
Desarrollo de un modelo de clasificación de imágenes de fondo de ojo con inteligencia artificial para la detección de anomalías oculares en Guatemala

Juan Andrés Galicia Reyes



UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**Desarrollo de un modelo de clasificación de imágenes de fondo
de ojo con inteligencia artificial para la detección de
anomalías oculares en Guatemala**

Trabajo de graduación presentado por Juan Andrés Galicia Reyes para
optar al grado académico de Licenciado en Ingeniería en Ciencia de los
Datos

Guatemala,

2025

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**Desarrollo de un modelo de clasificación de imágenes de fondo
de ojo con inteligencia artificial para la detección de
anomalías oculares en Guatemala**

Trabajo de graduación presentado por Juan Andrés Galicia Reyes para
optar al grado académico de Licenciado en Ingeniería en Ciencia de los
Datos

Guatemala,

2025

Vo. Bo.



(F) _____
Dr. Alejandro Campos Alvergue

Tribunal



(F) _____
Dr. Alejandro Campos Alvergue



(F) _____
Msc. Jose Antonio Medrano



(F) _____
Msc. Luis Pedro Flores

Fecha de aprobación.

22 de enero 2025

La motivación para esta tesis surgió de una inquietud personal y profesional sobre el impacto de la diabetes en la salud visual de los guatemaltecos. A través de conversaciones con mi oftalmólogo, el Dr. Alejandro Campos Alvergue, identificamos que la diabetes, un problema común en Guatemala debido a la alimentación y el estilo de vida de muchos, representa un riesgo significativo para la salud ocular. Aunque afortunadamente nadie en mi familia cercana ha padecido esta enfermedad, reconozco la importancia de cuidar nuestra salud para prevenir posibles complicaciones a futuro. Fue en conjunto con el Dr. Campos que nació la idea de desarrollar una herramienta que facilitara el acceso a un diagnóstico temprano de anomalías oculares relacionadas con la diabetes, especialmente considerando la necesidad de soluciones accesibles y adaptadas al contexto y características físicas nacionales. Las herramientas comerciales existentes suelen ser costosas, requieren acompañamiento profesional, y en algunos casos, no se ajustan a las necesidades específicas de la población guatemalteca y de los equipos que se poseen y utilizan comúnmente para obtener las imágenes de fondo de ojo de los pacientes.

Para materializar esta idea, presentamos la propuesta al Centro de Estudios de Informática Aplicada (CEIA) de la Universidad del Valle de Guatemala para poder trabajarla en la modalidad de trabajo de graduación. Este trabajo no habría sido posible sin el invaluable apoyo del Dr. Gabriel Barrientos, quien confió en esta iniciativa y ofreció su guía durante todo el proceso. Entre las principales limitaciones que enfrenté en el desarrollo de esta tesis estuvo la dificultad de acceder a imágenes de pacientes nacionales clasificadas, así como de la estandarización de un proceso de colaboración.

Con esta investigación, busco contribuir al acceso a la salud oftalmológica en Guatemala, acercando el diagnóstico temprano de retinopatía diabética a más personas y promoviendo el desarrollo de herramientas de inteligencia artificial que puedan tener un impacto positivo en el área de salud del país.

Quiero expresar mi más profundo agradecimiento, en primer lugar, a mi mamá, quien ha sido mi apoyo incondicional y me ha acompañado en cada paso de mi camino académico. Gracias a ella, hoy puedo decir que sigo persiguiendo y logrando mis sueños. A mi abuela, que ha sido como una segunda madre, le agradezco el amor y el impulso constante para superarme cada día. También agradezco a toda mi familia y a las personas importantes en

mi vida que siempre me han acompañado.

Al Dr. Alejandro Campos, le agradezco la confianza depositada en mí para desarrollar este aporte al área médica de Guatemala y por la oportunidad de innovar en el cruce entre la ciencia de datos y la salud en nuestro país.

Mi gratitud también va dirigida al Dr. Gabriel Barrientos, por su guía y apoyo constante durante todo el proceso de investigación, brindando orientación clave en cada etapa.

Finalmente, a mis amigos de la universidad, quienes me motivaron a siempre dar lo mejor de mí y nunca rendirme. Cada clase, risa y momento compartido fueron una aventura inolvidable, y les agradezco por haber hecho de estos años una experiencia tan enriquecedora. Gracias por hacer de cada día un gran día y siempre los llevaré en mi corazón.

Índice

Prefacio	IV
Lista de figuras	VIII
Lista de cuadros	IX
Lista de algoritmos	X
Resumen	XI
Abstract	XII
I Introducción	1
II Justificación	3
III Alcance	5
IV Limitaciones	7
V Objetivos	10
5.1 Objetivo general	10
5.2 Objetivos específicos	10
VI Marco teórico	11
6.1 Conceptos básicos de oftalmología	11
6.1.1 Anatomía y fisiología del ojo	11
6.1.2 Enfermedades oculares comunes	12
6.1.3 Clasificación internacional de severidad de la retinopatía diabética (ICDRS)	13
6.1.4 Clasificación de la retinopatía diabética según la organización mundial de la salud (OMS)	14
6.1.5 Clasificación ETDRS de la retinopatía diabética [22]	14
6.1.6 Descripción de signos claves	15
6.1.7 Importancia de la clasificación	16
6.2 Imágenes de fondo de ojo	19
6.2.1 Tipos de imágenes de fondo de ojo	19
6.2.2 Técnicas y equipos de captura	21
6.2.3 Calidad de las imágenes	21
6.3 Diagnóstico de anomalías oculares	21

6.4	Inteligencia artificial y aprendizaje automático	22
6.4.1	Conceptos básicos de IA y ML	22
6.4.2	Tipos de aprendizaje automático	23
6.4.3	Evaluación de modelos de clasificación	24
6.5	Técnicas de preprocesamiento de imágenes	30
6.5.1	Conversión de espacio de color	30
6.5.2	Corrección de sombras	31
6.5.3	Aumento de contraste adaptativo	31
6.5.4	Exclusión de fondo	31
6.5.5	Filtrado	32
6.5.6	Generación de máscaras	32
6.6	Redes neuronales	33
6.6.1	Redes neuronales convolucionales (CNN)	33
6.6.2	ResNet	38
6.6.3	VGG	40
6.6.4	EfficientNet	43
6.6.5	MDFSC	45
6.6.6	<i>Transfer learning</i>	47
VII	Metodología	50
7.1	Herramientas de ingeniería utilizadas	50
7.2	Recopilación de datos	51
7.3	Selección de arquitectura y algoritmos	52
7.4	Preprocesamiento de imágenes	53
7.4.1	Anonimización de la imagen	54
7.4.2	Preprocesamiento macular	55
7.5	Evaluación del modelo	55
7.6	Conjunto de datos	57
7.7	Entrenamiento de modelos	59
7.8	Modelos	60
7.8.1	VGG-19	60
7.8.2	ResNet152	62
7.8.3	EfficientNetB7	64
7.8.4	MDFSC	65
7.8.5	<i>Callbakcs</i>	67
7.9	Etiquetador de imágenes nuevas	68
7.10	<i>Dashboard</i> de diagnóstico	70
VIII	Resultados	72
8.1	Enfoque binario	72
8.2	Enfoque multiclase	74
8.2.1	Análisis comparativo de arquitecturas	74
8.3	Modelo en producción	84
IX	Análisis de resultados	87
X	Conclusiones	92
XI	Recomendaciones	94

XII Bibliografía	96
XIII Anexos	101
13.1 Cuadros de diagnóstico	101

Lista de figuras

1	Visualización de una retina sana y enferma [25]	16
2	Representación visual de la arquitectura de una red neuronal convolucional (CNN) [47]	34
3	Campo receptivo y padding en redes neuronales convolucionales [48]	35
4	Graficas de funciones de activación importantes en clasificación [50]	37
5	<i>Max Pooling</i> 3x3 [51]	38
6	Distribución de imágenes de dataset Kaggle	51
7	Distribución de imágenes binarizadas de dataset Kaggle	52
8	Familias de modelos utilizados para predecir RD [57]	52
9	Imágen de fondo de ojo descartada por corte	58
10	Imágen de fondo de ojo descartada por no contener información	58
11	Imágen de fondo de ojo descartada por mal enfoque y luz	59
12	Pantalla de inicio de sesión de la herramienta de clasificación	68
13	Diagrama de entidad-relación de la base de datos del clasificador	69
14	Interfaz de clasificación de imágenes	70
15	Dashboard de diagnóstico preliminar	71
16	Distribución de error de reconstrucción imágenes sanas (entrenamiento)	73
17	Distribución de error de reconstrucción imágenes con anomalías oculares (<i>test</i>)	74
18	Matriz de confusión VGG-19 multiclase	76
19	Matriz de confusión VGG-19 binario	77
20	Curva AUC PR VGG-19	78
21	Evolución de AUC PR en conjunto de validación de VGG-19	79
22	Evolución de AUC PR en test VGG-19	79
23	Evolución de pérdida en conjunto de validación VGG-19	80
24	Etapas de <i>transfer learning</i> gradual de modelo VGG-19	81
25	Curva AUC PR ResNet152	82
26	Curva AUC PR EfficientNetB7	84
27	Visualización de imágenes de cara al usuario del <i>dashboard</i> de diagnóstico	85
28	Gráficas de probabilidad binaria y multiclase de tenencia de enfermedad y severidad	85
29	PDF preliminar generado por el <i>dashboard</i> de diagnóstico	86
30	Limitantes de la variedad de criterios [61]	90

Lista de cuadros

1	Síntomas de la retinopatía diabética según su estadio	17
2	Tratamientos comunes para la retinopatía diabética	18
3	Resultados AUC PR de clasificación modelos binarios	72
4	Métricas de clasificación VGG-19 <i>output</i> binario (AUC-ROC: 0.53)	73
5	Reporte de clasificación VGG-19 multiclase	75
6	Reporte de clasificación VGG-19 (modelo binario)	75
7	Resultados de quadratic weighted kappa para diferentes modelos	81
8	Reporte de clasificación ResNet152 multiclase	82
9	Reporte de clasificación ResNet152 (modelo binario)	82
10	Reporte de clasificación EfficientNetB7 multiclase	83
11	Reporte de clasificación EfficientNetB7 (modelo binario)	83
12	Escala internacional de severidad de la retinopatía diabética y edema macular diabético	103

Lista de algoritmos

1	Anonimización de la imagen de fondo de ojo	54
2	Preprocesamiento macular para retinopatía diabética	55

Se desarrolló un modelo de clasificación de imágenes de fondo de ojo basado en inteligencia artificial (IA) para la detección de anomalías oculares en Guatemala. En específico, la retinopatía diabética, con el objetivo de facilitar el acceso a poblaciones con acceso limitado a servicios oftalmológicos. Partiendo de una base de datos de imágenes oftalmológicas de pacientes en centros especializados y proporcionada por EyePACS, el modelo emplea arquitecturas de redes neuronales profundas como ResNet152, VGG19 y EfficientNetB7, optimizadas para clasificar tanto la presencia como la severidad de las anomalías.

Se evaluó y entrenó múltiples modelos de IA, empleando algoritmos de aprendizaje automático y redes neuronales profundas, específicamente redes neuronales convolucionales, por sus siglas en inglés (CNN). El proceso de validación y evaluación se realizó con métricas de sensibilidad, precisión, AUC-ROC, y AUC-PR, alcanzando óptimos resultados con VGG19 (AUC-PR de 0.8020 y un índice de concordancia ponderado de 0.600 en modo binario y 0.680 en modo multclasificación). La precisión promedio superior al 80% en pruebas sugiere que el modelo puede superar métodos convencionales, brindando una herramienta accesible y de bajo costo para el diagnóstico temprano en entornos clínicos de bajos recursos. Además, su aplicación permite reducir el impacto de la ceguera asociada a estas patologías, estableciendo una base para futuras mejoras en la detección automatizada de enfermedades oculares y sistémicas.

An artificial intelligence (AI)-based fundus image classification model was developed for detecting ocular anomalies in Guatemala, specifically diabetic retinopathy, with the aim of improving access for populations with limited ophthalmology services. Using a database of ophthalmologic images from patients at specialized centers, provided by EyePACS, the model employs deep neural network architectures such as ResNet152, VGG19, and EfficientNetB7, optimized to classify both the presence and severity of anomalies.

Multiple AI models were trained and evaluated, using machine learning algorithms and deep neural networks, specifically Convolutional Neural Networks (CNN). The validation and evaluation process used metrics including sensitivity, precision, AUC-ROC, and AUC-PR, achieving optimal results with VGG19 (AUC-PR of 0.8020 and a weighted Kappa score of 0.600 in binary mode and 0.680 in multiclass mode). An average accuracy exceeding 80% in testing suggests that the model could outperform traditional methods, providing an accessible, low-cost tool for early diagnosis in low-resource clinical settings. Additionally, its application could reduce the impact of blindness associated with these conditions, establishing a foundation for future advancements in automated detection of ocular and systemic diseases.

En la era actual, el campo de la inteligencia artificial (IA) se ha convertido en una herramienta poderosa y prometedora para abordar diversos desafíos médicos, particularmente en el diagnóstico y la detección temprana de enfermedades [1]. Diversas anomalías oculares pueden estar bastante avanzadas antes de afectar la visión, un diagnóstico y tratamiento oportunos pueden reducir el riesgo de pérdida visual en aproximadamente un 57% comentan estudios publicados en la revista *Nature* [2]. Por lo tanto, una cita rutinaria y el seguimiento regular son esenciales para los pacientes con diabetes u otras enfermedades que hagan propensos a las personas a padecer de alguna anomalía, especialmente las personas de mediana edad y ancianas [3]. Sin embargo, un estudio ha indicado que una cantidad significativa de pacientes con diabetes no se somete a los exámenes oculares anuales recomendados, esto debido al largo tiempo de examen, la falta de síntomas y el acceso limitado a especialistas en retina [2]. Es por esto que el presente estudio se enfocó en el desarrollo de un modelo de clasificación basado en IA para la identificación de anomalías oculares en imágenes de fondo de ojo en Guatemala.

El propósito principal de este trabajo es implementar un sistema de IA capaz de detectar anomalías en imágenes de fondo de ojo, contribuyendo así a la prevención y el tratamiento oportuno de esta afectación ocular por la cual está siendo afectado el paciente, evitando diagnósticos incorrectos causados por la baja calidad de las imágenes, la experiencia individual y otros factores [4]. A través de la aplicación de algoritmos de aprendizaje automático sobre una amplia colección de imágenes oftalmológicas provenientes de pacientes atendidos en centros especializados de Guatemala, se busca generar un modelo preciso y confiable para la detección automática de esta irregularidad. Este enfoque permite explorar la capacidad de la inteligencia artificial para identificar patrones sutiles asociados con anomalías oculares, problemas que pueden llevar a la pérdida de la visión si no se detectan y tratan a tiempo. Mundialmente se estima que las principales enfermedades oculares que presentan las personas con anomalías oculares son: cataratas, glaucoma, conjuntivitis, orzuelo, ojo seco, desprendimiento de retina; degeneración macular asociada a la edad, por sus siglas en inglés (AMD), y retinopatía diabética (RD). Donde según un estudio de concienciación sobre las enfermedades oculares más comunes y sus riesgos, realizado en Polonia, obtuvo que un

83.6% de los adultos conocen sobre las cataratas, pero solo un 16.4% sobre la retinopatía diabética [5].

La metodología utilizada comprende la recopilación de una extensa base de datos de imágenes oftalmológicas y la implementación de múltiples arquitecturas para la clasificación de las mismas. Se evalúan y comparan diferentes modelos de IA utilizando métricas como sensibilidad, especificidad, precisión y F1-score, garantizando así un equilibrio adecuado entre la capacidad de detección de anomalías oculares y la minimización de los falsos positivos.

La motivación detrás de este estudio radica en una preocupación personal y generalizada: muchas personas, incluidos mis familiares cercanos, han enfrentado problemas oculares que solo se han detectado y tratado cuando la enfermedad ya estaba avanzada. Esto subraya la necesidad urgente de desarrollar métodos que permitan diagnósticos tempranos y accesibles. Por ello, el objetivo de este trabajo es contribuir a que más personas puedan recibir un diagnóstico temprano y asequible de enfermedades oculares.

Las anomalías oculares, como la RD, el glaucoma y la AMD, representan un desafío significativo para la salud pública debido a su potencial para causar pérdida de visión y ceguera. La detección temprana y el manejo adecuado de estas condiciones son esenciales para prevenir daños severos en la visión. La retinopatía diabética es una de las principales complicaciones oculares asociadas con la diabetes. Según un estudio sobre tendencias en la prevalencia de la ceguera y la discapacidad visual, se estima que 43.3 millones de personas son ciegas, donde 295 millones de personas tienen algún tipo de discapacidad visual moderada o grave [6]. En Guatemala, como en muchos otros países en desarrollo, la prevalencia de la diabetes está aumentando rápidamente debido a cambios en el estilo de vida y la dieta, lo que incrementa el riesgo de desarrollar retinopatía diabética [7].

La preocupación por este problema específico radica en la necesidad urgente de contar con herramientas efectivas y accesibles para la detección temprana de cualquier anomalía ocular. En el caso específico de la retinopatía diabética. A pesar de que la prevalencia global de cualquier tipo fue del 27.0% durante el período comprendido entre 2015 y 2019, en las etapas iniciales de la RD no existen síntomas distintivos. Esta condición puede estar bastante avanzada antes de afectar la visión, lo que resalta la importancia de un diagnóstico y tratamiento oportunos, los cuales pueden reducir el riesgo de pérdida visual hasta en un 57%. Por consiguiente, la detección rutinaria y el seguimiento regular son cruciales para los pacientes con diabetes, en especial para aquellos de mediana edad y personas mayores. Sin embargo, diversos estudios han señalado que una proporción significativa de pacientes con diabetes no se someten al examen ocular anual recomendado debido al prolongado tiempo de examinación, la ausencia de síntomas y la limitada accesibilidad a especialistas en retina. Para abordar estas barreras, se han propuesto distintos enfoques, entre ellos la aplicación de técnicas de inteligencia artificial (IA) para la detección y diagnóstico de la RD [8].

Actualmente, el acceso a servicios oftalmológicos especializados y a equipos médicos de alta tecnología es limitado en muchas regiones de Guatemala, especialmente en áreas rurales. En la encuesta realizada por el Ministerio de Salud Pública y Asistencia Social de Guatemala, en conjunto con entidades oftalmológicas del país e internacionales, se determinó

que muchos de los pacientes afectados, no conocían de la posibilidad de tratamiento, así como desconfianza del resultado, o bien sin acceso al mismo [9]. Esto significa que muchos pacientes oftalmológicos corren el riesgo de no recibir un diagnóstico oportuno y adecuado de esta condición. La prevalencia de ceguera en adultos mayores es más alta en Guatemala que en la mayoría de los países centroamericanos. La mayoría de los casos de ceguera y discapacidad visual eran prevenibles o tratables [9]. Es una condición urgente que traería un gran beneficio el poder apoyar a su diagnóstico temprano. La retinopatía diabética es la primera causa de ceguera en países en vías de desarrollo, según un estudio de la Dra. Loza. En este se muestra que la evaluación periódica, para su detección temprana en pacientes diabéticos de larga evolución, es importante para preservar la visión en estos pacientes [10].

El desarrollo de un modelo de inteligencia artificial para la clasificación automática de imágenes de fondo de ojo podría tener una importancia potencial significativa. Este modelo podría permitir una detección rápida y precisa de anomalías oculares, especialmente retinopatía diabética, utilizando tecnología de bajo costo y accesible, como cámaras oftalmológicas estándar y computadoras convencionales, como mencionan en la encuesta de ceguera realizada en Guatemala en 2015. El aumento en la disponibilidad de tratamientos de cataratas asequibles y de alta calidad tendría un impacto sustancial en la prevención de la ceguera [9].

La relevancia de este trabajo radica en varios aspectos:

1. **Reduce la brecha de accesibilidad:** al permitir el acceso a este modelo de IA, pacientes podrían obtener un primer diagnóstico sin la necesidad de un médico especialista en la zona. Únicamente con el equipo adecuado.

2. **Prevención de complicaciones:** la detección temprana y el tratamiento oportuno de la retinopatía diabética pueden prevenir complicaciones graves como la pérdida de visión y la ceguera, mejorando la calidad de vida de los pacientes y reduciendo la carga económica y social asociada con estas complicaciones.

3. **Aplicación de tecnología de vanguardia:** la implementación de inteligencia artificial en el campo de la salud ocular representa un avance significativo en la aplicación práctica de la tecnología para mejorar la atención médica y diagnosticar enfermedades de manera eficiente.

4. **Impacto a largo plazo:** la investigación y desarrollo en este campo sentará las bases para futuros avances en la detección automatizada de otras anomalías oculares y enfermedades sistémicas relacionadas.

El estudio cuenta con acceso a computadoras de alto rendimiento equipadas con unidad de procesamiento de graficos (GPU, por sus siglas en inglés), esenciales para el entrenamiento de los modelos de inteligencia artificial. Se utilizan herramientas de software avanzadas como TensorFlow y Keras para el desarrollo y entrenamiento de los modelos de clasificación. Además, se dispone de una amplia base de datos de alrededor de 35 mil imágenes oftalmológicas de pacientes internacionales, provenientes de un *dataset* de uso libre otorgado por EyePac. Este es un programa de telemedicina utilizado para crear programas sostenibles de detección de retinopatía diabética en clínicas comunitarias de todo Estados Unidos [11]. Fue patrocinado por California Healthcare Foundation [12]. Estas imágenes fueron dadas con el objetivo de realizar una competencia para la creación de un modelo de estas características pero especializado para sus imágenes en Kaggle. Kaggle es una subsidiaria de Google LLC, una comunidad en línea de científicos de datos y profesionales del aprendizaje automático. Permite a los usuarios encontrar y publicar conjuntos de datos, explorar y crear modelos en un entorno de ciencia de datos basado en la web; trabajar con otros científicos de datos e ingenieros de aprendizaje automático y participar en concursos para resolver desafíos de ciencia de datos. Estas imágenes serán fundamentales para entrenar y validar los modelos de IA. El estudio se ha planificado cuidadosamente, asignando tiempo suficiente para cada fase del proyecto, desde la recolección de datos hasta la implementación y evaluación de los modelos de IA.

Las imágenes de fondo de ojo provienen directamente de clínicas oftalmológicas en todo el mundo, asegurando datos relevantes y específicos para la población objetivo. Se cuenta con acceso a estudios y publicaciones científicas relevantes, a través de bases de datos académicas como PubMed, IEEE Xplore y Google Scholar, para fundamentar el desarrollo del modelo y contextualizar los resultados obtenidos, asegurando la confidencialidad y anonimato de los pacientes. Esto permite compartir los hallazgos con la comunidad científica y contribuir al avance del diagnóstico ocular mediante IA. También se liberan los derechos de autor y el libre uso de la plataforma construida.

Asimismo, se diseña un flujo para poder incorporar nuevas imágenes de cualquier clínica

de Guatemala. Esto con el objetivo de realizar *transfer learning*, es una técnica en la que se reutiliza el conocimiento adquirido en un dominio (llamado dominio fuente) para mejorar el rendimiento en otro dominio diferente (denominado dominio objetivo). Esta estrategia es especialmente útil cuando se dispone de datos limitados en el dominio objetivo, como suele ocurrir en la imagen médica. Las restricciones en la adquisición de datos, la rareza de algunas enfermedades, y las cuestiones legales y éticas que limitan el tamaño de los conjuntos de datos en el dominio médico, hacen que el transfer learning sea una herramienta valiosa para mejorar los resultados en este campo [13]. Esto abre la puerta a poder incorporar nuevas imágenes en estudios futuros, logrando especializar el modelo al tipo de imágenes que se tengan en la clínica, hospital o país, siempre que se respete el preprocesamiento utilizado en el modelo inicial.

El estudio se realiza partiendo de tener los datos previamente preprocesados con diferentes algoritmos. Se encarga de los pasos de extracción de imágenes preprocesadas, diseño de modelo de datos, visualización de datos, modelo predictivo, desarrollo de producto de datos y presentación final.

La aplicación final que podrán utilizar los doctores, clínicas y hospitales se encuentra implementada en un *dashboard* de Streamlit. Este es en formato de página web de uso libre, donde se podrá utilizar subiendo una imagen de fondo de ojo de un paciente y este retornará su diagnóstico. El propósito es que el paciente pueda obtener un diagnóstico temprano luego de haber obtenido su imagen de fondo de ojo, como un diagnóstico preliminar. Y este pueda ser dado por cualquier persona capacitada en realizar la toma de estas imágenes, sin necesidad que exista en ese momento un oftalmólogo. El objetivo es que, ya teniendo una imagen de fondo de ojo de un paciente, poder obtener un diagnóstico temprano de manera gratuita e inmediata únicamente mediante internet.

Las imágenes del *dataset* utilizado para el entrenamiento del modelo, es de Kaggle. Estas imágenes ya contienen un *label* asociado. Este fue colocado por un grupo de oftalmólogos que se encargaron de realizar los *labels* a todas las imágenes del *dataset*. Lograron establecer una clasificación objetiva, mediante un análisis de acuerdo entre evaluadores, una medida que evalúa el grado de concordancia o consistencia entre dos o más evaluadores o jueces cuando clasifican, evalúan o categorizan un conjunto de ítems o sujetos. Es especialmente relevante en estudios donde se necesita asegurar que las observaciones o juicios realizados por diferentes personas sean consistentes [14]. Por lo que, utilizando este índice, se pueden dar los *labels* para las imágenes del entrenamiento. Y de la misma manera se pueden utilizar nuevas imágenes para reentrenar el modelo. Pero se debe tener en cuenta que el modelo aprenderá o tendrá el mismo criterio médico que el planteado por las etiquetas colocadas a las imágenes de entrenamiento. Esto definirá el alcance de las predicciones logradas por el modelo.

Es importante establecer las limitaciones que un modelo de estas características puede tener en el contexto que se utiliza y la manera y tiempo en el que se logra su entrenamiento. Entre las limitaciones se incluyen factores relacionados con la calidad y disponibilidad de los datos, los recursos computacionales, la generalización del modelo, y las consideraciones éticas y regulatorias.

Variabilidad de las imágenes

La calidad y variabilidad de las imágenes de fondo de ojo pueden afectar significativamente la precisión del modelo. Imágenes de baja calidad, tomadas con diferentes equipos o en condiciones subóptimas, pueden introducir ruido en el entrenamiento del modelo. Esto lleva a que el modelo tenga dificultades para identificar correctamente las características que diferencian las imágenes normales de las que presentan anomalías.

Distribución de anomalías

Otro factor crucial es la distribución de las anomalías oculares dentro del conjunto de datos. La base de datos puede no contener una distribución equilibrada de todas las posibles condiciones oculares, lo que lleva a sesgar el modelo hacia la detección de las anomalías más comunes, limitando su capacidad para identificar condiciones menos frecuentes o más raras. Esto puede afectar negativamente la capacidad del modelo para generalizar a una variedad más amplia de casos clínicos.

Aplicabilidad a otras poblaciones

Aunque el modelo se entrena con un conjunto de imágenes específico, por ejemplo, imágenes de pacientes estadounidenses o indios, su aplicabilidad y precisión en otras poblaciones pueden ser limitadas. Las características demográficas, epidemiológicas y genéticas pueden variar significativamente entre regiones, lo que podría afectar el desempeño del modelo en nuevos entornos. Esto plantea la necesidad de realizar validaciones adicionales en poblaciones distintas para asegurar que el modelo sea robusto en diferentes contextos clínicos.

Limitaciones de hardware

A pesar de contar con acceso a computadoras de altas capacidades de procesamiento (CPU

y GPU), el entrenamiento de modelos de inteligencia artificial en imágenes médicas sigue siendo intensivo en recursos. Los modelos más complejos requieren grandes cantidades de tiempo de entrenamiento y el acceso a estos equipos puede estar limitado, lo que puede afectar la capacidad para optimizar el modelo completamente dentro del tiempo disponible.

Validación clínica y regulación

La implementación del modelo en un entorno clínico real requiere validaciones adicionales y la obtención de aprobaciones regulatorias. Aunque el modelo puede mostrar buenos resultados en un entorno controlado de laboratorio, su uso en la práctica médica diaria requiere una evaluación exhaustiva para garantizar que cumpla con los estándares clínicos y regulatorios. Sin una validación clínica completa, su aplicación en escenarios médicos reales sigue siendo limitada.

Transparencia y explicabilidad

Los modelos de inteligencia artificial, especialmente aquellos basados en redes neuronales profundas, suelen ser considerados como “caja negra”, debido a la dificultad para explicar las decisiones que toma el modelo. Esta falta de transparencia puede generar desconfianza entre los profesionales de la salud y los pacientes, quienes pueden dudar en adoptar una herramienta cuyos resultados no se pueden interpretar fácilmente. La explicabilidad es un desafío importante que debe abordarse para fomentar la confianza en el uso clínico de estos modelos.

Privacidad y consentimiento

El uso de datos médicos implica consideraciones éticas, especialmente en términos de privacidad y consentimiento informado. La recopilación de imágenes de fondo de ojo y su uso en el entrenamiento del modelo deben adherirse a estrictas normas de protección de datos y confidencialidad, garantizando que se respeten las leyes locales y las normativas internacionales, como la Ley de Protección de Datos Personales. Además, es fundamental que los pacientes comprendan el propósito del estudio y den su consentimiento informado para el uso de sus datos. Para su uso, se requerirá acceder al uso de una licencia de uso libre y bajo propias responsabilidades. Accediendo a poder subir los datos de los pacientes al modelo, donde no se estará guardando ninguna información ni imagen subida, pero esta se guardará temporalmente para poder ser procesada por el modelo.

Equidad y accesibilidad

Es esencial garantizar que las herramientas basadas en inteligencia artificial estén disponibles para todas las comunidades, independientemente de su nivel socioeconómico. Pero se entiende que el contexto es complejo y aunque esta herramienta sea de libre uso mediante internet, el acceso a este y otros servicios está fuera del alcance y limitaciones de este proyecto. El proyecto busca facilitar el acceso, más no lo garantiza.

Dependencia de instituciones locales

La aplicación del modelo en un conjunto de imágenes específico de pacientes guatemaltecos depende de la colaboración con instituciones oftalmológicas del país, así como de la conclusión exitosa de alianzas entre las instituciones y la universidad. Sin estas colaboraciones, la validación del modelo en pacientes locales no sería factible.

Dependencia de servicios de visualización

La visualización y utilización del modelo a través de aplicaciones como Streamlit depende

de la permanencia de las políticas de uso gratuito de sus servicios. En caso de que estas políticas cambien, el *dashboard* de visualización ya no sería utilizable sin incurrir en costos adicionales. Además, la visualización solo estará disponible si ya se cuenta con la imagen de fondo de ojo del paciente, lo que puede limitar el uso del modelo en situaciones donde el acceso a la imagen es limitado o no está disponible.

Dependencia de acceso al modelo

El modelo se encuentra con licencia de código abierto en un repositorio público. Pero la página de acceso al modelo mediante un *dashboard* interactivo depende del costo adquirido mediante una interfaz de programación de aplicaciones (API, por sus siglas en inglés) en Azure ML. Por lo que la superación del presupuesto corresponde con una la desactivación del mismo, y únicamente se podrá acceder incurriendo en costos extra o mediante el repositorio correspondiente.

5.1. Objetivo general

Desarrollar un modelo de clasificación de imágenes de fondo de ojo basado en inteligencia artificial para la detección temprana de anomalías oculares en pacientes de Guatemala, contribuyendo a un diagnóstico preliminar.

5.2. Objetivos específicos

- Implementar un modelo de IA para detectar indicios de anomalías oculares en imágenes de fondo de ojo.
- Comparar la eficacia de cuatro algoritmos de *deep learning* basados en CNN y MDFSC, utilizando métricas como AUC, matriz de confusión, sensibilidad, especificidad, precisión y F1-score, con el fin de optimizar la precisión del modelo y minimizar los falsos negativos.
- Desarrollar un *dashboard* gratuito en Streamlit utilizando Python, que permita a clínicas, hospitales y oftalmólogos cargar imágenes y obtener un diagnóstico preliminar.
- Diseñar un *pipeline* para poder incorporar nuevas fuentes de datos al modelo mediante *transfer learning*.

6.1. Conceptos básicos de oftalmología

La oftalmología es la rama de la medicina que se encarga del estudio y tratamiento de las enfermedades del ojo [15]. La oftalmología ha evolucionado en el tiempo. La primera constancia de la misma viene desde el periodo romano, siendo este entre 218 a. c. y el 507 d. c. Desde el principio de los tiempos el hombre ha tenido la necesidad de salvaguardar su salud general, incluyendo su capacidad de visión. La especialidad de la medicina como tal, llegando al nivel de la oftalmología se vió desde el siglo XIX. Siendo una de las especializaciones mejor consolidadas de la medicina [16].

Para abordar correctamente las patologías oculares, es fundamental comprender en detalle la anatomía y fisiología del ojo humano, así como las técnicas diagnósticas utilizadas en el campo oftalmológico.

Las técnicas diagnósticas en oftalmología han avanzado significativamente, permitiendo una detección temprana y precisa de diversas condiciones oculares. Métodos como la tomografía de coherencia óptica (OCT, por sus siglas en inglés), la angiografía fluorescénica, y las pruebas de campo visual, entre otros, son herramientas cruciales en la evaluación de la salud ocular. Estas técnicas permiten a los oftalmólogos visualizar y medir con precisión las diferentes partes del ojo, ayudando en la planificación de tratamientos efectivos y personalizados para cada paciente.

6.1.1. Anatomía y fisiología del ojo

El ojo humano es un órgano complejo compuesto por varias estructuras interdependientes. La córnea y el cristalino son responsables de la refracción de la luz. La retina, que contiene fotorreceptores (bastones y conos), convierte la luz en señales eléctricas. El nervio óptico transmite estas señales al cerebro, donde se procesan como imágenes visuales. La comprensión detallada de estas estructuras es esencial para identificar y tratar enfermedades

oculares.

El ojo no es una esfera simple, sino una combinación de una pequeña parte de una esfera fuertemente curvada y una gran parte de una esfera menos curvada. La porción pequeña y transparente se llama córnea, mientras que el resto, que es opaco y conocido como esclera. La unión de estas dos áreas se denomina limbo. Al mirar de frente el ojo, se ve la esclera blanca rodeando la córnea, que permite ver el iris, el cual determina el color del ojo, y en el centro, la pupila, que parece oscura debido a la poca reflexión de la luz. Utilizando un oftalmoscopio, se puede iluminar y observar el interior del ojo, conocido como el fondo de ojo, donde destacan grandes vasos sanguíneos que atraviesan el disco óptico. Está compuesto por tres capas que contienen el humor acuoso, el cristalino y el cuerpo vítreo. La capa externa incluye la córnea y la esclera; la capa media contiene el principal suministro de sangre y consiste en la coroides, el cuerpo ciliar y el iris; y la capa interna es la retina. Dentro de estas capas se encuentran el cristalino, el humor acuoso y el cuerpo vítreo. La cámara anterior del ojo está entre la córnea y el iris, mientras que la cámara posterior está entre el iris y el cuerpo ciliar. Ambas cámaras contienen humor acuoso y están conectadas a través de la pupila [17].

Las técnicas diagnósticas en oftalmología han avanzado significativamente, permitiendo una detección temprana y precisa de diversas condiciones oculares. Métodos como la OCT, la angiografía fluoresceínica, y las pruebas de campo visual, entre otros, son herramientas cruciales en la evaluación de la salud ocular. Estas técnicas permiten a los oftalmólogos visualizar y medir con precisión las diferentes partes del ojo, ayudando en la planificación de tratamientos efectivos y personalizados para cada paciente [18].

6.1.2. Enfermedades oculares comunes

Las enfermedades oculares más comunes incluyen cataratas, glaucoma y retinopatía diabética. Las cataratas son opacidades del cristalino que afectan la visión. El glaucoma es una neuropatía óptica caracterizada por la pérdida de células ganglionares y puede llevar a la ceguera si no se trata. La retinopatía diabética es una complicación de la diabetes que afecta los vasos sanguíneos de la retina y es una causa importante de ceguera en adultos.

En Guatemala, la retinopatía diabética representa una preocupación creciente en el ámbito de la salud pública, especialmente debido al incremento sostenido en la prevalencia de diabetes mellitus. Esta complicación ocular afecta considerablemente a la población diabética del país. Un estudio descriptivo realizado en el Hospital Nacional de Chiquimula entre octubre de 2015 y marzo de 2016 evaluó a 169 pacientes con diagnóstico previo de diabetes, encontrando que más del 50 % presentaba algún grado de retinopatía diabética. La mayoría de los casos se observaron en pacientes mayores de 50 años, con al menos cinco años de evolución de la enfermedad, lo cual evidencia una correlación significativa entre el tiempo de diagnóstico y la aparición de complicaciones oculares. Estos datos reflejan una alta prevalencia de RD en poblaciones del oriente del país y subrayan la necesidad de estrategias de tamizaje visual, control glucémico riguroso y educación continua para prevenir la progresión de esta patología y evitar la discapacidad visual asociada [19].

El estudio realizado por Kaminska et al. en Polonia, destaca la importancia del conocimiento público sobre las enfermedades oculares para fomentar comportamientos que favorezcan la prevención y el uso de servicios de atención ocular. El estudio reveló que la mayoría

de los encuestados había oído hablar de cataratas (83.6 %) y glaucoma (80.7 %), mientras que la conciencia sobre la retinopatía diabética era significativamente menor (16.4 %). Además, se identificaron factores de riesgo modificables para estas enfermedades, como el uso de tabaco, la obesidad y la exposición excesiva a la luz solar, que pueden influir negativamente en la salud ocular. Estos hallazgos subrayan la necesidad de campañas educativas personalizadas que aumenten el conocimiento sobre las enfermedades oculares y promuevan comportamientos preventivos, especialmente en poblaciones con menor nivel de conciencia sobre estas condiciones [5].

Una investigación nacional sobre ceguera y deficiencia visual en Guatemala, realizada en 2015, ofreció datos importantes sobre la prevalencia de enfermedades oculares en adultos mayores. Los resultados indicaron que la prevalencia ajustada por edad y sexo de ceguera en Guatemala fue del 2.9 % (intervalo de confianza del 95 %, 2.0 %-3.8 %). La deficiencia visual severa afectó al 5.2 % (4.0 %-6.4 %) de los participantes, mientras que la deficiencia visual moderada afectó al 27.6 % (23.3 %-32.0 %). Las cataratas fueron la causa principal de ceguera (77.6 %) y de deficiencia visual severa (79.4 %), seguidas de otras enfermedades del segmento posterior del ojo y de opacidades corneales no tracomatosas. Los errores refractivos no corregidos fueron la principal causa de deficiencia visual moderada (67.9 %). Este estudio subraya la necesidad urgente de aumentar la disponibilidad de tratamientos de cataratas que sean asequibles y de alta calidad, ya que el costo fue la barrera principal para la cirugía de cataratas (56.7 %). Mejorar el acceso a estos tratamientos podría tener un impacto significativo en la prevención de la ceguera en Guatemala [9].

6.1.3. Clasificación internacional de severidad de la retinopatía diabética (ICDRS)

La clasificación internacional de severidad de la retinopatía diabética y del edema macular (ICDRS, por sus siglas en inglés) fue propuesta en 2003 por un comité de expertos para unificar y simplificar el diagnóstico de la retinopatía diabética (RD) en contextos clínicos y epidemiológicos. Esta clasificación permite una valoración estandarizada de la enfermedad a partir de hallazgos clínicos obtenidos por oftalmoscopia o retinografía, dividiendo la RD en cinco niveles:

- **Sin retinopatía:** no hay signos visibles de RD.
- **RD no proliferativa leve:** únicamente presencia de microaneurismas.
- **RD no proliferativa moderada:** hallazgos más significativos que microaneurismas pero sin criterios para RD severa.
- **RD no proliferativa severa:** presencia de al menos una de las siguientes condiciones:
 - Hemorragias intrarretinianas en cuatro cuadrantes.
 - Arrosariamiento venoso en dos o más cuadrantes.
 - Anomalías microvasculares intrarretinianas (IRMAs) prominentes en al menos un cuadrante.

- **RD proliferativa:** presencia de neovascularización y/o hemorragia vítrea o prerretiniana.

Además, este sistema se complementa con una escala para el edema macular diabético (EMD), clasificado en leve, moderado y severo, según la localización y la extensión del engrosamiento retiniano [20].

6.1.4. Clasificación de la retinopatía diabética según la organización mundial de la salud (OMS)

La organización mundial de la salud (OMS) ha propuesto una clasificación funcional de la retinopatía diabética (RD) orientada a programas de salud pública y estudios epidemiológicos. A diferencia de otras escalas clínicas, esta clasificación se centra en el impacto visual y el grado de afectación retiniana observado durante el tamizaje o las campañas de detección.

Los grados establecidos por la OMS son los siguientes:

- **Grado 0:** retina normal, sin signos de RD.
- **Grado 1 (leve):** presencia de microaneurismas y hemorragias puntiformes sin afectación visual.
- **Grado 2 (moderada):** hemorragias más extensas y aparición de exudados duros.
- **Grado 3 (grave):** signos avanzados de RD no proliferativa o presencia de signos tempranos de proliferación.
- **Grado 4 (muy grave o ceguera funcional):** presencia de neovascularización, hemorragias vítreas o desprendimiento de retina.

Esta clasificación permite estimar el grado de pérdida funcional asociada a la retinopatía y es especialmente útil en zonas con recursos limitados donde el acceso a equipos de diagnóstico avanzados es restringido.

[21]

6.1.5. Clasificación ETDRS de la retinopatía diabética [22]

6.1.5.1. Retinopatía diabética no proliferativa (RDNP)

RDNP leve características: presencia de microaneurismas únicamente [23].

RDNP moderada características: presencia de más que microaneurismas, pero menos que la RDNP severa. Incluye microaneurismas, exudados duros, hemorragias intrarretinianas, y leves signos de dilatación venosa o anomalías microvasculares intrarretinianas (IRMA) [22].

RDNP severa características: uno de los siguientes (criterios de los 4-2-1):

- hemorragias intrarretinianas en los cuatro cuadrantes.
- IRMA en al menos un cuadrante.
- dilatación venosa en al menos dos cuadrantes.

6.1.5.2. Retinopatía diabética proliferativa (RDP)

RDP no de alto riesgo características: neovascularización que no cumple con los criterios de alto riesgo [23].

RDP de alto riesgo características: neovascularización de disco óptico (NVD) mayor o igual al 1/4 del área del disco óptico, cualquier neovascularización con hemorragia vítrea o prerretinal, o neovascularización en otros lugares (NVE) con hemorragia vítrea o prerretinal [22].

6.1.5.3. Edema macular diabético (EMD)

EMD leve características: engrosamiento retiniano o exudados duros en la mácula, pero lejos del centro de la fovea [23].

EMD moderado características: engrosamiento retiniano o exudados duros cercanos, pero no involucrando el centro de la fovea [22].

EMD severo características: engrosamiento retiniano que involucra el centro de la fovea [23].

6.1.6. Descripción de signos claves

- **Microaneurismas:** pequeños puntos rojos en la retina debido a la dilatación de los capilares.
- **Hemorragias intrarretinianas:** puntos o manchas rojas causadas por el sangrado de los capilares.
- **Exudados duros:** depósitos de lípidos que aparecen como manchas amarillas brillantes.
- **IRMA (anormalidades microvasculares intrarretinianas):** vasos sanguíneos anormales en la retina.
- **Dilatación venosa:** ensanchamiento de las venas retinianas.
- **Neovascularización:** crecimiento de nuevos vasos sanguíneos anormales en la retina o en la superficie del disco óptico.
- **Hemorragia vítrea o prerretinal:** sangrado en el vítreo o delante de la retina.

La retinopatía diabética presenta diversas manifestaciones en la retina, cada una indicativa de un posible grado de daño ocular. Los microaneurismas son los primeros signos, visibles como puntos rojos debido a la dilatación capilar. A medida que avanza la enfermedad, pueden aparecer hemorragias intraretinianas y exudados duros, que son depósitos de lípidos visibles como manchas amarillas. También pueden observarse anomalías microvasculares intraretinianas (IRMA) y dilatación venosa, que reflejan un compromiso en la circulación sanguínea. En etapas más avanzadas, la neovascularización genera vasos anormales que crecen en la retina y el disco óptico, mientras que hemorragias vítreas o prerretinales indican sangrados internos que pueden afectar gravemente la visión [22].

6.1.7. Importancia de la clasificación

La clasificación ETDRS ayuda a los oftalmólogos a evaluar la gravedad de la retinopatía diabética, a planificar el seguimiento y tratamiento adecuado, y a prevenir la progresión de la enfermedad y la pérdida de visión. Es crucial para la toma de decisiones clínicas sobre intervenciones como la fotocoagulación con láser y la terapia con medicamentos intravítreos [23].

Y estas diferencias en una imagen de fondo de ojo sana y una que presenta estas anomalías previamente descritas es lo que podemos observar en la Figura 1.

Una forma estándar de evaluar la agudeza visual es mediante la notación fraccionaria como 20/20 o 20/400. En esta escala, el primer número (20) representa la distancia a la que el paciente se encuentra del cartel optométrico, típicamente 20 pies (aproximadamente 6 metros), mientras que el segundo número indica la distancia a la que una persona con visión normal puede leer la misma línea. Por ejemplo, una persona con visión 20/20 puede leer a 20 pies lo que una persona promedio también puede leer a esa distancia, mientras que alguien con visión 20/400 solo puede leer a 20 pies lo que una persona con visión normal puede leer a 400 pies. Este sistema permite cuantificar el grado de pérdida visual de forma funcional, siendo 20/200 o peor uno de los criterios para ceguera legal en muchos países [24].

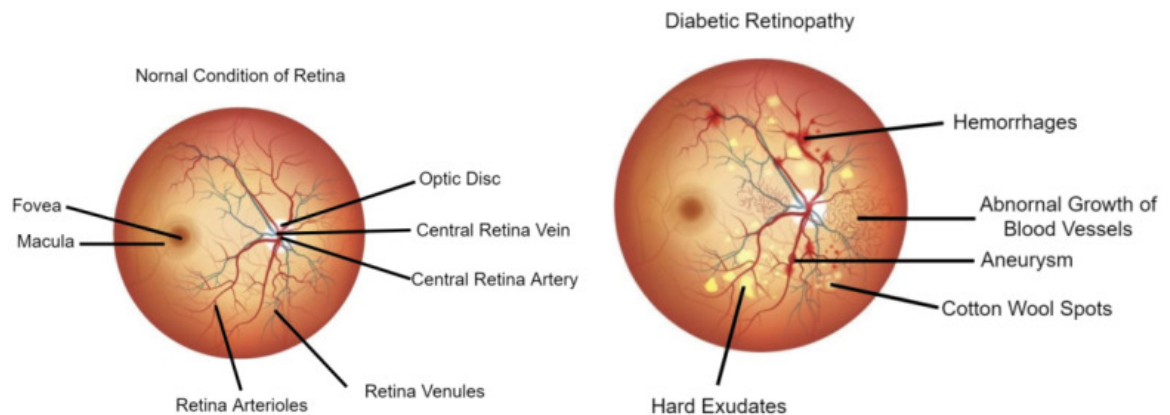


Figura 1: Visualización de una retina sana y enferma [25]

6.1.7.1. Causas y factores de riesgo

La retinopatía diabética se debe a cambios en los vasos sanguíneos de la retina causados por la diabetes. Estos cambios, junto con el tiempo que una persona ha tenido diabetes, aumentan el riesgo de desarrollar esta condición. Otros factores de riesgo incluyen fumar, presión arterial alta y embarazo en mujeres con diabetes [26]. El principal factor de riesgo para desarrollar retinopatía diabética es la diabetes mellitus, una enfermedad metabólica crónica caracterizada por hiperglucemia persistente debido a defectos en la secreción o acción de la insulina. La hiperglucemia prolongada causa daño progresivo a los vasos sanguíneos de la retina, lo que con el tiempo conduce a microaneurismas, hemorragias, exudados y eventualmente neovascularización patológica [27].

El riesgo de presentar retinopatía diabética aumenta con la duración de la diabetes, especialmente si no se mantiene un control glicémico adecuado. Otros factores que exacerban este riesgo incluyen la hipertensión arterial, dislipidemias, el tabaquismo, el embarazo en mujeres con diabetes, y la nefropatía diabética. La detección temprana de estos factores y su adecuada gestión es esencial para prevenir la aparición o progresión de la enfermedad ocular diabética [28].

6.1.7.2. Síntomas

En las primeras etapas, la retinopatía diabética puede no mostrar síntomas. Conforme la enfermedad avanza, se pueden presentar visión borrosa, visión doble, manchas flotantes, dolores en uno o ambos ojos, y pérdida parcial o total de la visión [26].

La retinopatía diabética es una enfermedad insidiosa, en sus etapas iniciales puede cursar de forma asintomática. A medida que progresa, el paciente puede experimentar alteraciones visuales de diversa gravedad. En el Cuadro 1 se describen los síntomas más comunes según el estadio de la enfermedad, basada en la descripción clínica estandarizada de la retinopatía diabética según Kanski [29].

Cuadro 1: Síntomas de la retinopatía diabética según su estadio

Estadio	Síntomas comunes
Etapas temprana (RDNP leve)	Generalmente asintomática.
Etapas intermedia (RDNP moderada-severa)	Visión borrosa, dificultad para enfocar, aparición de manchas flotantes (miodesopsias).
Etapas avanzadas (RDP)	Pérdida súbita de visión, visión distorsionada, hemorragia vítrea, dificultad para percibir colores o detalles finos.

6.1.7.3. Diagnóstico y tratamiento

Para diagnosticar la retinopatía diabética, se utilizan exámenes visuales como la prueba de agudeza visual, tonometría (para medir la presión ocular), dilatación de pupila, oftal-

moscopia, angiografía con fluoresceína y tomografía de coherencia óptica. El tratamiento depende de la gravedad de la enfermedad e incluye:

- **Cirugía láser:** utilizada para reducir o sellar los vasos anormales.
- **Vitrectomía:** extracción del humor vítreo para tratar hemorragias o tejido cicatricial.
- **Inyecciones:** medicamentos para inhibir el crecimiento de vasos anormales o reducir el edema macular.

Estos tratamientos buscan abordar las complicaciones de la retinopatía diabética y preservar la visión en etapas avanzadas de la enfermedad. Cada opción se adapta a las necesidades específicas del paciente y el estado de progresión de la retinopatía [26].

El diagnóstico oportuno es clave en la gestión de la retinopatía diabética. Existen múltiples métodos diagnósticos por imagen para evaluar el estado de la retina. La angiografía con fluoresceína permite observar con gran detalle la perfusión vascular retiniana, pero es un procedimiento más invasivo y costoso. En contraste, la retinografía o imágenes de fondo de ojo son más accesibles, menos invasivas y ampliamente utilizadas en programas de tamizaje, lo que permite, con ayuda de modelos de inteligencia artificial, estratificar el riesgo de progresión de la enfermedad [30].

El tratamiento depende del grado de afectación y puede dividirse según se detalla en el Cuadro 2.

Cuadro 2: Tratamientos comunes para la retinopatía diabética

Modalidad terapéutica	Descripción
Fotocoagulación con láser	Sella o reduce vasos sanguíneos anómalos, eficaz en RDNP severa y RDP.
Inyecciones intravítreas	Anti-VEGF o corticosteroides que reducen el edema macular y neovascularización.
Vitrectomía	Cirugía para remover el humor vítreo y tejido fibroso en casos avanzados con hemorragia vítrea o desprendimiento de retina.

El tratamiento es personalizado, dependiendo de la severidad y tipo de retinopatía diabética, y tiene como objetivo preservar la visión y prevenir complicaciones mayores.

6.1.7.4. Prevención

Si bien no siempre es posible prevenir la retinopatía diabética, el control adecuado de los niveles de azúcar en sangre, la presión arterial y los niveles de colesterol, junto con exámenes oculares regulares, pueden reducir significativamente el riesgo de desarrollar esta condición o evitar su progresión [26].

La prevención de la retinopatía diabética se basa en el control estricto de los factores sistémicos de riesgo, especialmente el nivel de glucosa en sangre, presión arterial y lípidos.

Sin embargo, incluso con un buen control metabólico, la enfermedad puede desarrollarse. Por ello, la detección temprana mediante exámenes periódicos de fondo de ojo es esencial.

Las imágenes de retina constituyen una herramienta diagnóstica accesible y no invasiva, ampliamente utilizada en campañas de tamizaje. Con el apoyo de tecnologías basadas en aprendizaje automático, es posible automatizar la detección y estratificación de pacientes en riesgo, optimizando el uso de recursos en contextos con acceso limitado a especialistas [31].

6.2. Imágenes de fondo de ojo

El fondo de ojo nos referimos a la superficie interior del ojo que se encuentra frente al cristallino e incluye la retina, el disco óptico, la mácula, la fovea y el polo posterior. Este se puede examinar a través de oftalmoscopia y/o fotografía del fondo de ojo. Este procedimiento permite observar diversas estructuras internas del globo ocular, como la mácula, la retina y la papila óptica, además de posibilitar la visualización directa de los vasos sanguíneos de la retina y detectar cualquier anomalía en ellos [32].

Las imágenes de fondo de ojo son esenciales para el diagnóstico de diversas patologías oculares. Estas imágenes proporcionan una vista detallada de la retina, el nervio óptico y otras estructuras internas del ojo, permitiendo la detección de anomalías y enfermedades en etapas tempranas.

6.2.1. Tipos de imágenes de fondo de ojo

Existen varios tipos de imágenes de fondo de ojo, cada una con aplicaciones específicas en el diagnóstico oftalmológico:

- **Tomografía de coherencia óptica (OCT, por sus siglas en inglés)**
 - **Ventajas**
 - No invasiva y fácil de interpretar.
 - Muy útil para medir el grosor de la retina y detectar acumulación de líquidos en enfermedades como el edema macular diabético (DME, por sus siglas en inglés) y la AMD.
 - Proporciona imágenes de alta resolución, revelando niveles de isquemia macular no visibles con métodos anteriores.
 - **Desventajas**
 - No identifica bien la presencia de sangre en la retina.
 - **Evolución**
 - Evolución hacia OCT de fuente de barrido (SS-OCT, por sus siglas en inglés), que permite una penetración más profunda y una adquisición más rápida de imágenes, mejorando la comprensión de condiciones como la coroidopatía serosa central.
- **Fotografía de fondo de ojo**

- **Ventajas**
 - Ampliamente disponible y fácil de usar.
 - Documenta bien la apariencia del nervio óptico y la presencia de sangre en el ojo.
- **Desventajas**
 - No es comúnmente utilizada para la toma de decisiones en tratamientos.
- **Evolución**
 - Incorporación de capacidades mejoradas para crear montajes de fotografías en color del polo posterior y la periferia utilizando software automatizado.
- **Angiografía con fluoresceína (FA, por sus siglas en inglés)**
 - **Ventajas**
 - Buena para encontrar lesiones focales que requieren tratamiento con láser.
 - Herramienta básica para el diagnóstico de trastornos coroideos y degeneración macular neovascular.
 - **Desventajas**
 - Requiere la inyección de un colorante fluorescente, lo cual puede ser invasivo.
 - **Evolución**
 - Evolución hacia la angiografía de campo amplio, que captura un campo de visión de 100 a 200 grados, revelando patologías que no pueden identificarse con la angiografía estándar.
- **Angiografía con verde de indocianina (ICGA, por sus siglas en inglés)**
 - **Ventajas**
 - Permite una mejor penetración de la RPE y la sangre debido a la mayor longitud de onda utilizada.
 - Complementa la FA al capturar imágenes de la circulación coroidea debajo del nivel de la RPE.
 - **Desventajas**
 - Menos uso clínico en comparación con la FA.
- **Autofluorescencia del fondo de ojo (FAF, por sus siglas en inglés)**
 - **Ventajas**
 - Revela la salud de la RPE y la acumulación de lipofuscina, útil en enfermedades maculares hereditarias y en el monitoreo de toxicidades por medicamentos.
 - **Desventajas**
 - Aún no está ampliamente en uso, aunque ha ganado tracción en la última década.

Cada tipo de imagen de fondo de ojo presenta ventajas y desventajas específicas que las hacen adecuadas para diferentes aplicaciones en el diagnóstico oftalmológico. Mientras que algunas técnicas como la OCT y la FA son fundamentales para detectar y monitorear enfermedades comunes, otras como la ICGA y la FAF proporcionan información adicional valiosa en casos más complejos. La evolución de estas tecnologías continúa mejorando la precisión y eficacia del diagnóstico, aunque también plantea desafíos como la necesidad de capacitación especializada y el manejo de datos complejos [33].

6.2.2. Técnicas y equipos de captura

Las técnicas y equipos utilizados para la captura de imágenes de fondo de ojo varían en función de la patología a detectar y la precisión requerida. Los equipos modernos incluyen cámaras de alta resolución y dispositivos OCT que permiten una visualización detallada de las estructuras internas del ojo. Se habla que la tendencia es que los fabricantes buscan un dispositivo con muchas funcionalidades y poder centralizar todo en un solo dispositivo [33].

6.2.3. Calidad de las imágenes

La calidad de las imágenes de fondo de ojo es crucial para un diagnóstico preciso. Factores como la resolución, el enfoque y la iluminación deben ser cuidadosamente controlados durante la captura de imágenes. La normalización de las imágenes y la reducción del ruido son pasos importantes en el preprocesamiento de los datos, así como el tratamiento del color. Ya que son características esenciales para poder brindar un buen diagnóstico [33].

6.3. Diagnóstico de anomalías oculares

El diagnóstico de anomalías oculares requiere una evaluación exhaustiva que combine historia clínica, examen físico y pruebas especializadas. Los oftalmólogos comienzan con una entrevista detallada para identificar síntomas, antecedentes familiares y factores de riesgo. Posteriormente, realizan un examen físico que incluye la evaluación de la agudeza visual, la inspección del fondo de ojo con un oftalmoscopio, y la medición de la presión intraocular. Pruebas complementarias como la tomografía de coherencia óptica (OCT), la angiografía con fluoresceína y el análisis de campo visual permiten una evaluación más precisa de estructuras internas del ojo y la detección de patologías como la degeneración macular, el glaucoma y las retinopatías. La combinación de estos métodos diagnósticos ayuda a los médicos a identificar anomalías oculares de manera eficaz y a desarrollar un plan de tratamiento adecuado.

Para la clasificación de retinopatía diabética y edema macular diabético debemos saber que esta anomalía está ligada a la diabetes. Es frecuente causa de ceguera entre pacientes de 20 a 64 años. Y su clasificación está basada en varios hallazgos clínicos. Su clasificación depende de la gravedad y avance de la enfermedad. Según un estudio epidemiológico de Wisconsin sobre la misma, al utilizar la fotografía de fondo de ojo, medida de niveles de hemoglobina glicosilada y tomas de agudeza visual corregida, se puede establecer la progresión de la diabetes. La causa es desconocida, pero lo que se cree según la comunidad de medicina

es que se debe a la exposición prolongada a hiperglucemia lo que provoca un daño del endotelio vascular. Cambios específicos en los capilares retinianos incluyen pérdida selectiva de pericitos y engrosamiento de la membrana basal, cambios que favorecen la oclusión vascular y la mala perfusión retiniana, como también la descompensación de la barrera funcional endotelial, la cual permite una fuga del contenido intravascular y genera edema retiniano. Lo recomendado es referirse a la escala internacional de severidad de la retinopatía diabética. Esto lo podemos verificar en el anexo.

6.4. Inteligencia artificial y aprendizaje automático

La inteligencia artificial (IA) y el aprendizaje automático (ML) están revolucionando el campo de la medicina, ofreciendo nuevas herramientas para el análisis y diagnóstico de enfermedades mediante el procesamiento y análisis de grandes volúmenes de datos.

Se ha visto como el uso de inteligencia artificial ha trascendido, pasando de ser utilizada en juegos como ajedrez o cartas, a poder resolver problemas complejos llevando a un diagnóstico médico, que ha llegado a igualar o incluso superar al de un especialista. En el ámbito de la oftalmología, específicamente en cuestión de retinopatía diabética, la FDA ya ha aprobado dos modelos de IA para el uso de diagnósticos clínicos. Que son IDx [34] y EyeArt [35], ya que han demostrado su habilidad para poder diferenciar esta anomalía.

Se ha hablado sobre lo que se le llama Oftalmología 2.0. Y lo que se puede llegar a esperar en el campo con la incorporación de lleno de estas tecnologías y el refinamiento de equipos y conocimientos por parte de los especialistas. Llegando incluso a predecir mediante información genética o metabólica para poder descubrir y diferenciar pacientes con condiciones degenerativas, mucho antes que estas sean visibles.

Esto involucra de igual manera un nuevo reto, el cual es crear un continuo aprendizaje y familiarización con estas herramientas del lado de la analítica hacia el lado médico. Pudiendo mejorar cada vez más esta fusión entre dos áreas muy importantes para lograr estos avances y objetivos. Poder entender qué es lo que se está haciendo, cómo se está haciendo y poder apoyar a la ingeniería inversa sobre el modelo, para poder mejorar el proceso de aprendizaje de la computadora de una mejor manera. Esto ayudará a que el diagnóstico sea mucho más costo efectivo y eficiente, llevando a que la salud ocular sea más accesible a la población.

La IA tiene la capacidad de mejorar la atención, hacer que los médicos sean mejores observadores clínicos y permitirles tomar decisiones médicas más rentables en la asignación de cuidados. También existe la posibilidad de que la Oftalmología con IA v2.0 trabaje para ayudar a nuestros colegas en el manejo de otras enfermedades neurodegenerativas y neuropsiquiátricas, como la enfermedad de Alzheimer y la esquizofrenia, ya que la retina es un aspecto accesible del sistema nervioso central [36].

6.4.1. Conceptos básicos de IA y ML

La inteligencia artificial se refiere a la capacidad de las máquinas para realizar tareas que normalmente requieren inteligencia humana, como el reconocimiento de patrones y la

toma de decisiones. El aprendizaje automático es una subdisciplina de la IA que se enfoca en la creación de algoritmos que permiten a las máquinas aprender de los datos y mejorar su rendimiento con el tiempo. Personalmente podría decir que la *inteligencia* se refiere al poder aprender

Se debe entender que *Machine learning* se refiere al proceso en el que la computadora aprende sobre experiencias pasadas, es decir datos que tiene acceso a poder ver, para poder predecir fuera de ese conjunto de datos hacia el futuro. Al proporcionar a los algoritmos de aprendizaje automático una gran cantidad de datos, podemos entrenarlos para examinar los datos, construir modelos y anticipar automáticamente el resultado deseado. La función de costo se puede usar para estimar qué tan bien funciona el algoritmo de aprendizaje automático dada la cantidad de datos. Podemos ahorrar tiempo y dinero utilizando el aprendizaje automático. Sus casos de uso hacen que sea fácil entender la importancia del aprendizaje automático. Actualmente, el aprendizaje automático se emplea en autos autónomos, detección de fraude cibernético, reconocimiento facial y sugerencias de amigos en Facebook, entre otras aplicaciones. Varias empresas líderes, incluidas Netflix y Amazon, han desarrollado algoritmos de aprendizaje automático que utilizan una gran cantidad de datos para analizar el interés de los clientes y hacer recomendaciones de productos. [37]

Tipos de modelos de clasificación

existen varios tipos de modelos de clasificación, cada uno con sus propias características y aplicaciones:

- Clasificación binaria: $Y = 0, 1$ o $Y = -1, 1$
- Clasificación multiclase: $Y = 1, 2, \dots, K$, donde $K > 2$
- Clasificación multilabel: cada instancia puede pertenecer a múltiples clases simultáneamente

Cada uno depende del tipo de problema y objetivo de nuestro modelo de predicción. Así como de los datos disponibles y su etiqueta [38].

6.4.2. Tipos de aprendizaje automático

El aprendizaje automático se divide en tres categorías principales:

- **Aprendizaje supervisado** el modelo se entrena con datos etiquetados, donde se conoce la salida deseada, como la clasificación de imágenes médicas.
- **Aprendizaje no supervisado:** el modelo busca patrones en datos no etiquetados, como la segmentación de imágenes médicas.
- **Aprendizaje por refuerzo:** el modelo aprende mediante la interacción con el entorno, optimizando una función de recompensa, útil en la optimización de procesos médicos.

Cada tipo de aprendizaje ofrece enfoques únicos para el procesamiento de datos médicos, facilitando la clasificación, segmentación y optimización de procesos en diversas aplicaciones clínicas. La elección del enfoque depende del objetivo específico del modelo y de la disponibilidad de datos etiquetados [37].

6.4.3. Evaluación de modelos de clasificación

En el caso de este estudio nos centraremos en los modelos de clasificación. Debido a la naturaleza del problema, donde el objetivo es poder clasificar una imagen en dos o más categorías.

Un modelo de clasificación es un tipo de modelo predictivo que asigna etiquetas o categorías a datos nuevos basándose en patrones aprendidos de un conjunto de datos de entrenamiento. Se utilizan en diversas áreas como detección de fraudes, diagnóstico médico, reconocimiento de imágenes, entre otros.

6.4.3.1. Matriz de confusión

	Predicción positiva	Predicción negativa
Clase positiva	TP (True positives)	FN (False negatives)
Clase negativa	FP (False positives)	TN (True negatives)

La matriz de confusión es fundamental para evaluar los resultados de un modelo de clasificación. Proporciona una visión detallada de los aciertos y errores del modelo, permitiendo analizar en profundidad la precisión y el error de las predicciones en cada clase.

6.4.3.2. Sensibilidad (recall o tasa de verdaderos positivos)

$$\text{Sensibilidad} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

La sensibilidad es crucial en modelos de clasificación de imágenes, especialmente cuando se trata de detectar correctamente instancias positivas (por ejemplo, identificar imágenes con ciertas características relevantes). Una alta sensibilidad indica que el modelo tiene un buen rendimiento al capturar las verdaderas instancias positivas.

6.4.3.3. Especificidad (tasa de verdaderos negativos)

$$\text{Especificidad} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

La especificidad es importante cuando es crítico minimizar los falsos positivos, es decir, evitar clasificar incorrectamente una imagen negativa como positiva. En contextos como la detección de enfermedades, una alta especificidad es esencial para reducir alarmas falsas.

6.4.3.4. Precisión (precision)

$$\text{Precisión} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

La precisión indica qué tan confiables son las predicciones positivas del modelo. En clasificación de imágenes, una alta precisión asegura que las imágenes clasificadas positivamente realmente pertenecen a la clase positiva, lo cual es crucial para mantener la confianza en el sistema de clasificación.

6.4.3.5. F1-Score

$$F1\text{-Score} = 2 \times \frac{\text{Precisión} \times \text{sensibilidad}}{\text{Precisión} + \text{sensibilidad}}$$

El F1-Score es una métrica balanceada que combina precisión y sensibilidad. Es particularmente útil en escenarios de clasificación de imágenes donde existe un desbalance de clases, ya que proporciona una visión equilibrada del rendimiento del modelo.

6.4.3.6. Área bajo la curva ROC (AUC - ROC)

$$\text{AUC-ROC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x)) dx$$

El AUC-ROC mide la capacidad del modelo para diferenciar entre clases. Un AUC alto indica que el modelo tiene un buen rendimiento general en la clasificación, siendo capaz de distinguir correctamente entre imágenes positivas y negativas.

6.4.3.7. Curva precision-recall

La curva precision-recall muestra la relación entre la precisión y la recuperación a diferentes umbrales de clasificación. Dependiendo del umbral seleccionado, el modelo puede optimizar la precisión o la recuperación, pero a menudo hay un compromiso entre ambas.

6.4.3.8. AUC-PR: área bajo la curva precision-recall

El AUC-PR es el área bajo la curva Precision-Recall y proporciona una métrica única que resume el rendimiento del modelo a través de todos los umbrales posibles.

El valor del AUC-PR varía entre 0 y 1:

- Si $\text{AUC-PR} = 1$, el modelo es perfecto: identifica todos los positivos correctamente sin ningún falso positivo.

- Si $AUC-PR = 0.5$, el modelo es casi aleatorio.
- Un $AUC-PR$ bajo (cercano a 0) indica que el modelo predice peor que un clasificador aleatorio.

6.4.3.9. AUC-PR vs. AUC-ROC

El $AUC-PR$ es preferido sobre el $AUC-ROC$ cuando las clases están desbalanceadas, ya que se enfoca en los casos positivos y no es influenciado por la abundancia de verdaderos negativos como en el caso del $AUC-ROC$ [39].

6.4.3.10. Quadratic weighted kappa

La métrica *Quadratic Weighted Kappa* (QWK) mide el grado de concordancia entre dos clasificadores en tareas de clasificación ordinal, penalizando más los desacuerdos mayores. Se utiliza principalmente cuando las categorías tienen un orden natural.

Sea O la matriz de confusión observada, donde $O_{i,j}$ representa el número de veces que el calificador A ha asignado una instancia a la categoría i mientras que el calificador B la ha asignado a la categoría j . Además, sea E la matriz esperada bajo independencia, donde $E_{i,j}$ es el número esperado de veces que el calificador A asignaría la categoría i y el calificador B asignaría la categoría j si sus distribuciones fueran independientes.

La métrica Kappa ponderada cuadráticamente se define como:

$$\kappa = 1 - \frac{\sum_{i,j} W_{i,j} O_{i,j}}{\sum_{i,j} W_{i,j} E_{i,j}}$$

donde:

$W_{i,j}$ es una matriz de pesos cuadráticos, que penaliza más fuertemente las discrepancias entre las categorías ordinales. Se calcula de la siguiente manera:

$$W_{i,j} = \frac{(i-j)^2}{(N-1)^2}$$

Aquí, N es el número total de categorías.

Cálculo paso a paso

1. **Matriz de confusión observada (O):** se construye una matriz O , donde $O_{i,j}$ representa cuántas veces la categoría i fue asignada por el calificador A y la categoría j fue asignada por el calificador B .
2. **Matriz esperada (E):** la matriz esperada E se calcula utilizando las frecuencias marginales de las calificaciones de A y B , asumiendo que ambas distribuciones son independientes.

3. **Matriz de pesos (W):** se asigna un peso $W_{i,j}$ a cada par (i, j) que penaliza más las discrepancias mayores entre categorías.
4. **Cálculo de kappa:** se compara la matriz observada O con la matriz esperada E , ponderando las discrepancias utilizando la matriz W . El valor de κ varía entre 1 (acuerdo perfecto) y 0 (sin mejor concordancia que el azar).

Con las matrices de confusión observada O y esperada E , el valor de κ se calcula como:

$$\kappa = 1 - \frac{\sum_{i,j} W_{i,j} O_{i,j}}{\sum_{i,j} W_{i,j} E_{i,j}}$$

Quadratic Weighted Kappa penaliza más fuertemente los errores que se alejan más del valor real, haciendo que sea útil para tareas donde el orden de las categorías es importante [40].

Al evaluar un modelo de clasificación de imágenes, es crucial considerar estas métricas porque cada una proporciona una perspectiva diferente sobre el rendimiento del modelo. La matriz de confusión ofrece una visión completa de los aciertos y errores, mientras que métricas como la precisión y la sensibilidad nos permiten comprender la eficacia del modelo en contextos específicos, como la detección de clases minoritarias o la minimización de errores críticos. Si nuestro objetivo principal es reducir los falsos negativos, es decir, asegurar que todas las instancias positivas sean correctamente identificadas, deberíamos priorizar la sensibilidad (recall). Una alta sensibilidad es vital en aplicaciones donde los falsos negativos pueden tener consecuencias graves, como en este caso de detección de anomalías en imágenes de fondo de ojo, donde no detectar una condición podría ser más perjudicial que una falsa alarma. Por tanto, enfocar la optimización del modelo en maximizar la sensibilidad será esencial para cumplir este objetivo [41].

6.4.3.11. Fleiss' Kappa para acuerdo entre múltiples evaluadores

Introducción a Fleiss' Kappa

El coeficiente de Fleiss' Kappa es una medida estadística utilizada para evaluar el grado de acuerdo entre múltiples evaluadores o clasificadores cuando se asignan categorías a un conjunto de elementos. A diferencia de otros coeficientes de Kappa, como el de Cohen, que solo se aplica a dos evaluadores, Fleiss' Kappa es útil para situaciones donde tres o más evaluadores intervienen, lo que lo hace ideal para proyectos donde se busca evaluar la concordancia en estudios de clasificación, etiquetado de datos o diagnósticos médicos [42].

El coeficiente Fleiss' Kappa proporciona una manera de cuantificar el nivel de acuerdo observado entre evaluadores más allá del mero azar, lo que permite diferenciar entre un acuerdo genuino y uno que podría esperarse por casualidad.

Definición formal

Matemáticamente, Fleiss' Kappa se define para una matriz de datos donde cada fila

corresponde a un elemento clasificado, y cada columna representa una categoría o etiqueta que los evaluadores pueden asignar. Supongamos que tenemos:

- n : el número de ítems o elementos a clasificar,
- k : el número de categorías posibles,
- m : el número de evaluadores asignados por ítem.

Cada evaluador asigna una categoría a cada uno de los n ítems. El acuerdo se evalúa mediante dos componentes principales:

1. **Acuerdo observado** (P_o): la proporción de veces que los evaluadores coinciden al clasificar un ítem.
2. **Acuerdo esperado** (P_e): la proporción de acuerdo que se esperaría por azar, basada en la distribución de las categorías.

Fleiss' Kappa está definido como:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

donde P_o es el acuerdo observado y P_e es el acuerdo esperado por azar. Si todos los evaluadores están en total acuerdo, entonces $\kappa = 1$. Si el nivel de acuerdo es igual al esperado por azar, entonces $\kappa = 0$, y si el acuerdo es peor que el azar, κ puede ser negativo.

Cálculo del acuerdo observado (P_o)

El acuerdo observado (P_o) mide la proporción de coincidencias entre los evaluadores más allá del azar. Para calcular P_o , primero se determina el grado de acuerdo entre los evaluadores para cada ítem. Para cada ítem i , se calcula el acuerdo entre evaluadores como:

$$P_o^{(i)} = \frac{1}{m(m-1)} \sum_{j=1}^k n_{ij}(n_{ij} - 1)$$

donde n_{ij} es el número de evaluadores que asignaron la categoría j al ítem i , y m es el número de evaluadores. El término $P_o^{(i)}$ representa el grado de acuerdo para el ítem i .

El valor total de P_o se calcula promediando los $P_o^{(i)}$ sobre todos los ítems:

$$P_o = \frac{1}{n} \sum_{i=1}^n P_o^{(i)}$$

Cálculo del acuerdo esperado (P_e)

El acuerdo esperado por azar P_e representa la probabilidad de que dos evaluadores coincidan en una categoría simplemente por azar. Para calcular P_e , primero se determina la proporción total de asignaciones a cada categoría j en todos los ítems, denotado como p_j :

$$p_j = \frac{1}{n \cdot m} \sum_{i=1}^n n_{ij}$$

donde n_{ij} es el número de veces que la categoría j fue asignada al ítem i .

El acuerdo esperado P_e se calcula como la suma de los cuadrados de las proporciones de cada categoría:

$$P_e = \sum_{j=1}^k p_j^2$$

Interpretación de Fleiss' Kappa

El valor de Fleiss' Kappa permite interpretar el grado de acuerdo entre los evaluadores:

- $\kappa = 1$: indica un acuerdo perfecto entre los evaluadores.
- $\kappa = 0$: indica que el nivel de acuerdo observado es equivalente al acuerdo esperado por azar.
- $\kappa < 0$: indica un acuerdo peor que el esperado por azar.

Según la convención de Landis y Koch, los valores de κ pueden interpretarse de la siguiente manera:

- $\kappa < 0$: Pobre.
- $0.01 \leq \kappa \leq 0.20$: ligero.
- $0.21 \leq \kappa \leq 0.40$: moderado.
- $0.41 \leq \kappa \leq 0.60$: aceptable.
- $0.61 \leq \kappa \leq 0.80$: sustancial.
- $0.81 \leq \kappa \leq 1.00$: casi perfecto.

Ventajas y limitaciones

Una de las principales ventajas de Fleiss' Kappa es que permite medir el acuerdo entre múltiples evaluadores, lo que lo hace aplicable a una amplia gama de estudios. Además, es una medida que tiene en cuenta el acuerdo más allá del azar, lo que le otorga robustez en la evaluación de clasificaciones.

Sin embargo, Fleiss' Kappa tiene ciertas limitaciones. Por ejemplo, el coeficiente puede verse afectado por la prevalencia de las categorías. Si una categoría es mucho más común que las demás, el valor de κ puede subestimar el nivel de acuerdo entre los evaluadores. Además, es importante tener en cuenta que esta medida supone que todos los evaluadores asignan la misma cantidad de ítems, lo que puede no ser el caso en situaciones reales.

Para garantizar la consistencia y fiabilidad de las clasificaciones, se aplica el coeficiente Fleiss Kappa, un coeficiente para evaluar el acuerdo interobservador entre múltiples médicos en la interpretación de las imágenes. Este evalúa qué tan consistentes son las calificaciones obtenidas en pruebas que involucran a más de dos observadores. Un valor de K por debajo de 0 indica un acuerdo pobre, mientras que valores entre 0.01 y 0.2 indican un acuerdo leve, entre 0.21 y 0.4 un acuerdo justo, entre 0.41 y 0.6 un acuerdo moderado, entre 0.61 y 0.8 un acuerdo sustancial, y entre 0.81 y 1 un acuerdo casi perfecto [42]. Además, se tiene en cuenta que en cualquier modelo de clasificación se utilizan clases o labels colocados por expertos del tema. Pero estos no siempre son consistentes debido a sesgos expertos inherentes, juicios y errores, entre otros factores, lo que puede llevar a un resultado pobre al predecir por parte del modelo. Sin embargo, un estudio realizado sobre el impacto que estas inconsistencias pueden provocar en los modelos enfocados en tomar decisiones clínicas de diagnóstico obtiene resultados que sugieren que puede que no siempre haya un superexperto en entornos clínicos agudos, y la búsqueda de consenso estándar (como la votación mayoritaria) conduce consistentemente a modelos subóptimos. Sin embargo, evaluar la capacidad de aprendizaje de las anotaciones y usar solo conjuntos de datos anotados 'aprendibles' para determinar el consenso logra modelos óptimos en la mayoría de los casos. En esta investigación se define aprendibilidad como la medida que indica qué tan bien se pueden aprender las asociaciones entre las variables de entrada y las anotaciones proporcionadas, y a su vez qué tan fácilmente se puede reproducir el razonamiento clínico del anotador [43].

6.5. Técnicas de preprocesamiento de imágenes

6.5.1. Conversión de espacio de color

La conversión de espacio de color es esencial en el análisis de imágenes de fondo de ojo. Los modelos RGB y HSI son intercambiables, lo que facilita el procesamiento de la intensidad de los píxeles sin alterar sus valores de color. El modelo HSI separa la intensidad del componente cromático, lo que facilita la mejora del contraste local y permite la conversión de nuevo a RGB sin modificar el color de los píxeles. Además, la normalización del color se realiza combinando los componentes R, G y B normalizados, mientras que la conversión de RGB a YIQ potencia la saturación del color y el contraste entre las lesiones y el fondo. La ecualización adaptativa del histograma en el espacio CMY y la conversión a CIELab permiten separar la intensidad de los demás componentes de color, mejorando así la calidad de la imagen para análisis posteriores. Estos procesos garantizan que el preprocesamiento se enfoque en el fondo circular de la imagen sin afectar los píxeles del fondo [44].

Técnicas adicionales como la ecualización adaptativa del histograma en el espacio HSV, la corrección de brillo en el modelo HSV, y el uso de algoritmos gaussianos y CLAHE, mejoran la claridad y el contraste de las imágenes de fondo de ojo, contribuyendo a diagnósticos más precisos. Comparaciones entre la normalización en escala de grises y la especificación del histograma ayudan a identificar los métodos más efectivos para el análisis de estas imágenes [44].

6.5.2. Corrección de sombras

Las imágenes de fondo de ojo en color a menudo presentan ruido a diferentes niveles. Para abordar esto, se han utilizado técnicas de filtrado de mediana y convolución con núcleos de suavizado. Las pequeñas variaciones de intensidad en el canal verde se eliminan en el paso de preprocesamiento, lo que resulta en una imagen corregida de sombras. Este proceso se realiza restando la imagen de fondo de la imagen del canal verde.

La imagen de fondo se estima mediante el suavizado de la imagen del canal verde con un filtro de mediana. En estudios previos, se ha utilizado un filtro de mediana de 56x56 para obtener una imagen con corrección de sombras. También se han empleado filtros de paso bajo o filtros de media o mediana de mayor tamaño que la característica retiniana más grande para mejorar la corrección de sombras. El objetivo principal de este paso es limpiar el ruido en las imágenes, seguido de un suavizado y normalización como método de corrección de sombras [44].

6.5.3. Aumento de contraste adaptativo

El aumento de contraste adaptativo es una técnica basada en histogramas que optimiza el brillo en áreas oscuras de la imagen sin causar sobreexposición en las áreas brillantes. Mientras que la ecualización de histograma simple puede deteriorar la imagen, el aumento de contraste adaptativo resuelve este problema al dividir la imagen en zonas pequeñas con contraste similar y mejorar cada una de ellas de manera individual.

Técnicas como CLAHE (Equalización Adaptativa de Histograma Limitada en Contraste) y BPDFHE (Ecualización Dinámica de Histograma Borroso Preservadora del Brillo) son comúnmente usadas en procesamiento de imágenes biomédicas. CLAHE, por ejemplo, mejora partes específicas de la imagen al dividirla en regiones disjuntas y aplicar ecualización de histograma local, seguida de la interpolación para eliminar los límites entre regiones. Estas técnicas también contribuyen a la eliminación del ruido presente en la imagen.

Diversas estrategias para mejorar el contraste de las imágenes de fondo de ojo incluyen la aplicación de aumento de contraste local adaptativo en áreas sub-imagen, utilizando la media y desviación estándar de intensidades para detectar áreas brillantes de la retinopatía diabética. Otras técnicas emplean la transformación del brillo similar a la corrección gamma y el uso de filtros de Gabor para igualar la iluminación [44].

6.5.4. Exclusión de fondo

El principal objetivo de la exclusión de fondo es eliminar las variaciones de iluminación en el fondo de una imagen, permitiendo que los objetos en primer plano sean más fácilmente analizados. Este proceso se logra restando la imagen original de intensidad de la imagen filtrada promedio. Las imágenes de fondo de ojo a menudo tienen fondos que cambian debido a factores como los ángulos de iluminación, el resplandor del flash y la variación natural de la retina.

Algunas investigaciones han mejorado esta técnica utilizando modelos matemáticos para

estimar la iluminación de fondo y han observado que la normalización del contraste puede afectar negativamente los algoritmos de segmentación de lesiones. La eliminación de fondo mediante filtrado promedio, filtrado de mediana para eliminar ruido y AHE ha sido propuesta en varios estudios. Otros métodos incluyen la separación de primer plano y fondo utilizando la estimación de luminosidad uniforme y desviación estándar, y la exclusión de fondo mediante técnicas de filtrado de mediana, CLAHE y estiramiento de contraste [44].

6.5.5. Filtrado

Las técnicas de preprocesamiento empleadas tienen como objetivo mejorar la precisión en la detección de lesiones de diversas maneras. Un primer paso común es el suavizado, que se realiza para suprimir el ruido o las pequeñas fluctuaciones en la imagen, usando el filtrado de mediana. Esta técnica reemplaza un punto dado en la imagen por la mediana de los valores de su vecindad, eliminando el ruido impulsivo y preservando los bordes.

El filtrado de mediana es útil especialmente para eliminar el ruido de sal y pimienta sin comprometer la nitidez de la imagen. Métodos como el filtrado gaussiano y el uso de perfiles de secciones transversales con múltiples orientaciones también mejoran el contraste y la calidad de las imágenes de fondo de ojo. Estos métodos se aplican por separado a los componentes rojo, verde y azul de la imagen RGB.

La combinación de ecualización de histograma y filtrado de suavizado mejora el contraste de las imágenes retinianas. Los filtros de mediana y CLAHE aplicados al canal verde de la imagen ayudan a reducir el ruido y mejorar la calidad general, facilitando la detección de características en los sistemas de análisis automatizado. Además, otras técnicas como el filtrado de paso bajo y el uso de filtros Gabor ayudan a mejorar el patrón vascular [44].

6.5.6. Generación de máscaras

Las imágenes de fondo de ojo presentan un fondo circular rodeado por un fondo oscuro. La exclusión de los píxeles de fondo y el procesamiento solo de los píxeles del fondo de ojo se realiza mediante el uso de una máscara de fondo. Esta máscara es una imagen binaria de la misma resolución que la imagen de fondo de ojo, en la que los píxeles positivos corresponden al área de primer plano. El tamaño y forma de la máscara varían según los ajustes de la cámara de fondo de ojo.

La generación automática de máscaras se utiliza para evitar el procesamiento de los bordes negros en las imágenes. Técnicas como la umbralización y las operaciones morfológicas pueden fallar cuando la imagen no está bien expuesta. Se han propuesto herramientas de procesamiento como ImageJ para generar máscaras y métodos automáticos para definir la región de interés utilizando filtros gaussianos. La combinación de técnicas como el filtrado de mediana, CLAHE y binarización mejora y normaliza la imagen, facilitando la extracción de características relevantes para la detección de candidatos mediante algoritmos de clasificación [44].

6.6. Redes neuronales

Las redes neuronales son modelos computacionales inspirados en la estructura del cerebro humano, diseñados para reconocer patrones complejos y realizar tareas de clasificación o regresión. Están compuestas por capas de neuronas artificiales, organizadas en una estructura de nodos conectados. Cada neurona recibe varias entradas ponderadas, realiza una suma ponderada de las mismas y aplica una función de activación no lineal para determinar su salida.

El aprendizaje en una red neuronal se basa en el ajuste de los pesos sinápticos mediante un algoritmo de optimización, generalmente utilizando el método de *backpropagation* y un optimizador como el gradiente descendente. Este proceso busca minimizar una función de pérdida que mide el error entre la salida esperada y la salida generada por la red.

Una red neuronal típica consta de tres tipos de capas: una capa de entrada, una o más capas ocultas, y una capa de salida. Cada capa oculta permite que la red capture características complejas en los datos, mejorando su capacidad para aprender representaciones profundas.

[45]

6.6.1. Redes neuronales convolucionales (CNN)

Las redes neuronales convolucionales (CNN) son una clase de redes neuronales profundas que han demostrado ser especialmente efectivas en la clasificación de imágenes debido a su capacidad para capturar características espaciales y estructurales.

Las Redes Neuronales Convolucionales (CNNs) son una clase de redes neuronales artificiales diseñadas para procesar datos estructurados en forma de cuadrícula, como las imágenes. Las CNNs se inspiran en la estructura y el funcionamiento de la corteza visual de los animales, donde las neuronas responden a estímulos en una región limitada del campo visual, conocida como campo receptivo. Este principio se aplica en las CNNs mediante el uso de filtros o kernels que recorren la imagen de entrada para detectar características locales, como bordes, texturas y formas.

Una CNN típica se compone de varias capas, incluyendo capas convolucionales, capas de agrupamiento (pooling) y capas completamente conectadas. Las capas convolucionales aplican filtros sobre la imagen de entrada, generando mapas de características que destacan patrones locales. Las capas de agrupamiento reducen la dimensionalidad de estos mapas de características, conservando la información más relevante y reduciendo el costo computacional. Finalmente, las capas completamente conectadas actúan como un clasificador que toma las características extraídas y produce la salida final, como la categoría de una imagen, como se observa en su arquitectura convencional 2. Durante el entrenamiento, la CNN ajusta los parámetros de los filtros a través del aprendizaje basado en gradientes, permitiendo que la red aprenda a identificar automáticamente las características más relevantes para la tarea específica [46].

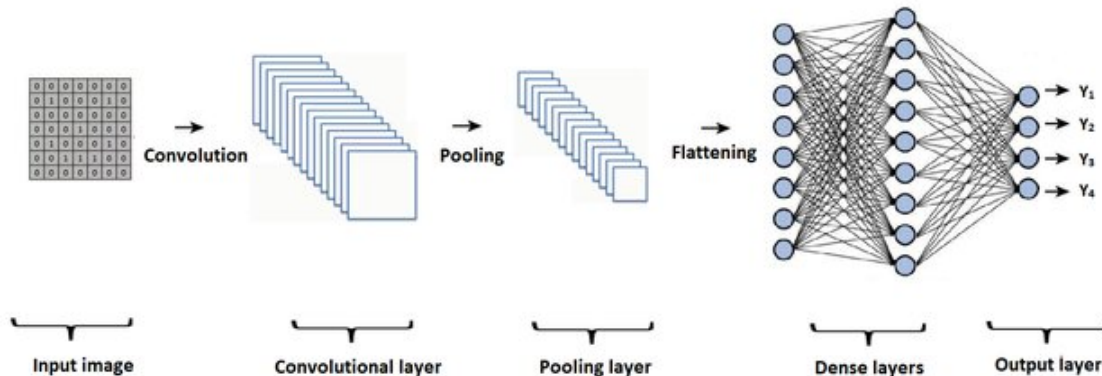


Figura 2: Representación visual de la arquitectura de una red neuronal convolucional (CNN) [47]

6.6.1.1. *Padding*

El *padding*, o relleno, es una técnica utilizada en las redes neuronales convolucionales (CNN) para mantener las dimensiones espaciales de la salida después de aplicar la operación de convolución. Cuando una imagen o un mapa de características se procesa a través de un filtro convolucional, los píxeles en los bordes de la imagen se ven afectados en menor medida que los píxeles en el centro, lo que puede generar una pérdida de información en las fronteras. El *padding* agrega una capa de píxeles alrededor de la entrada para compensar esta pérdida y preservar las dimensiones espaciales.

Existen dos tipos principales de *padding*:

- ***Valid Padding***: no se agrega ningún píxel adicional, lo que resulta en una reducción de las dimensiones espaciales de la salida. Como el que se observa en la figura 3
- ***Same Padding***: se agregan píxeles para mantener las dimensiones de la salida iguales a las de la entrada.

Este concepto es fundamental para evitar la reducción excesiva de las dimensiones y mejorar la capacidad de la red para aprender características de las imágenes. *Padding* permite preservar la resolución espacial en las capas convolucionales, facilitando el aprendizaje de características a lo largo de las fronteras [45].

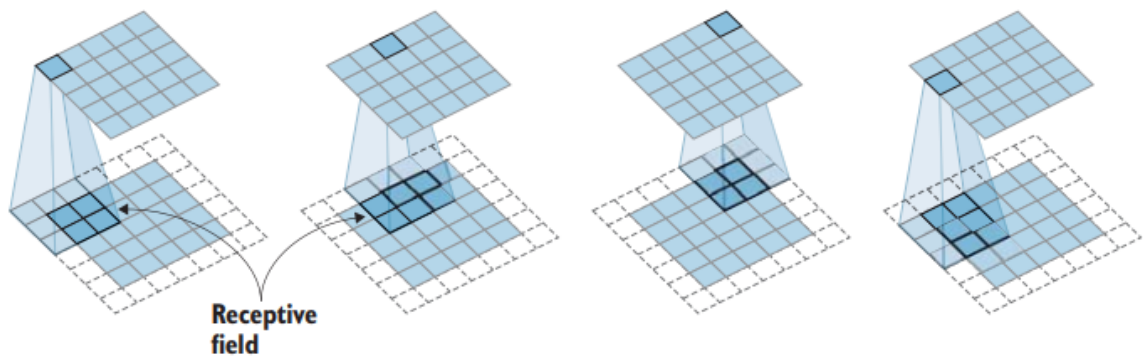


Figura 3: Campo receptivo y padding en redes neuronales convolucionales [48]

6.6.1.2. *Batch normalization*

La normalización por lotes, o *Batch Normalization* (BN), es una técnica utilizada para acelerar el entrenamiento de las redes neuronales profundas. El objetivo de BN es normalizar las activaciones de cada capa para que tengan media cero y varianza unitaria, lo que ayuda a estabilizar el proceso de entrenamiento. Durante el entrenamiento, BN calcula la media y la varianza de cada mini-lote de datos y ajusta las activaciones para que tengan una distribución más uniforme.

Al aplicar BN, la red es capaz de entrenar con tasas de aprendizaje más altas y puede mejorar la precisión del modelo. Sin embargo, hay ciertas consideraciones al utilizar BN en el contexto de *fine-tuning* o *transfer learning*, en donde se transfieren los pesos de una red preentrenada a un nuevo problema.

Reentrenamiento de las capas de *batch normalization*

Al realizar *fine-tuning* o *transfer learning*, una de las decisiones clave es si se debe o no reentrenar las capas de BN. Si las capas de BN no se reentrenan, los parámetros de media y varianza preentrenados en el modelo original se mantienen constantes, y se usan para normalizar las activaciones de la nueva tarea. Sin embargo, este enfoque puede no ser óptimo si las distribuciones de los datos en la nueva tarea son significativamente diferentes de las de la tarea original. En este caso, el reentrenamiento de las capas de BN puede ser necesario para que la red se ajuste mejor a las características del nuevo conjunto de datos.

Por otro lado, si se decide reentrenar las capas de BN durante el *fine-tuning*, estas capas actualizarán los parámetros de media y varianza en función de las activaciones de los datos de la nueva tarea. Este enfoque es útil cuando se espera que las distribuciones de activaciones cambien significativamente entre las tareas, lo que permite que el modelo se ajuste a las nuevas características.

- **Reentrenar BN:** se recomienda si el conjunto de datos de la nueva tarea es significativamente diferente del conjunto de datos original. Esto asegura que los parámetros de BN se ajusten a las nuevas distribuciones.
- **No reentrenar BN:** se puede optar por no reentrenar BN si se desea utilizar la red

preentrenada como está, lo que puede ser útil cuando la tarea es muy similar o la distribución de los datos no cambia mucho.

Modos de operación: entrenamiento vs inferencia

Durante el entrenamiento, BN calcula la media y la varianza a partir de un mini-lote de datos, utilizando estas estadísticas para normalizar las activaciones. Sin embargo, durante la inferencia, las estadísticas de media y varianza deben ser constantes, por lo que se utilizan las medias y varianzas acumuladas durante el entrenamiento. De esta manera, la inferencia es más estable y no depende de las estadísticas del mini-lote actual, lo que evitaría variaciones en la salida del modelo durante la inferencia. *Batch Normalization* puede acelerar el entrenamiento y mejorar la precisión del modelo, pero su comportamiento varía entre el modo de entrenamiento y el modo de inferencia [49].

6.6.1.3. Funciones de activación

Las funciones de activación son componentes clave en las redes neuronales, ya que permiten que la red aprenda relaciones no lineales. Sin una función de activación, las redes neuronales no serían capaces de aprender relaciones complejas, y se reducirían a una simple combinación lineal de entradas.

Las funciones de activación más comunes incluyen las siguientes y las podemos observar en la figura 4:

- **ReLU (*Rectified Linear Unit*)**: la función ReLU es una de las más utilizadas en CNN debido a su simplicidad y efectividad. Se define como $f(x) = \max(0, x)$, donde los valores negativos se establecen en cero.
- **Sigmoid**: la función sigmoide produce una salida en el rango de $(0, 1)$, útil para problemas de clasificación binaria.
- **Softmax**: se utiliza en la capa de salida para problemas de clasificación multiclase. Normaliza los valores para que sumen 1, proporcionando probabilidades para cada clase.

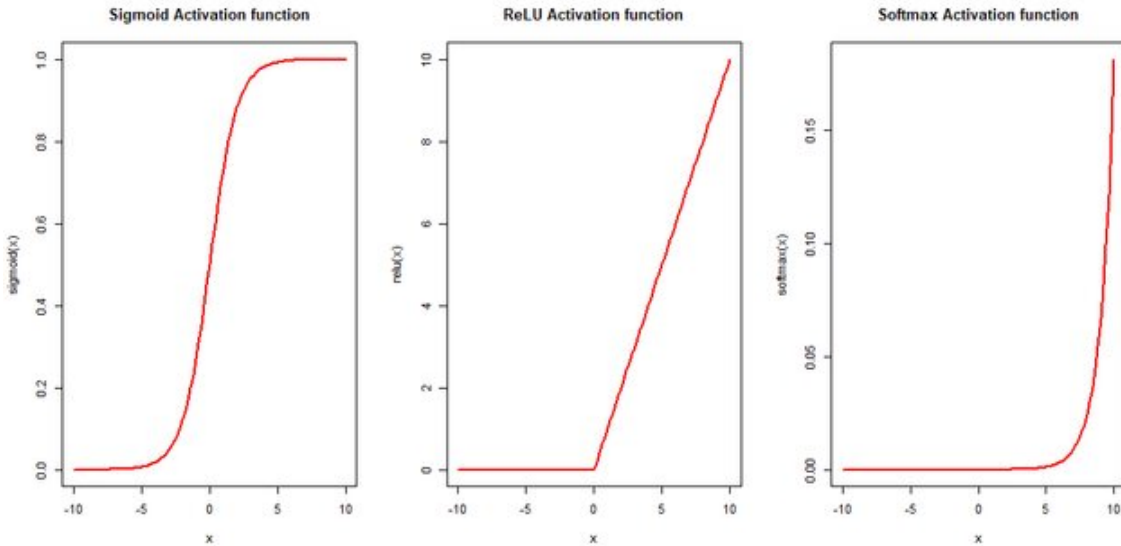


Figura 4: Graficas de funciones de activación importantes en clasificación [50]

La elección de la función de activación puede afectar significativamente el rendimiento de la red. Las funciones de activación son esenciales para que las redes neuronales modelen relaciones no lineales y aprendan representaciones complejas [45].

6.6.1.4. *Max pooling*

El *max pooling* es una operación de reducción de dimensionalidad utilizada en las redes neuronales convolucionales. Este proceso consiste en dividir una imagen en regiones pequeñas y seleccionar el valor máximo de cada región. El objetivo es reducir el tamaño espacial de la representación, lo que ayuda a disminuir el número de parámetros, reducir la carga computacional y prevenir el sobreajuste.

El *max pooling* también proporciona invariancia a la traslación, lo que significa que el modelo es capaz de reconocer una característica en diferentes posiciones de la imagen.

Por lo general, el *max pooling* se implementa con una ventana deslizante (por ejemplo, de tamaño 3×3) como la que se observa en el ejemplo 5, que se mueve sobre la imagen, seleccionando el valor máximo en cada área. El *max pooling* reduce la complejidad computacional y proporciona invariancia a la traslación al seleccionar las características más importantes [45].

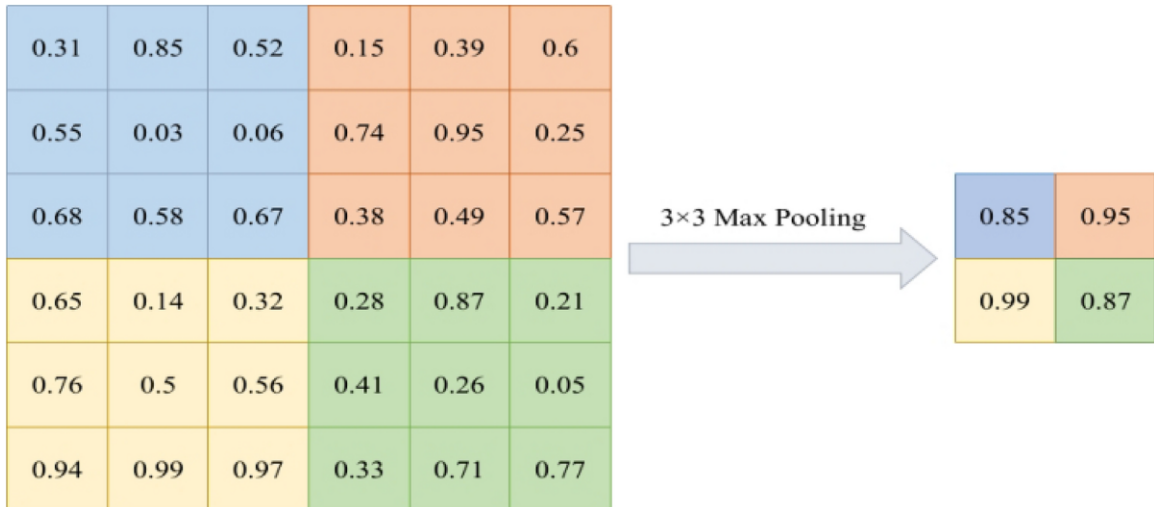


Figura 5: *Max Pooling* 3x3 [51]

6.6.2. ResNet

La ResNet-152 es una arquitectura de redes neuronales profundas propuesta en 2015 por He et al. [52], como parte de la familia de redes neuronales ResNet (*Residual Networks*), que ha demostrado un rendimiento sobresaliente en diversas tareas de visión por computadora, incluyendo la clasificación de imágenes, detección de objetos y segmentación semántica. A diferencia de las redes neuronales tradicionales, ResNet introduce la idea de *bloques residuales*, lo que permite entrenar redes extremadamente profundas sin caer en los problemas comunes de redes profundas, como la degradación del gradiente.

6.6.2.1. Arquitectura de ResNet-152

La arquitectura ResNet-152 está compuesta por 152 capas, que incluyen capas convolucionales, de normalización por lotes (*Batch Normalization*), activaciones ReLU, capas de pooling y una capa de clasificación final totalmente conectada. La clave de la arquitectura ResNet es el uso de *conexiones residuales* o *skip connections*, que permiten el flujo directo de información y gradientes a través de la red, facilitando el entrenamiento de redes mucho más profundas en comparación con las arquitecturas previas.

6.6.2.2. Bloques residuales

Un bloque residual está compuesto por dos o tres capas convolucionales consecutivas, cada una seguida de normalización por lotes y activaciones ReLU. Lo que distingue a ResNet es la adición de una conexión directa entre la entrada del bloque y la salida, lo que se conoce como *skip connection*. Matemáticamente, la salida de un bloque residual se puede describir de la siguiente manera:

$$y = F(x, \{W_i\}) + x$$

Donde x es la entrada al bloque, $F(x, \{W_i\})$ es la función de la red neuronal (normalmente compuesta por varias capas convolucionales), y y es la salida del bloque. La idea central es que la red no necesita aprender la transformación completa $F(x)$, sino la diferencia residual entre la entrada y la salida. Esto mitiga los problemas de vanishing gradients, comunes en redes profundas.

6.6.2.3. Profundidad y capacidad de ResNet-152

La versión ResNet-152 consta de 152 capas, distribuidas en bloques residuales de la siguiente manera:

- 1 capa convolucional inicial
- 49 bloques residuales, organizados en 4 conjuntos con 3, 8, 36 y 3 bloques residuales, respectivamente
- 1 capa totalmente conectada al final para la clasificación

La red comienza con una capa convolucional 7×7 seguida de un max-pooling, lo que reduce significativamente las dimensiones espaciales de la imagen. Luego, los bloques residuales se organizan en cuatro conjuntos, donde el tamaño de los filtros convolucionales varía para capturar características de diferentes tamaños.

Gracias a su gran profundidad, la ResNet-152 puede aprender representaciones jerárquicas y detalladas de las características presentes en los datos de entrada, lo que la convierte en una de las arquitecturas más poderosas para tareas de visión por computadora.

6.6.2.4. Modelo preentrenado ResNet-152

Los modelos preentrenados de ResNet-152 se han entrenado en grandes conjuntos de datos, como ImageNet, que contiene millones de imágenes etiquetadas en más de 1,000 clases. Estos modelos preentrenados ofrecen una excelente capacidad de generalización y son comúnmente utilizados para *transfer learning*, donde el modelo se ajusta a una nueva tarea con un conjunto de datos más pequeño.

En el proceso de *fine-tuning*, las capas iniciales del modelo preentrenado, que aprenden características genéricas como bordes y texturas, se conservan, mientras que las capas finales, que capturan características específicas del conjunto de datos de origen, se reemplazan o ajustan para la nueva tarea. ResNet-152 es particularmente adecuado para *transfer learning* debido a la riqueza de las características aprendidas gracias a su gran profundidad.

6.6.2.5. Ventajas de ResNet-152

ResNet-152 presenta varias ventajas clave:

- **Profundidad extrema:** con 152 capas, puede capturar representaciones muy complejas y profundas de los datos.
- **Bloques residuales:** las conexiones residuales permiten un entrenamiento más efectivo de redes muy profundas, resolviendo problemas de degradación del gradiente.
- **Transfer learning:** gracias a los modelos preentrenados en grandes conjuntos de datos, ResNet-152 es altamente efectivo para reutilizarse en nuevas tareas mediante el ajuste fino.

A pesar de sus ventajas, la mayor profundidad de ResNet-152 implica también mayores requisitos computacionales en términos de memoria y tiempo de procesamiento, en comparación con versiones más pequeñas de ResNet como ResNet-50.

6.6.2.6. Desempeño y uso en la práctica

La ResNet-152 ha sido ampliamente utilizada en competiciones de visión por computadora y ha logrado resultados sobresalientes en conjuntos de datos como ImageNet y COCO, que son conjuntos de imágenes muy grandes. Su capacidad de generalizar a nuevas tareas la hace útil en aplicaciones como la clasificación de imágenes médicas, la detección de objetos y la segmentación semántica. Además, el uso de modelos preentrenados reduce el tiempo de entrenamiento para nuevos conjuntos de datos, lo que es crítico en aplicaciones prácticas. ResNet-152 es un modelo de referencia en visión por computadora que ha demostrado ser altamente eficaz en tareas complejas de clasificación y detección de imágenes [52].

6.6.3. VGG

La red VGG es una arquitectura de redes neuronales convolucionales propuesta por Simonyan y Zisserman en 2014 [53], que ha sido ampliamente utilizada en la clasificación de imágenes y otras tareas de visión por computadora. VGG se caracteriza por su simplicidad en el diseño de la arquitectura, donde se utilizan capas convolucionales de tamaño fijo de 3×3 , seguidas de capas de pooling, y el aumento de profundidad para mejorar el rendimiento. Las versiones más comunes de VGG son VGG-16 y VGG-19, que cuentan con 16 y 19 capas de peso respectivamente.

6.6.3.1. Arquitectura de VGG

La arquitectura de VGG se organiza en una serie de capas convolucionales, seguidas de operaciones de pooling y capas totalmente conectadas. Su estructura es mucho más simple en comparación con otras arquitecturas más recientes como ResNet, pero ha demostrado

un alto rendimiento en tareas de clasificación de imágenes, especialmente en competiciones como ImageNet.

6.6.3.2. Capas convolucionales

Una de las características más importantes de VGG es que todas sus capas convolucionales utilizan filtros de tamaño fijo 3×3 , con un paso (*stride*) de 1 y padding adecuado para mantener las dimensiones espaciales después de cada convolución. Esta configuración permite capturar las características locales de las imágenes mientras mantiene la simplicidad en el diseño. Además, cada bloque convolucional está seguido de una función de activación ReLU (*Rectified Linear Unit*), que introduce no linealidad en la red.

6.6.3.3. Capas de *pooling*

Después de cada conjunto de capas convolucionales, se aplica una operación de *max-pooling* con un tamaño de filtro de 2×2 y un paso de 2. Esta operación reduce las dimensiones espaciales de la imagen, manteniendo las características más importantes y reduciendo el costo computacional de las capas subsiguientes. El uso de max-pooling también ayuda a introducir cierta invariancia a la translación en la red.

6.6.3.4. Capas completamente conectadas

Al final de la arquitectura de VGG, se encuentran tres capas completamente conectadas (*fully connected*). Estas capas reciben las características extraídas por las capas convolucionales y de *pooling*, y se encargan de la tarea de clasificación final. La última capa completamente conectada es una capa softmax que genera una distribución de probabilidad sobre las clases de salida, dependiendo de la tarea.

6.6.3.5. Modelos VGG-16 y VGG-19

Los modelos VGG-16 y VGG-19 se diferencian por el número de capas convolucionales que poseen: VGG-16 tiene 16 capas con pesos (13 convolucionales y 3 completamente conectadas), mientras que VGG-19 tiene 19 capas con pesos (16 convolucionales y 3 completamente conectadas). A pesar de que VGG-19 es más profunda, ambas arquitecturas comparten un rendimiento similar en tareas de clasificación de imágenes, con ligeras diferencias en términos de precisión y eficiencia.

El diseño profundo de VGG hace que los modelos puedan aprender representaciones muy complejas de los datos de entrada, lo que ha contribuido a su éxito en competiciones de visión por computadora. Sin embargo, esta profundidad también conlleva un alto costo computacional y un gran consumo de memoria, lo que ha sido una desventaja frente a arquitecturas más eficientes como ResNet.

6.6.3.6. Modelo preentrenado VGG

Los modelos preentrenados VGG-16 y VGG-19 han sido entrenados en grandes conjuntos de datos, como *ImageNet*, y han mostrado un rendimiento sobresaliente en tareas de clasificación de imágenes. Los modelos preentrenados de VGG se utilizan comúnmente en *transfer learning*, donde se transfieren las características aprendidas en una tarea de origen a una nueva tarea con un conjunto de datos más pequeño.

En el proceso de *fine-tuning*, las capas convolucionales preentrenadas, que capturan características generales de bajo nivel como bordes y texturas, suelen congelarse, mientras que las capas completamente conectadas al final se reentrenan para adaptarse a la nueva tarea. Este enfoque es especialmente útil en tareas donde el conjunto de datos objetivo es pequeño o cuando no hay suficiente capacidad computacional para entrenar una red desde cero.

6.6.3.7. Ventajas de VGG

A pesar de su diseño simple, VGG ofrece varias ventajas:

- **Simplicidad en la arquitectura:** el uso uniforme de convoluciones 3×3 y operaciones de pooling facilita la implementación y comprensión del modelo.
- **Gran capacidad de representación:** la profundidad de VGG-16 y VGG-19 permite aprender representaciones detalladas de los datos, lo que se traduce en un alto rendimiento en tareas de clasificación de imágenes.
- **Transfer Learning:** los modelos preentrenados de VGG son ampliamente utilizados en transfer learning, lo que permite reutilizar los pesos preentrenados en nuevas tareas con diferentes conjuntos de datos.

Sin embargo, uno de los mayores inconvenientes de VGG es su alto costo computacional debido a la gran cantidad de parámetros y la falta de técnicas más modernas como las skip connections, que facilitan el entrenamiento de redes más profundas.

6.6.3.8. Desempeño y aplicaciones

A lo largo de los años, VGG ha sido ampliamente adoptada en diversas aplicaciones de visión por computadora, como la clasificación de imágenes, la detección de objetos y la segmentación semántica. A pesar de ser superada por arquitecturas más modernas como ResNet y EfficientNet en términos de eficiencia, VGG sigue siendo una referencia importante, especialmente en escenarios donde la simplicidad y el uso de modelos preentrenados son clave. La arquitectura VGG es un ejemplo clásico de redes profundas, ampliamente utilizada en tareas de clasificación de imágenes, gracias a su simplicidad y capacidad para el aprendizaje de características jerárquicas [53].

6.6.4. EfficientNet

La arquitectura EfficientNet, propuesta por Tan y Le en 2019 [54], es una familia de redes neuronales convolucionales que introduce una novedosa técnica de escalado para mejorar la eficiencia y el rendimiento de los modelos. A diferencia de otras arquitecturas más convencionales, como ResNet o VGG, que escalan las redes aumentando principalmente la profundidad, EfficientNet emplea un enfoque de escalado compuesto que equilibra las tres dimensiones clave: profundidad, ancho y resolución de entrada. Esto permite obtener redes más precisas y eficientes a nivel computacional.

6.6.4.1. Arquitectura de EfficientNet

La arquitectura de EfficientNet se basa en un modelo básico llamado EfficientNet-B0, y luego escala hacia arquitecturas más grandes (EfficientNet-B1 a EfficientNet-B7). La base de la arquitectura está compuesta por bloques de convolución con convoluciones de profundidad (*depthwise separable convolutions*) y capas de normalización por lotes, que permiten reducir el número de parámetros y la complejidad computacional. La característica clave de EfficientNet es su estrategia de escalado compuesto, que se explica a continuación.

6.6.4.2. Escalado compuesto

El escalado compuesto es un enfoque sistemático para escalar la red, optimizando tres factores:

- **Profundidad** (*depth*): número de capas en la red.
- **Ancho** (*width*): número de canales en cada capa.
- **Resolución** (*resolution*): tamaño de la entrada de la imagen.

El escalado compuesto es expresado mediante un coeficiente de escalado único ϕ , que ajusta estos tres factores de manera conjunta. Matemáticamente, la estrategia de escalado se define por las siguientes ecuaciones:

$$d = \alpha^\phi, \quad w = \beta^\phi, \quad r = \gamma^\phi$$

donde d es la profundidad, w es el ancho, r es la resolución de entrada, y α , β , y γ son hiperparámetros que determinan cómo se escala cada factor. La idea detrás de este enfoque es que el rendimiento de una red no mejora simplemente aumentando una de estas dimensiones, sino que deben ajustarse de manera conjunta para obtener mejoras óptimas.

6.6.4.3. Bloques MBConv

Una de las piezas clave de la arquitectura EfficientNet es el uso de bloques MBConv (*Mobile Inverted Bottleneck Convolution*). Estos bloques se componen de:

- Convoluciones de profundidad (*depthwise separable convolutions*), que descomponen la convolución estándar en dos pasos más eficientes: convoluciones de profundidad y convoluciones puntuales (*pointwise convolutions*).
- *Squeeze-and-Excitation* (SE) Blocks: estos bloques permiten que la red pondere la importancia de los canales de características, lo que mejora la capacidad de la red para enfocarse en las características más relevantes.

Este tipo de convoluciones, junto con los bloques SE, permiten a EfficientNet lograr un gran rendimiento con una cantidad significativamente menor de parámetros en comparación con otras redes profundas.

6.6.4.4. Modelos preentrenados EfficientNet

Los modelos preentrenados de EfficientNet han sido entrenados en grandes conjuntos de datos como ImageNet y han mostrado un rendimiento superior en términos de precisión y eficiencia computacional. Los modelos EfficientNet-B0 a EfficientNet-B7 están disponibles como modelos preentrenados para diversas tareas de visión por computadora, y son ampliamente utilizados en escenarios de transfer learning debido a su capacidad para generalizar bien en nuevos conjuntos de datos.

Durante el proceso de *fine-tuning*, las primeras capas convolucionales de los modelos EfficientNet, que capturan características genéricas como bordes y texturas, suelen congelarse, mientras que las capas superiores pueden reentrenarse para adaptarse a una tarea específica. Esto reduce el tiempo y los recursos necesarios para entrenar el modelo en una nueva tarea.

6.6.4.5. Ventajas de EfficientNet

EfficientNet ofrece varias ventajas significativas en comparación con otras arquitecturas de redes neuronales convolucionales:

- **Eficiencia computacional:** gracias a la estrategia de escalado compuesto y al uso de convoluciones separables, EfficientNet logra un mejor rendimiento con menos parámetros y menos costos computacionales en comparación con redes como ResNet o VGG.
- **Escalabilidad óptima:** la técnica de escalado compuesto permite ajustar el modelo para diferentes necesidades, desde modelos pequeños como EfficientNet-B0 para dispositivos móviles, hasta modelos más grandes como EfficientNet-B7 para servidores de alto rendimiento.
- **Rendimiento SOTA (*State-of-the-Art*):** EfficientNet ha alcanzado resultados de vanguardia en tareas de clasificación de imágenes, superando a arquitecturas más profundas y grandes, gracias a su diseño eficiente.

6.6.4.6. Desempeño y aplicaciones

Desde su introducción, EfficientNet ha sido utilizado en numerosas aplicaciones de visión por computadora, como la clasificación de imágenes, la detección de objetos y la segmentación semántica. Además, su capacidad para ser utilizado en dispositivos con recursos limitados, como teléfonos móviles, lo hace muy versátil para implementaciones en el mundo real. Gracias a su alta eficiencia y capacidad de escalado, ha sido adoptado en aplicaciones que requieren un balance entre precisión y eficiencia computacional. EfficientNet es una arquitectura innovadora que ha establecido un nuevo estándar en la relación entre precisión y eficiencia computacional, y es ampliamente utilizada en *transfer learning* y aplicaciones de visión por computadora [54].

6.6.5. MDFSC

El método *Multi-Scale Deep Feature Sparse Coding* (MDFSC) es una técnica novedosa para la detección de anomalías en imágenes de retina. Propone una combinación de autoencoders y codificación dispersa (*sparse coding*) para identificar características en múltiples escalas, logrando un rendimiento superior en datasets médicos. La principal motivación detrás de MDFSC es superar las limitaciones de los métodos supervisados de detección de anomalías, que dependen de grandes cantidades de datos etiquetados, lo cual es difícil de obtener en el ámbito médico. Y mucho más en contextos como los de Guatemala, o países latinos donde el tener un *dataset* como tal se vuelve una tarea complicada [55].

6.6.5.1. Entrenamiento de *autoencoder*

El método MDFSC comienza con el entrenamiento de un *autoencoder* basado en la arquitectura VGG-16. Este *autoencoder* está diseñado para aprender representaciones detalladas de imágenes normales en diferentes escalas. El *encoder* del *autoencoder* se compone de 13 capas convolucionales y 5 capas de *max-pooling*, seguidas de una capa completamente conectada que produce un vector latente de longitud 256. El objetivo es que el *autoencoder* aprenda características específicas de las imágenes normales para luego ser capaz de identificar aquellas imágenes que no se ajusten a estas características (anómalas).

La función de pérdida utilizada para entrenar el *autoencoder* es la pérdida L_2 , que mide la diferencia entre la imagen original y su reconstrucción. Esto permite que el *autoencoder* aprenda a capturar tanto las características macro como micro de las imágenes normales.

6.6.5.2. Codificación dispersa de características multi-escala

Después de entrenar el *autoencoder*, se extraen características de múltiples escalas desde diferentes capas del *encoder*, obteniendo representaciones más ricas de las imágenes. Las características se extraen de capas a distintas profundidades del *encoder*, lo que permite capturar información tanto de bajo nivel como de alto nivel.

La codificación dispersa se utiliza para modelar las características extraídas, utilizando

una representación esparsa basada en un diccionario de características entrenado solo con imágenes normales. La codificación dispersa permite representar las características multi-escala como combinaciones lineales de un pequeño número de vectores base (*basis vectors*) en el diccionario. Si una imagen es normal, su reconstrucción utilizando el diccionario tendrá un error bajo; si es anómala, el error será alto debido a la incapacidad del diccionario para representar características no vistas en las imágenes normales.

Matemáticamente, el objetivo de la codificación dispersa se define como:

$$\min_{D \in C, (w_i)_{i=1}^m} \sum_{i=1}^m \left(\frac{1}{2} \|f_i - Dw_i\|_2^2 + \alpha \|w_i\|_1 \right)$$

Donde:

- f_i son las características extraídas,
- D es el diccionario de vectores base,
- w_i son los coeficientes dispersos que representan las características como combinaciones lineales de los vectores base,
- α es un coeficiente de regularización que controla la esparsidad.

6.6.5.3. Inferencia

Durante la fase de inferencia, se extraen parches de la imagen de entrada y se pasan por el *encoder* del *autoencoder* entrenado para obtener características multi-escala. Luego, estas características se utilizan para calcular el error de reconstrucción mediante la codificación dispersa. La suma de los errores de los k parches con mayor error se utiliza como puntaje de anomalía. Este enfoque asegura que solo las regiones más anómalas contribuyan significativamente al puntaje final, descartando pequeñas zonas con errores normales, lo cual es crítico para imágenes con múltiples zonas normales y pocas áreas anómalas.

6.6.5.4. Ventajas del método MDFSC

El método MDFSC presenta varias ventajas significativas en comparación con otros métodos de detección de anomalías en imágenes médicas:

- **Detección a múltiples escalas:** al extraer características de diferentes profundidades del encoder, MDFSC captura información tanto de bajo como de alto nivel, lo que le permite detectar anomalías de diferentes tamaños y formas.
- **Codificación dispersa:** el uso de codificación dispersa asegura que solo las características relevantes de las imágenes normales sean utilizadas para la reconstrucción, lo que resulta en una mejor diferenciación entre imágenes normales y anómalas.

- **Capacidad generalizada:** MDFSC es robusto frente a diversos tipos de datos médicos y ha demostrado su eficacia en varios conjuntos de datos de imágenes de retina, superando a otros métodos del estado del arte.

6.6.5.5. Resultados y comparación con el estado del arte

El método MDFSC ha demostrado un rendimiento sobresaliente en la detección de anomalías en imágenes de retina en tres conjuntos de datos distintos: Eye-Q, IDRiD y OCTID. En comparación con métodos del estado del arte como AnoGAN, Deep-Geo y SPADE, MDFSC logra mejoras significativas en las métricas AP y AUC, con una mejora de hasta el 12% en algunos casos. Estos resultados indican que el método propuesto no solo es más preciso, sino también más eficiente en términos computacionales [55]. El método MDFSC logra una mejora significativa en la detección de anomalías en imágenes médicas, combinando la eficacia de los *autoencoders* y la codificación dispersa [55].

6.6.6. *Transfer learning*

Transfer learning, o aprendizaje por transferencia, es una técnica clave en el ámbito del aprendizaje profundo, en la que se transfiere el conocimiento adquirido en una tarea o dominio (*source domain*) para mejorar el desempeño en otra tarea o dominio diferente (*target domain*) [13]. Esta técnica ha demostrado ser particularmente útil en situaciones donde los datos del dominio objetivo son escasos, lo que es común en áreas como la imagen médica, donde la recolección de datos etiquetados es costosa y enfrenta barreras éticas y legales. En este contexto, el uso de modelos preentrenados en grandes conjuntos de datos, como ImageNet, ha sido la estrategia preferida.

En este capítulo, exploramos los conceptos básicos de *transfer learning*, sus aplicaciones en imágenes médicas y las razones por las cuales los modelos preentrenados en ImageNet son efectivos en este campo, basándonos en los hallazgos de Matsoukas et al. (2022) [13].

6.6.6.1. Fundamentos de *transfer learning*

El objetivo del *transfer learning* es reutilizar características aprendidas en un dominio fuente (por ejemplo, el conjunto de datos ImageNet) para mejorar el desempeño en un dominio objetivo (por ejemplo, imágenes médicas). En general, los modelos preentrenados son ajustados (mediante fine-tuning) en el nuevo conjunto de datos para adaptarse a las peculiaridades del nuevo dominio. El *fine-tuning* ajusta los pesos del modelo preentrenado para que se adapten a las nuevas clases o características presentes en los datos del dominio objetivo.

El proceso típico de *transfer learning* incluye los siguientes pasos:

1. Entrenar un modelo en un conjunto de datos grande y generalizado, como ImageNet.
2. Transferir los pesos de ese modelo al modelo que se usará en el nuevo dominio.

3. Realizar *fine-tuning* o ajustes en las capas superiores del modelo con datos específicos del dominio objetivo.

6.6.6.2. Aplicaciones en imágenes médicas

El uso de *transfer learning* ha sido extremadamente útil en imágenes médicas, donde las limitaciones en los datos son particularmente pronunciadas debido a la rareza de muchas enfermedades y la dificultad de adquisición de datos etiquetados. Los modelos preentrenados en ImageNet se utilizan ampliamente debido a su capacidad para aprender representaciones generales que luego pueden ser ajustadas para tareas médicas más específicas [13]. A pesar de las diferencias entre imágenes naturales (como las de ImageNet) y las imágenes médicas (por ejemplo, rayos X, resonancias magnéticas), las características aprendidas en las capas inferiores de los modelos convolucionales son suficientemente generales como para ser útiles en ambos contextos.

6.6.6.3. Ventajas del uso de modelos preentrenados en ImageNet

A pesar de las diferencias entre los datos de ImageNet y los datos médicos, existen varias razones por las que utilizar modelos preentrenados en ImageNet es efectivo en el ámbito médico:

6.6.6.4. Reutilización de características (*feature reuse*)

Uno de los principales hallazgos de Matsoukas et al. (2022) es que la reutilización de características juega un papel crucial en el éxito del *transfer learning* en imágenes médicas. Las características generales aprendidas en ImageNet, como bordes, texturas y formas simples, pueden reutilizarse en imágenes médicas, ya que estas características de bajo nivel son comunes entre ambos dominios [13]. Este fenómeno es particularmente efectivo en redes convolucionales, donde las capas iniciales capturan características genéricas, mientras que las capas más profundas se especializan en las particularidades del conjunto de datos objetivo.

6.6.6.5. Eficacia en conjuntos de datos pequeños

El estudio de Matsoukas et al. (2022) muestra que el transfer learning es más efectivo cuando el conjunto de datos del dominio objetivo es pequeño. En tareas médicas, donde la disponibilidad de datos etiquetados suele ser limitada, los modelos preentrenados en ImageNet proporcionan un punto de partida robusto. Cuando los datos del dominio médico son escasos, el modelo se beneficia más de las características ya aprendidas en ImageNet, reduciendo la necesidad de aprender desde cero y acelerando el proceso de entrenamiento [13].

6.6.6.6. Distancia entre el dominio fuente y el dominio objetivo

La distancia entre el dominio fuente y el dominio objetivo también es un factor clave en la efectividad del *transfer learning*. Aunque las imágenes médicas y las imágenes naturales pueden parecer visualmente diferentes, Matsoukas et al. encontraron que la reutilización de características es efectiva incluso cuando hay una distancia significativa entre los dominios [13]. Sin embargo, cuanto más similares sean los datos del dominio fuente y el dominio objetivo, mayor será el beneficio del *transfer learning*.

6.6.6.7. Factores que afectan el éxito de *transfer learning* en imágenes médicas

Matsoukas et al. identifican varios factores que determinan el éxito del *transfer learning* en imágenes médicas:

- **Tamaño del conjunto de datos:** como se mencionó anteriormente, el tamaño del conjunto de datos del dominio objetivo es crucial. Los conjuntos de datos más pequeños tienden a beneficiarse más del *transfer learning*, ya que el modelo preentrenado proporciona características que serían difíciles de aprender desde cero con pocos datos.
- **Capacidad del modelo:** los modelos con mayor capacidad (es decir, aquellos con más parámetros) tienden a beneficiarse más del *transfer learning*, especialmente en dominios con conjuntos de datos pequeños. Modelos como ResNet o *Vision Transformers* (ViTs) pueden adaptarse mejor cuando se transfieren características generales aprendidas en ImageNet [13].
- **Bias inductivo del modelo:** los modelos con menos *bias inductivo*, como los *Vision Transformers*, dependen más de la reutilización de características, ya que no tienen sesgos estructurales que favorezcan características específicas del dominio. Por otro lado, los modelos con un fuerte *bias inductivo*, como ResNet, tienden a beneficiarse menos de la transferencia en grandes conjuntos de datos médicos, pero aún muestran mejoras cuando los datos son limitados.

El *transfer learning* es una técnica extremadamente poderosa en el análisis de imágenes médicas, especialmente en situaciones donde los datos del dominio objetivo son limitados. Los beneficios del *transfer learning* son mayores en dominios con conjuntos de datos pequeños, como es típico en tareas médicas, donde la reutilización de características permite acelerar el entrenamiento y mejorar la precisión [13]. El uso de modelos preentrenados en ImageNet ha demostrado ser efectivo debido a la reutilización de características generales que son aplicables tanto en imágenes naturales como médicas. El estudio de Matsoukas et al. (2022) destaca que la eficacia del *transfer learning* en el ámbito médico depende de varios factores, como el tamaño del conjunto de datos, la capacidad del modelo y el bias inductivo, pero en la mayoría de los casos, la reutilización de características desempeña un papel fundamental en el éxito del proceso.

Para abordar la detección de retinopatía diabética como objetivo principal, se desarrollaron varios modelos utilizando diferentes algoritmos y arquitecturas, incluyendo técnicas avanzadas como *autoencoders* y *multi scale deep feature space coding* (MDFSC) [55]. Y modelos preentrenados basados en *Convolutional Neural Networks* CNN de la familia *Visual Geometry Group* (VGG), *Residual Neural Network* (ResNet) y EfficientNet. Con estas arquitecturas como la Resnet-34, VGG-16 y MobileNetV2, enfocadas en la clasificación de imágenes con un grado de detalle alto, se pueden llegar a obtener resultados como los vistos en el estudio realizado por la Universidad de North South, donde obtienen una precisión de hasta 97.23 % [56]. Por lo que cualquiera de estas arquitecturas es capaz de lograr resultados considerablemente buenos.

7.1. Herramientas de ingeniería utilizadas

Para la creación y entrenamiento del modelo predictivo, así como el preprocesamiento de las imágenes y la creación de los *dashboards* para la clasificación por parte de los especialistas y el uso en producción del modelo de clasificación.

- Lenguaje de Programación Python: la implementación de los modelos y el preprocesamiento de las imágenes se realizarán en Python, aprovechando sus bibliotecas avanzadas para el manejo de datos y aprendizaje automático.
- Bibliotecas de *deep learning*:
 - Tensorflow y Keras: estas bibliotecas serán fundamentales para construir, entrenar y evaluar los modelos de CNN como Resnet y VGG.

- OpenCV: se utilizó para aplicar los preprocesamientos sugeridos a las imágenes, como la normalización y el aumento de datos.
- Plataformas de computación de alto rendimiento:
 - Google Colab y Kaggle Notebooks: se utilizó para el entrenamiento de los modelos, aprovechando el acceso a Unidades de Procesamiento Gráfico, (GPU, por sus siglas en inglés) P100, lo que permitió manejar el entrenamiento intensivo en recursos de las arquitecturas de *deep learning*.
- Herramientas de evaluación y validación:
 - *Scikit-learn*: para la implementación de métricas de evaluación como sensibilidad, especificidad, precisión y F1-score, permitiendo una evaluación exhaustiva del rendimiento del modelo.
 - Fleiss Kappa: para evaluar la consistencia y fiabilidad de las clasificaciones manuales realizadas por los médicos oftalmólogos.
- Desarrollo de interfaz de usuario:
 - *Streamlit*: se utilizó para desarrollar una interfaz interactiva en Python que permita a los médicos y otros usuarios probar el modelo en tiempo real, facilitando la detección temprana de anomalías oculares en entornos clínicos.

7.2. Recopilación de datos

1. **Recopilación de datos:** se obtuvieron los datos mediante el dataset publicado en la competencia de Kaggle [12]. En el cual se contiene un total de 35,126 imágenes para el entrenamiento del modelo. De las cuáles 25,810 (73.48 %) no contenían retinopatía diabética, 5,292 (15.06 %) con retinopatía leve, 2,443 (6.96 %) con una presencia moderada, 873 (2.49 %) con una presencia severa y 708 (2.01 %) con retinopatía diabética proliferativa como se puede observar en 6. Por lo que se observa un desbalance de clases bastante importante. Siendo un 73.48 % contra un 26.52 % de cada clase binaria como observamos en 7.

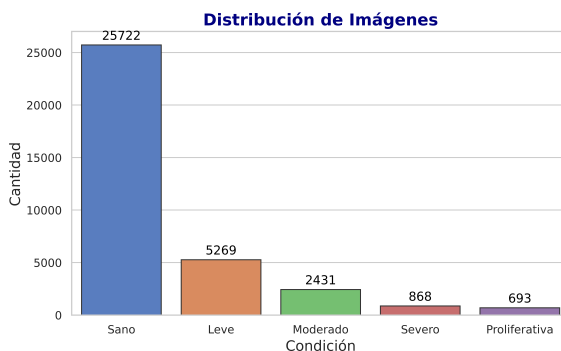


Figura 6: Distribución de imágenes de dataset Kaggle

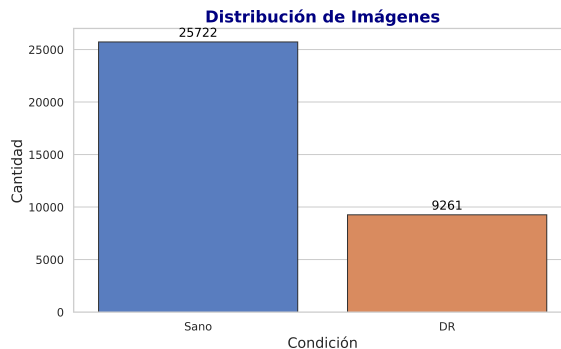


Figura 7: Distribución de imágenes binarizadas de dataset Kaggle

7.3. Selección de arquitectura y algoritmos

Basado en los trabajos previos realizados con diferentes conjuntos de datos y arquitecturas para predecir la enfermedad. Se eligieron los más robustos de los tres grupos con mejor rendimiento según los resultados obtenidos por [57]. Estos son de la familia VGG, ResNet y EfficientNet. Por lo que se comparan los modelos VGG-19, ResNet152 y EfficientNetB7 para la misma tarea de predicción binaria. El objetivo de tener un modelo como VGG-19 es poder comparar un modelo mucho más simple y por consiguiente rápido frente a otros modelos más pesados.

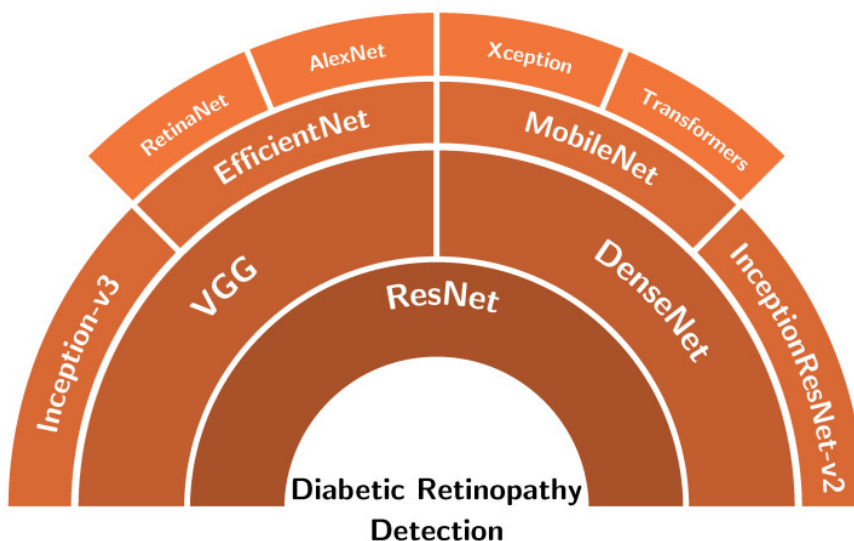


Figura 8: Familias de modelos utilizados para predecir RD [57]

Se separaron en dos grupos los modelos, ya que los modelos VGG-19, ResNet152 y EfficientNetB7 se enfocan en una respuesta binaria o multiclase, mientras que el modelo MDFSC es un tipo de *autoencoder* que busca detectar anomalías de cualquier tipo. Por lo que detectó cualquier tipo de cambio de una imagen sana como una anomalía. Por lo que no necesariamente se estuvo prediciendo retinopatía diabética únicamente.

7.4. Preprocesamiento de imágenes

Ya que el objetivo principal no era lograr un preprocesamiento perfecto de las imágenes, aunque debemos tomar en cuenta que es de las partes más importantes, más bien centrarse en la selección y entrenamiento de modelos para la predicción de retinopatía diabética. Para ello, se adoptó un enfoque práctico en el preprocesamiento de las imágenes, siguiendo un flujo específico que se ajusta a las necesidades del modelo. En primer lugar, se verificó que las imágenes cumplieran con los canales requeridos (rojo, azul y verde o RGB, por sus siglas en inglés) para asegurar su compatibilidad con los algoritmos utilizados. Posteriormente, se implementó una binarización de las imágenes, un paso crucial para preservar la privacidad de los pacientes y para facilitar la comprensión del modelo al trabajar con imágenes simplificadas. Finalmente, se aplicaron transformaciones adicionales, como la detección y anonimización del área macular, y el ajuste de la imagen a una escala estándar para ser procesada por el modelo de clasificación. En este enfoque, la eficiencia y la claridad en la interpretación del modelo fueron priorizadas sobre la perfección técnica del preprocesamiento. El preprocesamiento utilizado se describe como el siguiente:

1. Ajuste del tamaño de la imagen

- a) Calcular el radio de la retina midiendo la suma de intensidades en una línea horizontal que pase por el centro de la imagen.
- b) Escalar la imagen para que el radio calculado coincida con el valor deseado de escala.

2. Mejora del canal verde

- a) Extraer el canal verde de la imagen, que contiene la mayor cantidad de información relevante para el análisis de fondo de ojo.
- b) Aplicar ecualización de histograma adaptativa (CLAHE) al canal verde para mejorar el contraste y resaltar detalles importantes.

3. Corrección del color local

- a) Aplicar un filtro de desenfoque gaussiano a la imagen procesada y restar el color promedio local, para mejorar el contraste en las áreas de interés.

4. Eliminación de bordes externos

- a) Crear una máscara circular centrada que cubra aproximadamente el 90% del área de la imagen, eliminando el borde exterior que podría contener ruido o información irrelevante.
- b) Multiplicar la imagen por la máscara para conservar solo la región central; asignar un valor constante (128) al área eliminada.

5. Ajuste de dimensiones y canales de salida

- a) Asegurar que la imagen tenga tres canales (RGB), si es necesario.
- b) Redimensionar la imagen a 600x600 píxeles

El Algoritmo de anonimización de la imagen de fondo de ojo 1 permite identificar y extraer únicamente la región circular de la retina en la imagen, eliminando el entorno circundante y cualquier información potencialmente identificable. Este proceso de anonimización es crucial para preservar la confidencialidad de los pacientes al omitir detalles fuera de la zona de interés. Además, al mantener solo el 'círculo' de la retina, el algoritmo asegura la coherencia con los datos usados en el entrenamiento del modelo de predicción, lo que mejora la precisión de las predicciones. La imagen resultante se adecua a los requisitos del modelo, el cual fue diseñado para trabajar con imágenes ya binarizadas y centradas en la región retiniana.

7.4.1. Anonimización de la imagen

Algorithm 1 Anonimización de la imagen de fondo de ojo

```

1: Entrada: img_array (imagen de fondo de ojo)
2: if channels = 4 (RGBA) then
3:   img_array ← convertir_RGBA_a_RGB(img_array)
4: end if
5: Convertir a escala de grises:
6: gray ← cv2.cvtColor(
7:   img_array, cv2.COLOR_RGB2GRAY)
8: Detectar círculos:
9: circles ← cv2.HoughCircles(gray,
10:  método=cv2.HOUGH_GRADIENT)
11: if circles ≠ None then
12:   for cada (x, y, r) en circles do
13:     Crear máscara circular:
14:     mask ← np.zeros_like(img_array)
15:     cv2.circle(mask, centro=(x,y),
16:       radio=r, color=(255,255,255),
17:       grosor=-1)
18:     Aplicar máscara:
19:     img_array ← cv2.bitwise_and(
20:       img_array, mask)
21:   end for
22:   Salida: img_array (imagen anonimizada)
23: else
24:   Advertencia: No se detectó círculo.
25:   Advertencia: Procesando sin anonimización.
26:   Salida: img_array (imagen sin anonimizar)
27: end if

```

El procedimiento descrito se puede ver representado en el Algoritmo 2, representa una metodología estructurada para el preprocesamiento de imágenes de fondo de ojo, diseñado para optimizar la detección de características relevantes en el análisis ocular. A través de los pasos de ajuste de tamaño, mejora de contraste y filtrado de ruido, este algoritmo permite resaltar detalles esenciales, mejorando la precisión en las etapas de análisis posteriores.

7.4.2. Preprocesamiento macular

Algorithm 2 Preprocesamiento macular para retinopatía diabética

```
1: Entrada: img (imagen de fondo de ojo)
2: Parámetro: scale (tamaño deseado de la imagen, por defecto 600)
3: function SCALERADIUS(img, scale)
4:    $x \leftarrow$  Suma de los píxeles de la fila central
5:    $r \leftarrow$  (Píxeles  $>$  10% de la media)/2
6:    $s \leftarrow$  scale  $\times$  (1.0/ $r$ )
7:   Retornar: cv2.resize(img, (0, 0), fx=s, fy=s)
8: end function
9: function PREPROCESS_IMAGE(img)
10:  Seleccionar canal verde:
11:  img_green = img[:, :, 1]
12:  Aplicar CLAHE:
13:  clahe = cv2.createCLAHE(clipLimit=2.0,
14:    tileGridSize=(8, 8))
15:  img_clahe = clahe.apply(img_green)
16:  Restar color promedio local:
17:  blur = cv2.GaussianBlur(img_clahe,
18:    (0, 0), scale/30)
19:  img_clahe = cv2.addWeighted(
20:    img_clahe, 4,
21:    blur, -4, 128)
22:  Retornar: img_clahe
23: end function
24: if img es None then
25:   Salida: None
26: end if
27: Intentar:
28: Escalar imagen:
29: img = scaleRadius(img, scale)
30: Aplicar preprocesamiento:
31: img = preprocess_image(img)
32: Crear y aplicar máscara circular (90% del radio)
33: Verificar 3 canales y tamaño 224x224
34: Salida: img (imagen preprocesada)
35: Excepto:
36: Mostrar: Error en el preprocesamiento
37: Salida: None
```

7.5. Evaluación del modelo

Para la evaluación del modelo en su entrenamiento se buscó maximizar las diversas métricas. El objetivo es garantizar que los modelos no solo sean precisos, sino que también

sean robustos y generalizables a nuevos datos. Para lograr esto, se utilizarán una serie de métricas de evaluación y procedimientos de validación que se detallan a continuación. Como expresa Vujović en su paper en cuanto a las métricas necesarias para medir un algoritmo de clasificación [41].

- **Selección de métricas de evaluación:**

- **Sensibilidad (*Recall*):** Sensibilidad = $\frac{TP}{TP+FN}$
- **Especificidad:** Especificidad = $\frac{TN}{TN+FP}$
- **Precisión:** Precisión = $\frac{TP}{TP+FP}$
- **F1-Score:** F1-Score = $2 \times \frac{\text{Precisión} \times \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$
- **AUC PR:**
- **Quadratic Weighted Kappa:**

$$\kappa = 1 - \frac{\sum_{i,j} W_{i,j} O_{i,j}}{\sum_{i,j} W_{i,j} E_{i,j}}$$

Además se emplearon “*class weights*”. La cual es una técnica utilizada para abordar el problema del desbalanceo de clases en problemas de clasificación. En situaciones donde las clases están desproporcionadamente representadas en los datos, el modelo tiende a favorecer la clase mayoritaria, lo que puede llevar a un rendimiento deficiente en la predicción de la clase minoritaria. El uso de pesos de clase permite asignar mayor importancia a las clases menos representadas durante el entrenamiento del modelo, ajustando así el impacto de cada clase en la función de pérdida. De esta forma, se busca que el modelo aprenda a identificar las clases desbalanceadas con mayor precisión, mejorando el desempeño general sin la necesidad de modificar el conjunto de datos o utilizar técnicas complejas de muestreo. Estos se asignan generalmente de manera inversamente proporcional a la frecuencia de las clases en el conjunto de entrenamiento, aunque también pueden ajustarse manualmente para reflejar la severidad de los errores de clasificación en las clases minoritarias. En este caso fue de manera inversamente proporcional a la frecuencia de las clases. Teniendo los siguientes pesos:

- Clase 0 (Sano): 0.2720
- Clase 1 (RD leve): 2.5481
- Clase 2 (Rd moderado): 1.3279
- Clase 3 (RD grave): 8.0606
- Clase 4 (RD proliferativa): 10.0961

7.6. Conjunto de datos

El conjunto de datos utilizado contiene imágenes de fondo de ojo de pacientes con y sin retinopatía diabética, clasificadas en diferentes niveles de gravedad. Se dividió en tres subconjuntos:

- Entrenamiento (70 %): utilizado para ajustar los pesos del modelo.
- Validación (15 %): utilizado para ajustar los hiperparámetros y prevenir el sobreajuste.
- Prueba (15 %): usado para la evaluación final del modelo.

Estos tres subconjuntos se crearon *estratificando* la variable objetivo. Esto para asegurar que cada grupo (o subconjunto) tenga una proporción representativa de cada clase o categoría. La estratificación garantiza que cada subconjunto (como el de entrenamiento, validación y prueba) tenga la misma proporción de cada clase que el conjunto de datos original. Al tener representatividad de todas las clases en cada subconjunto, se previene que el modelo entrene o evalúe en datos que no reflejan la distribución real de las clases. Mejora la generalización, ayudando a que los modelos sean más robustos y generalizables a nuevos datos, ya que aprenden con subconjuntos que son representativos de la población total. Durante la validación y prueba, se garantiza que las métricas de evaluación no estén sesgadas por una mala distribución de las clases.

Al realizar una exploración de las imágenes proporcionadas por el *dataset* se observó que varias imágenes no contenían la calidad necesaria, ya que existían imágenes totalmente negras, como se puede observar en la Figura 10, cortadas horizontalmente como se puede apreciar en la Figura 9, mal enfocadas como se puede apreciar en la Figura 11 que pierde toda la información sobre el fondo del ojo, entre otros defectos. Lo cuál podría inducir sesgos y dificultar el aprendizaje de alguna clase del modelo. Por lo que se decide depurar el dataset para poder obtener el mejor entrenamiento y calidad de datos posible. Por lo que la solución planteada es mediante una observación totalmente arbitraria tratar de depurar las imágenes buscando estas imágenes. Para esto se crea otro archivo CSV (del inglés *comma-separated-values*) que contiene las imágenes que se juzgaron como que no cumplían estas características. En total fueron 144 imágenes del total de 35,126. Lo cual representa un total de 0.41% aproximadamente. Entre las imágenes que se marcaron como no aptas para el entrenamiento fueron las siguientes:

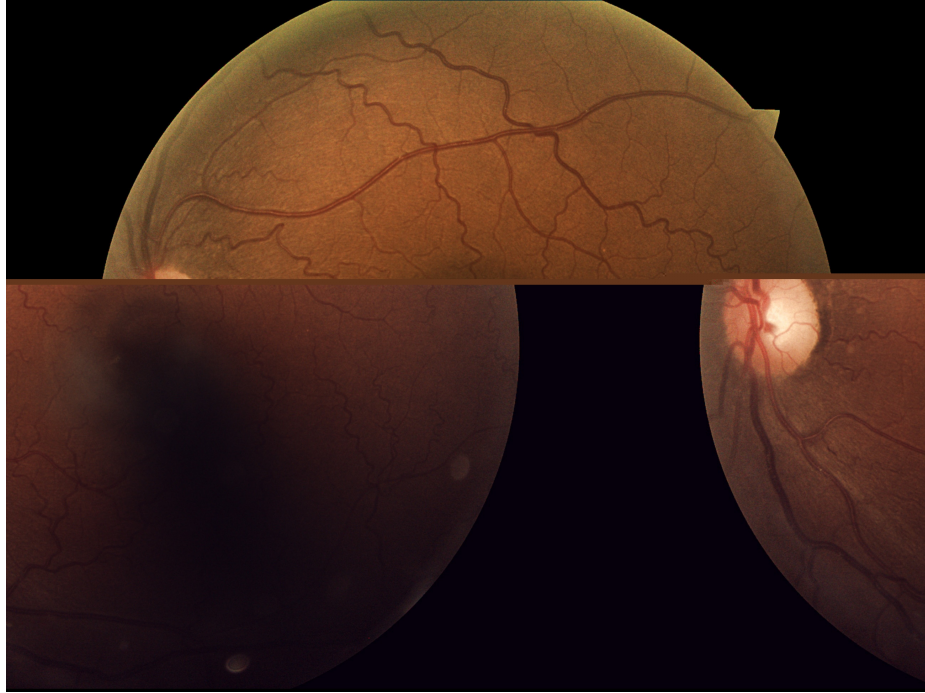


Figura 9: Imágen de fondo de ojo descartada por corte



Figura 10: Imágen de fondo de ojo descartada por no contener información

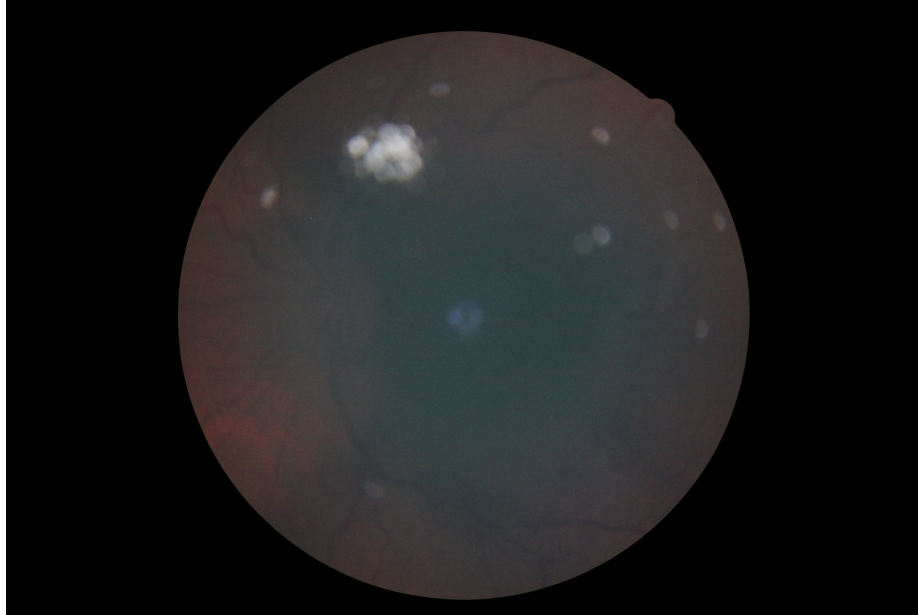


Figura 11: Imágen de fondo de ojo descartada por mal enfoque y luz

Lo cual resultó en una pequeña disminución en la cantidad de imagen y distribución de las clases. Con un total de 25,722 (73.53 %) para las imágenes libres de retinopatía, 5,269 (15.06 %) para la clase de retinopatía leve, 2,431 (6.95 %) para la retinopatía moderada, 868 (2.48 %) para la severa y 693 (1.98 %) para la retinopatía proliferativa. Por lo que el efecto observado sobre las clases fue pequeño.

7.7. Entrenamiento de modelos

Para el entrenamiento de los modelos se utiliza una técnica de *Fine Tuning* y *Transfer Learning* llamada *congelación progresiva* o *fine-tuning gradual*. Esta técnica se utiliza comúnmente en *transfer learning* y es una variación del proceso de *fine-tuning*. Consiste en entrenar inicialmente solo las nuevas capas (normalmente las capas superiores o finales añadidas a un modelo preentrenado), mientras se mantienen congeladas las capas base del modelo preentrenado. Luego, una vez que las nuevas capas han convergido y están bien entrenadas, se comienzan a desbloquear progresivamente las capas más profundas del modelo base, entrenándolas de manera gradual [58].

Pasos típicos de esta técnica:

- **Inicialización con capas congeladas:** las capas del modelo preentrenado (por ejemplo, EfficientNetB7) se congelan, es decir, sus pesos no se actualizan durante el entrenamiento. Se agregan nuevas capas personalizadas (generalmente capas densas o de clasificación) al final del modelo, que sí son entrenables desde el inicio. Se entrena únicamente estas capas superiores durante un número de épocas, permitiendo que el modelo ajuste su capacidad de clasificar los nuevos datos basados en las representaciones aprendidas del modelo preentrenado.

- **Descongelación gradual:** una vez que el modelo ha alcanzado una buena convergencia en las nuevas capas, se desbloquean progresivamente algunas capas del modelo preentrenado. El modelo se entrena nuevamente, ajustando no solo las capas añadidas sino también las capas progresivamente descongeladas del modelo base. Esto permite al modelo refinar las representaciones preentrenadas en función del nuevo conjunto de datos. Este proceso puede repetirse varias veces, desbloqueando más capas en cada fase de entrenamiento.
- **Ajuste fino completo:** finalmente, todas o la mayoría de las capas del modelo original se descongelan y se entrenan juntas, utilizando una tasa de aprendizaje muy pequeña para evitar sobreajustar o desestabilizar los pesos previamente aprendidos.

Al entrenar inicialmente solo unas pocas capas, se reduce el costo computacional al inicio del proceso. Además, esto ayuda a prevenir el sobreajuste. Al entrenar primero las capas añadidas y luego desbloquear progresivamente las capas profundas, se evita modificar de forma abrupta los pesos preentrenados, lo cual puede llevar a un sobreajuste si se hace demasiado rápido. La descongelación gradual permite mejorar el rendimiento del modelo refinando las representaciones más abstractas en las capas inferiores. Esta técnica es especialmente útil cuando el conjunto de datos con el que se va a entrenar es más pequeño o diferente del conjunto de datos en el que se entrenó inicialmente el modelo preentrenado, permitiendo aprovechar las representaciones aprendidas sin sobreescribirlas abruptamente [59].

Por lo tanto, para el entrenamiento de los tres modelos se realiza una ronda de 12 a 15 épocas con una *inicialización con capas congeladas*. Verificando el rendimiento del mismo, luego se buscó ir desbloqueando gradualmente mientras va convergiendo el modelo en cada una de sus métricas y si el rendimiento al descongelar las capas sigue mejorando.

7.8. Modelos

Se presentan cuatro arquitecturas de redes neuronales convolucionales que se utilizaron para abordar el problema de clasificación de imágenes médicas, en particular en el contexto de la retinopatía diabética. Las arquitecturas seleccionadas son VGG-19, ResNet152, EfficientNetB7 y MDFSC, cada una con características y capacidades únicas que las hacen aptas para tareas complejas de visión por computadora.

VGG-19, conocida por su simplicidad y efectividad, con 19 capas entrenables y el uso de filtros pequeños para la extracción de características. Por otro lado, ResNet152, introduce conexiones residuales que permiten entrenar redes más profundas sin que se pierda información relevante debido al problema del desvanecimiento del gradiente. Finalmente, EfficientNetB7, optimiza la precisión y eficiencia computacional, lo que lo convierte en una opción ideal para clasificación en grandes volúmenes de datos.

7.8.1. VGG-19

La arquitectura VGG-19 es una red neuronal convolucional profunda que fue desarrollada por Visual Geometry Group (VGG) de la Universidad de Oxford en 2014. Esta arquitectura

es una extensión del modelo VGG16, participando en la competición *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)*, donde logró excelentes resultados en la tarea de clasificación de imágenes.

Características principales

- **Profundidad:** VGG-19 consta de 19 capas entrenables (capas convolucionales y capas completamente conectadas).
- **Filtros pequeños:** utiliza filtros pequeños de 3×3 píxeles para las operaciones convolucionales.
- **Max Pooling:** aplica capas de max pooling de 2×2 para reducir las dimensiones espaciales.
- **Función de activación:** cada capa convolucional está seguida de una función de activación *ReLU* (*Rectified Linear Unit*).
- **Tamaño de entrada:** el tamaño de entrada de las imágenes es de $224 \times 224 \times 3$ (RGB).

Estructura de VGG-19

La arquitectura completa de VGG-19 sigue la siguiente secuencia de capas:

1. **Entrada:** Imagen de tamaño $224 \times 224 \times 3$.
2. **Bloque 1:**
 - Convolución 3×3 (64 filtros) + *ReLU*.
 - Convolución 3×3 (64 filtros) + *ReLU*.
 - Max Pooling 2×2 .
3. **Bloque 2:**
 - Convolución 3×3 (128 filtros) + *ReLU*.
 - Convolución 3×3 (128 filtros) + *ReLU*.
 - Max Pooling 2×2 .
4. **Bloque 3:**
 - Convolución 3×3 (256 filtros) + *ReLU*.
 - Convolución 3×3 (256 filtros) + *ReLU*.
 - Convolución 3×3 (256 filtros) + *ReLU*.
 - Convolución 3×3 (256 filtros) + *ReLU*.
 - Max Pooling 2×2 .
5. **Bloque 4:**

- Convolución 3×3 (512 filtros) + *ReLU*.
- Convolución 3×3 (512 filtros) + *ReLU*.
- Convolución 3×3 (512 filtros) + *ReLU*.
- Convolución 3×3 (512 filtros) + *ReLU*.
- Max Pooling 2×2 .

6. Bloque 5:

- Convolución 3×3 (512 filtros) + *ReLU*.
- Convolución 3×3 (512 filtros) + *ReLU*.
- Convolución 3×3 (512 filtros) + *ReLU*.
- Convolución 3×3 (512 filtros) + *ReLU*.
- Max Pooling 2×2 .

7. Capas completamente conectadas:

- Capa totalmente conectada con 4096 neuronas + *ReLU*.
- Capa totalmente conectada con 4096 neuronas + *ReLU*.
- Capa completamente conectada con 1000 neuronas (para clasificación) + *softmax*.

A esta arquitectura se le ha realizado la modificación para poder adaptarla a nuestro problema. Añadiendo capas para poder adaptarlo a las clasificaciones que se requieren en el modelo [53].

7.8.2. ResNet152

La arquitectura ResNet152 es una red neuronal convolucional profunda perteneciente a la familia de ResNet (*Residual Networks*) es una de las variantes más profundas de esta familia, diseñada para solucionar el problema del *vanishing gradient*, que afecta el entrenamiento de redes profundas. Esto lo logra mediante la introducción de conexiones residuales o *skip connections*, que permiten que el gradiente fluya mejor hacia las capas iniciales de la red.

Características principales

- **Conexiones residuales:** las conexiones *skip* permiten que la información pase directamente entre capas, sin pasar por las convoluciones, lo que facilita el entrenamiento de redes muy profundas.
- **Profundidad:** tiene 152 capas entrenables, lo que la hace una de las redes más profundas en la arquitectura original de ResNet.
- **Bloques residuales:** utiliza bloques residuales compuestos por tres capas convolucionales en lugar de dos, como en versiones más simples como ResNet50.
- **Normalización por lotes (*Batch Normalization*):** se aplica después de cada operación de convolución para estabilizar y acelerar el entrenamiento.

- **Activación ReLU:** cada operación convolucional es seguida de una activación no lineal ReLU.
- **Tamaño de entrada:** la red toma imágenes de tamaño $224 \times 224 \times 3$ (RGB).

Estructura de ResNet152

La arquitectura de ResNet152 sigue un diseño modular basado en bloques residuales compuestos por varias capas convolucionales:

1. **Entrada:** imagen de tamaño $224 \times 224 \times 3$.
2. **Capa inicial:**
 - Convolución 7×7 (64 filtros, stride 2) + *ReLU*.
 - Max Pooling 3×3 (stride 2).
3. **Bloques residuales:**
 - **Bloque Residual 1:** tres capas de convolución 3×3 con 64, 64 y 256 filtros respectivamente.
 - **Bloque Residual 2:** tres capas de convolución 3×3 con 128, 128 y 512 filtros respectivamente.
 - **Bloque Residual 3:** tres capas de convolución 3×3 con 256, 256 y 1024 filtros respectivamente.
 - **Bloque Residual 4:** tres capas de convolución 3×3 con 512, 512 y 2048 filtros respectivamente.
4. **Capa de *pooling* promedio global:**
 - *Pooling* promedio global para reducir el tamaño de las características a 1×1 .
5. **Capa completamente conectada:**
 - capa completamente conectada con 1000 neuronas (para clasificación en ImageNet) + *softmax*.

Conexiones residuales

Las conexiones residuales son el componente fundamental de la arquitectura ResNet. En lugar de simplemente apilar capas convolucionales, cada bloque introduce una conexión directa (*skip connection*) que permite que la salida de una capa se sume a la salida de una capa posterior. Matemáticamente, esto se puede representar como:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}$$

donde \mathbf{x} es la entrada del bloque residual, $\mathcal{F}(\mathbf{x}, \{W_i\})$ es la función de la operación convolucional, y \mathbf{y} es la salida final del bloque. Esta adición directa ayuda a mitigar el problema de la desaparición del gradiente (*vanishing gradient*) y facilita el entrenamiento de redes muy profundas [52].

7.8.3. EfficientNetB7

EfficientNetB7 es una de las variantes más grandes de la familia de modelos EfficientNet, introducida por Tan y Le en 2019. EfficientNet fue diseñado para optimizar tanto la precisión como la eficiencia computacional, empleando una estrategia de escalado compuesta que balancea de manera uniforme la profundidad, el ancho y la resolución de entrada. EfficientNetB7 ha demostrado ser altamente eficiente en términos de precisión computacional en tareas de clasificación de imágenes.

Características principales

- **Escalado compuesto:** EfficientNet introduce una metodología de escalado compuesta que ajusta la profundidad, el ancho y la resolución de entrada de la red de forma coordinada.
- **Bloques MBConv:** utiliza bloques convolucionales invertidos con convoluciones separables en profundidad (bloques MBConv), lo que reduce la cantidad de parámetros sin comprometer la precisión.
- **Normalización por lotes (Batch Normalization):** se aplica después de cada convolución para estabilizar el entrenamiento.
- **Función de activación Swish:** utiliza la activación *Swish*, que ha demostrado ser más eficaz que ReLU en arquitecturas profundas.
- **Tamaño de entrada:** EfficientNetB7 toma imágenes con un tamaño de entrada de $600 \times 600 \times 3$ (RGB).

Estructura de EfficientNetB7

EfficientNetB7 se basa en bloques MBConv, que combinan convoluciones separables en profundidad y convoluciones 1×1 , permitiendo una arquitectura más eficiente. La arquitectura también implementa una combinación de técnicas de escalado para maximizar la precisión.

A continuación, se describe la arquitectura general de EfficientNetB7:

1. **Entrada:** Imagen de tamaño $600 \times 600 \times 3$.
2. **Bloque inicial:**
 - Convolución 3×3 (64 filtros, stride 2) + *Swish*.
3. **Bloques MBConv:**
 - **Bloques MBConv 1:** un bloque MBConv con 32 filtros, kernel 3×3 , expansión de 1.
 - **Bloques MBConv 2:** dos bloques MBConv con 48 filtros, kernel 3×3 , expansión de 6.

- **Bloques MBConv 3:** tres bloques MBConv con 80 filtros, kernel 5×5 , expansión de 6.
- **Bloques MBConv 4:** cuatro bloques MBConv con 160 filtros, kernel 3×3 , expansión de 6.
- **Bloques MBConv 5:** seis bloques MBConv con 224 filtros, kernel 5×5 , expansión de 6.
- **Bloques MBConv 6:** nueve bloques MBConv con 384 filtros, kernel 5×5 , expansión de 6.
- **Bloques MBConv 7:** tres bloques MBConv con 640 filtros, kernel 3×3 , expansión de 6.

4. Capa de *pooling* promedio global:

- Pooling promedio global para reducir el tamaño de las características a 1×1 .

5. Capa completamente conectada:

- Capa completamente conectada con 1000 neuronas (para clasificación en ImageNet) + *softmax*.

Escalado compuesto

EfficientNet introduce el concepto de **escalado compuesto**, que consiste en balancear tres dimensiones clave de la arquitectura: la profundidad, el ancho y la resolución de entrada. El escalado compuesto puede describirse de la siguiente manera:

$$\text{Profundidad: } d = \alpha^\phi, \quad \text{Ancho: } w = \beta^\phi, \quad \text{Resolución: } r = \gamma^\phi$$

donde α , β y γ son constantes de escalado, y ϕ es un factor que determina la cantidad de recursos computacionales disponibles. En el caso de EfficientNetB7, $\phi = 7$, lo que significa que es la variante más grande de la familia, con mayor profundidad, ancho y resolución en comparación con otras variantes [54].

7.8.4. MDFSC

El proceso empleado para la creación de un modelo *Multi-Scale Deep Feature Sparse Coding* (MDFSC) de *Deep Learning* orientado a la detección de anomalías oculares, con un enfoque específico en la retinopatía diabética fue utilizar un autoencoder basado en la arquitectura VGG-16, como se explica en [55]. Este modelo fue diseñado con el propósito de entrenarse exclusivamente con imágenes de ojos sanos, bajo la hipótesis de que las imágenes que presenten retinopatía diabética generarán un error de representación considerablemente mayor, lo cual permitiría su identificación.

Conjunto de datos el conjunto de datos fue dividido en dos categorías principales: imágenes sanas e imágenes enfermas. Con el fin de optimizar el proceso de entrenamiento y evaluar el desempeño del modelo, las imágenes sanas se subdividieron en dos subconjuntos:

entrenamiento (80 %) y validación (20 %). Esta división permitió entrenar el autoencoder con una proporción representativa de imágenes sanas y evaluar su capacidad de generalización.

Entrenamiento del *autoencoder* el modelo fue entrenado empleando la arquitectura VGG-16 como base, con el objetivo de minimizar el error de representación en el conjunto de datos de entrenamiento. Durante el entrenamiento, se ajustaron los parámetros del modelo a lo largo de múltiples épocas hasta que el error de representación convergió. Por lo que las partes entrenables de esta arquitectura fueron el espacio latente y el *decoder*.

Umbral de clasificación una vez concluido el entrenamiento, se utilizó la distribución de los errores de representación en las imágenes sanas para establecer un umbral adecuado. Este umbral permitió definir un valor límite de error de representación, a partir del cual las imágenes podrían ser clasificadas como sanas o enfermas. Al recibir una nueva imagen, el modelo comparaba su error de representación con el umbral definido, permitiendo identificarla como una imagen sana o, en caso de superar dicho umbral, como una imagen que potencialmente presentaba retinopatía diabética o alguna anomalía ocular no representada por el *autoencoder*.

El *threshold* se calculó usando la desviación estándar de los errores de reconstrucción en el conjunto de entrenamiento. Primero, se obtuvo la media y la desviación estándar de estos errores. Luego, el *threshold* se fijó sumando un múltiplo de la desviación estándar a la media, lo cual permite definir un límite superior para los errores de reconstrucción aceptables. En este caso específico, se definió un *threshold* a 2 desviaciones estándar por encima de la media. Esto implica que los errores que superen este valor se consideran potencialmente anómalos, ya que representan una desviación significativa respecto al comportamiento típico de los datos de entrenamiento. Sin embargo, para el conjunto de test, se aplicó directamente el valor de *threshold* obtenido en el entrenamiento, sin recalcularlo, asegurando así que la evaluación del modelo sea consistente.

Evaluación del modelo para evaluar el desempeño del modelo, se aplicaron métricas de clasificación sobre el conjunto de validación y el conjunto de imágenes enfermas. Este proceso permitió analizar la efectividad del modelo en la detección de anomalías oculares y ajustar los parámetros necesarios para mejorar su precisión en la clasificación de imágenes con retinopatía diabética.

Cálculo del umbral de clasificación para clasificar correctamente las imágenes como sanas o enfermas, se calculó un umbral de clasificación basado en la distribución de los errores de representación de las imágenes sanas. Este umbral fue determinado para maximizar la sensibilidad del modelo a las imágenes con retinopatía diabética, minimizando las clasificaciones incorrectas de imágenes sanas como enfermas. Sea x_i una imagen en el conjunto de datos de validación, y \hat{x}_i la reconstrucción de x_i generada por el autoencoder. El error de representación e_i de la imagen x_i se define como la norma L_2 (o la distancia euclidiana) entre la imagen original y su reconstrucción:

$$e_i = \|x_i - \hat{x}_i\|_2 = \sqrt{\sum_j (x_{i,j} - \hat{x}_{i,j})^2}$$

donde $x_{i,j}$ y $\hat{x}_{i,j}$ representan los valores de cada píxel j en las imágenes x_i y \hat{x}_i , respectivamente. Una vez calculados los errores de representación e_i para todas las imágenes sanas en el conjunto de validación, se construyó una distribución de estos valores de error. Esta

distribución representa el comportamiento esperado del error de representación para imágenes sanas, permitiendo definir un umbral que diferencie las imágenes sanas de las imágenes enfermas. Para determinar el umbral τ , se utilizó el percentil p de la distribución de los errores de representación de las imágenes sanas. Comúnmente, se selecciona un valor alto de percentil (por ejemplo, $p = 95\%$), para asegurarse de que el umbral permita captar la mayoría de las imágenes sanas y minimizar los falsos positivos. Así, el umbral de clasificación τ se define como:

$$\tau = \text{percentil}_p(e_1, e_2, \dots, e_n)$$

donde e_1, e_2, \dots, e_n son los errores de representación de las n imágenes en el conjunto de validación sano. Este umbral τ se utilizó como criterio de clasificación: dado un nuevo ejemplo de imagen x_{nuevo} , se calculaba su error de representación e_{nuevo} . Si $e_{\text{nuevo}} \leq \tau$, la imagen se clasificaba como sana; de lo contrario, se clasificaba como enferma:

$$\text{Clasificación}(x_{\text{nuevo}}) = \begin{cases} \text{sana} & \text{si } e_{\text{nuevo}} \leq \tau \\ \text{enferma} & \text{si } e_{\text{nuevo}} > \tau \end{cases}$$

7.8.5. *Callbacks*

Para entrenar los modelos se utilizaron diversos *callbacks* proporcionados directamente del *backend* de Tensorflow. Estos son funciones especiales que se ejecutan automáticamente en ciertos momentos durante el entrenamiento de un modelo. Sirven para personalizar y controlar el proceso de entrenamiento, permitiendo realizar acciones específicas en momentos clave (como al final de una época o tras cada lote de datos). Esto permite monitorear el desempeño del modelo, ajustar hiperparámetros, guardar el modelo en su mejor estado, o terminar el entrenamiento si ciertas condiciones se cumplen. Esto para poder optimizar el entrenamiento del modelo y lograr los mejores resultados sin tener que realizarlo por fuerza bruta. Los *callbacks* utilizados fueron:

- ***Reduce LR on Plateau:*** ajusta dinámicamente la tasa de aprendizaje durante el entrenamiento, como reducirla si el modelo deja de mejorar o incrementarla si se espera mayor exploración.
- ***Model Checkpoint:*** guarda el modelo en ciertos puntos (por ejemplo, cuando tiene el mejor desempeño en el conjunto de validación), permitiendo restaurarlo posteriormente.
- ***Early Stopping:*** detiene el entrenamiento si el modelo deja de mejorar en el conjunto de validación durante cierto número de épocas. Esto evita el sobreajuste.
- ***CSV Logger:*** realiza un seguimiento de métricas como pérdida, precisión u otras métricas personalizadas, y las registra o visualiza en herramientas como TensorBoard o en este caso un CSV.

Esto aprovechando las capacidades computacionales de GPU y Unidad de Procesamiento de Tensores (TPU) proporcionadas por la plataforma de Kaggle y Colab, ambas proporcionadas por Google [60].

7.9. Etiquetador de imágenes nuevas

Con el fin de etiquetar nuevas imágenes y facilitar el reentrenamiento del modelo en futuras iteraciones, se diseñó un sistema que permite incorporar nueva información sin depender exclusivamente de conjuntos de datos públicos, sus criterios y tipos de imágenes. Esta herramienta permite, una vez obtenidas nuevas imágenes de alguna organización, realizar un proceso de anonimización y carga en una base de datos. Mediante una interfaz web sencilla y privada, los médicos pueden clasificar las imágenes en categorías preestablecidas, o en nuevas categorías que se vayan definiendo según se requiera. Como se muestra en la Figura 12 de la pantalla de inicio de sesión de la plataforma.



Figura 12: Pantalla de inicio de sesión de la herramienta de clasificación

La herramienta cuenta con un sistema de inicio de sesión para garantizar la privacidad de la información. Al ser una página de acceso controlado, se previene que personas externas obtengan acceso a datos sensibles, aun si las imágenes están anonimizadas. Asimismo, se minimiza el uso innecesario de recursos, optimizando el consumo de la herramienta en términos de costos y velocidad.

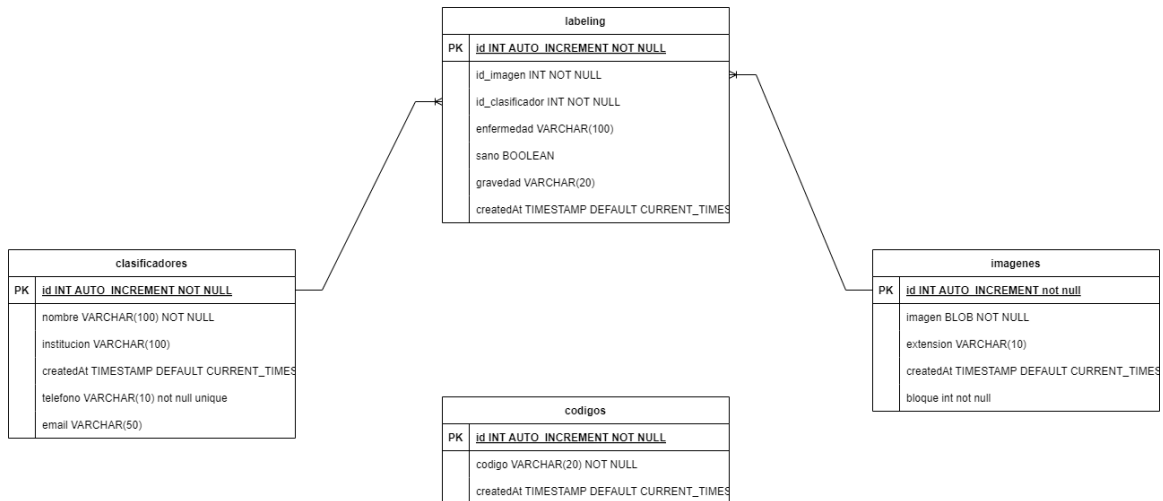


Figura 13: Diagrama de entidad-relación de la base de datos del clasificador

La base de datos se aloja en un servidor flexible de Azure Database for MySQL. Esta infraestructura permite una gestión escalable y segura de la información generada por el clasificador. Esta se planteó como cuatro Cuadros principales donde se almacenan los clasificadores, las imágenes y sus clasificaciones. Como se observa en el diagrama de entidad relación de la Figura 13.

Para acceder al sistema de clasificación, los médicos deben estar registrados previamente mediante su número de teléfono, el cual es ingresado en el sistema por un administrador. Como posible mejora, se sugiere implementar una funcionalidad de autogestión que permita a los médicos registrarse y someter su solicitud a la aprobación de un administrador. Actualmente, una vez registrado, el médico recibe un código único de acceso, evitando la necesidad de memorizar una contraseña específica para este sitio. Este código actúa como una llave privada para ingresar al sistema.

Dentro de la herramienta de clasificación, la interfaz agrupa las imágenes en lotes de 10 para facilitar la tarea a los médicos y evitar una carga de trabajo excesiva. Esta estrategia ofrece dos ventajas: en primer lugar, los usuarios experimentan una sensación de progreso al completar cada bloque, y en segundo lugar, se reduce la frecuencia de las operaciones de escritura y lectura en la base de datos, lo cual optimiza la experiencia de clasificación y disminuye los tiempos de respuesta. La plataforma se plantea como una interfaz sencilla y minimalista como se observa en su interfaz en la Figura 14.

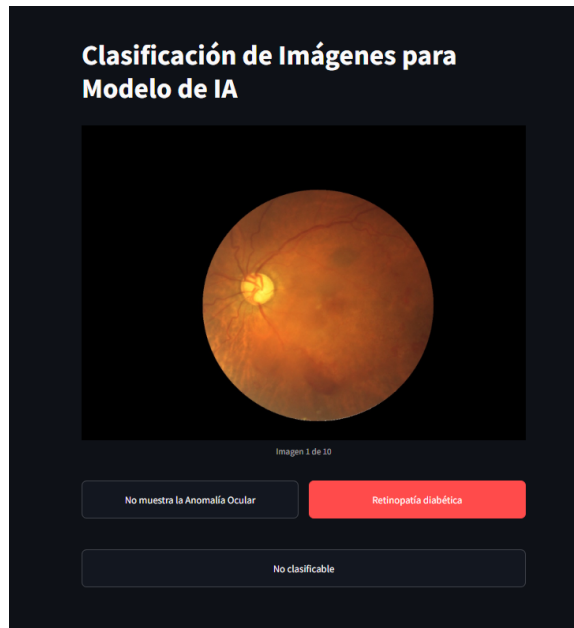


Figura 14: Interfaz de clasificación de imágenes

La organización por bloques mejora significativamente la experiencia de usuario y optimiza la eficiencia del sistema, permitiendo que los médicos retomen la clasificación en cualquier momento y continúen con el siguiente bloque. Este diseño minimiza el impacto en la velocidad de la base de datos y contribuye a una clasificación ágil y sin interrupciones, alineada con los objetivos de la plataforma.

Esta herramienta se deja a disposición y como un planteamiento para futuros proyectos de cómo se podrían incluir nuevas imágenes de otras fuentes si estas no tienen una clasificación previa.

Puedes ingresar a ver el etiquetador en el siguiente *link*: [Etiquetador Imágenes](#)

7.10. *Dashboard* de diagnóstico

Se desarrolla una herramienta interactiva para el diagnóstico preliminar de retinopatía diabética mediante un *dashboard* accesible y fácil de utilizar, utilizando Streamlit. El sistema permite a médicos y personal capacitado cargar imágenes de fondo de ojo obtenidas previamente y obtener una predicción preliminar sobre la presencia y gravedad de la retinopatía diabética. Este *dashboard*, implementado como un apoyo en la toma de decisiones clínicas, es capaz de procesar imágenes de hasta 200 MB en formatos JPG, JPEG o PNG, devolviendo un resultado en pocos segundos, lo cual facilita la identificación temprana de la enfermedad. Además, se incluyen advertencias y recomendaciones para que los usuarios comprendan que el sistema es un complemento del diagnóstico profesional y que no reemplaza la valoración realizada por un especialista en oftalmología. La interfaz contiene secciones de información sobre el modelo, advertencias y los logotipos de las instituciones participantes, destacando su colaboración en el desarrollo de esta herramienta.

El propósito es facilitar su uso y acceso para poder aportar en un diagnóstico preliminar a bajo o nulo costo con únicamente obtener su imagen de fondo de ojo e internet. La interfaz se muestra de una manera directa y sencilla de utilizar, como se muestra en la Figura 15.



Figura 15: Dashboard de diagnóstico preliminar

8.1. Enfoque binario

La primera fase del desarrollo del modelo se centra en la implementación de un enfoque de clasificación binaria para la detección de retinopatía diabética. Por lo que estos fueron los resultados de plantear los modelos con un resultado binario.

En el Cuadro 3 se presentan los resultados de clasificación obtenidos mediante tres modelos binarios: VGG-19, ResNet152 y EfficientNetB7. Para cada modelo, se muestra el valor de la métrica AUC PR, que evalúa la calidad de la clasificación. Los resultados indican que el modelo VGG-19 obtiene un valor de 0.53, mientras que ResNet152 y EfficientNetB7 tienen valores de 0.47 y 0.46, respectivamente.

Cuadro 3: Resultados AUC PR de clasificación modelos binarios

Modelo	Métrica	Resultado
VGG-19	AUC PR	0.53
ResNet152	AUC PR	0.47
EfficientNetB7	AUC PR	0.46
MDFSC	AUC PR	0.26

El Cuadro 4 presenta las métricas de clasificación obtenidas para el modelo VGG-19 con salida binaria. Se muestran los valores de precisión, recall y f1-score para ambas clases, con una precisión de 0.74 y recall de 0.72 para la clase 0, y una precisión de 0.31 y recall de 0.33 para la clase 1. Además, se incluyen la exactitud total, que es de 0.61, así como los promedios macro y ponderado para precisión, recall y f1-score, con valores de 0.52 y 0.62 respectivamente.

Cuadro 4: Métricas de clasificación VGG-19 *output* binario (AUC-ROC: 0.53)

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
0	0.74	0.72	0.73	1223
1	0.31	0.33	0.32	458
accuracy			0.61	1681
macro avg	0.52	0.53	0.52	1681
weighted avg	0.62	0.61	0.62	1681

Asimismo, en los resultados obtenidos por el único algoritmo totalmente binario que fue el MDFSC, en el Cuadro 3 se presentan sus métricas de clasificación obtenidas. Basadas únicamente en el threshold establecido para el error de reconstrucción de las imágenes que fue

En la Figura 16 se muestra la distribución de los errores de reconstrucción generados por el *autoencoder* durante el entrenamiento. La curva azul representa la estimación de densidad de kernel (KDE), la cual sigue la forma de la distribución de los errores de reconstrucción. En el gráfico, una línea verde punteada marca el valor medio de los errores de reconstrucción, mientras que una línea roja punteada indica el valor del umbral definido, basado en el análisis de la desviación estándar de la distribución. Este umbral se utilizó para diferenciar entre reconstrucciones con errores aceptables y aquellas consideradas anómalas, siendo aquellas que superan dicho umbral.

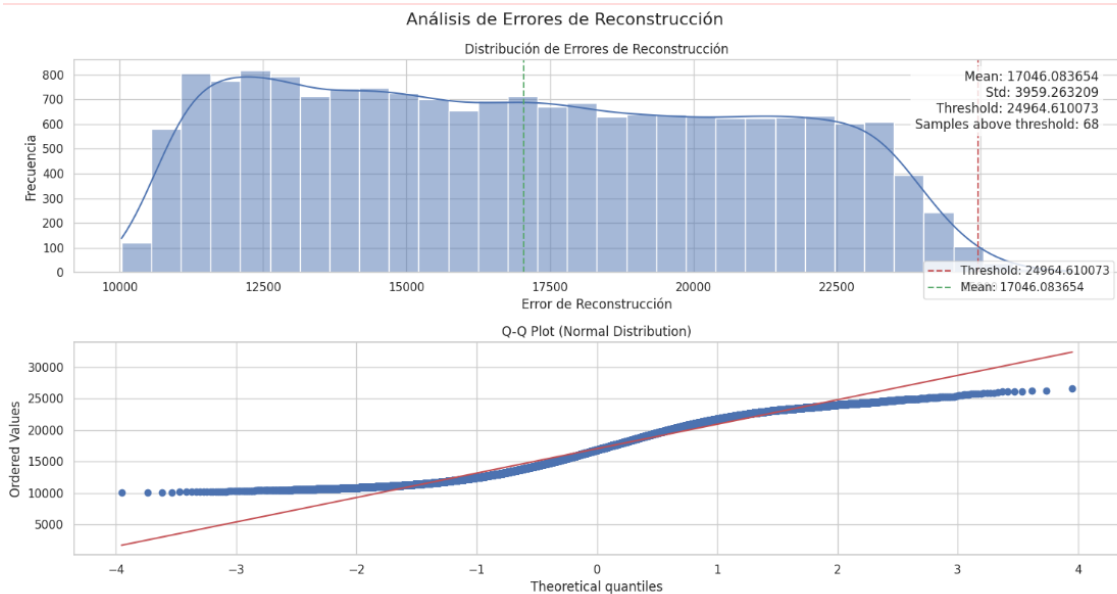


Figura 16: Distribución de error de reconstrucción imágenes sanas (entrenamiento)

En la Figura se muestra la distribución de los errores de reconstrucción del *autoencoder* en el conjunto de prueba, el cual contiene únicamente imágenes con anomalías. La primera gráfica ilustra el histograma de los errores, con una línea verde punteada que indica el

valor medio y una línea roja punteada que representa el umbral de detección de anomalías (*threshold*), previamente calculado a partir del conjunto de entrenamiento. La segunda gráfica presenta un Q-Q plot para evaluar la normalidad de la distribución de los errores de reconstrucción, mostrando una ligera desviación de la normalidad en los valores más extremos.

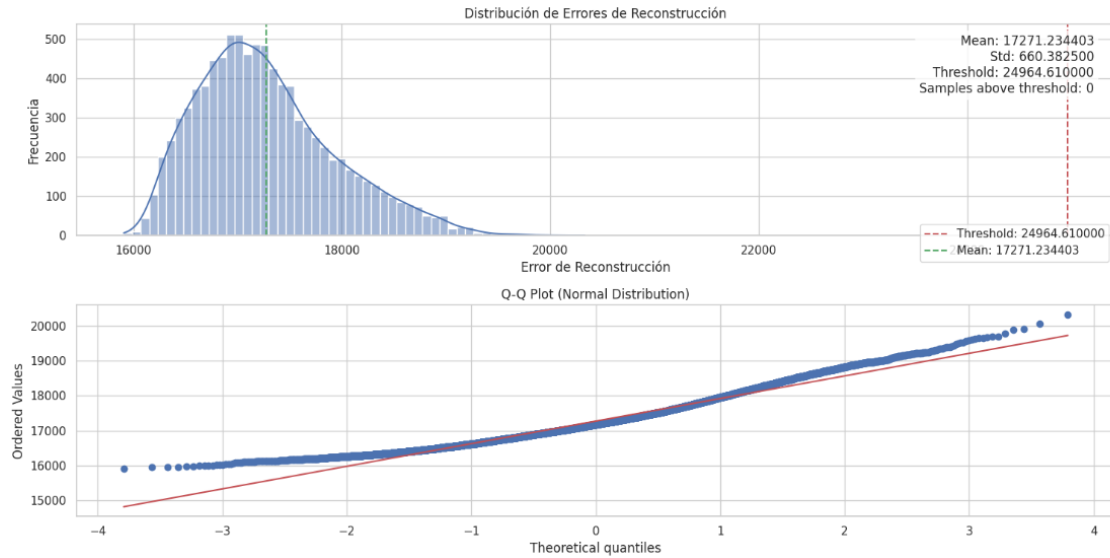


Figura 17: Distribución de error de reconstrucción imágenes con anomalías oculares (*test*)

8.2. Enfoque multiclase

Ante las limitaciones observadas en el enfoque binario, se implementa una estrategia de clasificación multiclase que aprovecha la granularidad de las etiquetas proporcionadas en el conjunto de datos. Esta aproximación considera cinco niveles de clasificación: sano (0), leve (1), moderado (2), grave (3) y retinopatía proliferativa (4). La hipótesis subyacente plantea que permitir al modelo distinguir entre los diferentes niveles de severidad podría mejorar la capacidad de discriminación general del sistema.

Se evalúan tres arquitecturas de redes neuronales convolucionales: *VGG-19* (20.5M de parámetros), *ResNet152* (73.6M de parámetros) y *EfficientNetB7* (87.6M de parámetros). Los resultados revelan que la arquitectura *VGG-19*, a pesar de contar con la menor cantidad de parámetros totales, alcanza el mejor rendimiento, logrando un AUC PR de 80.20% en la clasificación binaria derivada.

8.2.1. Análisis comparativo de arquitecturas

8.2.1.1. Rendimiento de VGG-19

El Cuadro 5 presenta las métricas de clasificación obtenidas para el modelo VGG-19 con salida multiclase. Se muestran los valores de precisión, *recall* y *f1-score* para cada clase,

desde la clase 0 (sano) con una precisión de 0.87 y recall de 0.93, hasta la clase 4 (RD proliferativa) con una precisión de 0.75 y recall de 0.48. También se incluyen la exactitud total, que es de 0.80, así como los promedios macro y ponderado para precisión, *recall* y *f1-score*, con valores de 0.60 y 0.79 respectivamente.

Cuadro 5: Reporte de clasificación VGG-19 multiclase

Clase	Precision	Recall	F1-score	Support
0 (Sano)	0.87	0.93	0.90	3859
1 (Leve)	0.25	0.23	0.24	364
2 (Moderado)	0.67	0.54	0.60	791
3 (Grave)	0.48	0.35	0.41	130
4 (RD Proliferativa)	0.75	0.48	0.58	104
Accuracy		0.80		5248
Macro Avg	0.60	0.51	0.55	5248
Weighted Avg	0.79	0.80	0.79	5248

El Cuadro 6 presenta las métricas de clasificación para el modelo VGG-19 con salida binaria, evaluando dos clases: "Sano"(Clase 0) y RD"(Retinopatía Diabética, Clase 1). Se incluyen las métricas de precisión, recall, F1-score y soporte para cada clase, así como los promedios macro y ponderado, junto con la precisión general (*accuracy*) del modelo. Cada fila muestra los valores específicos de estas métricas para las clases individuales y los promedios globales del modelo.

Cuadro 6: Reporte de clasificación VGG-19 (modelo binario)

Clase	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Clase 0 (Sano)	0.87	0.93	0.90	3859
Clase 1 (RD)	0.77	0.63	0.69	1389
Accuracy		0.85		5248
Macro Avg	0.82	0.78	0.80	5248
Weighted Avg	0.85	0.85	0.85	5248

La Figura 18 presenta la matriz de confusión de los resultados multiclase del modelo VGG-19. En esta matriz, las filas corresponden a las clases reales y las columnas a las clases predichas por el modelo. Los valores en las celdas indican la cantidad de instancias clasificadas correctamente o incorrectamente para cada combinación de clase real y predicha. La matriz proporciona una visión clara de cómo el modelo clasificó las diferentes clases, mostrando la precisión en la asignación de las categorías y el desempeño en la clasificación de cada una de ellas.

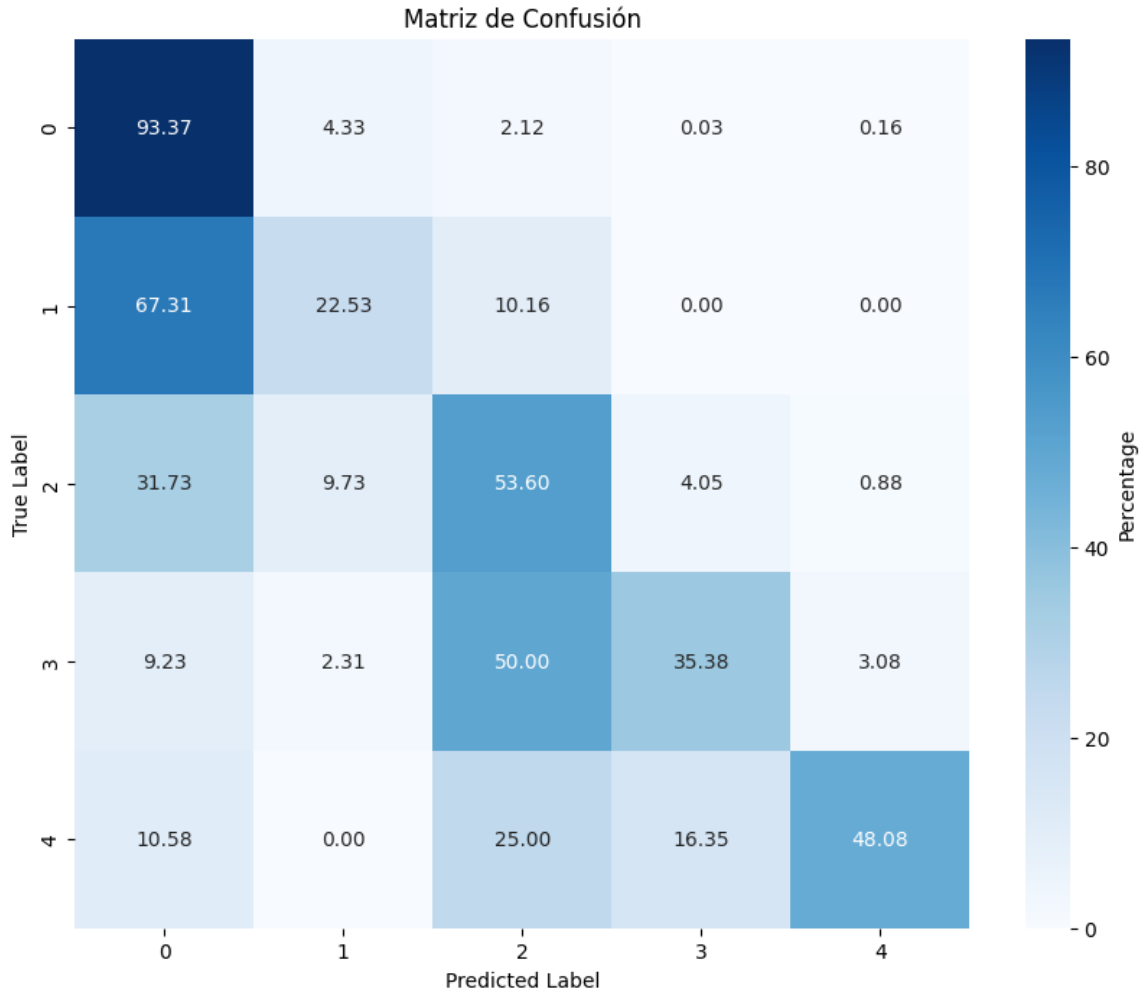


Figura 18: Matriz de confusión VGG-19 multiclase

La Figura 19 muestra la matriz de confusión de los resultados binarios del modelo VGG-19. En esta matriz, las filas representan las clases reales (sano y RD), y las columnas muestran las clases predichas por el modelo. Los valores en las celdas indican el número de instancias clasificadas correctamente o incorrectamente, distribuidas en las categorías de Verdaderos Positivos, Falsos Positivos, Verdaderos Negativos y Falsos Negativos. Esta matriz ayuda a visualizar el rendimiento del modelo en la clasificación binaria, mostrando cuántas instancias fueron correctamente identificadas y cuántas fueron mal clasificadas.

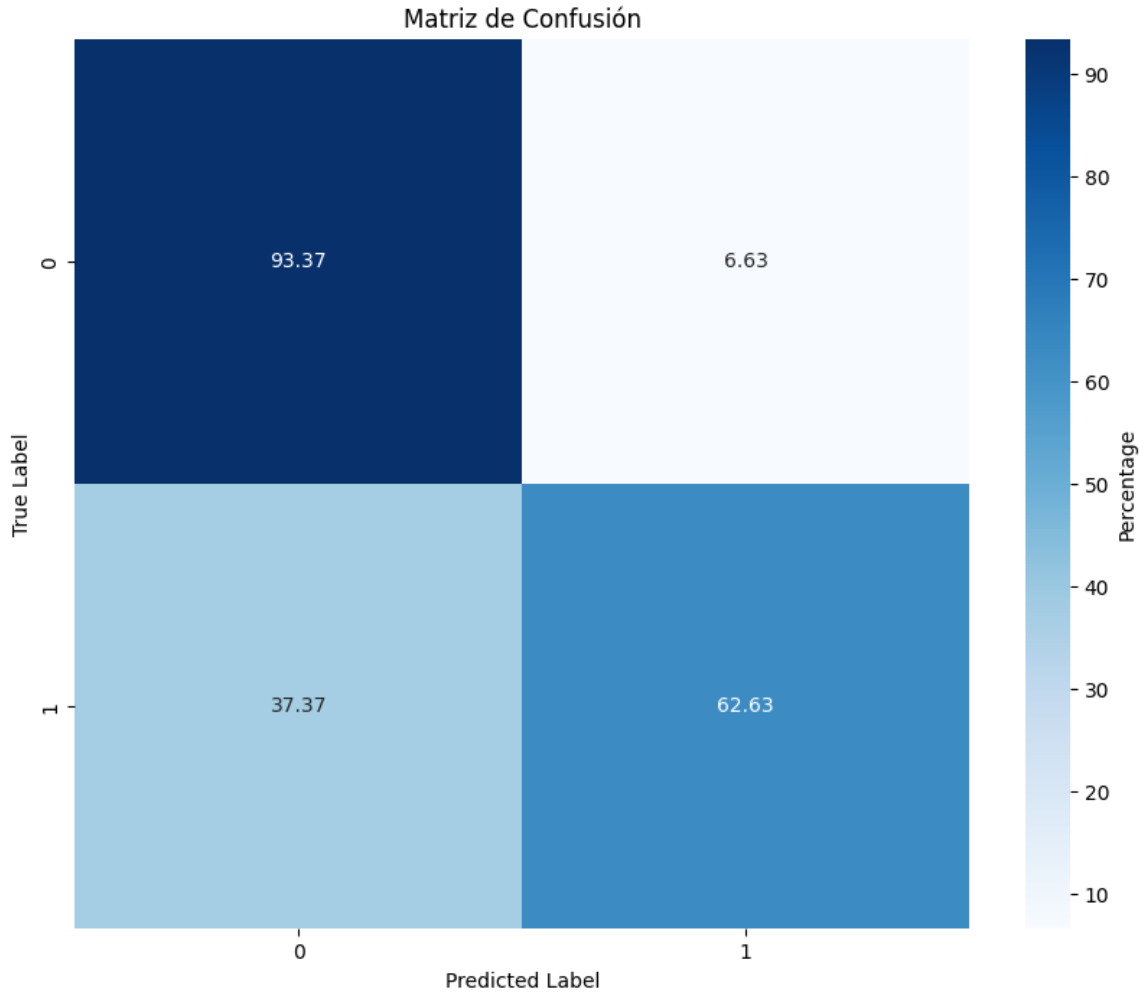


Figura 19: Matriz de confusión VGG-19 binario

Como se muestra en la Figura 20, la curva AUC PR (Área Bajo la Curva Precision-Recall) del rendimiento del modelo VGG-10. Esta curva ilustra la relación entre la precisión y el recall a medida que se varía el umbral de decisión del modelo. En el eje horizontal se representa el recall, mientras que en el eje vertical se muestra la precisión. El área bajo la curva (AUC PR) es una métrica que proporciona una evaluación global del desempeño del modelo en tareas de clasificación desbalanceadas, donde una mayor área bajo la curva indica un mejor rendimiento en términos de precisión y recall combinados.

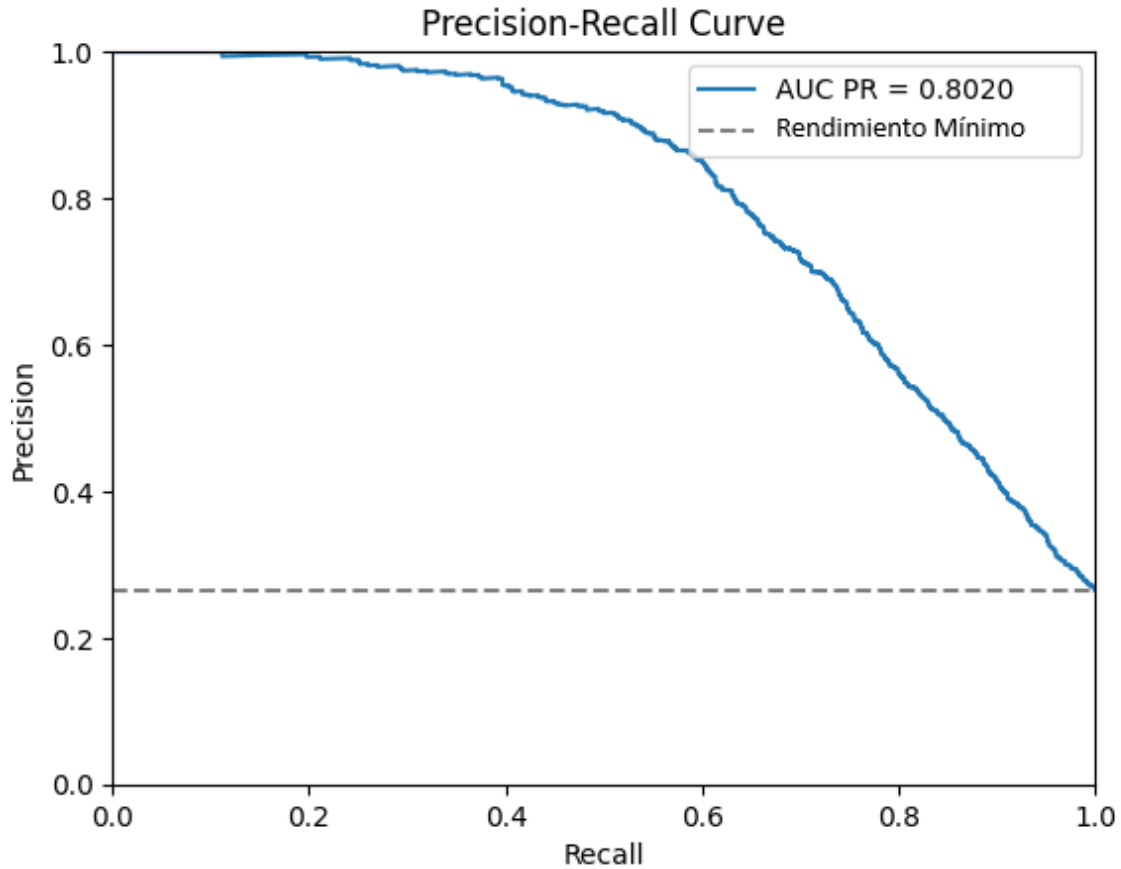


Figura 20: Curva AUC PR VGG-19

La Figura 21 muestra el rendimiento del modelo en términos de AUC PR (Área Bajo la Curva Precision-Recall) en el conjunto de validación a lo largo del entrenamiento. En el eje horizontal se representa el número de pruebas o las épocas de entrenamiento, mientras que en el eje vertical se muestra el valor de AUC PR obtenido en el conjunto de validación. Esta gráfica permite observar cómo evolucionó el rendimiento del modelo a medida que avanzaba el entrenamiento, proporcionando información sobre la estabilidad y mejora de las métricas de precisión y recall durante el proceso.

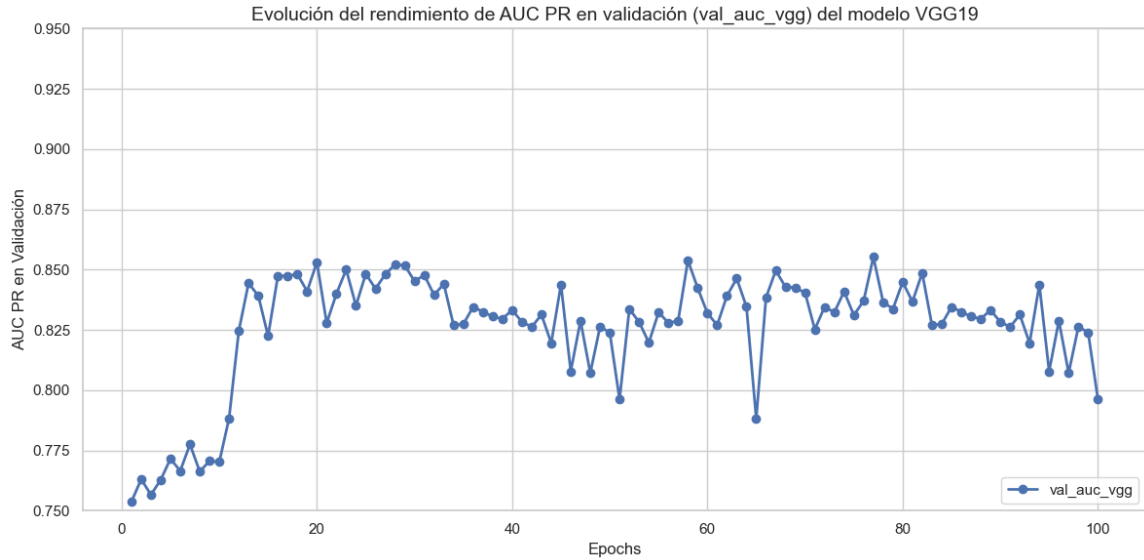


Figura 21: Evolución de AUC PR en conjunto de validación de VGG-19

La Figura 22 ilustra el desempeño del modelo en términos de AUC PR a lo largo de cada test realizado durante el entrenamiento. En el eje horizontal se presenta el número de test, mientras que el eje vertical muestra los valores de AUC PR alcanzados en cada uno. Esta gráfica permite visualizar la evolución del rendimiento del modelo durante el entrenamiento, ofreciendo información sobre la estabilidad y los cambios en las métricas de precisión y recall a lo largo del proceso. Así como su poder de generalización en el tiempo.

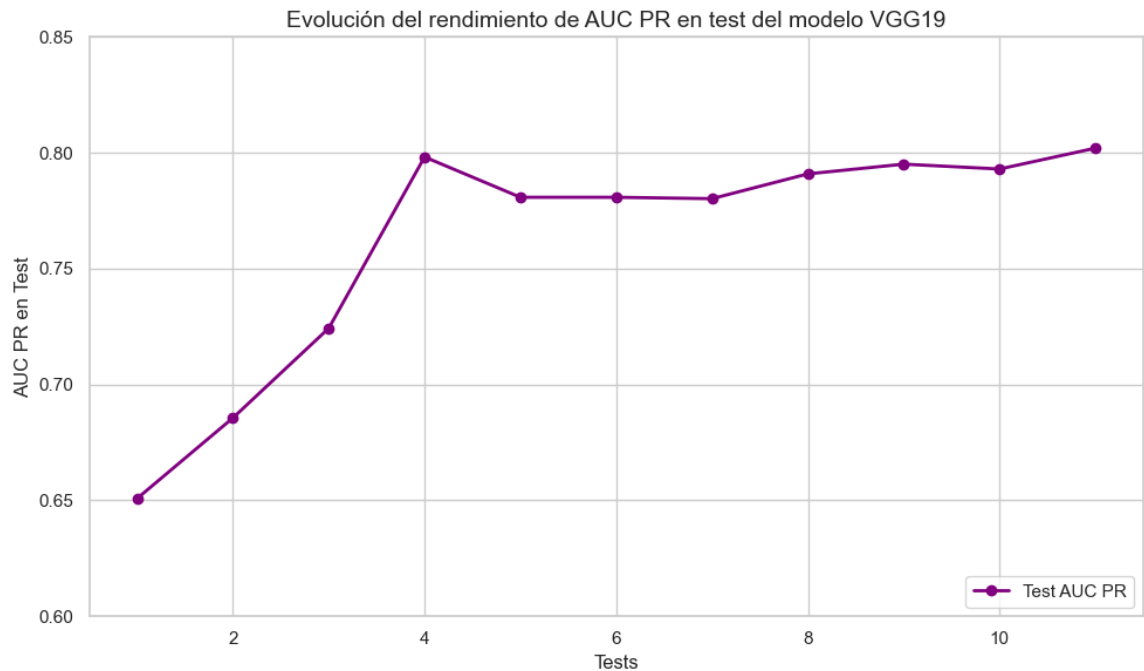


Figura 22: Evolución de AUC PR en test VGG-19

La Figura 23 muestra el comportamiento del error de pérdida en el conjunto de validación a lo largo del tiempo durante el entrenamiento. En el eje horizontal se representa el número de épocas, mientras que en el eje vertical se muestra el valor del error de pérdida en el conjunto de validación en cada época, es decir en este caso el *Categorical Cross-Entropy*. Esta visualización permite observar cómo se comporta la pérdida del modelo en el conjunto de validación durante el proceso de entrenamiento, proporcionando información sobre la convergencia y la estabilidad del modelo a medida que avanza el entrenamiento.

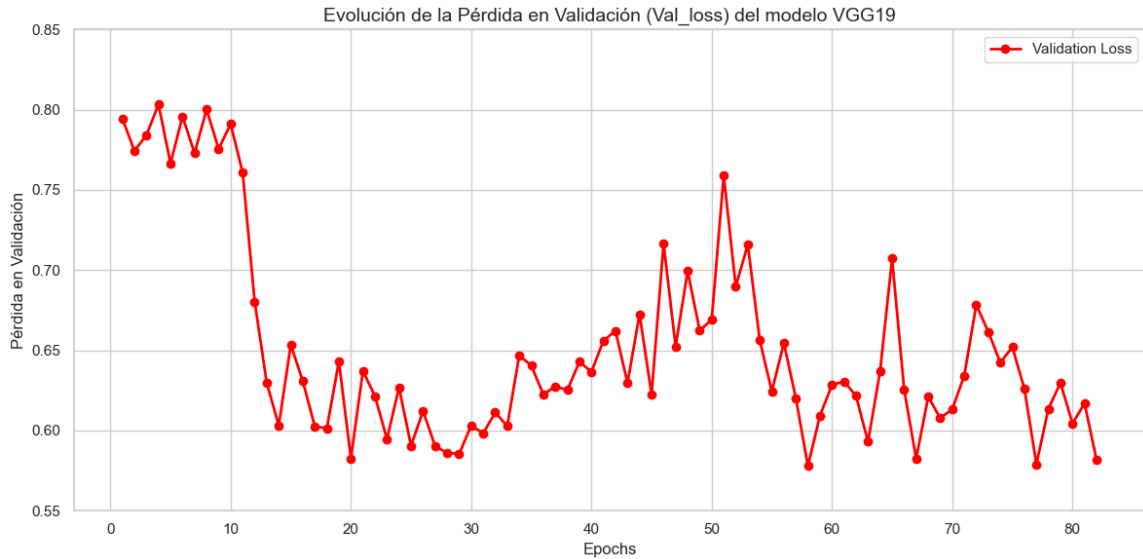


Figura 23: Evolución de pérdida en conjunto de validación VGG-19

En la Figura 24 se muestra el número de parámetros entrenables del modelo a lo largo de 100 épocas, con el proceso dividido en diferentes etapas. En el eje vertical se representa el número de parámetros entrenables en millones, mientras que el eje horizontal indica las épocas de entrenamiento. Durante las primeras épocas, se observa un número bajo de parámetros entrenables, que aumenta notablemente a medida que avanza el entrenamiento, reflejando las distintas etapas del proceso de Gradual Transfer Learning. La gráfica está dividida en áreas sombreadas que corresponden a diferentes fases del entrenamiento.

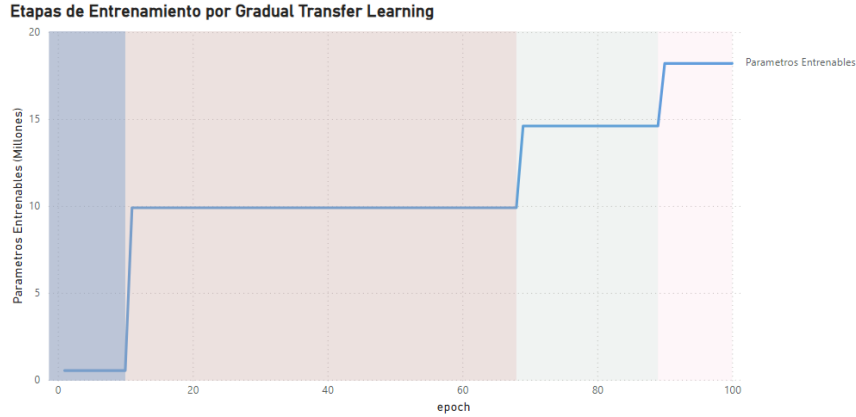


Figura 24: Etapas de *transfer learning* gradual de modelo VGG-19

Asimismo se calcularon los resultados de la métrica de concordancia entre el modelo y el criterio de clasificación *Quadratic Weighted Kappa*. Y se obtuvieron los resultados siguientes para cada modelo, como se puede observar en el Cuadro 7.

Cuadro 7: Resultados de quadratic weighted kappa para diferentes modelos

Modelo	QWK Binario	QWK Multiclase
VGG-19	0.600	0.680
ResNet152	0.3845	0.5667
EfficientNetB7	0.4281	0.5726

8.2.1.2. Rendimiento de ResNet152

El Cuadro 8 presenta las métricas de clasificación obtenidas para el modelo ResNet152 con salida multiclase. Se muestran los valores de precisión, recall y f1-score para cada clase, desde la clase 0 (Sano), hasta la clase 4 (RD Proliferativa). También se incluyen la exactitud total, así como los promedios macro y ponderado para precisión, recall y f1-score.

Además se puede se presentan las métricas obtenidas al convertir los resultados multiclase en las clasificaciones binarias del modelo ResNet152, esto en la Figura ??.

Cuadro 8: Reporte de clasificación ResNet152 multiclase

Clase	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
0 (Sano)	0.84	0.82	0.83	3859
1 (Leve)	0.13	0.26	0.17	364
2 (Moderado)	0.44	0.26	0.33	791
3 (Grave)	0.27	0.47	0.34	130
4 (RD proliferativo)	0.52	0.42	0.47	104
<i>Accuracy</i>		0.68		5248
Macro Avg	0.44	0.44	0.43	5248
Weighted Avg	0.71	0.68	0.69	5248

Cuadro 9: Reporte de clasificación ResNet152 (modelo binario)

Clase	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Clase 0 (Sano)	0.84	0.82	0.83	3859
Clase 1 (RD)	0.53	0.57	0.55	1389
<i>Accuracy</i>		0.75		5248
Macro Avg	0.69	0.70	0.69	5248
Weighted Avg	0.76	0.75	0.76	5248

Como se puede observar en la Figura ??, se presenta la curva AUC PR que refleja el rendimiento del modelo ResNet152. Esta curva muestra la relación entre la precisión y el recall al modificar el umbral de decisión del modelo. El eje horizontal representa el recall, mientras que el eje vertical muestra la precisión. El área bajo la curva (AUC PR) es una métrica que ofrece una evaluación global del desempeño del modelo en tareas de clasificación desequilibradas, donde un área mayor indica un mejor desempeño en cuanto a la combinación de precisión y recall.

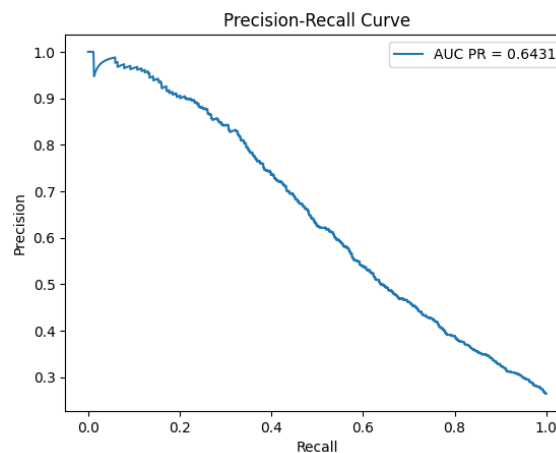


Figura 25: Curva AUC PR ResNet152

8.2.1.3. Rendimiento de EfficientNetB7

El Cuadro 10 muestra las métricas de clasificación obtenidas para el modelo EfficientNetB7 en su configuración multiclase. Se presentan los valores de precisión, recall y f1-score para cada clase, desde la clase 0 (Sano) hasta la clase 4 (RD Proliferativa). También se incluyen la exactitud general, así como los promedios macro y ponderado para precisión, recall y f1-score.

Adicionalmente, en la Figura ?? se exhiben las métricas obtenidas al convertir los resultados multiclase a clasificaciones binarias del modelo EfficientNetB7.

Cuadro 10: Reporte de clasificación EfficientNetB7 multiclase

Clase	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
0 (Sano)	0.82	0.93	0.87	3859
1 (Leve)	0.05	0.00	0.01	364
2 (Moderado)	0.43	0.20	0.27	791
3 (Grave)	0.26	0.60	0.36	130
4 (RD proliferativa)	0.36	0.55	0.43	104
Accuracy		0.74		5248
Macro Avg	0.38	0.46	0.39	5248
Weighted Avg	0.68	0.74	0.70	5248

Cuadro 11: Reporte de clasificación EfficientNetB7 (modelo binario)

Clase	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Clase 0 (Sano)	0.82	0.93	0.87	3859
Clase 1 (RD)	0.70	0.43	0.53	1389
Accuracy		0.80		5248
Macro Avg	0.76	0.68	0.70	5248
Weighted Avg	0.79	0.80	0.78	5248

La Figura 26 presenta la curva AUC PR, la cual representa el rendimiento del modelo EfficientNetB7. Esta curva muestra cómo varían la precisión y el recall al ajustar el umbral de decisión del modelo. En el eje horizontal se encuentra el recall, mientras que en el eje vertical se muestra la precisión. El área bajo la curva (AUC PR) sirve como una métrica para evaluar globalmente el rendimiento del modelo en problemas de clasificación desbalanceada, donde una mayor área indica un mejor desempeño en la combinación de precisión y recall.

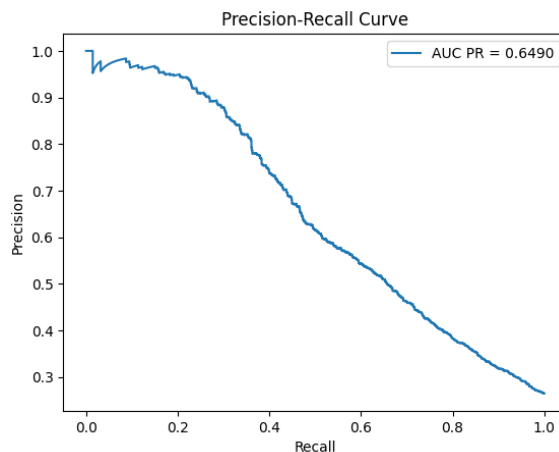


Figura 26: Curva AUC PR EfficientNetB7

8.3. Modelo en producción

El *dashboard* de producción para el diagnóstico preliminar de retinopatía diabética es una herramienta interactiva diseñada para apoyar a los profesionales de la salud en la detección temprana de esta enfermedad mediante el uso de inteligencia artificial. En la parte superior del panel, se presenta información relevante del modelo, como su versión actual y detalles del desarrollador. También incluye la licencia de uso, donde se especifican los términos y responsabilidades, y una sección de actualizaciones, que muestra los cambios o mejoras en cada nueva versión.

En la sección principal, el usuario tiene la opción de subir una imagen de fondo de ojo en formatos compatibles como JPG, JPEG o PNG. Una vez cargada la imagen, el sistema presenta dos visualizaciones: la imagen original y la imagen preprocesada, que es la que se ingresa al modelo para realizar el análisis. Esta comparación permite a los usuarios entender el proceso de preparación de la imagen antes de la evaluación por el modelo, como se puede observar en la Figura 27.

En la sección de resultados, el dashboard muestra dos gráficos importantes. El primero indica la probabilidad binaria de que el paciente tenga o no retinopatía diabética. Este resultado proporciona una predicción inicial que ayuda al profesional médico a decidir si es necesario realizar un análisis más profundo. El segundo gráfico desglosa la probabilidad de cada severidad de retinopatía, permitiendo identificar el nivel de gravedad en caso de que la enfermedad esté presente. Esta evaluación de severidad es útil para orientar las decisiones de tratamiento y seguimiento, ofreciendo un apoyo valioso en la gestión del paciente. Esto lo podemos observar en la Figura 28.

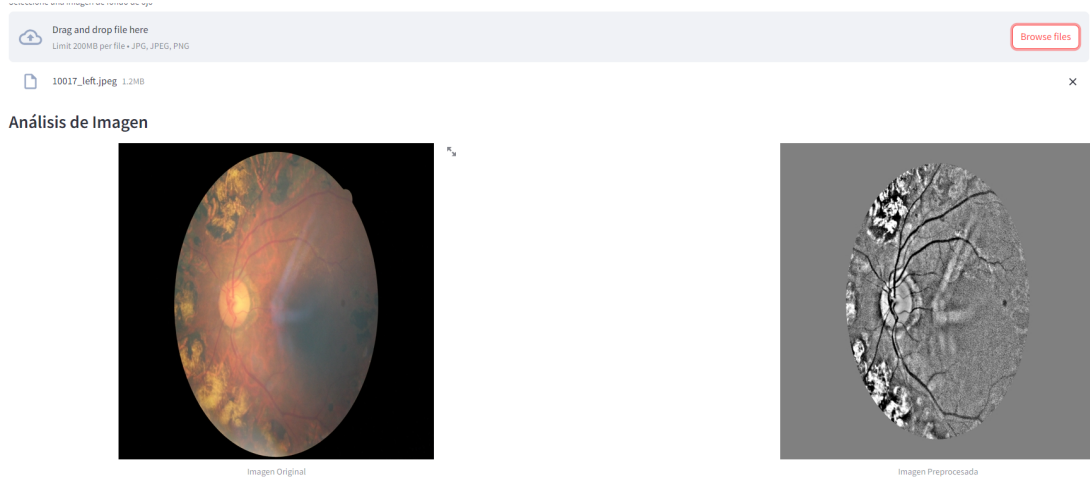
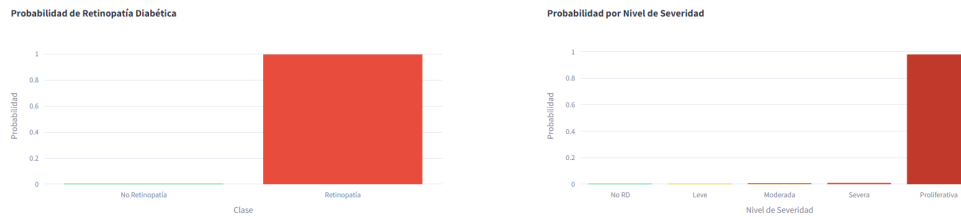


Figura 27: Visualización de imágenes de cara al usuario del *dashboard* de diagnóstico

Resultados del Análisis



Generar Reporte PDF

▲ Se detectaron signos de Retinopatía Diabética. Se recomienda consultar con un especialista para una evaluación detallada.

▲ IMPORTANTE: Este es un sistema de apoyo al diagnóstico. Las predicciones son estimaciones y no reemplazan el diagnóstico profesional. Siempre consulte a un especialista en oftalmología.

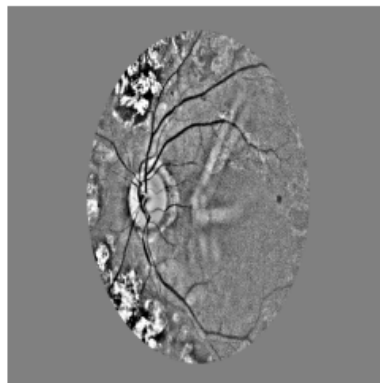
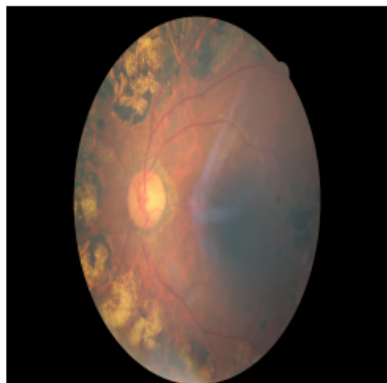


Figura 28: Gráficas de probabilidad binaria y multiclase de tenencia de enfermedad y severidad

Finalmente, como se ve en la Figura 29, si es necesario el usuario puede exportar los resultados obtenidos en un formato de PDF para poder guardar y llevar un control del diagnóstico preliminar obtenido en la plataforma para cada paciente.

Reporte de Diagnóstico - Retinopatía Diabética

Fecha: 2024-11-09 07:02:11



Resultados del Análisis:

Diagnóstico: Se detectaron signos de Retinopatía

Probabilidad de Retinopatía: 99.55%

Probabilidades por nivel de severidad:

No RD: 0.45%

Leve: 0.01%

Moderada: 0.61%

Severa: 1.09%

Proliferativa: 97.83%

Este reporte es generado automáticamente y debe ser validado por un profesional médico.

Figura 29: PDF preliminar generado por el *dashboard* de diagnóstico

Puedes ingresar al *dashboard* de diagnóstico preliminar en el siguiente enlace: [Etiquetador Imágenes](#)

Análisis de resultados

Al finalizar el proceso de entrenamiento fue realizado de manera paralela para las cuatro arquitecturas diferentes: *VGG-19*, *ResNet152*, *MDFSC* y *EfficientNetB7*, empleando unidades de procesamiento gráfico (GPU) y aplicando la técnica de *Fine Tuning Gradual*. Los resultados obtenidos durante esta fase no alcanzaron los niveles esperados de rendimiento para ninguna de las arquitecturas implementadas. Después de 50 épocas de entrenamiento, y ante la ausencia de mejoras significativas en el rendimiento, se procede a finalizar esta fase para realizar un análisis detallado de los resultados.

El análisis de los resultados muestra que la arquitectura *VGG-19* alcanza el mejor rendimiento entre los tres modelos evaluados, con un valor máximo de AUC PR de 0.53 en el conjunto de validación, como se observa en la Cuadro 3.

Las métricas detalladas del mejor modelo (*VGG-19*) se presentan en la Cuadro 4, donde se observa un rendimiento desequilibrado entre las clases. El modelo alcanza una precisión general de 0.61, con un mejor desempeño en la detección de casos negativos (clase 0) en comparación con los casos positivos (clase 1).

Los resultados obtenidos a través de los diferentes enfoques de modelado traen varios hallazgos relevantes. En el caso de los modelos binarios iniciales, se observó una dificultad para agrupar adecuadamente las imágenes con diferentes RD. Particularmente, los casos leves y moderados de la enfermedad presentaron una tendencia a ser clasificados erróneamente como sanos por estos modelos. Principalmente del modelo *MDFSC*, ya que con las condiciones actuales y métricas obtenidas, se observa que no se cumplen los supuestos de que el error de reconstrucción de una imagen anómala va a ser muy superior al error de reconstrucción de una imagen sana. Ya que como se observa en las Figuras 16 y 17. Por lo que, como se observa, no existe un rango de error de representación de las imágenes sanas que logre excluir a los errores de representación de las imágenes anómalas. Por lo que no se pudo conseguir identificar casos anómalos con esta metodología en la arquitectura actual.

Este comportamiento puede atribuirse a la distribución desequilibrada del conjunto de datos, donde los casos leves y moderados representaban la mayoría de las muestras con retinopatía. Ya que, aunque se empleó un balanceo mediante pesos por clase, la diferencia de cantidad y de pocos ejemplos de las clases positivas era importante. Al estar sesgados hacia estas categorías intermedias, los modelos binarios tuvieron dificultades para establecer límites de decisión claros entre las clases, confundiendo con frecuencia los casos leves y moderados con la clase sana. Esto también un factor importante puede ser la resolución de la imagen. Ya que mientras más resolución tendremos mejores características para poder observar pero un mayor costo computacional.

Al migrar a un enfoque multiclase, que permite una categorización más granular de los diferentes niveles de severidad, se observó una mejora significativa en el rendimiento general de los modelos. Esta aproximación multiclase facilitó que el modelo pudiera aprender a discriminar mejor entre las diversas etapas de la retinopatía diabética, incluso si existía cierta confusión entre las clases adyacentes (por ejemplo, entre leve y moderado). Al agrupar posteriormente los resultados multiclase en una clasificación binaria (sano vs. retinopatía), el modelo fue capaz de generalizar de manera más efectiva.

La arquitectura *VGG-19* demuestra un rendimiento superior, alcanzando un accuracy global de 80 % en la clasificación multiclase. El modelo exhibe particular efectividad en la identificación de casos sanos y casos de retinopatía proliferativa, como se observa en el Cuadro 5. Al transformar los resultados multiclase en una clasificación binaria (Sano vs. RD), el modelo mantiene un rendimiento robusto, como se evidencia en la Cuadro 6.

Un hallazgo interesante fue que la arquitectura *VGG-19*, a pesar de tener una menor cantidad de parámetros (20.5M) en comparación con *ResNet152* (60M) y *EfficientNetB7* (66M), demostró ser la más efectiva en este contexto. Esto sugiere que, en ciertos problemas, un modelo más sencillo pero bien entrenado puede llegar a superar a arquitecturas más complejas.

La arquitectura *ResNet152* muestra un rendimiento inferior al *VGG-19*, alcanzando un accuracy de 68 % en la clasificación multiclase y 75 % en la clasificación binaria derivada (Cuadros 8 y 9). Asimismo, la arquitectura *EfficientNetB7* presenta el rendimiento más bajo entre las tres arquitecturas evaluadas en la clasificación multiclase, aunque mantiene un rendimiento aceptable en la clasificación binaria con un accuracy de 80 % (Cuadros 10 y 11).

En este caso, es posible que el *VGG-19* haya logrado aprender de manera más eficiente las características relevantes para la clasificación de retinopatía diabética, al requerir un menor número de parámetros y, por lo tanto, una carga computacional y de entrenamiento más manejable. En contraste, las arquitecturas más profundas y complejas, como *ResNet152* y *EfficientNetB7*, pueden haber requerido un mayor esfuerzo de entrenamiento para alcanzar un nivel de rendimiento similar, sin llegar a superar al *VGG-19* en este escenario específico.

Como podemos observar en la Figura 21 las métricas obtenidas por el mejor modelo en el entrenamiento podemos observar como tuvo un crecimiento bastante abrupto.

Además, es importante considerar que la resolución de las imágenes utilizada en este estudio (600x600 píxeles) puede haber limitado la capacidad de los modelos más complejos para capturar detalles sutiles, especialmente en los casos de retinopatía leve. Es posible que al

aumentar la resolución de las imágenes, las arquitecturas más avanzadas puedan aprovechar mejor su mayor capacidad de modelado y mejorar su rendimiento en la clasificación de los casos leves y moderados, que representan la mayoría de las muestras con retinopatía en el conjunto de datos.

Por otro lado, la confusión entre las clases leves y moderadas puede deberse a que las diferencias entre estos niveles de severidad son más sutiles y difíciles de discriminar, incluso para un ojo humano experto. En este sentido, el enfoque multiclase demostró ser más efectivo, ya que permitió al modelo aprender a distinguir entre estos matices, aunque aún se observó cierta confusión entre las clases adyacentes. Al observar las diferentes etapas del entrenamiento en la Figura 24 podemos notar que alrededor del cuarto batch de prueba, el modelo comenzó a mostrar signos de estancamiento en su rendimiento. Aunque todavía se observó una mejora, esta fue significativamente más lenta en comparación con las primeras etapas del entrenamiento. Este cambio en el ritmo de mejora sugiere que el modelo pudo haber alcanzado un punto en el que las mejoras se volvieron menos pronunciadas, lo cual es común en muchos procesos de aprendizaje automático cuando el modelo se acerca a su capacidad máxima para aprender de los datos disponibles.

Los hallazgos de este estudio destacan la importancia de considerar enfoques multiclase para la clasificación de retinopatía diabética, ya que permiten una mejor discriminación entre los diferentes niveles de severidad de la enfermedad. Además, subrayan la relevancia de evaluar arquitecturas de redes neuronales convolucionales de complejidad variable, ya que un modelo más simple como *VGG-19* puede llegar a ser una opción efectiva y eficiente en determinados contextos. Asimismo, la exploración de estrategias de preprocesamiento y aumento de la resolución de las imágenes podría ser un área de interés para mejorar aún más el rendimiento de los modelos, especialmente en la clasificación de los casos leves y moderados de retinopatía.

En el análisis de los resultados de un modelo de clasificación para la retinopatía diabética, es importante considerar la variabilidad inherente a las evaluaciones humanas en esta tarea. La imagen de Google muestra de manera clara cómo la interpretación de los grados de retinopatía por parte de los oftalmólogos presenta una gran inconsistencia. Este aspecto es crucial, ya que define el límite entre qué se considera un diagnóstico leve, moderado, severo o proliferativo, y sugiere que, incluso entre expertos, la subjetividad juega un papel importante.

Los resultados obtenidos con los modelos *VGG-19*, *ResNet152* y *EfficientNetB7* revelan diferencias significativas en términos de rendimiento, reflejadas en los valores del Quadratic Weighted Kappa (QWK) tanto para las predicciones binarias como para las predicciones multiclase. Estos valores indican la concordancia entre las predicciones del modelo y las etiquetas reales, lo que es particularmente relevante en el contexto de un sistema de diagnóstico.

El modelo *VGG-19* muestra un rendimiento relativamente alto en comparación con los otros modelos, con un QWK de 0.600 para el modelo binario y de 0.68 para el multiclase. Esto sugiere que *VGG-19* es capaz de capturar patrones en las imágenes que corresponden de manera más precisa a las distintas categorías de severidad de la retinopatía diabética, particularmente en el enfoque multiclase. Esto indica que los resultados del modelo multiclase se alinean mejor con las evaluaciones de los expertos, mostrando una concordancia razonable.

Por otro lado, el modelo ResNet152 arroja los QWK más bajos, con 0.3845 en la clasificación binaria y 0.5667 en la multiclase. Este bajo desempeño podría estar relacionado con una menor capacidad del modelo para capturar características específicas de las imágenes de fondo de ojo asociadas a la retinopatía diabética, sugiriendo que ResNet152 no es el modelo óptimo para esta tarea en particular.

Finalmente, el modelo EfficientNetB7 alcanza valores intermedios de QWK, con 0.4281 para el binario y 0.5725 para el multiclase. Aunque su rendimiento es superior al de ResNet152, sigue siendo menor al de VGG-19. Esto indica que, si bien EfficientNetB7 logra capturar características relevantes para la clasificación, no alcanza el nivel de precisión que VGG-19 muestra en el análisis de las distintas etapas de la enfermedad. Como se observa en la Cuadro 7.

Este Cuadro refleja que el enfoque multiclase mejora la capacidad de los modelos para alinearse con las evaluaciones clínicas en todos los casos, pero particularmente en VGG-19, que muestra el rendimiento más alto y parece ser el modelo más adecuado para el diagnóstico de retinopatía diabética en este contexto.

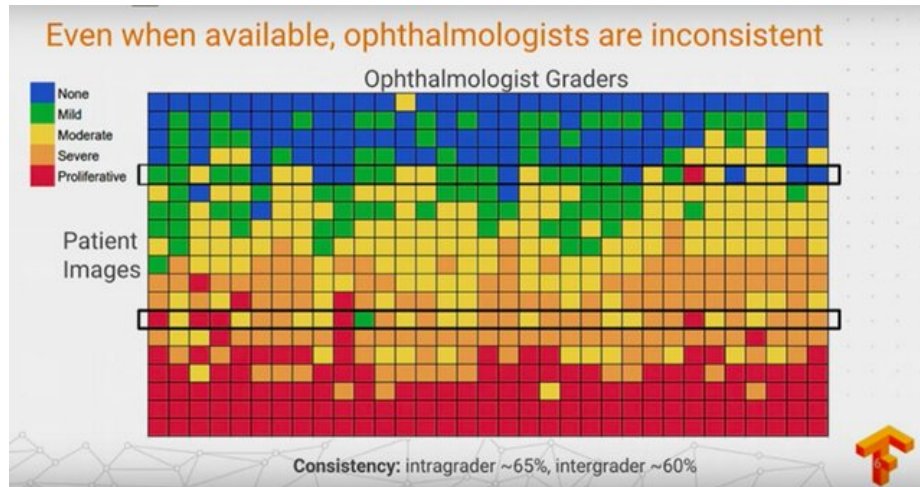


Figura 30: Limitantes de la variedad de criterios [61]

Google destaca que la evaluación de los oftalmólogos tiende a ser más consistente en los casos extremos (es decir, en los extremos de *sin retinopatía* y *retinopatía diabética proliferativa*) que en los casos intermedios (leve, moderado o grave). Esto se debe a que los signos de enfermedad en los extremos son más evidentes y reconocibles, mientras que las diferencias entre las etapas leves y moderadas son más sutiles y, por tanto, propensas a ser interpretadas de distintas maneras. Ya que es difícil poder estar de acuerdo y transmitir eso entre el equipo técnico y un grupo médico y viceversa, ya que no hay una manera objetiva y definitiva de definir leve, medio o grave. En la imagen, esto se representa con un alto grado de variabilidad en las categorías centrales, y solo se observa una relativa estabilidad en las categorías de los extremos.

Según la información proporcionada por Google 30, los oftalmólogos en el estudio presentaron una consistencia de aproximadamente 65% en sus propias evaluaciones (*intra-grader*) y de alrededor del 60% entre diferentes evaluadores (*inter-grader*). Estos porcentajes in-

dican que, en promedio, los expertos no siempre coinciden en sus diagnósticos y que sus evaluaciones pueden diferir considerablemente entre ellos y consigo mismos en evaluaciones repetidas. Este hallazgo refuerza la idea de que las evaluaciones son susceptibles a variaciones interpretativas y al juicio subjetivo de cada médico [61].

Este fenómeno tiene implicaciones directas en el desarrollo de modelos de inteligencia artificial para la clasificación de retinopatía diabética. Si bien los modelos pueden ser entrenados para realizar una clasificación multicategoría que distinga entre las distintas etapas de la enfermedad, los resultados obtenidos podrían reflejar la misma falta de consistencia observada entre los evaluadores humanos. En este sentido, es posible que el modelo *aprenda* no solo a identificar características de la enfermedad, sino también los sesgos e inconsistencias inherentes a los datos de entrenamiento generados a partir de diagnósticos humanos.

Si el rendimiento del modelo alcanza niveles similares de precisión (alrededor de un 60-65%) al evaluar las categorías intermedias, es importante contextualizar este resultado: no necesariamente indica un fallo del modelo, sino que refleja las limitaciones intrínsecas de la tarea misma. La ausencia de un *gold standard* o criterio universalmente aceptado para distinguir entre los niveles leves y moderados significa que cualquier modelo diseñado para hacer esta distinción estará, inevitablemente, sujeto a los mismos factores de variabilidad que afectan a los expertos humanos.

Además, este aspecto sugiere que los desarrolladores de modelos de diagnóstico de RD deben prestar especial atención a la definición de los criterios de clasificación en cada etapa de la enfermedad. Establecer definiciones claras y reproducibles es fundamental para reducir el sesgo en los resultados y para mejorar la consistencia del modelo. De lo contrario, el modelo podría replicar o incluso amplificar los sesgos inherentes en los datos de entrenamiento, lo que podría afectar su generalización y su aplicabilidad clínica.

La inconsistencia observada entre los evaluadores humanos subraya la importancia de establecer criterios objetivos y precisos para la clasificación de las etapas de la retinopatía diabética. Aunque el modelo de IA puede alcanzar niveles de rendimiento similares a los humanos, estos resultados deben interpretarse en el contexto de la variabilidad humana y la falta de un consenso claro en los criterios de clasificación. Así, el desarrollo de modelos de clasificación multicategoría para la retinopatía diabética requiere no solo un enfoque técnico, sino también un análisis crítico de los datos y las definiciones de las categorías clínicas para reducir el sesgo y mejorar la precisión diagnóstica.

Finalmente, el uso de un dashboard para visualizar y analizar los resultados del modelo de clasificación de retinopatía diabética puede ser de gran ayuda para los profesionales de la salud, ya que permite una interpretación rápida y una priorización de los casos de mayor gravedad. Sin embargo, para que este sistema sea aplicable en la práctica clínica real, es fundamental incorporar un filtro que evalúe la calidad de las imágenes de entrada y descarte aquellas que no cumplen con los estándares necesarios. Sin este control, el modelo podría intentar analizar imágenes de baja calidad, lo cual afectaría significativamente la precisión y fiabilidad de los resultados. La implementación de un filtro de calidad es, por lo tanto, un requisito urgente para asegurar que solo se utilicen imágenes válidas y adecuadas en el diagnóstico asistido, reduciendo así el riesgo de diagnósticos erróneos y mejorando la confianza en el sistema en un entorno clínico.

1. Se desarrolló un modelo de inteligencia artificial capaz de clasificar imágenes de fondo de ojo para detectar signos de retinopatía diabética, alineándose con el objetivo general de contribuir al diagnóstico temprano de esta enfermedad en pacientes de Guatemala. Los resultados evidenciaron que el modelo multiclase supera a los modelos binarios al discriminar entre niveles de severidad, mostrando una mayor precisión en la clasificación y mejorando la capacidad del sistema para proporcionar un diagnóstico preliminar más detallado.
2. Se implementó un modelo de inteligencia artificial diseñado específicamente para identificar retinopatía diabética en imágenes de fondo de ojo. Este modelo permitió detectar la presencia y severidad de la misma, contribuyendo significativamente a la identificación temprana de posibles problemas de salud ocular. Logrando en producción un desempeño de 81 % de precisión para la severidad y 86 % para la presencia. Así como un AUC ROC de 87 %, un AUC PR de 80 % cumpliendo las expectativas y un QWK de 60 y 68 para la presencia y severidad respectivamente.
3. La evaluación de las arquitecturas VGG19, ResNet152 y EfficientNetB7 demostró que VGG19, con menor cantidad de parámetros, logró un rendimiento superior en este contexto. Esta arquitectura obtuvo los valores de Kappa ponderado más altos, tanto en la clasificación binaria como en la multiclase, sugiriendo que un modelo más sencillo puede ser más efectivo y eficiente para esta tarea específica. Este hallazgo subraya la importancia de evaluar la complejidad del modelo en función de la tarea y del conjunto de datos utilizado.
4. Se implementó un *dashboard* en Streamlit que facilita la carga de imágenes de fondo de ojo por parte de clínicas, hospitales y profesionales de la salud. Esta plataforma mejora la accesibilidad del diagnóstico al permitir que cualquier persona con acceso a Internet pueda realizar una evaluación preliminar, independientemente de la ubicación o disponibilidad de especialistas, lo que es crucial para áreas rurales y de difícil acceso en Guatemala.

5. El diseño de un pipeline adaptable para incorporar nuevas fuentes de datos mediante *transfer learning* fortalece la capacidad del modelo para actualizarse y mejorar su rendimiento con futuros datos de mayor calidad o de diferentes contextos clínicos, aumentando su potencial de generalización y relevancia a largo plazo. Adicionalmente, se propone el uso de una herramienta de etiquetado de imágenes (*labeler*) para facilitar la clasificación de imágenes por parte de los médicos. Esta herramienta permitiría a los doctores colaborar en la anotación de datos, logrando una base de datos con etiquetas más precisas y adaptadas al contexto guatemalteco. Con este enfoque, se abre la posibilidad de establecer alianzas con organizaciones nacionales y de salud, para especializar el modelo con imágenes locales y convertirlo en una herramienta totalmente nativa y adaptada a las necesidades del país.
6. La resolución de las imágenes (600x600 píxeles) limitó la capacidad de los modelos más complejos para capturar detalles sutiles, especialmente en los casos leves de retinopatía diabética. Incrementar la resolución podría potencialmente mejorar la precisión del diagnóstico en las etapas tempranas, pero implica un mayor costo computacional. Este balance entre resolución y capacidad computacional es crucial para optimizar la eficiencia del sistema en entornos con recursos limitados.
7. La variabilidad observada entre las evaluaciones de los expertos humanos, particularmente en los niveles intermedios (leve, moderado y grave), indica que la retinopatía diabética carece de un criterio universalmente aceptado para diferenciar entre estas categorías. Por su significado de carácter subjetivo. Esta inconsistencia sugiere que el modelo también refleja esta falta de consenso, lo cual debe considerarse al interpretar su desempeño en las categorías intermedias. No obstante, los resultados obtenidos son comparables a la precisión alcanzada por evaluadores humanos, lo que refuerza el valor del modelo como herramienta de apoyo diagnóstico.
8. El desarrollo del *dashboard* también permitió la automatización del proceso diagnóstico, generando resultados preliminares de manera rápida y eficiente. Al integrar un modelo de inteligencia artificial, esta herramienta ofrece un diagnóstico automático de anomalías oculares, lo que reduce la necesidad de intervención manual y acelera la detección temprana de condiciones como la retinopatía diabética, optimizando el tiempo de respuesta en el ámbito clínico.

1. Explorar el uso de imágenes de mayor resolución para mejorar la detección en etapas tempranas, evaluando el impacto en el costo computacional y buscando un balance entre precisión y eficiencia.
2. Generar conjuntos de datos más equilibrados y definir criterios de clasificación más consistentes para reducir la variabilidad en las etiquetas, especialmente en las clases intermedias. Involucrar a un mayor número de oftalmólogos y generar consenso en las anotaciones podría ayudar a reducir el sesgo inherente a los datos de entrenamiento.
3. Validar el modelo en entornos clínicos con pacientes reales en Guatemala, evaluando su aplicabilidad y utilidad como herramienta de apoyo en el diagnóstico temprano de la retinopatía diabética en distintas poblaciones.
4. Dado el éxito de VGG19 en esta investigación, se recomienda explorar otras arquitecturas de menor complejidad que puedan ofrecer resultados similares, especialmente para su implementación en áreas con recursos computacionales limitados.
5. Dado que la tecnología y las necesidades médicas evolucionan, es recomendable realizar actualizaciones periódicas en el modelo y el dashboard para incorporar avances en el procesamiento de imágenes y mejorar la experiencia del usuario en el uso clínico de la herramienta.
6. Al no obtener buenos resultados con los modelos más complejos, se recomienda poder continuar el entrenamiento desbloqueando un mayor porcentaje de parámetros entrenables del total. Esto para darle la oportunidad a los modelos complejos de adaptarse mejor al problema.
7. Se recomienda integrar al dashboard de diagnóstico un modelo de verificación de calidad de imagen, con el fin de garantizar que las imágenes cargadas en la plataforma cumplan con los estándares necesarios de iluminación, enfoque y resolución. Esto permitiría obtener diagnósticos más precisos y confiables, asegurando que las imágenes sean de calidad adecuada para un análisis óptimo.

8. Se recomienda recolectar información adicional como sexo, edad y antecedentes médicos relevantes (duración de la diabetes, presencia de hipertensión y niveles de hemoglobina glucosilada) para robustecer el modelo de detección de retinopatía diabética. La plataforma de etiquetado puede extenderse para incluir esta nueva información durante la recolección de datos, permitiendo la creación de un dataset representativo de la población guatemalteca y mejorando la precisión del modelo al integrar factores de riesgo clínicos adicionales.

- [1] M. Khalifa y M. Albadawy, “AI in diagnostic imaging: Revolutionising accuracy and efficiency,” *Computer Methods and Programs in Biomedicine Update*, vol. 5, pág. 100 146, 2024, ISSN: 2666-9900. DOI: <https://doi.org/10.1016/j.cmpbup.2024.100146>. dirección: <https://www.sciencedirect.com/science/article/pii/S2666990024000132>.
- [2] “Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images,” *Scientific Reports*, vol. 11, 1 2021, ISSN: 20452322. DOI: 10.1038/s41598-021-81539-3.
- [3] B. A. Cooper, “Ten Key Elements of a Diabetes-Related Eye Examination,” *Compendia*, vol. 2022, n.º 3, págs. 4-7, jul. de 2022, ISSN: 2771-6872. DOI: 10.2337/db20223-4. eprint: <https://diabetesjournals.org/compendia/article-pdf/2022/3/4/686363/db202234.pdf>. dirección: <https://doi.org/10.2337/db20223-4>.
- [4] Y. Pan, J. Liu, Y. Cai et al., “Fundus image classification using Inception V3 and ResNet-50 for the early diagnostics of fundus diseases,” *Frontiers in Physiology*, vol. 14, 2023, ISSN: 1664042X. DOI: 10.3389/fphys.2023.1126780.
- [5] A. Kamińska, J. Pinkas, I. Wrześniewska-Wal, J. Ostrowski y M. Jankowski, “Awareness of Common Eye Diseases and Their Risk Factors—A Nationwide Cross-Sectional Survey among Adults in Poland,” *International Journal of Environmental Research and Public Health*, vol. 20, 4 2023, ISSN: 16604601. DOI: 10.3390/ijerph20043594.
- [6] R. R. Bourne, J. D. Steinmetz, S. Flaxman et al., “Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the Global Burden of Disease Study,” *The Lancet Global Health*, vol. 9, 2 2021, ISSN: 2214109X. DOI: 10.1016/S2214-109X(20)30425-3.
- [7] M. I. Orantes, J. M. Rodriguez, V. E. Villacinda et al., “Estilo de Vida y Retinopatía Diabética,” jul. de 2015.
- [8] K. Oh, H. M. Kang, D. Leem, H. Lee, K. Y. Seo y S. Yoon, “Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images,” *Scientific Reports*, vol. 11, 1 2021, ISSN: 20452322. DOI: 10.1038/s41598-021-81539-3.

- [9] G. M. S. Chávez, A. R. S. de Barrios, O. L. F. Pojoy et al., “National survey of blindness and visual impairment in Guatemala, 2015,” *Arquivos Brasileiros de Oftalmologia*, vol. 82, págs. 91-97, 2 2019, ISSN: 16782925. DOI: 10.5935/0004-2749.20190029.
- [10] M. V. Loza, “Caracterización de la retinopatía diabética en pacientes diabéticos en la unidad nacional de oftalmología,” *Universidad de San Carlos de Guatemala Facultad de Ciencias Médicas y Escuela de Estudio de Postgrado*, 2015.
- [11] EyePACS, *Data Analysis - EyePACS*, Accedido: 9 de noviembre de 2024, 2024. dirección: <https://www.eyepacs.com/data-analysis>.
- [12] D. Emma, Jared, Jorge y C. Will, *Diabetic Retinopathy Detection*. Kaggle, feb. de 2015. dirección: <https://www.kaggle.com/competitions/diabetic-retinopathy-detection/data>.
- [13] C. Matsoukas, J. F. Haslum, M. Sorkhei, M. Soderberg y K. Smith, “What Makes Transfer Learning Work for Medical Images: Feature Reuse Other Factors,” en *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2022-June, 2022. DOI: 10.1109/CVPR52688.2022.00901.
- [14] N. Gisev, J. S. Bell y T. F. Chen, *Interrater agreement and interrater reliability: Key concepts, approaches, and applications*, 2013. DOI: 10.1016/j.sapharm.2012.04.004.
- [15] R. A. Española, *Diccionario de la lengua española*, 2024. dirección: <https://dle.rae.es/oftalmolog%C3%ADa?m=form>.
- [16] J. García-Sánchez, “What is ophthalmology? Genesis of a specialty,” *ANALES RANM*, vol. 136, págs. 72-76, 01 mayo de 2019, ISSN: 00340634. DOI: 10.32440/ar.2019.136.01.dle03. dirección: https://analesranm.es/revista/2019/136_01/13601dle03.
- [17] H. Davson y E. S. Perkins. “Human eye.” *Encyclopedia Britannica*. (jul. de 2024), dirección: <https://www.britannica.com/science/human-eye>.
- [18] J. P. Berdahl, F. J. Martin y A. Y. Maa, “Advances in Eye Care,” *Eyenet Magazine*, dic. de 2020. dirección: <https://www.aao.org/eyenet/article/advances-in-eye-care>.
- [19] E. L. L. Cuéllar, “Caracterización de la Retinopatía Diabética,” Tesis de mtría., Universidad de San Carlos de Guatemala, Centro Universitario de Oriente, Chiquimula, Guatemala, 2016. dirección: https://cunori.edu.gt/descargas/Caracterizacion_de_la_Retinopatia_diabetica.pdf.
- [20] C. P. Wilkinson, F. L. Ferris, R. E. Klein et al., “Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales,” *Ophthalmology*, vol. 110, n.º 9, págs. 1677-1682, 2003.
- [21] World Health Organization, *Diabetes Mellitus: Report of a WHO Study Group* (Technical Report Series). Geneva: World Health Organization, 1985, vol. 727.
- [22] T. Y. Wong y C. Sabanayagam, “Strategies to tackle the global burden of diabetic retinopathy: from screening to treatment,” *The Lancet Diabetes & Endocrinology*, vol. 9, n.º 4, e5-e7, 2021. DOI: 10.1016/S2213-8587(20)30411-4.
- [23] Frontiers, “Classification of diabetic retinopathy: Past, present and future,” *Frontiers in Ophthalmology*, 2023, Available online: <https://www.frontiersin.org/articles/10.3389/fopht.2023.1062785/full>. dirección: <https://www.frontiersin.org/articles/10.3389/fopht.2023.1062785/full>.

- [24] A. Colenbrander, “Visual standards: aspects and ranges of vision loss with emphasis on population surveys,” *Report prepared for the International Council of Ophthalmology at the 29th International Congress of Ophthalmology, Sydney, 2002*.
- [25] D. R. R. P, *Diabetic Retinopathy: Can It Be Reversed?* Accessed: 2024-11-09, 2024. dirección: <https://neoretina.com/blog/diabetic-retinopathy-can-it-be-reversed/>.
- [26] M. Sachdeva, *Diabetic Retinopathy*, Accessed: 2024-11-01, 2024. dirección: <https://www.hopkinsmedicine.org/health/conditions-and-diseases/diabetes/diabetic-retinopathy#:~:text=Proliferative%20retinopathy%20is%20the%20advanced,retinopathy%20typically%20requires%20urgent%20treatment>.
- [27] A. D. Association, “Standards of medical care in diabetes—2022,” *Diabetes Care*, vol. 45, n.º Supplement_1, 2022.
- [28] R. Klein, M. D. Knudtson, K. E. Lee, R. Gangnon y B. E. K. Klein, “The epidemiology of diabetic retinopathy: a population-based study,” *Ophthalmology*, vol. 116, n.º 11, págs. 2182-2189, 2009.
- [29] J. J. Kanski y B. Bowling, *Clinical Ophthalmology: A Systematic Approach*, 8.^a ed. Elsevier Health Sciences, 2020.
- [30] M. D. Abramoff, P. T. Lavin, M. Birch, S. Shah y J. C. Folk, “Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices,” *npj Digital Medicine*, vol. 1, n.º 1, pág. 39, 2018.
- [31] D. S. Ting, C. Y. Cheung, G. Lim et al., “Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes,” *JAMA*, vol. 318, n.º 22, págs. 2211-2223, 2017.
- [32] J. J. Kanski, *Oftalmología Clínica*, 6ta. Elsevier España, 2009, pág. 916, ISBN: 8480864419, 9788480864411.
- [33] M. Karmel, J. S. Duker, D. Sarraf y B. Freund, *Retinal Imaging: Choosing the Right Method*, 2014. dirección: www.aaopt.org/retinavideos..
- [34] U. Food y D. Administration, *FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems*, 2023. dirección: <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye>.
- [35] R. Vought, V. Vought, M. Shah, B. Szirth y N. Bhagat, “EyeArt artificial intelligence analysis of diabetic retinopathy in retinal screening events,” *International Ophthalmology*, vol. 43, n.º 12, págs. 4851-4859, 2023, Epub 2023 Oct 17. DOI: 10.1007/s10792-023-02887-9.
- [36] L. R. Pasquale, *What will ai ophthalmology v2.0 look like?* 2021. DOI: 10.35119/maio.v3i1.120.
- [37] Shaveta, “A review on machine learning,” *International Journal of Science and Research Archive*, vol. 9, 1 2023. DOI: 10.30574/ijrsra.2023.9.1.0410.
- [38] M. Taboga, “Classification models”, *Lectures on probability theory and mathematical statistics*, Kindle Direct Publishing. *Online appendix*, 2021.

- [39] J.-G. Gaudreault, P. Branco y J. Gama, “An analysis of performance metrics for imbalanced classification,” en *International Conference on Discovery Science*, Springer, 2021, págs. 67-77.
- [40] S. Chilukoti, L. Shan y V. e. a. Tida, “A reliable diabetic retinopathy grading via transfer learning and ensemble learning with quadratic weighted kappa metric,” *BMC Medical Informatics and Decision Making*, vol. 24, pág. 37, 2024. DOI: 10.1186/s12911-024-02446-x.
- [41] Ž. Vujović, “Classification Model Evaluation Metrics,” *International Journal of Advanced Computer Science and Applications*, vol. 12, págs. 599-606, 6 2021, ISSN: 21565570. DOI: 10.14569/IJACSA.2021.0120670.
- [42] R. Obuchowicz, M. Oszust y A. Piorkowski, “Interobserver variability in quality assessment of magnetic resonance images,” *BMC Medical Imaging*, vol. 20, 1 2020, ISSN: 14712342. DOI: 10.1186/s12880-020-00505-z.
- [43] A. Sylolypavan, D. Sleeman, H. Wu y M. Sim, “The impact of inconsistent human annotations on AI driven clinical decision making,” *npj Digital Medicine*, vol. 6, 1 2023, ISSN: 23986352. DOI: 10.1038/s41746-023-00773-3.
- [44] S. Joshi y P. T. Karule, “Review of Preprocessing Techniques for Fundus Image Analysis,” *AMSE JOURNALS-AMSE IETA publication, Series: Advances B*, vol. 60, n.º 3, págs. 593-612, 2017. DOI: 10.18280/ama_b.600306.
- [45] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*. MIT Press, 2016.
- [46] Y. LeCun, L. Bottou, Y. Bengio y P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, 11 1998, ISSN: 00189219. DOI: 10.1109/5.726791.
- [47] V. Maeda-Gutiérrez, C. Galván Tejada, L. Zanella Calzada et al., “Comparison of Convolutional Neural Network Architectures for Classification of Tomato Plant Diseases,” *Applied Sciences*, vol. 10, pág. 1245, feb. de 2020. DOI: 10.3390/app10041245.
- [48] I. Dandekar, *Introduction to Convolutional Neural Networks - Part 1*, <https://medium.com/@ishandandekar/introduction-to-convolutional-neural-networks-part-1-c02b9fa3bcf2>, Accessed: 2024-10-16, 2023. dirección: <https://medium.com/@ishandandekar/introduction-to-convolutional-neural-networks-part-1-c02b9fa3bcf2>.
- [49] S. Ioffe y C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, 2015. arXiv: 1502.03167 [cs.LG]. dirección: <https://arxiv.org/abs/1502.03167>.
- [50] H. Pfaff y V. Chaikovska, “Deep Convolutional Nets,” sep. de 2020. DOI: 10.13140/RG.2.2.35229.90088.
- [51] L. Zhao y Z. Zhang, “An improved pooling method for convolutional neural networks,” *Nature*, vol. 14, pág. 1589, 2024. DOI: 10.1038/s41598-024-51258-6. dirección: <https://doi.org/10.1038/s41598-024-51258-6>.
- [52] K. He, X. Zhang, S. Ren y J. Sun, “Deep Residual Learning for Image Recognition,” *CoRR*, vol. abs/1512.03385, 2015. arXiv: 1512.03385. dirección: <http://arxiv.org/abs/1512.03385>.

- [53] K. Simonyan y A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*, 2015. arXiv: 1409.1556 [cs.CV]. dirección: <https://arxiv.org/abs/1409.1556>.
- [54] M. Tan y Q. V. Le, *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*, 2020. arXiv: 1905.11946 [cs.LG]. dirección: <https://arxiv.org/abs/1905.11946>.
- [55] S. D. Das, S. Dutta, N. A. Shah, D. Mahapatra y Z. Ge, “Anomaly Detection in Retinal Images using Multi-Scale Deep Feature Sparse Coding,” ene. de 2022. dirección: <http://arxiv.org/abs/2201.11506>.
- [56] N. M. Dipu, S. A. Shohan y K. Salam, “Ocular Disease Detection Using Advanced Neural Network Based Classification Algorithms,” *ASIAN JOURNAL OF CONVERGENCE IN TECHNOLOGY*, vol. 7, 2 2021. DOI: 10.33130/ajct.2021v07i02.019.
- [57] A. Gour, H. Raj y P. Agrawal, “Role of Convolutional Neural Networks in Diabetic Retinopathy Detection: A Comprehensive Review,” *Diagnostics (Basel)*, vol. 13, pág. 345, 3 2023, Accessed: 2024-11-09. dirección: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9914068/#sec3-diagnostics-13-00345>.
- [58] L.-M. Yuan, L. He y J.-L. Zhang, “Gradual Transfer Learning for Edge Intelligence: A Case Study of Image Classification,” *arXiv preprint arXiv:2109.11806*, 2021. dirección: <https://arxiv.org/pdf/2109.11806>.
- [59] S. Ha, S. Jung y J. Lee, *Domain-Aware Fine-Tuning: Enhancing Neural Network Adaptability*, 2024. arXiv: 2308.07728 [cs.LG]. dirección: <https://arxiv.org/abs/2308.07728>.
- [60] Keras, *Callbacks API*, <https://keras.io/api/callbacks/>, Accedido: 30-oct-2024.
- [61] B. Tymchenko, P. Marchenko y D. Spodarets, *Deep Learning Approach to Diabetic Retinopathy Detection*, mar. de 2020. DOI: 10.48550/arXiv.2003.02261.

13.1. Cuadros de diagnóstico

Escala internacional de severidad de la retinopatía diabética	
Grado de severidad	Hallazgos en fondo de ojo
Sin retinopatía aparente	Ninguna alteración.
Retinopatía no proliferativa leve	Microaneurismas (Figura 22).
Retinopatía no proliferativa moderada	Más microaneurismas pero menos alteraciones que en la RDNP severa (Figuras 23, 24, 25, 26).
Retinopatía no proliferativa severa	Más de 20 microaneurismas en 4 cuadrantes, tortuosidades venosas en 2 o más cuadrantes, AMVEs en al menos 1 cuadrante (Figura 27).
Retinopatía diabética proliferativa	Neovascularización y hemorragia vítrea o prerretiniana (Figuras 28, 31).

Escala internacional de severidad del edema macular diabético	
Edema macular	Hallazgos
Edema macular ausente	Sin exudados o engrosamiento retiniano en polo posterior.
Edema macular presente	Exudados o engrosamiento retiniano en polo posterior (Figura 29).
Leve	Diámetro del disco (Figura 30).
Moderado	Proximidad de la fovea pero sin afectarle la mácula (Figura 31).
Severo	Englobando el centro de la mácula (Figura 32, 34).

Edema macular diabético	
Tipo	Descripción
No clínicamente significativo	-
Clínicamente significativo (EMCS)	<ul style="list-style-type: none"> ▪ Engrosamiento retiniano en áreas de menos de 500 micras del centro de la fovea. ▪ Engrosamiento retiniano de al menos un diámetro de disco del tamaño de un diámetro papilar del centro de la fovea. ▪ Exudados duros en el polo posterior.

Retinopatía diabética proliferativa	
Grado	Hallazgos
Moderada	Neovascularización retiniana y/o neovascularización papilar de menos de 1/3 de área papilar.
De alto riesgo	<ul style="list-style-type: none"> ▪ Neovascularización papilar mayor a 1/3 de área papilar y/o hemorragia prerretiniana y vítrea. ▪ Neovascularización papilar de alto riesgo. (Figura 35).
Avanzada	Proliferación fibrovascular, desprendimiento de retina, hemorragia vítrea que ocluye el fondo (Figura 36).

Retinopatía diabética no proliferativa, severa y muy severa (regla del 4-2-1)	
Tipo	Descripción
Severa	<ul style="list-style-type: none"> ▪ Microaneurismas/hemorragias severas: 4 cuadrantes ▪ Rosarios venosos en al menos: 2 cuadrantes ▪ AMVEs en al menos: 1 cuadrante
Muy severa	2 o 3 de las anteriores

Cuadro 12: Escala internacional de severidad de la retinopatía diabética y edema macular diabético