este sentido, el *F1-score* penaliza aquellos modelos que presentan un buen resultado en una de las métricas, pero un desempeño deficiente en la otra (Sokolova & Lapalme, 2009).

Dentro del alcance de este proyecto, esta métrica permite analizar de forma conjunta la capacidad del modelo para identificar correctamente los distintos niveles de retinopatía diabética y la fiabilidad de las predicciones realizadas. Por ello, el *F1-score* es ampliamente empleado en sistemas de diagnóstico asistido por computador, ya que proporciona una visión más completa del rendimiento del modelo en contextos clínicos (Powers, 2020).

2.9.6 Matriz de confusión

La matriz de confusión es una herramienta clave para evaluar de manera detallada el desempeño de un modelo de clasificación. A través de esta representación es posible observar cómo se distribuyen las predicciones correctas e incorrectas, identificando los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos en cada una de las clases. Esto facilita el análisis de los errores cometidos por el modelo y permite comprender mejor los patrones de confusión entre categorías (Visa et al., 2011).

En escenarios de clasificación multiclase, como la identificación de los diferentes grados de severidad de la retinopatía diabética, la matriz de confusión resulta especialmente útil, ya que permite determinar qué clases son más propensas a ser confundidas entre sí. Esta información no solo contribuye a una mejor interpretación desde el punto de vista clínico, sino que también orienta posibles ajustes y mejoras en el modelo para trabajos futuros.

2.9.7 Curva ROC y Área Bajo la Curva (AUC-ROC)

La curva ROC (*Receiver Operating Characteristic*) representa la relación entre la sensibilidad y la tasa de falsos positivos del modelo al variar el umbral de decisión. A partir de esta curva se obtiene el área bajo la curva (AUC-ROC), que resume el comportamiento del clasificador en un único valor numérico, donde valores próximos a uno reflejan una mayor capacidad del modelo para distinguir entre clases (Fawcett, 2006).

En el ámbito médico, el AUC-ROC es ampliamente empleado debido a que permite evaluar el desempeño global del modelo sin depender de un umbral específico de clasificación. En el caso de la retinopatía diabética, esta métrica resulta especialmente útil para analizar la capacidad del sistema de diferenciar entre imágenes normales y patológicas, así como entre distintos niveles de severidad de la enfermedad. Además, facilita la comparación objetiva entre distintos modelos y enfoques utilizados tanto en entornos clínicos como experimentales (Altman & Bland, 1994).

CAPITULO 3

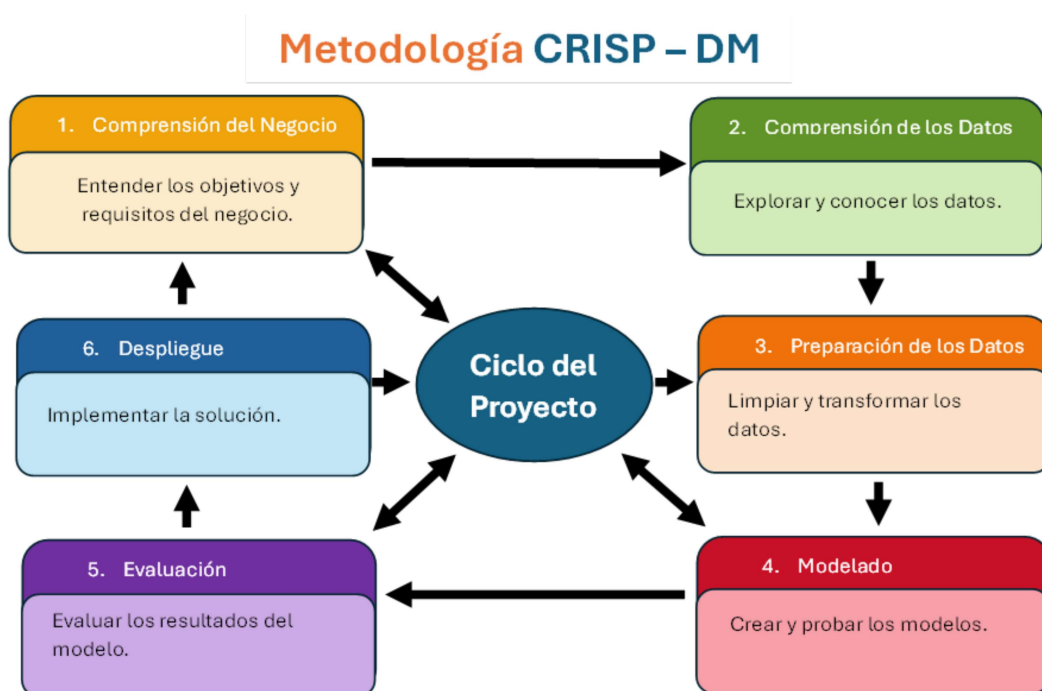
3. DESARROLLO

3.1 Metodología CRISP-DM

El desarrollo del presente trabajo se fundamenta en la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), la cual proporciona un marco estructurado y sistemático para proyectos de análisis de datos y aprendizaje automático. Esta metodología consta de seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue (Schröer et al., 2021). A continuación, se describe la aplicación de cada una de estas fases en el contexto del problema de clasificación de imágenes retinianas para la detección temprana de retinopatía diabética.

Figura 1

Diagrama esquemático de la metodología CRISP-DM



Nota. Adaptado de Wang et al. (2024)

3.1.1 Comprensión del negocio (Business Understanding)

La retinopatía diabética es una complicación ocular de la diabetes mellitus y constituye una de las principales causas de pérdida visual irreversible a nivel mundial. La detección temprana de esta enfermedad resulta clave para prevenir su progresión hacia estadios avanzados; sin embargo, los métodos tradicionales de diagnóstico dependen de la evaluación manual de imágenes de fondo de ojo por parte de especialistas, lo que limita su escalabilidad y aplicación masiva en programas de tamizaje.

Bajo este contexto, el objetivo del proyecto se define como el desarrollo de un modelo de clasificación automática, capaz de identificar la presencia y el grado de severidad de la retinopatía diabética a partir de imágenes retinianas. El sistema se concibe como una herramienta de apoyo al diagnóstico, orientada a complementar la evaluación clínica, sin sustituir el criterio médico ni constituir un sistema de diagnóstico definitivo.

3.1.2 Comprensión de los datos (Data Understanding)

Para el desarrollo del modelo se empleó un conjunto de datos público de imágenes retinianas, correspondiente al dataset Diabetic Retinopathy disponible en Mendeley Data. Este conjunto contiene imágenes de fondo de ojo capturadas mediante retinografía y etiquetadas según categorías clínicas de retinopatía diabética.

Durante esta fase se realizó una exploración inicial del dataset, analizando la distribución de las clases, la calidad visual de las imágenes y la presencia de variaciones en iluminación, enfoque y orientación. Este análisis permitió identificar desafíos inherentes al problema, como el desbalance entre clases y la variabilidad visual propia de las imágenes médicas, justificando el uso de técnicas avanzadas de procesamiento de imágenes y aprendizaje profundo.

3.1.3 Preparación de los datos (Data Preparation)

La preparación de los datos constituyó una etapa fundamental para garantizar el correcto desempeño del modelo de clasificación. Las imágenes se organizaron en una estructura de carpetas, donde cada carpeta representa una clase clínica específica, permitiendo su carga automática mediante las herramientas de procesamiento de imágenes.

Para adaptar el conjunto de datos al modelo preentrenado VGG16, todas las imágenes fueron redimensionadas a 224×224 píxeles, tamaño requerido por la arquitectura. Posteriormente, se aplicó la función `preprocess_input`, que normaliza los valores de los píxeles según el esquema utilizado durante el entrenamiento original en ImageNet.

Con el objetivo de mejorar la capacidad de generalización del modelo y reducir el riesgo de sobreajuste, se aplicaron técnicas de aumento de datos (data augmentation), incluyendo rotaciones, desplazamientos, zoom y volteos horizontales. El conjunto de datos se dividió en 80 % para entrenamiento y 20 % para validación, permitiendo evaluar el desempeño del modelo sobre datos no vistos durante el proceso de entrenamiento.

En este bloque se establece la ruta hacia el dataset en local storage y se definen los nombres de las clases correspondientes a las categorías de retinopatía diabética: Mild, Moderate, No_DR, Proliferate_DR y Severe.

También se configuran los parámetros básicos del proceso de entrenamiento, como el tamaño de las imágenes (224x224 píxeles, que es el tamaño requerido por VGG16), el tamaño del batch y el porcentaje de validación.

Figura 2

Parámetros de entrenamiento

```
class_names = ['Mild', 'Moderate', 'No_DR', 'Proliferate_DR', 'Severe']  
img_size = (224, 224)  
batch_size = 32  
val_split = 0.2  
seed = 1337
```

Nota. La imagen muestra la redimensión de las imágenes (224x224)

3.1.4 Modelado (Modeling)

Para abordar el problema de clasificación de imágenes retinianas, se seleccionó la arquitectura VGG16, una red neuronal convolucional desarrollada por el Visual Geometry Group de la Universidad de Oxford y preentrenada sobre el conjunto de datos ImageNet. Este modelo fue elegido por su estabilidad, simplicidad estructural y amplio uso en tareas de visión por computador y análisis de imágenes médicas.

En una primera etapa se aplicó Transfer Learning, congelando todas las capas convolucionales del modelo base y añadiendo una nueva cabeza de clasificación compuesta por una capa de Global Average Pooling, una capa Dropout con tasa de 0.3 y una capa densa con activación softmax, encargada de generar las probabilidades correspondientes a las cinco clases clínicas de retinopatía diabética.

Posteriormente, se implementó un proceso de Fine-Tuning, descongelando únicamente las capas del último bloque convolucional (block5_) y ajustando los pesos con una tasa de aprendizaje reducida. Se evaluaron diferentes configuraciones de entrenamiento, determinándose que 3 épocas con una tasa de aprendizaje de $1e-4$ ofrecían el mejor equilibrio entre mejora del rendimiento y control del sobreentrenamiento.

3.1.5 Evaluación (Evaluación)

La evaluación del modelo se realizó utilizando el conjunto de validación y métricas estándar de clasificación en contextos médicos. Se analizaron la exactitud (accuracy), la precisión macro, el recall macro y el F1-score macro, además de la matriz de confusión, que permitió examinar el comportamiento del modelo por cada categoría clínica.

Asimismo, se compararon los resultados obtenidos con los de un modelo entrenado desde cero. Los resultados evidenciaron que el modelo basado en Transfer Learning y Fine-Tuning alcanzó un mejor desempeño global y mayor estabilidad, al aprovechar el conocimiento previo aprendido en ImageNet y requerir menos datos para converger.

Adicionalmente, se incorporó la técnica Grad-CAM como herramienta de explicabilidad, permitiendo analizar las regiones de las imágenes que influyen en las decisiones del modelo. Este análisis contribuyó a validar que el clasificador se enfoca en estructuras clínicamente relevantes y a identificar las principales limitaciones del modelo.

- **Técnica Transfer Learning**

Figura 3

Métricas de evaluación y reporte de clasificación utilizando transfer learning

MÉTRICAS DE EVALUACIÓN	REPORTE DE CLASIFICACIÓN				
✓ ACCURACY: 0.7839 (78.39%)		precision	recall	f1-score	support
✓ F1-SCORE (Macro): 0.5310	Mild	0.7308	0.2568	0.3800	74
✓ F1-SCORE (Weighted): 0.7555	Moderate	0.6140	0.8794	0.7231	199
F1-Score por clase:	No_DR	0.9517	0.9834	0.9673	361
- Mild: 0.3800	Proliferate_DR	0.5556	0.3390	0.4211	59
- Moderate: 0.7231	Severe	0.3636	0.1053	0.1633	38
- No_DR: 0.9673					
- Proliferate_DR: 0.4211	accuracy			0.7839	731
- Severe: 0.1633	macro avg	0.6431	0.5128	0.5310	731
✓ MSE (Mean Squared Error): 0.061264	weighted avg	0.7749	0.7839	0.7555	731
✓ RMSE (Root MSE): 0.247515					

Nota. La imagen demuestra la evaluación del preprocesamiento VGG16

Dentro de la Figura 3. Se observan los resultados obtenidos muestran que el modelo alcanza una precisión global

accuracy) del 78.39%, lo que indica que, en términos generales, la mayoría de las muestras fueron clasificadas correctamente. No obstante, este valor debe interpretarse con cautela, ya que el conjunto de datos presenta un claro desbalance entre las clases. En este contexto, métricas como el F1-score macro y ponderado ofrecen una visión más completa del desempeño del modelo.

El F1-score macro obtenido fue de 0.5310, lo que evidencia un rendimiento desigual entre las diferentes clases. Este resultado sugiere que el modelo no logra un equilibrio adecuado al momento de clasificar todas las categorías, afectando principalmente a aquellas con menor número de muestras. Por otro lado, el F1-score ponderado alcanzó un valor de 0.7555, reflejando un mejor desempeño en las clases mayoritarias, lo que confirma que el

modelo tiende a favorecer aquellas categorías con mayor representación en el conjunto de datos.

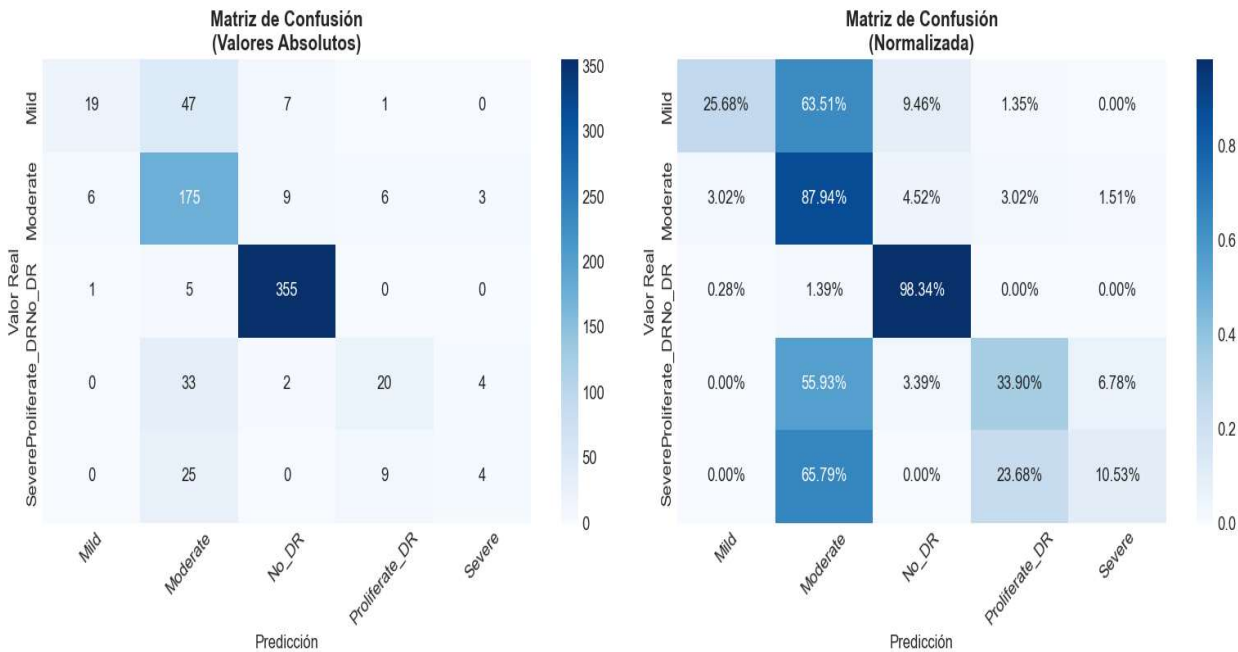
Al analizar el desempeño por clase, se observa que la categoría *No_DR* presenta resultados sobresalientes, con un F1-score de 0.9673, lo que indica una alta capacidad del modelo para identificar correctamente los casos sin presencia de la enfermedad. De manera similar, la clase *Moderate* muestra un rendimiento aceptable, con un F1-score de 0.7231, evidenciando un buen equilibrio entre precisión y recall.

En contraste, las clases *Mild*, *Proliferate_DR* y *Severe* presentan un desempeño considerablemente inferior. En el caso de *Mild*, el bajo valor de recall (0.2568) indica que una proporción significativa de los casos leves no fue correctamente identificada. Asimismo, la clase *Proliferate_DR* obtiene un F1-score de 0.4211, lo que sugiere dificultades del modelo para diferenciar esta condición de otras etapas de la enfermedad. La situación más crítica se observa en la clase *Severe*, donde el F1-score alcanza apenas 0.1633, acompañado de un recall extremadamente bajo, lo que implica que la mayoría de los casos severos no están siendo detectados adecuadamente.

En cuanto a las métricas de error, el modelo presenta un MSE de 0.0613 y un RMSE de 0.2475, valores que indican un error promedio relativamente bajo. Sin embargo, estas métricas no reflejan completamente las limitaciones del modelo en términos de clasificación de las clases minoritarias, por lo que su interpretación debe complementarse con métricas específicas de clasificación.

Figura 4

Matriz de confusión usando transfer learning

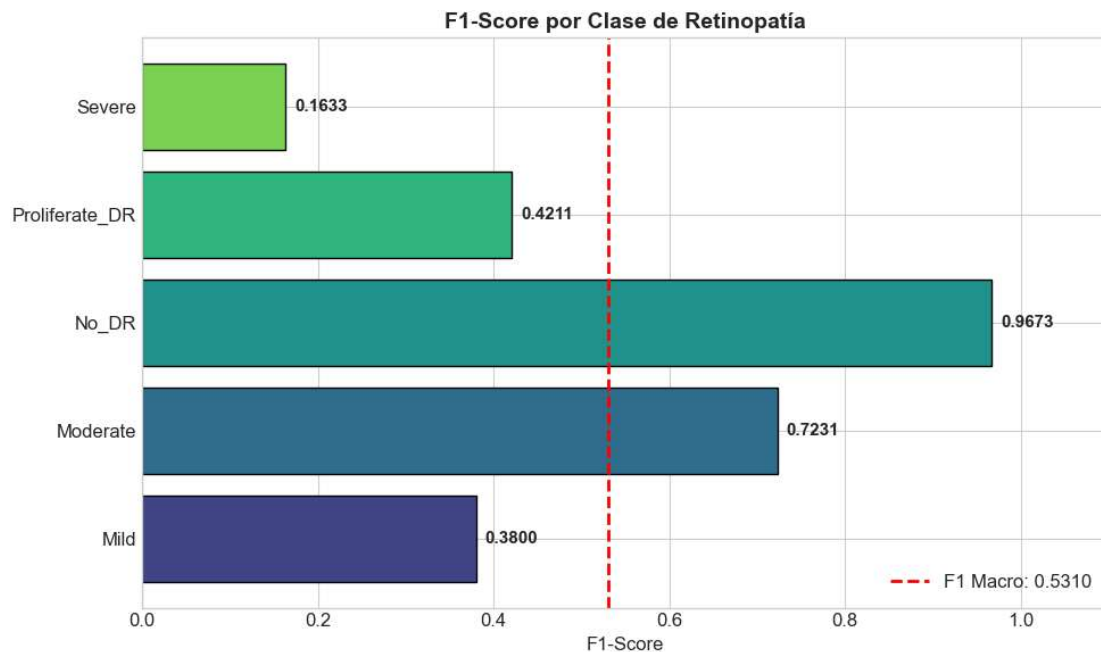


Nota. La imagen evidencia la comparativa de etiquetas reales vs predicciones del modelo

En la Figura 4. se observa la matriz de confusión evidencia que el modelo presenta un desempeño sólido en la clase *No_DR*, con un alto porcentaje de clasificaciones correctas y mínima confusión con otras categorías. Asimismo, la clase *Moderate* muestra una tasa de acierto elevada, aunque aún se observa confusión con *Mild*. En contraste, las clases *Proliferate_DR* y *Severe* presentan mayores errores de clasificación, siendo frecuentemente confundidas con *Moderate*. Esto confirma la tendencia del modelo a favorecer clases intermedias y mayoritarias, y pone de manifiesto la dificultad para identificar correctamente los estadios más avanzados de la enfermedad.

Figura 5

F1-Scores por Clase de Retinopatía

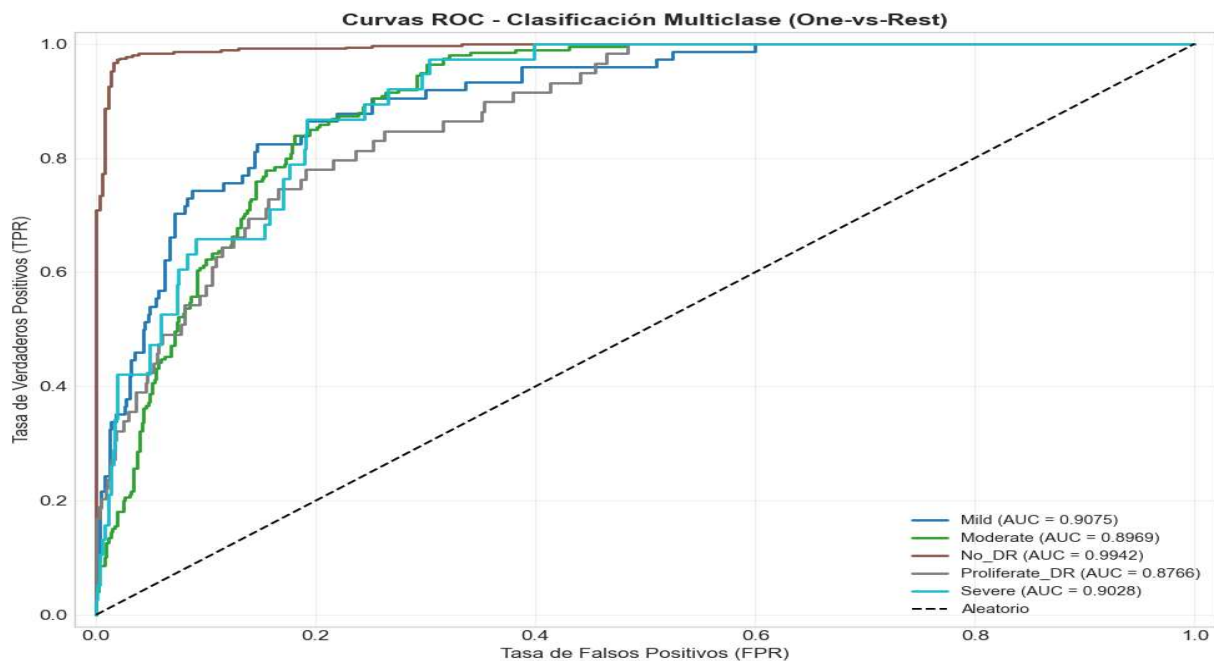


Nota. La imagen muestra el desempeño por clase de Retinopatía

El gráfico de F1-score por clase muestra un desempeño muy alto en *No_DR* y aceptable en *Moderate*, superando el promedio macro. En contraste, las clases *Mild*, *Proliferate_DR* y especialmente *Severe* presentan valores bajos, lo que evidencia dificultades del modelo para identificar correctamente los estadios menos representados y más críticos de la enfermedad.

Figura 6

Curvas Roc (One-vs-Rest)

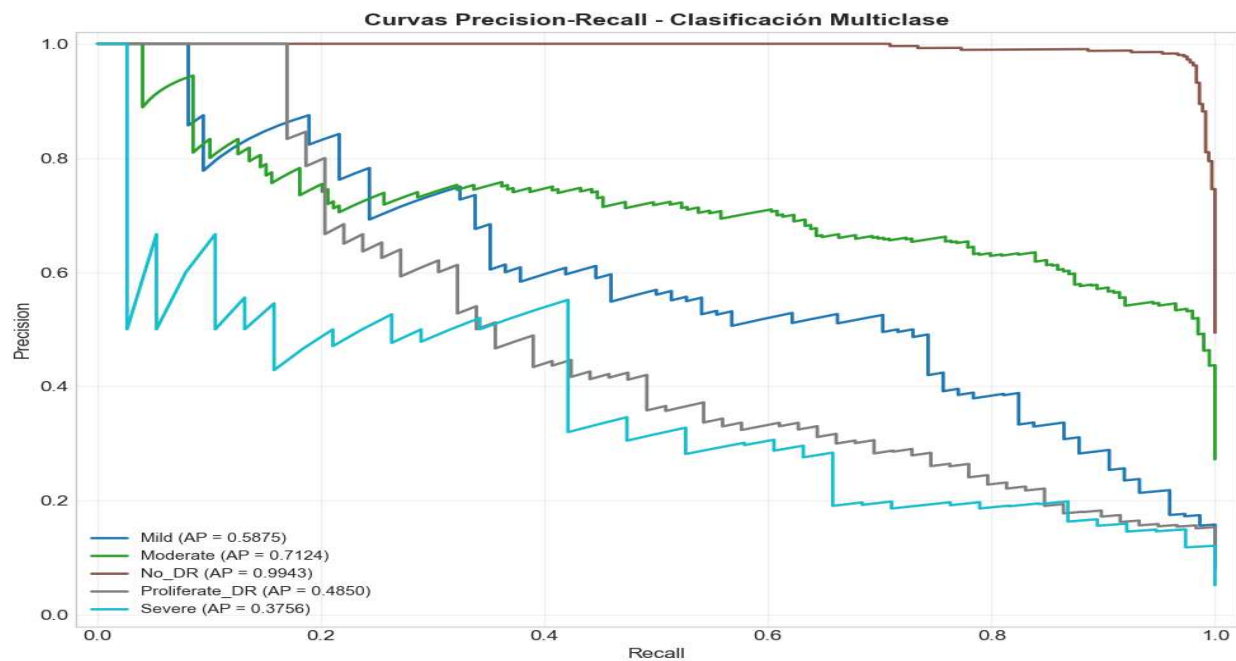


Nota. La imagen muestra la curva ROC multiclase usando One-vs-Rest

En la Figura 6. se observa el modelo de clasificación multiclase de retinopatía diabética presenta un alto poder discriminativo para todas las categorías clínicas, con valores AUC superiores a 0.87. Destaca especialmente la clase Sin Retinopatía (AUC \approx 0.99), mientras que las clases intermedias y avanzadas mantienen un desempeño robusto, consistente con la complejidad clínica del problema.

Figura 7

Curvas Precision, Recall and Clasificación Multiclase

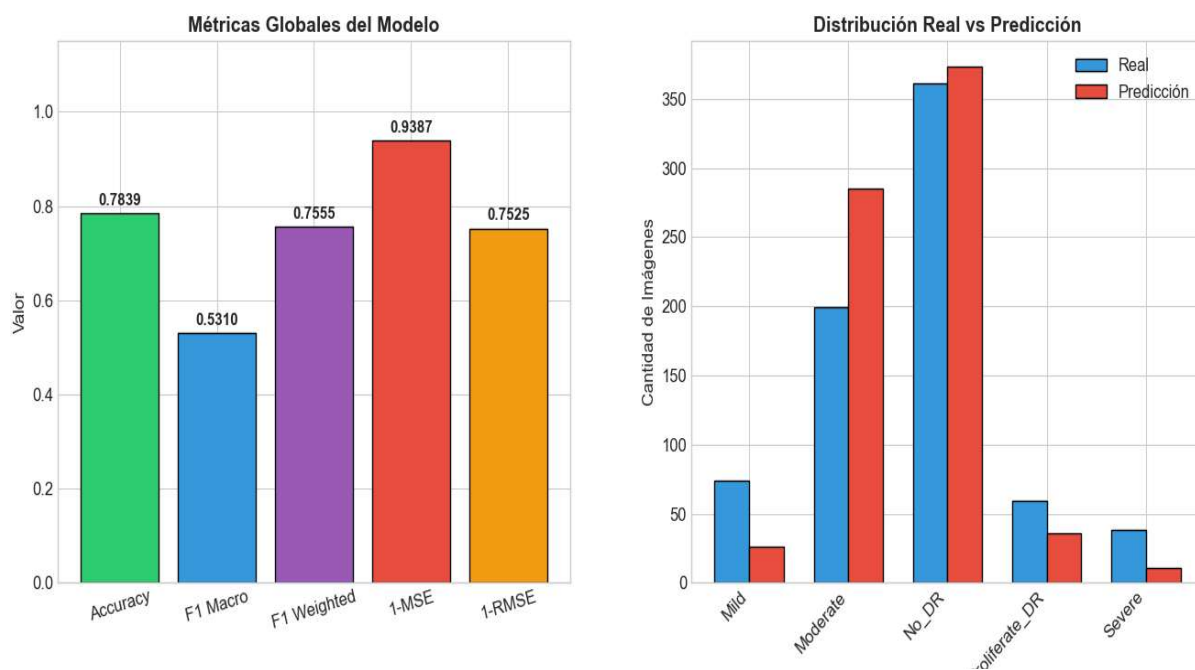


Nota. La imagen muestra el desempeño de la curva ROC

En la Figura 7. las curvas ROC muestran una adecuada capacidad discriminante del modelo, con áreas bajo la curva elevadas en la mayoría de las clases, lo que las clases buen balance entre sensibilidad y especificidad. Sin embargo, las curvas Precision–Recall evidencian variabilidad entre clases, sugiriendo que el desempeño disminuye en escenarios con desbalance, especialmente para categorías minoritarias.

Figura 8

Resumen de Métricas globales del modelo



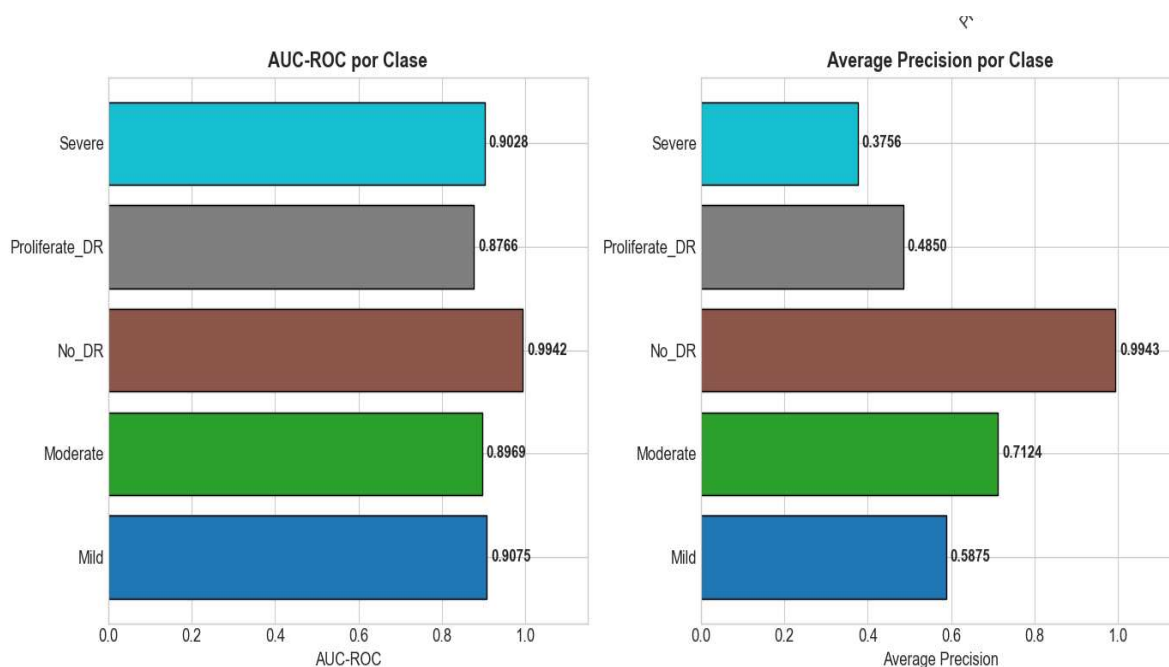
Nota. La imagen muestra el desempeño de la curva ROC

De acuerdo con la Figura 8. el modelo presenta un buen desempeño global, con alta estabilidad y buen ajuste a las clases mayoritarias, aunque muestra dificultades esperables en clases clínicas minoritarias, reflejado en un F1 Macro más bajo.

El modelo presenta una tendencia conservadora, redistribuyendo casos hacia clases intermedias (Moderate), lo cual reduce falsos negativos severos, pero penaliza la detección precisa de clases avanzadas poco representadas.

Figura 9

AUC-ROC por Clase y Average Precision por Clase

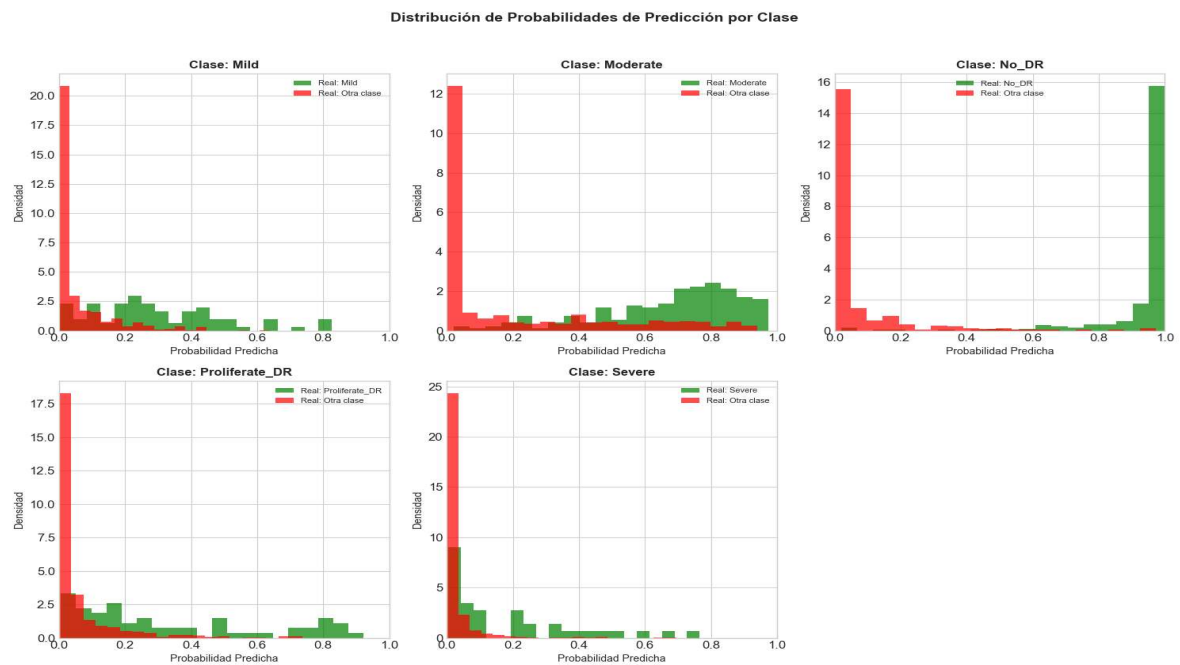


Nota. La imagen presenta el desempeño AUC-ROC por Clase

El modelo presenta un desempeño global sólido, con una accuracy cercana al 78% y valores consistentes de F1 macro y ponderado, lo que refleja un equilibrio razonable entre clases. El AUC-ROC es elevado en todas las categorías, destacando *No DR*, lo que evidencia alta capacidad discriminante. No obstante, la precisión promedio disminuye en clases más severas, indicando mayor dificultad en identificar correctamente casos menos frecuentes.

Figura 10

Distribución de Probabilidades de Predicción por Clase



Nota. La imagen presenta la distribución de probabilidades predichas por clase (One-vs-Rest)

Figura 11

Reporte de clasificación Fine-Tuning

Accuracy: 0.7907 | Precision(macro): 0.7179 | Recall(macro): 0.5530 | F1(macro): 0.5842

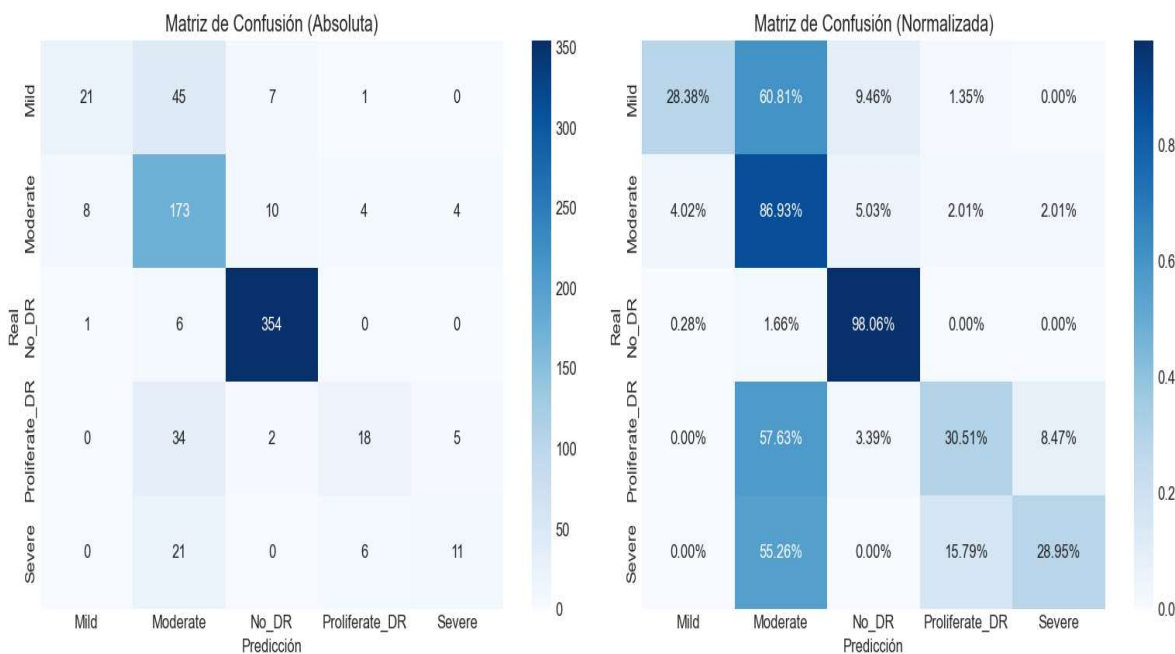
Reporte de clasificación:

	precision	recall	f1-score	support
Mild	0.74	0.27	0.40	74
Moderate	0.61	0.87	0.72	199
No_DR	0.95	0.98	0.96	361
Proliferate_DR	0.73	0.27	0.40	59
Severe	0.56	0.37	0.44	38
accuracy			0.79	731
macro avg	0.72	0.55	0.58	731
weighted avg	0.80	0.79	0.77	731

Nota. La imagen presenta el reporte de clasificación de Fite-Tuninig

Tras el fine-tuning con descongelamiento parcial del backbone, el modelo alcanza una accuracy de 0.79, evidenciando un desempeño global aceptable. La clase *No DR* presenta resultados sobresalientes, con alta precisión y recall, lo que indica una correcta identificación de casos negativos. En contraste, las clases *Mild*, *Proliferate_DR* y *Severe* muestran valores bajos de recall, reflejando dificultades para detectar correctamente casos menos representados. El F1 macro de 0.58 confirma un impacto del desbalance de clases en el rendimiento general. Estos resultados sugieren la necesidad de estrategias adicionales, como reponderación de clases o aumento de datos, para mejorar la sensibilidad en categorías clínicas más críticas.

Figura 12
Reporte de clasificación Fine-Tuning

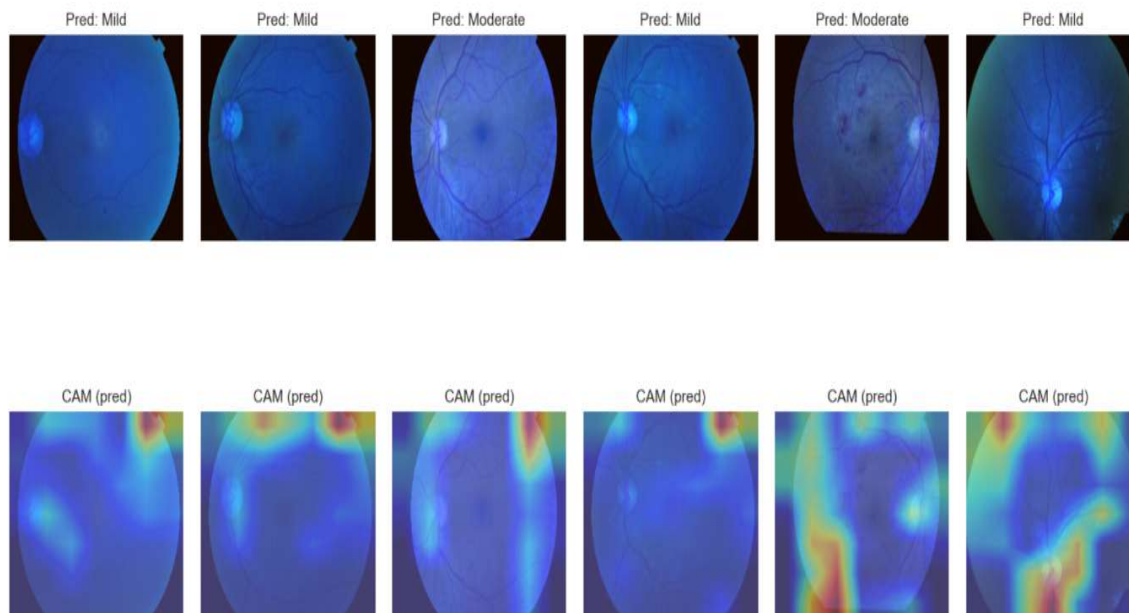


Nota. La imagen presenta la matriz de confusión Absoluta y Normalizada

En la Figura 12. Se muestra la matriz de confusión que confirma un desempeño robusto en la clase *No DR*, con alta tasa de clasificación correcta y mínimos errores. Sin embargo, se observa confusión relevante entre *Mild* y *Moderate*, así como en *Proliferate_DR* y *Severe*, donde predominan errores hacia clases intermedias. Esto evidencia el efecto del desbalance y la similitud clínica entre categorías adyacentes.

Figura 13

Explicabilidad con Grad-CAM



Nota. La imagen explica por qué una red neuronal convolucional tomó una decisión

3.1.6 Despliegue (Deployment)

En la fase de despliegue, se implementó un prototipo funcional del clasificador mediante el desarrollo de una API orientada a exponer el modelo entrenado como un servicio consumible. El objetivo de este despliegue es facilitar el acceso a la inferencia del modelo desde aplicaciones externas (por ejemplo, un cliente web o una interfaz de pruebas), permitiendo enviar una imagen de fondo de ojo y recibir como respuesta la predicción de clase junto con la probabilidad asociada, de acuerdo con el esquema de clasificación definido.

Para garantizar trazabilidad y buenas prácticas de ingeniería, el proyecto fue versionado y documentado en un repositorio de GitHub, donde se mantiene el código fuente de la API, el proceso de carga del modelo y las instrucciones de ejecución. Este enfoque permite asegurar reproducibilidad, control de versiones, colaboración y mantenimiento del proyecto, además de

habilitar futuras mejoras como pruebas automatizadas, integración continua y empaquetamiento del servicio.

El flujo general de despliegue considera: (1) la carga del modelo entrenado en el servidor, (2) la recepción de imágenes a través de un endpoint, (3) la aplicación del mismo preprocesamiento utilizado durante el entrenamiento (redimensionamiento, normalización y transformación requerida por VGG16), y (4) la generación de la salida del modelo en formato estructurado (por ejemplo, JSON) con la clase predicha y sus probabilidades. De forma complementaria, el despliegue puede incorporar la generación de salidas interpretables mediante Grad-CAM para auditoría visual del comportamiento del clasificador, como apoyo al análisis de los resultados.

Es importante destacar que este despliegue se plantea con fines académicos y demostrativos, y no constituye una herramienta clínica certificada. Por tanto, no se contempla su uso directo en entornos hospitalarios ni decisiones médicas reales, sino como un prototipo que evidencia la viabilidad técnica de integrar modelos de clasificación basados en aprendizaje profundo dentro de flujos de trabajo de salud digital y telemedicina.

Figura 14

Arquitectura del proyecto



La Figura 14. representa la arquitectura del sistema implementado para la inferencia de retinopatía diabética basada en aprendizaje profundo. El sistema sigue un esquema cliente–servidor compuesto por tres componentes principales.

En primer lugar, el cliente, desarrollado con Streamlit y ejecutado en el puerto 8501, actúa como la interfaz gráfica de usuario. Este componente permite la carga de imágenes de fondo de ojo y el envío de solicitudes al sistema mediante el protocolo HTTP.

Las solicitudes son recibidas por una API REST implementada con FastAPI, que se ejecuta en el puerto 8000. Este componente cumple el rol central del sistema, ya que se encarga de cargar el modelo entrenado y realizar el proceso de inferencia sobre las imágenes recibidas. La comunicación entre el cliente y la API se realiza de forma síncrona, permitiendo obtener respuestas en tiempo real.

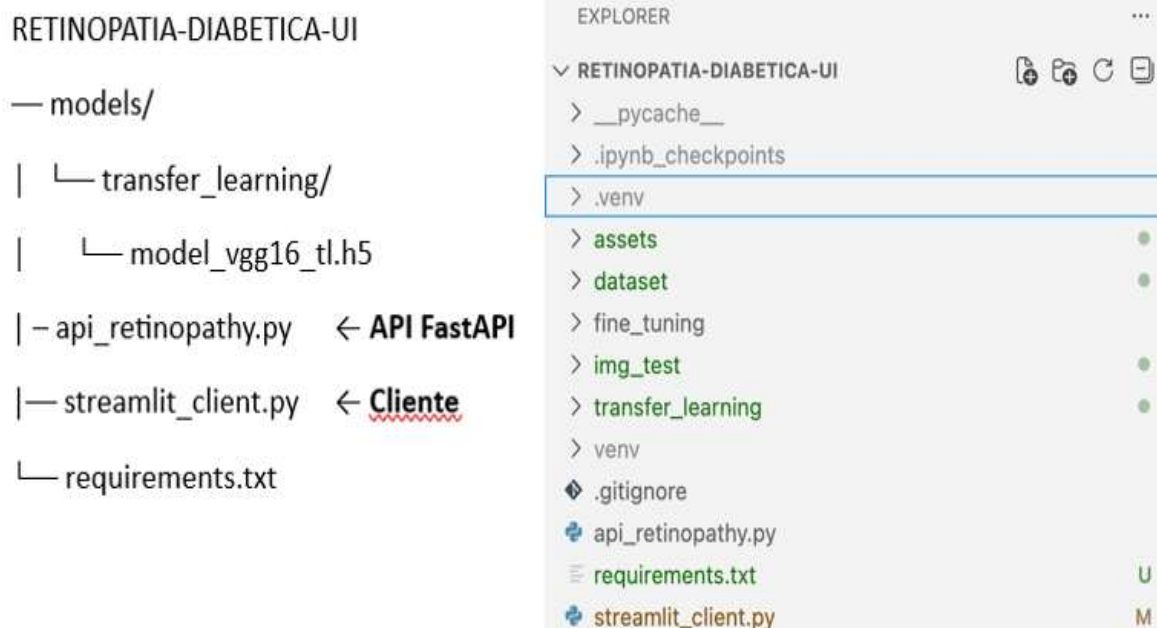
El tercer componente corresponde al modelo de aprendizaje profundo, almacenado en un archivo .h5, el cual ha sido previamente entrenado para la clasificación de imágenes retinianas según el grado de severidad de la retinopatía diabética. La API interactúa directamente con este modelo para generar las predicciones, las cuales son devueltas al cliente para su visualización.

Cabe destacar que Streamlit y FastAPI se ejecutan de manera simultánea como servicios independientes, mientras que Visual Studio Code actúa únicamente como entorno de desarrollo y orquestación durante la implementación del sistema, sin intervenir en el flujo de inferencia. Esta arquitectura permite una separación clara entre la interfaz de usuario, la lógica de negocio y el modelo de inferencia, facilitando la escalabilidad y mantenibilidad del sistema.

- **Arquitectura del Backend y Documentación de la API**

Figura 15

Arquitectura del Front- End y Back-End



Nota. La imagen muestra la arquitectura del proyecto

El motor de inferencia del proyecto se implementa a través de una API REST construida con el framework FastAPI, lo que permite una comunicación eficiente y asíncrona entre el modelo de Inteligencia Artificial y la interfaz de usuario. Como se observa en la documentación interactiva (generada bajo el estándar OpenAPI 3.1), el servidor expone diversos puntos de enlace o endpoints diseñados para la gestión del servicio. Entre ellos destaca el método POST /predict, el cual constituye el núcleo operativo del sistema al recibir las imágenes retinales y retornar la clasificación predicha por el modelo. Asimismo, se incluyen rutas de control como /health, destinadas a monitorear el estado operativo del servidor en tiempo real.

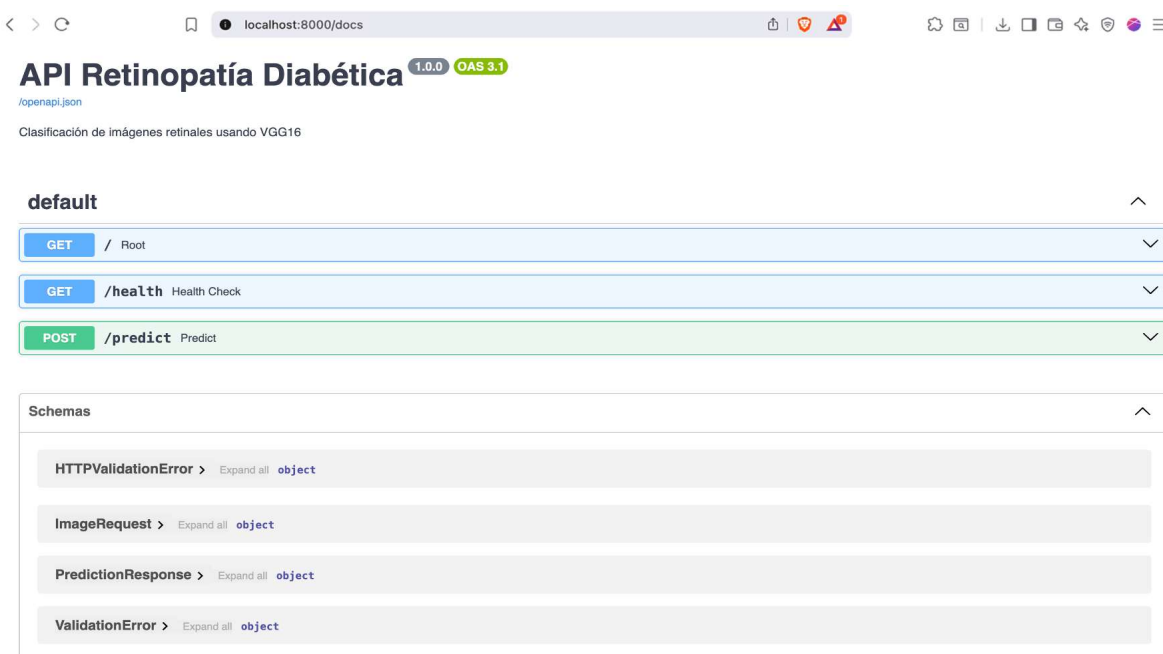
Para garantizar la integridad de los datos, la API define esquemas estructurados de entrada y salida, tales como ImageRequest y PredictionResponse. Estos componentes aseguran

que la información enviada por el cliente sea validada antes de ser procesada por la red neuronal, minimizando errores de ejecución. La arquitectura se complementa con mecanismos de manejo de excepciones que capturan fallos de validación, proporcionando una base robusta y escalable que cumple con los estándares modernos de desarrollo de software para aplicaciones de salud.

- **Estructura de Directorios y Organización del Proyecto**

La organización del código fuente sigue un patrón modular que separa claramente las responsabilidades del sistema, facilitando su mantenimiento y escalabilidad. En la raíz del directorio principal, denominado RETINOPATIA-DIABETICA-UI, se encuentran los archivos fundamentales para el despliegue: el script `api_retinopathy.py`, que contiene la lógica del servidor FastAPI, y `streamlit_client.py`, que gestiona la interfaz gráfica del usuario. Esta separación entre el cliente y el servidor permite que el sistema opere de manera distribuida, donde la interfaz actúa como un consumidor de los servicios de predicción alojados en el backend.

El proyecto incluye carpetas especializadas para cada fase del flujo de trabajo de ciencia de datos. El directorio `models/` almacena los pesos del modelo entrenado (`model_vgg16_tl.h5`), mientras que carpetas como `transfer_learning` y `fine_tuning` contienen los scripts y recursos utilizados durante la etapa de entrenamiento y optimización de la red neuronal VGG16. Finalmente, el archivo `requirements.txt` centraliza todas las dependencias necesarias, garantizando la reproducibilidad del entorno de ejecución, mientras que otros directorios como `dataset` e `img_test` aseguran una gestión organizada de los activos de datos utilizados para las pruebas de validación.

Figura 16**Application Programming Interface (API)**

Nota. Documentación visual e interactiva de la API

Esta imagen muestra la documentación interactiva de la Interfaz de Programación de Aplicaciones (API) diseñada para el proyecto de detección de retinopatía diabética. Basado en la captura, aquí tienes una descripción técnica organizada en párrafos para tu documento de titulación:

La arquitectura del sistema emplea una API REST desarrollada, según la interfaz, bajo el estándar OpenAPI 3.1 (comúnmente asociada con frameworks como FastAPI). Esta capa de backend es la encargada de gestionar la lógica de negocio y servir el modelo de Inteligencia Artificial para la clasificación de imágenes retinales utilizando la arquitectura VGG16. La documentación se presenta de forma estructurada, permitiendo la verificación y prueba de los diferentes puntos finales o endpoints que facilitan la comunicación entre la interfaz de usuario y el motor de análisis.

El núcleo de la API se compone de tres endpoints principales definidos en la sección "default":

- GET /Root: Utilizado generalmente para verificar la conexión inicial o mostrar un mensaje de bienvenida al servicio.
- GET /health: Un punto de control de salud (Health Check) que permite monitorear si el servicio y el modelo están operativos y listos para recibir peticiones.
- POST /predict: El componente crítico de la aplicación, diseñado para recibir los datos de la imagen y retornar la predicción del nivel de retinopatía calculada por el modelo de IA.

Adicionalmente, la imagen detalla los Esquemas (Schemas) de datos que rigen el intercambio de información, asegurando la integridad de las peticiones. Entre estos destacan ImageRequest, que define el formato de entrada de la imagen, y PredictionResponse, que estructura la respuesta del análisis. El sistema también incluye esquemas de validación de errores (HTTPValidationError y ValidationError), lo que garantiza que cualquier entrada incorrecta sea gestionada adecuadamente por el servidor, proporcionando una robustez técnica necesaria para aplicaciones de salud digital.

Figura 17

Instrucciones de uso del Sistema de Detección de Retinopatía diabética



UIDE
Powered by
Arizona State University

Menú
Instrucciones
Subir imágenes

API Conectada

API:
<http://localhost:8000>

Estudiantes:

- Bonilla Zarate, María Camila
- Carranza Villarroel, Carlos Daniel
- Lincango Simbaña, Betsy Belén
- Saguay Sagui, Bryan Alexander

Sistema de Detección de Retinopatía Diabética

Esta aplicación utiliza **Inteligencia Artificial** para clasificar imágenes del fondo de ojo y detectar signos de Retinopatía Diabética.

Clases de clasificación

Icono	Clase	Descripción
	No_DR	Sin retinopatía
	Mild	Leve
	Moderate	Moderada

Icono	Clase	Descripción
	Severe	Severa
	Proliferate_DR	Proliferativa

Instrucciones de uso

1. Selecciona "Subir imágenes" en el menú lateral
2. Arrastra o selecciona una imagen del fondo de ojo
3. Haz clic en "Clasificar imagen"
4. Espera el resultado del análisis

Modelo utilizado

- Arquitectura: VGG16 con Transfer Learning
- Entrada: Imágenes 224x224 RGB
- Salida: 5 clases de clasificación

Nota. Visualización del módulo de Instrucciones UI

El sistema presentado consiste en una plataforma web integral desarrollada bajo el respaldo académico de la Universidad Internacional del Ecuador (UIDE). La interfaz está diseñada con un enfoque de usabilidad clínica, facilitando la interacción entre el usuario y un modelo de Inteligencia Artificial especializado en el análisis de imágenes de fondo de ojo. El objetivo principal de la herramienta es la detección temprana y clasificación automática de la Retinopatía Diabética, permitiendo optimizar los tiempos de diagnóstico médico mediante el uso de algoritmos de visión computacional.

En el núcleo técnico de la aplicación, el procesamiento de las imágenes se fundamenta en una arquitectura de red neuronal convolucional VGG16, implementada mediante la técnica de Transfer Learning. Esta configuración permite que el sistema reciba imágenes en formato RGB con una resolución de 224×224 píxeles para extraer características patológicas

clave. Como resultado, el modelo es capaz de categorizar el estado del paciente en cinco niveles de severidad: desde la ausencia de la enfermedad (No_DR) hasta etapas críticas como la retinopatía proliferativa, utilizando una escala visual de colores para una interpretación rápida de los resultados.

La arquitectura de la solución se divide en un frontend intuitivo y un backend robusto. En la barra lateral, se observa un panel de control que incluye un indicador de estado de la API, el cual confirma la conectividad en tiempo real con el servidor local (puerto 8000) encargado de ejecutar las inferencias del modelo. Además, el flujo de trabajo para el usuario está claramente delimitado en cuatro pasos: la navegación hacia el módulo de carga, la selección de la imagen retiniana, la ejecución del proceso de clasificación y la recepción final del análisis, garantizando así una experiencia de usuario eficiente y simplificada para el personal de salud o investigadores.

Finalmente, la interfaz cumple con una función informativa al detallar los parámetros técnicos del modelo y los créditos de autoría de los estudiantes desarrolladores. Este diseño no solo busca la precisión técnica en el diagnóstico, sino también la transparencia tecnológica al mostrar las especificaciones de entrada y salida del modelo de IA, consolidándose como una herramienta de apoyo a la decisión médica dentro del marco de la transformación digital de la salud.

Figura 18

Sistema de Detección de Retinopatía módulo de clasificación



Nota. Visualización del módulo de clasificación UI

La Figura 18. muestra la interfaz principal del sistema web desarrollado para la detección automática de retinopatía diabética, implementado como una prueba de concepto del modelo de clasificación basado en aprendizaje profundo.

En la parte superior se visualiza el logotipo institucional de la UIDE – Universidad Internacional del Ecuador, lo que identifica el carácter académico del proyecto. A continuación, se presenta el título “Detección de Retinopatía Diabética”, acompañado de un ícono representativo del análisis visual, reforzando el enfoque del sistema en el procesamiento de imágenes médicas.

En el área central de la interfaz se encuentra el módulo de carga de imágenes, el cual permite al usuario subir una imagen de fondo de ojo mediante la opción drag and drop o

seleccionando un archivo desde el dispositivo local. Se especifican claramente las restricciones del sistema, indicando los formatos admitidos (PNG, JPG, JPEG) y el tamaño máximo del archivo, garantizando así la compatibilidad con el modelo entrenado.

En el panel lateral izquierdo se dispone un menú de navegación que incluye las secciones Menú, Instrucciones y Subir imágenes, facilitando la interacción del usuario con el sistema. Asimismo, se muestra el estado de conexión con la API, indicando que el servicio se encuentra activo y correctamente enlazado mediante una dirección local (<http://localhost:8000>), lo que evidencia la integración entre la interfaz gráfica y el modelo de clasificación desplegado.

Adicionalmente, en la parte inferior del panel lateral se presenta un listado de estudiantes participantes, lo que refuerza el carácter colaborativo y académico del desarrollo del sistema.

En conjunto, esta imagen ilustra la fase inicial del flujo de funcionamiento del sistema, correspondiente a la entrada de datos, donde el usuario proporciona la imagen retiniana que será posteriormente procesada por el modelo de aprendizaje profundo para la clasificación del grado de retinopatía diabética. Esta interfaz constituye un componente fundamental de la prueba de concepto, ya que demuestra la viabilidad de integrar el modelo entrenado en una aplicación interactiva de apoyo al diagnóstico.

Figura 19

Sistema de Detección de Retinopatía: Módulo subir imagen



Nota. Visualización del módulo de clasificación de una imagen

La Figura 19. se presenta la interfaz gráfica desarrollada para la evaluación del modelo de clasificación de retinopatía diabética, la cual permite cargar una imagen de fondo de ojo y visualizar de manera inmediata los resultados de la predicción generada por el modelo basado en la arquitectura VGG16 mediante técnicas de Transfer Learning y Fine-Tuning.

En el panel izquierdo de la interfaz se observa el módulo de carga de imágenes, donde el usuario puede seleccionar una retinografía en formato digital (PNG, JPG o JPEG). Una vez cargada la imagen, esta se visualiza en la sección central bajo el rótulo Imagen cargada, permitiendo verificar la calidad y el contenido de la retinografía antes de proceder con su clasificación. Este paso resulta fundamental para asegurar que la imagen corresponda a un fondo de ojo válido y cumpla con las condiciones mínimas para el análisis automático.

CAPITULO 4

3 ANÁLISIS DE RESULTADOS

4.1 Pruebas de Concepto

Como prueba de concepto, se implementó un modelo de clasificación multiclase de imágenes retinianas utilizando la arquitectura VGG16 bajo un enfoque de Transfer Learning y Fine-Tuning, con el objetivo de evaluar su viabilidad para la detección temprana de la retinopatía diabética.

El modelo fue entrenado a partir de un conjunto de imágenes de fondo de ojo pertenecientes a cinco categorías clínicas, representando distintos niveles de severidad de la enfermedad. Durante el proceso experimental se aplicaron técnicas de preprocesamiento y aumento de datos, con el fin de reducir el sobreajuste y mejorar la capacidad de generalización del modelo.

Los resultados obtenidos durante la fase de entrenamiento evidencian una convergencia progresiva del modelo, observándose una disminución consistente de la función de pérdida y un incremento en la exactitud tanto en el conjunto de entrenamiento como en el de validación. Este comportamiento indica que el modelo logró aprender patrones relevantes presentes en las imágenes retinianas sin incurrir en un sobreajuste significativo.

4.2 Análisis de Resultados

Lo que permitió evaluar de manera detallada el comportamiento del modelo de clasificación por clase, se lo realizó mediante métricas estándar de clasificación multiclase, incluyendo exactitud, precisión, recall y F1 -score, complementadas con el análisis de la matriz de confusión.

La arquitectura VGG16 presenta capacidad para extraer características discriminativas que son relevantes en imágenes médicas, esto se puede observar a través de los resultados

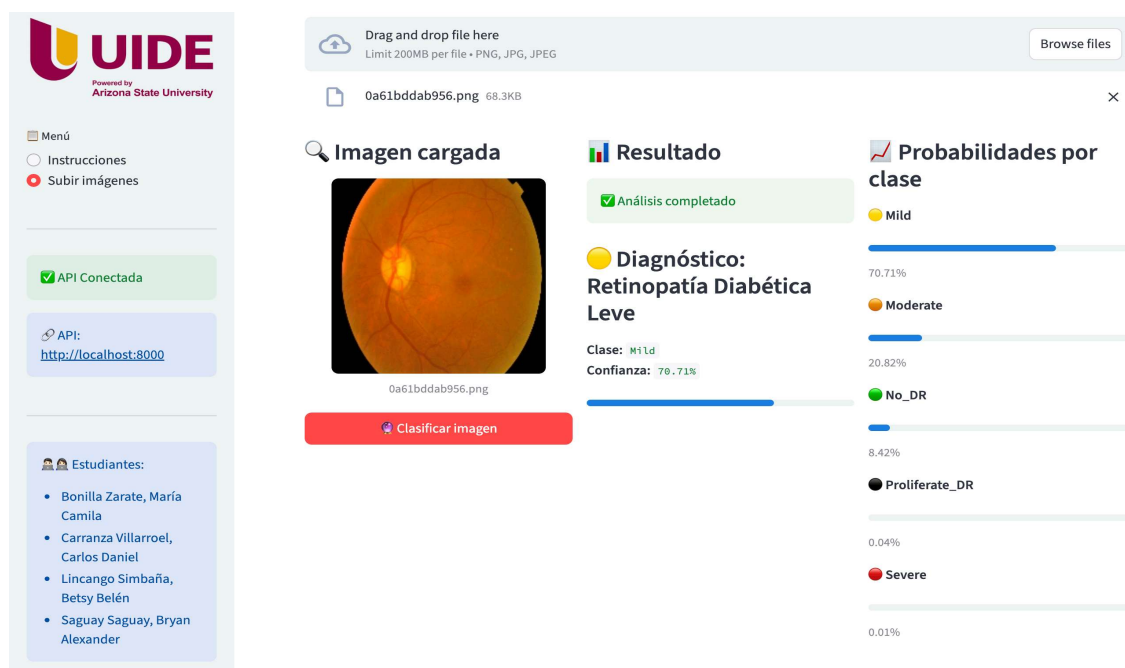
globales ya que muestran que el modelo alcanza un valor de exactitud satisfactorio aun cuando este fue entrenado originalmente sobre un dominio distinto.

A través del análisis por clase nos permitió observar un mayor desempeño en las categorías que presentan una mayor representación en el conjunto de datos, a su vez que las clases sobre estudios más avanzados de la enfermedad presentan mayor dificultad en la clasificación. Esta conducta puede deberse al desbalance del grupo de datos y de igual manera a la semejanza presente entre los niveles de severidad de la retinopatía diabética.

Con la ayuda de la matriz de confusión se pudo identificar confusiones que se presentan de manera recurrente entre clases clínicas cercanas, y esto resulta lógico desde la observación médica, dado que los cambios anómalos en la retina se van presentando de manera progresiva.

Además, a través de la integración de la técnica de aplicabilidad Grad-CAM se permite observar los sectores de las imágenes que influyeron en las decisiones del modelo. Las estimulaciones recibidas se concentraron principalmente en zonas relevantes de la retina, como son en este caso vasos sanguíneos y sectores con lesiones visibles, lo que reafirma la interpretabilidad del modelo y esto brinda confianza sobre la coherencia de las predicciones obtenidas.

En líneas generales, los resultados permiten observar que el enfoque planteado es técnicamente viable y presenta un desempeño adecuado para actividades de clasificación automática de imágenes de afecciones visuales en este caso retinopatía diabética, cumpliendo con los objetivos propuestos en el desarrollo del presente trabajo.

Figura 20**Clasificación Imagen: Retinopatía Leve**

Nota. Visualización del módulo de clasificación UI (retinopatía leve)

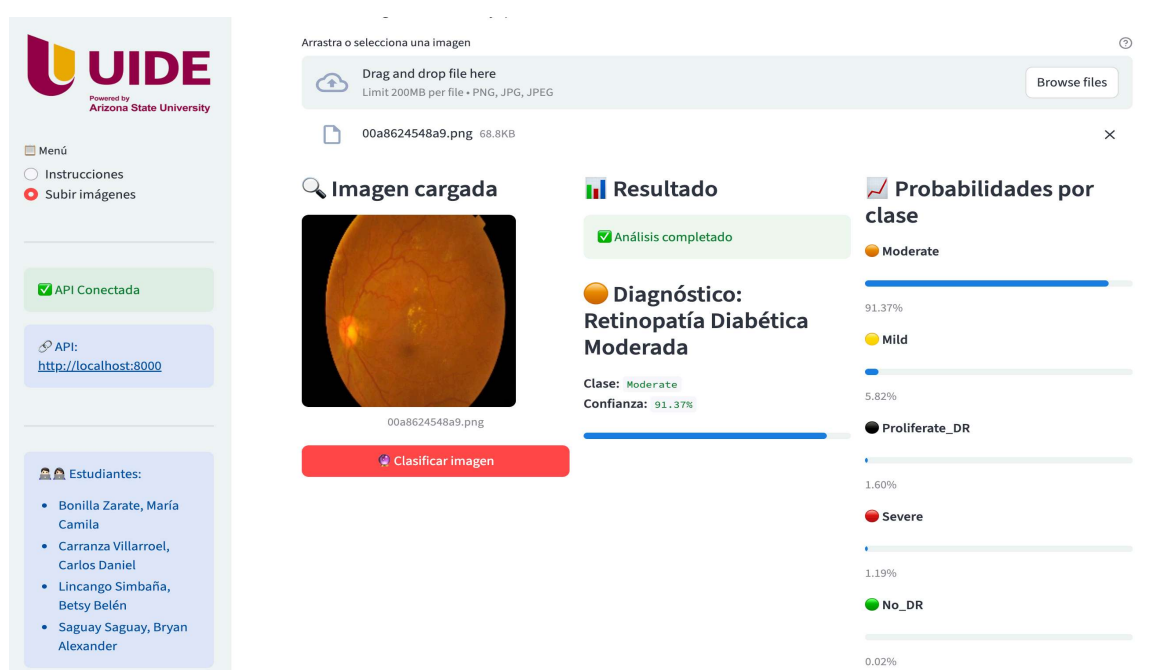
En la Figura 20, se puede observar que el modelo efectúa el proceso de clasificación de clases el cual despliega un diagnóstico de Retinopatía Diabética Leve (Mild) con un nivel de confianza de 70.71%, este valor permite cuantificar la certeza la predicción de este modelo. De igual manera, dentro del panel de probabilidades por clase detalla la distribución entre las cinco categorías analizadas (desde No_DR hasta Proliferative_DR), demostrando que tras la clase predominante “Mild”, la categoría “Moderate” presenta la segunda mayor probabilidad 20.82%, lo cual ayuda aportando transparencia al momento de la decisión en escenarios clínicos que tienen límites difusos.

Desde el enfoque analítico, esta prueba de concepto permite validar la capacidad del modelo para analizar patrones visuales en estudios tempranos de la enfermedad, aportando a la mejora de la interpretación de los resultados mediante las visualizaciones gráficas. Por lo que

se puede mencionar que la interfaz ratifica la factibilidad del uso de la herramienta como una herramienta que sirva de soporte experimental eficaz para el análisis automatizado de imágenes retinianas, consolidando su utilidad académica en la detección temprana de la patología.

Figura 21

Clasificación Imagen: Retinopatía Moderada



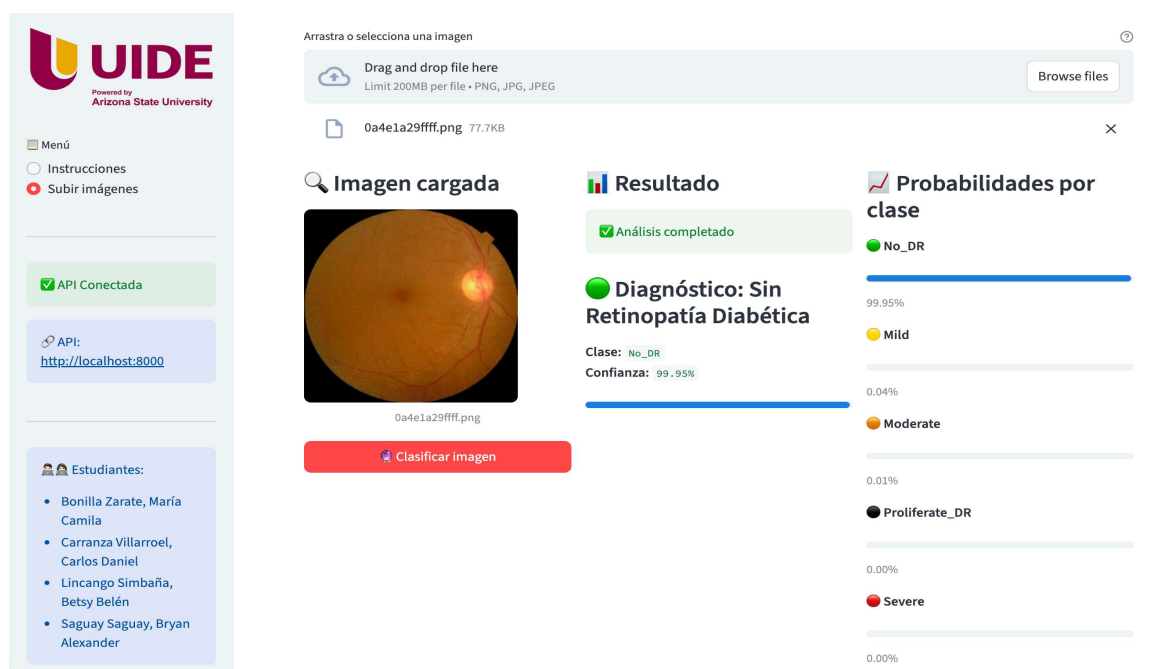
Nota. Visualización del módulo de clasificación UI (retinopatía Moderada)

En la Figura 21, se puede observar que el modelo efectúa el proceso de clasificación de clases el cual despliega un diagnóstico de Retinopatía Diabética Moderada (Moderate) con un nivel de confianza de 91.37%, este valor permite cuantificar la certeza la predicción de este modelo. De igual manera, dentro del panel de probabilidades por clase detalla la distribución entre las cinco categorías analizadas (desde No_DR hasta Proliferative_DR), demostrando que tras la clase predominante “Moderate”, la categoría “Mild” presenta la segunda mayor probabilidad 1.60%, lo cual ayuda aportando transparencia al momento de la decisión en

escenarios clínicos que tienen límites difusos.

Figura 22

Clasificación Imagen: Sin Retinopatía

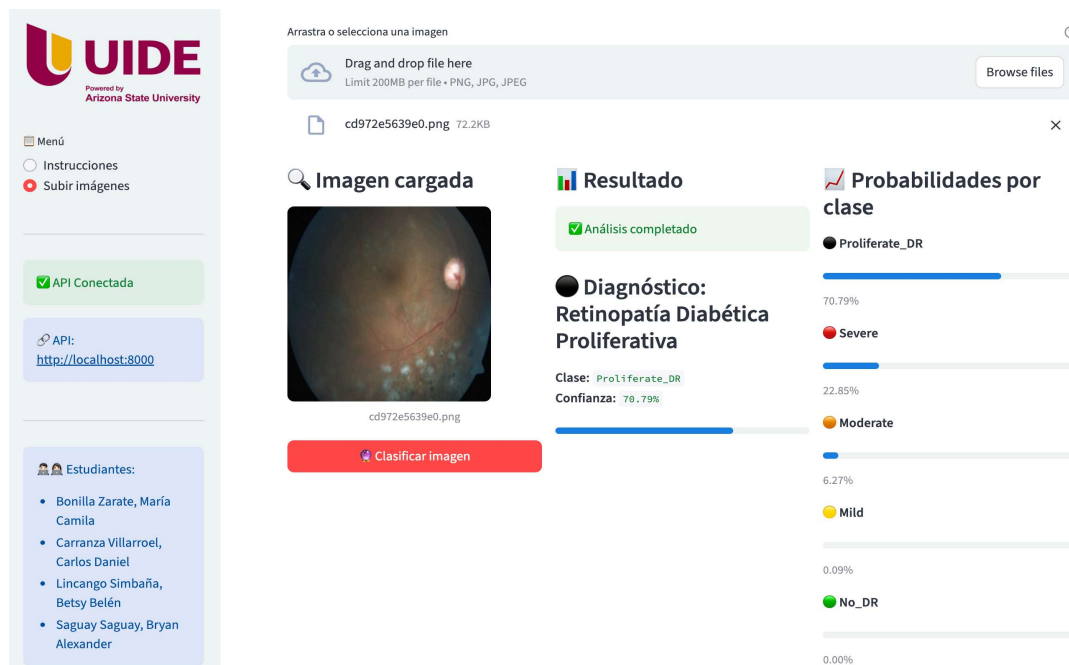


Nota. Visualización del módulo de clasificación UI (sin retinopatía)

En la Figura 22, se puede observar que el modelo efectúa el proceso de clasificación de clases el cual despliega un diagnóstico de la ausencia de Retinopatía (No_DR) con un nivel de confianza de 99.95%, este valor permite cuantificar la certeza la predicción de este modelo. De igual manera, dentro del panel de probabilidades por clase detalla la distribución entre las cinco categorías analizadas (desde No_DR hasta Proliferative_DR), demostrando que tras la clase predominante “No_DR”, la categoría “Mild” presenta la segunda mayor probabilidad 0.04%, lo cual ayuda aportando transparencia al momento de la decisión en escenarios clínicos que tienen límites difusos.

Figura 23

Clasificación Imagen: Retinopatía Proliferativa

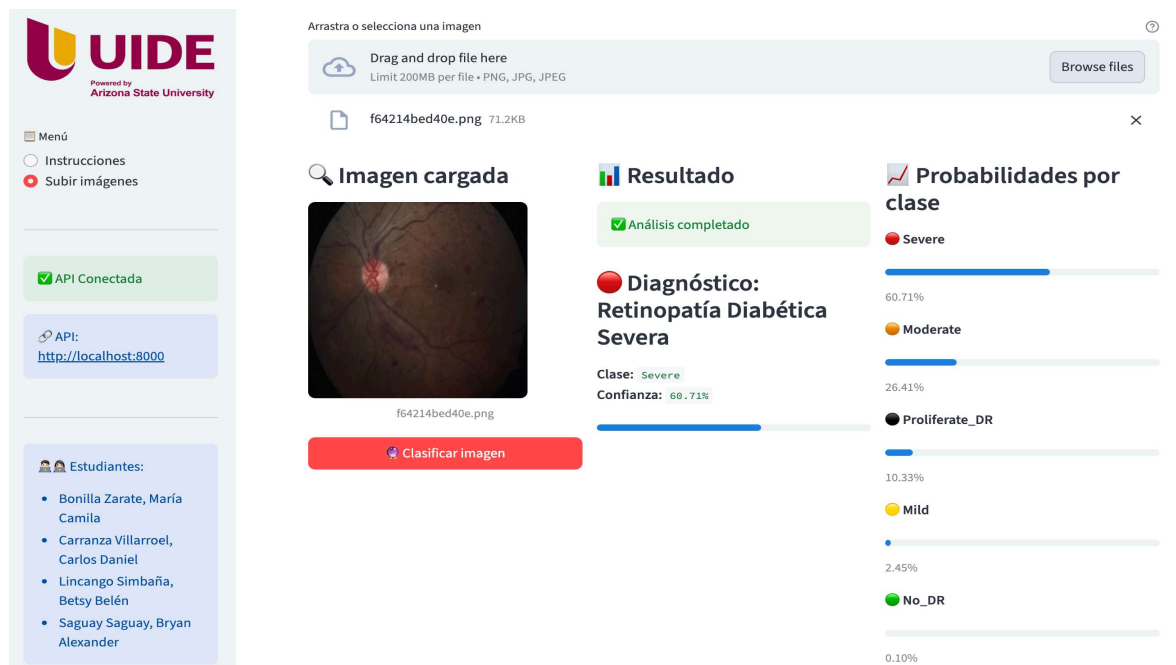


Nota. Visualización del módulo de clasificación UI (retinopatía proliferativa)

En la Figura 23, se puede observar que el modelo efectúa el proceso de clasificación de clases el cual despliega un diagnóstico de Retinopatía Diabética Proliferativa (Proliferate_DR) con un nivel de confianza de 70.79%, este valor permite cuantificar la certeza la predicción de este modelo. A su vez, dentro del panel de probabilidades por clase detalla la distribución entre las cinco categorías analizadas, demostrando que tras la clase predominante “Proliferate_DR”, la categoría “Severe” presenta la segunda mayor probabilidad 22.89%, lo cual ayuda aportando transparencia al momento de la decisión en escenarios clínicos que tienen límites difusos.

Figura 24

Clasificación Imagen: Retinopatía Severa




Nota. Visualización del módulo de clasificación UI (retinopatía severa)

En la Figura 24, se puede observar que el modelo efectúa el proceso de clasificación de clases el cual despliega un diagnóstico de Retinopatía Severa (Severe) con un nivel de confianza de 60.71%, este valor permite cuantificar la certeza la predicción de este modelo. De igual manera, dentro del panel de probabilidades por clase detalla la distribución entre las cinco categorías analizadas (desde No_DR hasta Proliferate_DR), demostrando que tras la clase predominante “Severe”, la categoría “Moderate” presenta la segunda mayor probabilidad 26.41% y seguido de “Proliferate_DR” con 10.33%, lo cual ayuda aportando transparencia al momento de la decisión en escenarios clínicos que tienen límites difusos.

Figura 25

Comparación con el modelo entrenado desde cero

 Resultados Transfer_Learning:
Accuracy: 0.7825 | Precision: 0.6809 | Recall: 0.5331 | F1-macro: 0.5640

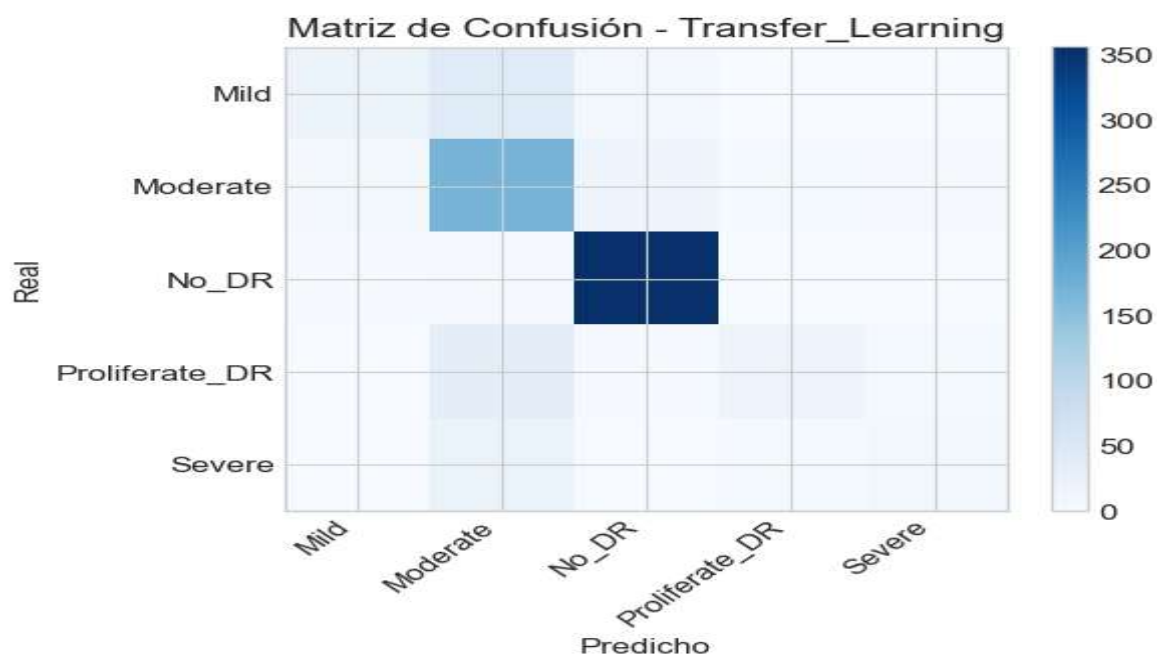
	precision	recall	f1-score	support
Mild	0.72	0.28	0.41	74
Moderate	0.62	0.84	0.71	199
No_DR	0.93	0.99	0.96	361
Proliferate_DR	0.61	0.29	0.39	59
Severe	0.53	0.26	0.35	38
accuracy			0.78	731
macro avg	0.68	0.53	0.56	731
weighted avg	0.78	0.78	0.76	731

Nota. Visualización de resultados de Transfer Learning

La Figura 25, se presenta un reporte de métricas de clasificación necesarias para entender el rendimiento del modelo implementado mediante Transfer Learning. En este reporte se detallan indicadores esenciales como son la precisión, el recall y el f1-score para cada una de las categorías. Al momento de analizar más detalladamente este reporte podemos darnos cuenta que el desempeño para clasificar en este caso la clase “No_DR” es sobresaliente teniendo una precisión de 0.93 y un recall de 0.99, lo que nos permite indicar una alta capacidad del modelo para indentificar imágenes de retinas sanas. Al analizar de manera más general podemos observar que el modelo alcanza un accuracy de 0.7825, demostrando una robustez sólida, aunque las métricas en clases de severidad intermedia sugieren áreas de optimización debido a la complejidad de los patrones visuales.

Figura 26

Matriz de confusión Transfer Learning



Nota. Visualización de la matriz de confusión usando transfer_learning

La Figura 26, muestra la Matriz de Confusión obtenida mediante el enfoque de Transfer Learning, la cual permite visualizar la relación entre los valores reales y las predicciones del modelo. Aquí podemos visualizar la diagonal principal que presenta colores más intensos, lo cual confirma que la mayoría de las muestras son clasificadas de manera correcta, especialmente en la categoría "No_DR" donde se concentra la mayor cantidad de aciertos. Pero a su vez se puede observar que la matriz revela ligeras confusiones entre clases conjuntas, como "Mild" y "Moderate", un fenómeno esperado en el diagnóstico médico asistido donde la transición entre estudios patológicos presenta características anómalas similares, validando así la necesidad de la visualización de probabilidades discutida anteriormente.

Figura 27

Resultados Fine Tuning

 Resultados Fine_Tuning:

Accuracy: 0.7907 | Precision: 0.7179 | Recall: 0.5530 | F1-macro: 0.5842

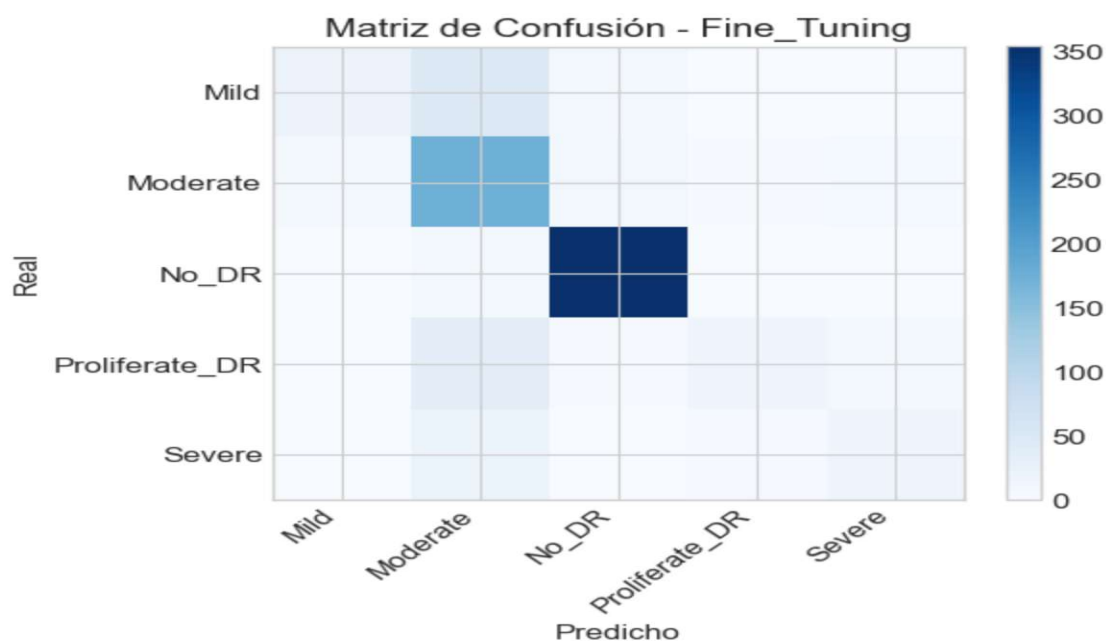
	precision	recall	f1-score	support
Mild	0.74	0.27	0.40	74
Moderate	0.61	0.87	0.72	199
No_DR	0.95	0.98	0.96	361
Proliferate_DR	0.73	0.27	0.40	59
Severe	0.56	0.37	0.44	38
accuracy			0.79	731
macro avg	0.72	0.55	0.58	731
weighted avg	0.80	0.79	0.77	731

Nota. Visualización de resultados de Fine Tuning

La Figura 27, se presenta un reporte de métricas de clasificación necesarias tras el proceso de Fine Tuning, evidencia que mejora el rendimiento general del modelo. En este reporte se detallan indicadores esenciales como son la precisión, el recall y el F1-score para cada una de las categorías. El modelo alcanza un accuracy general de 0.7907, superando el desempeño del enfoque inicial de Transfer Learning, al momento de analizar más detalladamente este reporte podemos darnos cuenta que el desempeño para clasificar en este caso la clase “No_DR” es sobresaliente teniendo un F1-score de 0.96, *mientras que otras categorías como Proliferative_DR muestran un incremento en su precisión, alcanzando el 0.73. Estos indicadores sugieren que el ajuste fino de las capas profundas de la red permitió una extracción de características más específica para las patologías de la retina, elevando el promedio ponderado de la precisión a 0.80.*

Figura 28

Matriz de confusión Fine Tuning



Nota. Visualización de la matriz de confusión usando fine_tuning

La Figura 28, muestra la Matriz de Confusión obtenida mediante el enfoque de Transfer Learning, la cual permite visualizar la relación entre los valores reales y las predicciones del modelo. Aquí podemos visualizar la diagonal principal que presenta colores más intensos, lo cual confirma que la mayoría de las muestras son clasificadas de manera correcta, especialmente en la categoría "No_DR" donde se concentra la mayor cantidad de aciertos. En comparación con la matriz anterior, se percibe una mayor nitidez en la clasificación de la categoría "Moderate", lo que reduce el ruido en las predicciones hacia las clases adyacentes.

Figura 29

Resultados comparación general

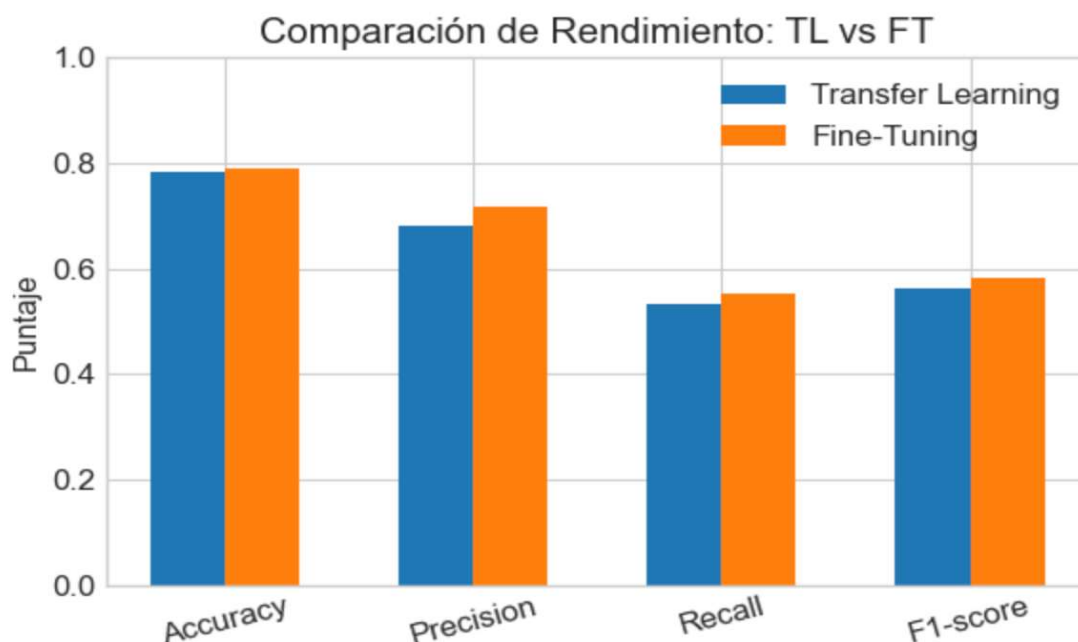
	Métrica	Transfer Learning	Fine-Tuning	Δ (FT - TL)
0	Accuracy	0.782490	0.790698	0.008208
1	Precision	0.680950	0.717876	0.036926
2	Recall	0.533090	0.552972	0.019882
3	F1-score	0.563963	0.584161	0.020198

Nota. Comparativa de los resultados Transfer Learning y Fine Tuning

En la Figura 29 se presenta una tabla comparativa que sintetiza el rendimiento de las dos metodologías empleadas: Transfer Learning y Fine Tuning. En esta se observa que el Fine Tuning supera al Transfer Learning en las métricas evaluadas, en este caso se observa un incremento en el accuracy de 0.0082 y una mejora más significativa en la Precisión de 0.0369. Este reporte de datos es fundamental para demostrar que el Fine Tuning permitió al modelo capturar de forma más eficiente las sutilezas del diagnóstico, logrando un F1-score final de 0.5842.

Figura 30

Comparación de Rendimiento: Transfer Learning vs Fine Tuning



Nota. La gráfica compara el desempeño de dos enfoques de entrenamiento.

En la Figura 30 se visualiza que el enfoque de Fine-Tuning supera al Transfer Learning, lo que demuestra que ajustar más capas del modelo permite capturar mejor las características específicas de las imágenes de retinopatía diabética. Esto indica que el ajuste de capas adicionales permite una mejor adaptación a las características específicas de las imágenes de retinopatía diabética, mejorando la precisión en la clasificación de los distintos niveles de severidad.

CAPITULO 5

5. Conclusiones y Recomendaciones

5.1 Conclusiones

- La revisión del estado del arte evidenció que los modelos de aprendizaje profundo, especialmente las redes convolucionales preentrenadas, constituyen una de las técnicas más utilizadas y efectivas en la clasificación de imágenes médicas, destacándose su aplicación en la detección de retinopatía diabética por su capacidad para extraer características complejas sin necesidad de ingeniería manual de variables.
- La correcta organización del conjunto de datos público de Mendeley Data permitió estructurar un problema de clasificación multiclase, facilitando el entrenamiento y evaluación del modelo, aunque se evidenció un desbalance entre clases que influyó en el desempeño diferencial del clasificador.
- La aplicación de técnicas de preprocesamiento y aumento de datos contribuyó a mejorar la capacidad de generalización del modelo, reduciendo el riesgo de sobreajuste y permitiendo un aprendizaje más robusto frente a la variabilidad presente en las imágenes retinianas.
- El uso de Transfer Learning mediante la arquitectura VGG16 permitió aprovechar características visuales previamente aprendidas en grandes volúmenes de datos, logrando un desempeño satisfactorio en la clasificación multiclase, incluso con un conjunto de datos limitado.
- La aplicación de Fine-Tuning sobre las capas superiores del modelo permitió especializar la red en patrones propios de las imágenes de fondo de ojo, mejorando el desempeño en determinadas clases y evidenciando la importancia del ajuste de hiperparámetros como la tasa de aprendizaje y el número de épocas.
- Las métricas de evaluación utilizadas, tales como exactitud, precisión, recall y F1-score, junto con el análisis mediante matrices de confusión, permitieron identificar un desempeño

global adecuado del modelo, así como diferencias en la capacidad de clasificación entre las distintas categorías de retinopatía diabética.

- La incorporación de Grad-CAM permitió visualizar las regiones de las imágenes que influyen en las decisiones del modelo, evidenciando que la red enfoca su atención en áreas clínicamente relevantes, lo que fortalece la interpretabilidad y confiabilidad del sistema propuesto en un contexto médico.

5.2 Recomendaciones

- Se recomienda ampliar el conjunto de datos, incorporando imágenes provenientes de diferentes fuentes y dispositivos, con el fin de mejorar la generalización del modelo en contextos clínicos reales.
- Se sugiere abordar el desbalance de clases mediante técnicas como ponderación de clases o estrategias avanzadas de aumento de datos, especialmente para las categorías con menor representación.
- Para trabajos futuros, se recomienda evaluar arquitecturas más recientes, como ResNet, EfficientNet o DenseNet, y comparar su desempeño con VGG16.
- Se aconseja profundizar en técnicas de explicabilidad adicionales, combinando Grad-CAM con otros métodos interpretables para fortalecer la confianza clínica en el modelo.
- Se recomienda realizar una validación externa con datos reales de pacientes, lo que permitiría evaluar la aplicabilidad del modelo como herramienta de apoyo al diagnóstico médico.
- Finalmente, se sugiere integrar el modelo en un sistema de apoyo a la decisión clínica, como una aplicación web o móvil, que permita su uso práctico en entornos de atención primaria.

Referencias

- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests 1: Sensitivity and specificity. *BMJ*, 308(6943), 1552. <https://doi.org/10.1136/bmj.308.6943.1552>
- Arias-Pacheco, L. T., Villacrés-Landeta, T. M., & Mora-Brito, E. V. (2024). Estimación de los costos integrales de la diabetes mellitus tipo 2 y sus complicaciones en el contexto ecuatoriano. *Revista de la ALAD*, 14(2), 57. <https://doi.org/10.24875/alad.25000003>
- Bhatia, K., Arora, S., & Tomar, R. (2017). Diagnosis of diabetic retinopathy using machine learning classification algorithm. In *Proceedings of the 2nd International Conference on Next Generation Computing Technologies (NGCT)* (pp. 347–351). IEEE. <https://doi.org/10.1109/ngct.2016.7877439>
- Bhimavarapu, U., & Battineni, G. (2022). Deep learning for the detection and classification of diabetic retinopathy with an improved activation function. *Healthcare*, 11(1), 97. <https://doi.org/10.3390/healthcare11010097>
- Dai, L., Wu, L., Li, H., Cai, C., Wu, Q., Kong, H., ... Jia, W. (2021). A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nature Communications*, 12, 3242. <https://doi.org/10.1038/s41467-021-23458-5>
- Das, D., Biswas, S. K., & Bandyopadhyay, S. (2022). A critical review on diagnosis of diabetic retinopathy using machine learning and deep learning. *Multimedia Tools and Applications*, 81(18), 25613–25655. <https://doi.org/10.1007/s11042-022-12642-4>
- Dávila-Flores, J. X., Montenegro-Morán, E. E., Macías-Gaytán, Á. M., & Tayupanda-Martínez, J. L. (2023). La diabetes mellitus y diabetes gestacional en adolescentes en el mundo y en el Ecuador: Manejo, prevención, tratamiento y mortalidad. *RECIMUNDO*, 7(2), 33–48. [https://doi.org/10.26820/recimundo/7.\(2\).jun.2023.33-48](https://doi.org/10.26820/recimundo/7.(2).jun.2023.33-48)

- De, M., Sánchez-Villa, J., & Martínez-Jiménez, E. O. (2025). Técnicas de inteligencia artificial para detección temprana de retinopatía diabética. *Pädi Boletín Científico de Ciencias Básicas e Ingenierías del ICBI*, 13(25), 12–17. <https://doi.org/10.29057/icbi.v13i25.12848>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Goutte, C., & Gaussier, E. (2005). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *Lecture Notes in Computer Science* (Vol. 3408, pp. 345–359). Springer. https://doi.org/10.1007/978-3-540-31865-1_25
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), Article 93. <https://doi.org/10.1145/3236009>
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22), 2402–2410. <https://doi.org/10.1001/jama.2016.17216>
- Herrero-Tudela, M. (2021). *Clasificación automática de la severidad de la retinopatía diabética mediante técnicas de deep learning* [Trabajo de fin de máster, Universidad de Valladolid]. <https://uvadoc.uva.es/handle/10324/50053>
- Jaderberg, M., Simonyan, K., & Zisserman, A. (2015). Spatial transformer networks. *Advances in Neural Information Processing Systems*, 28.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88. <https://doi.org/10.1016/j.media.2017.07.005>

- Martínez-López, E., De la Cruz-Gámez, E., Hernández-Hernández, M., Martínez-Arroyo, M., & Montero-Valverde, J. A. (2024). Avances en la detección de retinopatía diabética: El rol prometedor de la inteligencia artificial. *Ciencia Latina: Revista Multidisciplinar*, 8(1), 5744–5756. https://doi.org/10.37811/cl_rcm.v8i1.9925
- Math, L., & Fatima, R. (2020). Adaptive machine learning classification for diabetic retinopathy. *Multimedia Tools and Applications*, 80(4), 5173–5186. <https://doi.org/10.1007/s11042-020-09793-7>
- Mitchell, T. M. (1997). Does machine learning really work? *AI Magazine*, 18(3), 11. <https://doi.org/10.1609/aimag.v18i3.1303>
- World Health Organization. (2005). *Prevention of blindness from diabetes mellitus: Report of a WHO consultation in Geneva, Switzerland*. https://apps.who.int/iris/bitstream/handle/10665/43576/924154712X_eng.pdf
- Organización Panamericana de la Salud. (2019). *La diabetes, un problema prioritario de salud pública en el Ecuador y la región de las Américas*. <https://www.paho.org/es/temas/diabetes>
- Parikh, R., Mathai, A., Parikh, S., Chandra-Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Ophthalmology*, 56(1), 45–50. <https://doi.org/10.4103/0301-4738.37595>
- Patiño-Pérez, D., Armijos-Valarezo, L., Chóez-Acosta, L., & Burgos-Robalino, F. (2025). Redes neuronales convolucionales para detección de retinopatía diabética. *Ingenius. Revista de Ciencia y Tecnología*, 33, 91–101. <https://doi.org/10.17163/ings.n33.2025.08>
- Powers, D. M. W. (2020). *Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation*. arXiv. <https://doi.org/10.48550/arXiv.2010.16061>
- Rao, M., Zhu, M., & Wang, T. (2020). *Conversion and implementation of state-of-the-art deep learning algorithms for the classification of diabetic retinopathy*. arXiv. <https://doi.org/10.48550/arXiv.2010.11692>

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>

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 618–626).

Shukla, U. V., & Tripathy, K. (2023). *Diabetic retinopathy*. In *StatPearls*. StatPearls Publishing.

<https://www.ncbi.nlm.nih.gov/books/NBK560805/>

Wang, C., Stupina, A., & Bezhitskiy, S. (2024). Online sales prediction approach using methodology of CRISP-DM. *ITM Web of Conferences*, 59, 01006.

<https://doi.org/10.1051/itmconf/20245901006>

Wu, J. H., Liu, T. Y. A., Hsu, W. T., Ho, J. H. C., & Lee, C. C. (2021). Performance and limitation of machine learning algorithms for diabetic retinopathy screening: Meta-analysis. *Journal of Medical Internet Research*, 23(7), e23863. <https://doi.org/10.2196/23863>

Youldash, M., Rahman, A., Alsayed, M., Sebiany, A., Alzayat, J., Aljishi, N., Alshammari, G., & Alqahtani, M. (2024). Early detection and classification of diabetic retinopathy: A deep learning approach. *AI*, 5(4), 2586. <https://doi.org/10.3390/ai5040125>

Anexos

Repositorio Proyecto: <https://github.com/Syetb/retinopatia-diabetica-UI.git>