



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**CENTRO DE FÍSICA APLICADA Y TECNOLOGÍA  
AVANZADA**

**IMPLEMENTACIÓN DE UNA RED NEURONAL  
CONVOLUCIONAL PARA LA DETECCIÓN DE  
RETINOPATÍA DIABÉTICA EN IMÁGENES DE  
FONDO OCULAR**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**Licenciado en tecnología**

**P R E S E N T A :**

**LUIS ÁNGEL ESCOBAR HERNÁNDEZ**

**DIRECTOR DE TESIS:**

**M en C. ALEJANDRO DE LEÓN CUEVAS**

**Juriquilla, Querétaro 2020**





Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# Agradecimientos

A la Universidad Nacional Autónoma de México por la formación que recibí durante mi periodo universitario.

A mis padres Pedro Escobar Roque y Rosa Elena Hernández Martínez por el aprendizaje, apoyo y amor que me han ofrecido a lo largo de mi vida.

A mis hermanos Roberto Jesús Escobar Hernández y Paloma Escobar Hernández por llegar a mi vida y llenarme de momentos únicos y especiales; los cuales siempre recordaré.

Al Laboratorio de Visualización Científica Avanzada por darme la oportunidad de desarrollar mi tema de investigación y ofrecerme todo el apoyo necesario; en especial a Alejandro De León Cuevas por todo el seguimiento y asesoría brindada; sin él este trabajo nunca habría sido posible, y a Luis Alberto Aguilar Bautista por todo el soporte técnico y sobre todo personal que me ofreció durante mi estancia en el laboratorio.

Al Instituto Mexicano de Oftalmología por confiar en mí y brindarme la orientación y el apoyo necesario para realizar y mejorar mi entendimiento del tema de investigación.

A la doctora Lucia Guadalupe Reyes Morales por el apoyo personal brindado para continuar con mi formación académica.

Por último, pero no menos importante, a todas las personas que conocí y formaron parte de mi proceso universitario ya que directa o indirectamente aprendí algo de ellos que será muy valioso para el resto de mi vida.

## Resumen

La retinopatía diabética es una enfermedad ocular común en pacientes diabéticos debido al daño de los vasos sanguíneos de la retina; su complicación puede provocar la ceguera parcial o total en el portador. Actualmente el médico puede realizar el diagnóstico de retinopatía diabética con imágenes de grado médico como pueden ser imágenes de fondo ocular, tomografía de coherencia óptica (OCT) o angiografía con fluorescencia. Los avances registrados en inteligencia artificial y en técnicas de adquisición de imágenes de grado médico, han permitido el desarrollo de herramientas auxiliares para el diagnóstico automático de enfermedades; entre ellas las oculares. El objetivo de este trabajo es implementar una Red Neuronal Convolutiva [ConvNet] para la detección de retinopatía diabética no proliferativa en imágenes de fondo ocular. Las Redes Neuronales Convolutivas han superado el desempeño de otros algoritmos de clasificación debido a que emplean información de las imágenes para construir una jerarquía de filtros no lineales que permiten distinguir las características relevantes, logrando así su categorización. Se trabajó con diferentes arquitecturas para comparar sus resultados al clasificar imágenes de pacientes sanos y enfermos. Se usaron conjuntos de imágenes públicas y privadas, de fondo ocular, para entrenar y validar los modelos. Todas las imágenes fueron sometidas a diferentes técnicas de procesamiento buscando mejorar la clasificación del prediagnóstico. Los resultados obtenidos demuestran que las Redes Neuronales Convolutivas permiten un prediagnóstico satisfactorio de retinopatía diabética. Se validó que la transferencia de aprendizaje mejora significativamente el desempeño de los modelos aún cuando se realiza a partir de imágenes significativamente diferentes a las de grado médico. Se demostró que es necesario utilizar el método de regiones de interés para identificar lesiones que se consideran microscópicas debido a la relación existente entre el área que las contiene y el área total de la imagen; como lo son las microaneurismas. El desempeño de este tipo de modelos está totalmente relacionado a la cantidad y variedad de datos empleados ya que afectan directamente la generalización y exactitud registrada en etapas de validación.

**Palabras clave** retinopatía diabética, redes neuronales convolutivas, ConvNets, imágenes médicas, GPUs.

# Índice de contenido

<b>Agradecimientos</b> .....	<b>I</b>
<b>Resumen</b> .....	<b>II</b>
<b>Introducción</b> .....	<b>1</b>
<b>1.1 Imagen del fondo ocular</b> .....	<b>1</b>
<b>1.2 Implementación de Inteligencia Artificial en la medicina</b> .....	<b>5</b>
<b>1.3 Aprendizaje automático</b> .....	<b>7</b>
1.3.1 Aprendizaje supervisado .....	7
1.3.2 Aprendizaje no supervisado .....	7
1.3.3 Redes Neuronales Convolucionales .....	8
<b>1.4 Diabetes</b> .....	<b>11</b>
<b>1.5 Retinopatía Diabética</b> .....	<b>11</b>
1.5.1 Etapas de la retinopatía diabética.....	11
<b>1.6 Objetivos</b> .....	<b>14</b>
1.6.1 Objetivo principal.....	14
1.6.2 Objetivos secundarios .....	14
<b>1.7 Hipótesis</b> .....	<b>14</b>
<b>1.8 Justificación</b> .....	<b>14</b>
<b>1.9 Planteamiento del problema</b> .....	<b>15</b>
<b>Antecedentes</b> .....	<b>16</b>
<b>2.1 Publicaciones relevantes</b> .....	<b>16</b>
<b>2.2 Aplicaciones y casos de éxito</b> .....	<b>25</b>
<b>Marco teórico</b> .....	<b>28</b>
<b>3.1 Procesamiento de imágenes</b> .....	<b>28</b>
3.1.1 Normalización .....	28
3.1.2 Contraste de ecualización de histograma adaptativo limitado [CLAHE] .....	28
3.1.3 Anotación de imágenes.....	29
<b>3.2 Redes Neuronales</b> .....	<b>31</b>

3.2.1	Entropía cruzada .....	33
3.2.2	Modelos no lineales .....	34
3.2.3	Tasa de aprendizaje .....	35
3.2.4	Algoritmos de optimización .....	37
3.2.5	Propagación y retropropagación.....	47
<b>3.3</b>	<b>Redes Neuronales Convolucionales .....</b>	<b>50</b>
3.3.1	Imágenes digitales.....	50
3.3.2	Conectividad local .....	51
3.3.3	Operadores .....	52
3.3.4	Técnicas de regularización .....	58
3.3.5	Capas de Clasificación .....	61
3.3.6	Transferencia de aprendizaje.....	62
3.3.7	Arquitectura de una ConvNet .....	64
3.3.8	Entrenamiento de ConvNets.....	66
<b>3.4</b>	<b>Datos .....</b>	<b>67</b>
3.4.1	Conjunto de datos.....	67
3.4.2	División de datos .....	68
3.4.3	Métricas de evaluación .....	69
	<b>Metodología .....</b>	<b>71</b>
<b>4.1</b>	<b>Creación del conjunto de datos .....</b>	<b>71</b>
4.1.1	Colecciones utilizadas .....	72
4.1.2	División de los datos .....	76
4.1.3	Preprocesamiento.....	77
<b>4.2</b>	<b>Definición de la Red Neuronal Convolutional .....</b>	<b>82</b>
4.2.1	Selección de la arquitectura.....	82
4.2.2	Selección de transformaciones para aumento de datos.....	83
4.2.3	Transferencia de aprendizaje.....	84
4.2.4	Definición del optimizador .....	84
4.2.5	Hiperparámetros.....	85
<b>4.3</b>	<b>Entrenamiento .....</b>	<b>85</b>
4.3.1	Configuración para la selección del mejor modelo .....	86
4.3.2	Gráficas de desempeño del modelo de clasificación .....	86
4.3.3	Definición de los modelos empleados .....	86

<b>4.4</b>	<b>Ajustes .....</b>	<b>88</b>
4.4.1	Revisión de métricas de evaluación y ajuste de hiperparámetros.....	88
<b>4.5</b>	<b>Validación .....</b>	<b>89</b>
4.5.1	Evaluación de datos de reserva.....	89
4.5.2	Evaluación de datos adicionales .....	89
4.5.3	Visualización del modelo.....	90
<b>4.6</b>	<b>Hardware y Software empleado .....</b>	<b>90</b>
4.6.1	Configuración del equipo .....	90
4.6.2	Entorno de trabajo .....	91
	<b><i>Resultados y discusión.....</i></b>	<b>92</b>
	<b><i>Conclusiones.....</i></b>	<b>120</b>
	<b><i>Bibliografía .....</i></b>	<b>122</b>
	<b><i>Apéndice .....</i></b>	<b>130</b>
<b>1.</b>	<b>Tablas con valores utilizados en resultados .....</b>	<b>130</b>
	<b>Inicialización aleatoria.....</b>	<b>130</b>
	<b>Inicialización aleatoria.....</b>	<b>132</b>
	<b>Inicialización con transferencia de aprendizaje .....</b>	<b>138</b>

# Índice de Figuras

FIGURA 1. PRIMERA IMAGEN DOCUMENTADA DE RETINA HUMANA DIBUJADA POR VAN TRIGT EN 1853. ....	1
FIGURA 2. DEL LADO IZQUIERDO SE OBSERVA PARTE DEL PROCESO PARA LA CAPTURA DEL FONDO OCULAR. DEL LADO DERECHO UN DIBUJO DE LA IMAGEN OBTENIDA. ....	3
FIGURA 3. GRADOS DE CAPTURA DEL FONDO OCULAR. ....	4
FIGURA 4. FOTO GRUPAL DE LOS PARTICIPANTES DEL SIMPOSIO AIRLIE HOUSE EN 1968. ....	4
FIGURA 5. EVOLUCIÓN EN EL DIAGNÓSTICO DE RETINOPATÍA DIABÉTICA. ....	6
FIGURA 6. TAREAS REALIZADAS EN EL DESAFÍO ILSVRC. ....	9
FIGURA 7. RESULTADOS REGISTRADOS EN DIFERENTES TAREAS DE LA COMPETENCIA ILSVRC. ....	10
FIGURA 8. COMPARACIÓN DE VISIÓN ENTRE UN PACIENTE SANO [IZQUIERDA] Y UN PACIENTE CON RETINOPATÍA DIABÉTICA [DERECHA]. ..	12
FIGURA 9. LESIONES CARACTERÍSTICAS DE RETINOPATÍA DIABÉTICA ....	13
FIGURA 10. SISTEMA IDX-DR PARA EL DIAGNÓSTICO DE RETINOPATÍA DIABÉTICA [37] ....	25
FIGURA 11. REPORTE DEL DIAGNOSTICO REALIZADO POR EL SISTEMA EYENUK. [38]. ....	26
FIGURA 12. ANOMALÍAS RELACIONADAS CON RD DETECTADAS AUTOMÁTICAMENTE CON EL SISTEMA PEGASUS [39]. ....	27
FIGURA 13. ANOTACIONES REALIZADAS MEDIANTE CUADRO DELIMITADOR. ....	29
FIGURA 14. ANOTACIONES POLIGONALES REALIZADAS EN UNA IMAGEN DE RETINA. ....	30
FIGURA 15. SEGMENTACIÓN SEMÁNTICA REALIZADA EN UNA IMAGEN DE RETINA. ....	30
FIGURA 16. DIAGRAMA DE UN PERCEPTRÓN MULTICAPA. ....	31
FIGURA 17. GRÁFICA DE LA FUNCIÓN SIGMOID. ....	32
FIGURA 18. PROBABILIDADES DE UN CONJUNTO DE DATOS. ....	33
FIGURA 19. ENTROPÍA CRUZADA DE LOS MODELOS PRESENTADOS EN LA FIGURA 18. ....	34
FIGURA 20. MODELO NO LINEAL. ....	35
FIGURA 21. MODELOS GRÁFICOS DE RED NEURONAL. ....	35
FIGURA 22. CLASIFICADOR LINEAL. ....	36
FIGURA 23. CLASIFICADOR LINEAL CON TASA DE APRENDIZAJE. ....	37
FIGURA 24. DESCENSO DE GRADIENTE POR LOTE. ....	40
FIGURA 25. DESCENSO DE GRADIENTE ESTOCÁSTICO. ....	41
FIGURA 26. DESCENSO DE GRADIENTE POR MINI LOTES. ....	42
FIGURA 27. MÍNIMO LOCAL Y PUNTO DE INFLEXIÓN DE UNA FUNCIÓN. ....	42
FIGURA 28. DESCENDIENTE DE GRADIENTE Y DESCENDIENTE DE GRADIENTE CON MOMENTO. ....	43
FIGURA 29. PROPAGACIÓN HACIA DELANTE. ....	47
FIGURA 30. RETROPROPAGACIÓN DE RED NEURONAL. ....	47
FIGURA 31. RED NEURONAL CON VALORES ARBITRARIOS. ....	48
FIGURA 32. ACTUALIZACIÓN DEL VALOR DE LOS PESOS. ....	50



FIGURA 33. DEL LADO IZQUIERDO SE PUEDE VER UNA IMAGEN EN EL ESPACIO DE COLOR RGB. DEL LADO DERECHO SE REPRESENTA UNA SECCIÓN DE LOS PÍXELES DE UN CANAL DEL ESPACIO DE COLOR RGB PERTENECIENTES A LA MISMA IMAGEN. ....	51
FIGURA 34. ENTRADA TOTALMENTE CONECTADA A LA CAPA OCULTA. ....	51
FIGURA 35. CONECTIVIDAD LOCAL EN REDES NEURONALES. ....	52
FIGURA 36. OPERACIÓN DE CONVOLUCIÓN EN UN ARREGLO MATRICIAL. ....	53
FIGURA 37. FILTROS EMPLEADOS EN EL PROCESAMIENTO DE IMÁGENES. ....	53
FIGURA 38. DEL LADO IZQUIERDO REPRESENTACIÓN NUMÉRICA DE UN FILTRO. DEL LADO DERECHO REPRESENTACIÓN GRÁFICA DEL MISMO FILTRO. ....	54
FIGURA 39. DE IZQUIERDA A DERECHA. IMAGEN CON REGIÓN DE INTERÉS SEÑALADA. MATRIZ DE UN CANAL DE LA REGIÓN DE INTERÉS. FILTRO PARA LA DETECCIÓN DE CURVAS. ....	54
FIGURA 40. DE IZQUIERDA A DERECHA. IMAGEN CON REGIÓN DE INTERÉS SEÑALADA. MATRIZ DE UN CANAL DE LA REGIÓN DE INTERÉS. FILTRO PARA LA DETECCIÓN DE CURVAS. ....	55
FIGURA 41. DEL LADO IZQUIERDO CONVOLUCIÓN CON PASO IGUAL A DOS. DEL LADO DERECHO OPERACIÓN DE CONVOLUCIÓN CON PASO IGUAL A UNO. ....	55
FIGURA 42. RELLENO CON CEROS A UNA IMAGEN DE 32 X 32 X 3. ....	56
FIGURA 43. FUNCIÓN DE ACTIVACIÓN RELU. ....	56
FIGURA 44. SALIDAS DE LA OPERACIÓN DE AGRUPACIÓN. ....	57
FIGURA 45. MODELO SIMPLE Y MODELO COMPLEJO. ....	58
FIGURA 46. AUMENTO DE DATOS EN IMÁGENES. ....	60
FIGURA 47. GRÁFICA DETECCIÓN TEMPRANA. ....	60
FIGURA 48. CAPA TOTALMENTE CONECTADA. ....	61
FIGURA 49. CAPA DE AGRUPACIÓN PROMEDIO GLOBAL. ....	62
FIGURA 50. PRIMERAS CAPAS DE UNA CONVNET. ....	62
FIGURA 51. ÚLTIMAS CAPAS DE UNA CONVNET. ....	63
FIGURA 52. TABLA DE DECISIÓN PARA TRANSFERENCIA DE APRENDIZAJE. ....	63
FIGURA 53. ARQUITECTURA INCEPTION V3. ....	65
FIGURA 54. ARQUITECTURA RESNET. ....	65
FIGURA 55. RELACIÓN ENTRE EXACTITUD Y NÚMERO DE OPERACIONES DE ARQUITECTURAS ESTÁNDAR. ....	66
FIGURA 56. PROCESO DE PROPAGACIÓN EN UNA CONVNET. ....	67
FIGURA 57. PORCENTAJES COMÚNMENTE UTILIZADOS PARA LA DIVISIÓN DE DATOS. LA REGIÓN VERDE CORRESPONDE A LOS DATOS DE ENTRENAMIENTO, LA REGIÓN ROJA A LOS DATOS DE VALIDACIÓN Y RESERVA. ....	69
FIGURA 58. METODOLOGÍA UTILIZADA PARA LA DETECCIÓN DE RETINOPATÍA DIABÉTICA. ....	71
FIGURA 59. IMAGEN DEL OJO IZQUIERDO Y DERECHO DE UN MISMO PACIENTE. ....	72
FIGURA 60. ANOTACIONES REALIZADAS EN UNA IMAGEN DE LA COLECCIÓN DIARETDB. ....	73
FIGURA 61. EJEMPLO DE VARIACIONES PRESENTES EN LAS IMÁGENES DE RETINA. ....	77

FIGURA 62. PASOS DEL PREPROCESAMIENTO.....	78
FIGURA 63. IMÁGENES DE RETINA RECORTADAS.....	78
FIGURA 64. IMÁGENES DE RETINA NORMALIZADAS.....	79
FIGURA 65. IMÁGENES DE RETINA ECUALIZADAS.....	80
FIGURA 66. DIVISIÓN DE IMAGEN ORIGINAL.....	80
FIGURA 67. ÁRBOL DE DECISIÓN PARA LA LOCALIZACIÓN DE ANOTACIONES.....	81
FIGURA 68. ANOTACIONES BAJO FORMATO DE COORDENADAS.....	82
FIGURA 69. GRÁFICA DE LAS MÉTRICAS DE EXACTITUD Y PÉRDIDA DE UN MODELO CLASIFICADOR.....	86
FIGURA 70. EVALUACIÓN DE OPTIMIZADORES EN UN CONJUNTO LIMITADO DE IMÁGENES DE RETINA.....	92
FIGURA 71. EXACTITUD DE LOS MODELOS UTILIZANDO DIFERENTES PROCESAMIENTOS Y ARQUITECTURAS. LA TERMINACIÓN _ROI CORRESPONDE A LOS RESULTADOS DE MODELOS ENTRENADOS CON REGIONES DE INTERÉS. LA TERMINACIÓN _ORIGINAL CORRESPONDE A LOS MODELOS ENTRENADOS CON LAS IMÁGENES ORIGINALES. LA ABREVIACIÓN TL SE UTILIZA EN LOS MODELOS QUE EMPLEARON TRANSFERENCIA DE APRENDIZAJE.....	93
FIGURA 72. RESULTADOS DE LAS PÉRDIDAS DE LOS MODELOS EMPLEADOS. LA TERMINACIÓN _ROI CORRESPONDE A LOS RESULTADOS DE MODELOS ENTRENADOS CON REGIONES DE INTERÉS. LA TERMINACIÓN _ORIGINAL CORRESPONDE A LOS MODELOS ENTRENADOS CON LAS IMÁGENES ORIGINALES. LA ABREVIACIÓN TL SE UTILIZA EN LOS MODELOS QUE EMPLEARON TRANSFERENCIA DE APRENDIZAJE.	94
FIGURA 73. EVALUACIÓN DE LOS MODELOS CON MAYOR EXACTITUD Y MENOR PÉRDIDA EN LA MÉTRICA F1. LA TERMINACIÓN _ROI CORRESPONDE A LOS RESULTADOS DE MODELOS ENTRENADOS CON REGIONES DE INTERÉS. LA TERMINACIÓN _ORIGINAL CORRESPONDE A LOS MODELOS ENTRENADOS CON LAS IMÁGENES ORIGINALES. LA ABREVIACIÓN TL SE UTILIZA EN LOS MODELOS QUE EMPLEARON TRANSFERENCIA DE APRENDIZAJE.....	95
FIGURA 74. EXACTITUD DE CADA PREPROCESAMIENTO EVALUADO CON LOS DATOS DE RESERVA. LA ABREVIATURA TL INDICA QUE LOS MODELOS FUERON ENTRENADOS CON TRANSFERENCIA DE APRENDIZAJE.....	96
FIGURA 75. GRÁFICA CON LOS VALORES DE SENSIBILIDAD OBTENIDOS PARA CADA PREPROCESAMIENTO CON LOS DATOS DE RESERVA. LA ABREVIATURA TL INDICA QUE LOS MODELOS FUERON ENTRENADOS CON TRANSFERENCIA DE APRENDIZAJE.....	97
FIGURA 76. GRÁFICA CON LOS VALORES DE ESPECIFICIDAD OBTENIDOS PARA CADA PREPROCESAMIENTO CON LOS DATOS DE RESERVA. LA ABREVIATURA TL INDICA QUE LOS MODELOS FUERON ENTRENADOS CON TRANSFERENCIA DE APRENDIZAJE.....	97
FIGURA 77. VISUALIZACIÓN DE IMAGEN DE RETINA. EL ÁREA EN COLOR ROJO ES LA QUE MAYOR PESO TIENE AL MOMENTO DE REALIZAR LA DECISIÓN. [1] IMAGEN ORIGINAL DE RETINA [2] VISUALIZACIÓN DEL MODELO DE DECISIÓN INCEPTIONV3 [3] VISUALIZACIÓN DEL MODELO DE DECISIÓN ResNET50.....	98
FIGURA 78. VISUALIZACIÓN DE IMAGEN DE RETINA. EL ÁREA EN COLOR ROJO ES LA QUE MAYOR PESO TIENE AL MOMENTO DE REALIZAR LA DECISIÓN. [1] IMAGEN ORIGINAL DE RETINA [2] VISUALIZACIÓN DEL MODELO DE DECISIÓN INCEPTIONV3 [3] VISUALIZACIÓN DEL MODELO DE DECISIÓN ResNET50.....	99
FIGURA 79. VISUALIZACIÓN DE IMAGEN DE RETINA. EL ÁREA EN COLOR ROJO ES LA QUE MAYOR PESO TIENE AL MOMENTO DE REALIZAR LA DECISIÓN. [1] IMAGEN ORIGINAL DE RETINA [2] VISUALIZACIÓN DEL MODELO DE DECISIÓN INCEPTIONV3 [3] VISUALIZACIÓN DEL MODELO DE DECISIÓN ResNET50.....	100

FIGURA 80. VISUALIZACIÓN DE IMAGEN DE RETINA. EL ÁREA EN COLOR ROJO ES LA QUE MAYOR PESO TIENE AL MOMENTO DE REALIZAR LA DECISIÓN. [1] IMAGEN ORIGINAL DE RETINA [2] VISUALIZACIÓN DEL MODELO DE DECISIÓN INCEPTIONV3 [3] VISUALIZACIÓN DEL MODELO DE DECISIÓN ResNet50. ....	101
FIGURA 81. IMAGEN DE RETINA ORIGINAL.....	102
FIGURA 82. COMPARACIÓN DE LA RESPUESTA DE LOS MODELOS DE DECISIÓN ENTRENADOS AL UTILIZAR DIFERENTES PREPROCESAMIENTOS. SE UTILIZA LA IMAGEN DE LA FIGURA 81 PARA SU COMPARACIÓN. [1] IMAGEN ORIGINAL EVALUADA CON LA ARQUITECTURA INCEPTIONV3 [2] IMAGEN ORIGINAL EVALUADA CON LA ARQUITECTURA ResNet50 [3] IMAGEN NORMALIZADA EVALUADA CON LA ARQUITECTURA INCEPTIONV3 [4] IMAGEN NORMALIZADA EVALUADA CON LA ARQUITECTURA ResNet50 [5] IMAGEN CON ECUALIZACIÓN DE HISTOGRAMA EVALUADA CON LA ARQUITECTURA INCEPTIONV3 [6] IMAGEN CON ECUALIZACIÓN DE HISTOGRAMA EVALUADA CON LA ARQUITECTURA ResNet50 [7] IMAGEN NORMALIZADA Y CON ECUALIZACIÓN DEL HISTOGRAMA EVALUADA CON LA ARQUITECTURA INCEPTIONV3 [8] IMAGEN NORMALIZADA Y CON ECUALIZACIÓN DEL HISTOGRAMA EVALUADA CON LA ARQUITECTURA ResNet50.....	102
FIGURA 83. EXACTITUD CON RELACIÓN A LOS PREPROCESAMIENTOS EMPLEADOS PARA LAS ARQUITECTURAS INCEPTION Y ResNet EN DATOS DE RESERVA. LA ABREVIATURA TL INDICA QUE LOS MODELOS FUERON ENTRENADOS CON TRANSFERENCIA DE APRENDIZAJE. ....	103
FIGURA 84. GRÁFICA CON LOS VALORES DE SENSIBILIDAD OBTENIDOS PARA CADA PREPROCESAMIENTO CON LOS DATOS DE RESERVA. LA ABREVIATURA TL INDICA QUE LOS MODELOS FUERON ENTRENADOS CON TRANSFERENCIA DE APRENDIZAJE. ....	104
FIGURA 85. GRÁFICA CON LOS VALORES DE ESPECIFICIDAD OBTENIDOS PARA CADA PREPROCESAMIENTO CON LOS DATOS DE RESERVA. LA ABREVIATURA TL INDICA QUE LOS MODELOS FUERON ENTRENADOS CON TRANSFERENCIA DE APRENDIZAJE. ....	105
FIGURA 86. VISUALIZACIÓN DE REGIÓN DE INTERÉS [ROI] OBTENIDA A PARTIR DE UNA IMAGEN ORIGINAL DE RETINA. [1] ROI ORIGINAL [2] ROI PREPROCESADA [3] VISUALIZACIÓN DEL MODELO DE DECISIÓN INCEPTIONV3 [4] VISUALIZACIÓN DEL MODELO DE DECISIÓN ResNet50.....	106
FIGURA 87. VISUALIZACIÓN DE REGIÓN DE INTERÉS [ROI] OBTENIDA A PARTIR DE UNA IMAGEN ORIGINAL DE RETINA. [1] ROI ORIGINAL [2] ROI PREPROCESADA [3] VISUALIZACIÓN DEL MODELO DE DECISIÓN INCEPTIONV3 [4] VISUALIZACIÓN DEL MODELO DE DECISIÓN ResNet50.....	106
FIGURA 88. VISUALIZACIÓN DE REGIÓN DE INTERÉS [ROI] OBTENIDA A PARTIR DE UNA IMAGEN ORIGINAL DE RETINA. [1] ROI ORIGINAL [2] ROI PREPROCESADA [3] VISUALIZACIÓN DEL MODELO DE DECISIÓN INCEPTIONV3 [4] VISUALIZACIÓN DEL MODELO DE DECISIÓN ResNet50.....	107
FIGURA 89. VISUALIZACIÓN DE REGIÓN DE INTERÉS [ROI] OBTENIDA A PARTIR DE UNA IMAGEN ORIGINAL DE RETINA. [1] ROI ORIGINAL [2] ROI PREPROCESADA [3] VISUALIZACIÓN DEL MODELO DE DECISIÓN INCEPTIONV3 [4] VISUALIZACIÓN DEL MODELO DE DECISIÓN ResNet50.....	108
FIGURA 90. REGIÓN DE INTERÉS EXTRAÍDA DE UNA IMAGEN DE RETINA.....	109
FIGURA 91. COMPARACIÓN DE LA RESPUESTA DE LOS MODELOS DE DECISIÓN ENTRENADOS AL UTILIZAR DIFERENTES PREPROCESAMIENTOS. SE UTILIZA LA IMAGEN DE LA FIGURA 92 PARA SU COMPARACIÓN. [1] ROI EVALUADA CON LA ARQUITECTURA INCEPTIONV3 [2] ROI EVALUADA CON LA ARQUITECTURA ResNet50 [3] ROI NORMALIZADA EVALUADA CON LA ARQUITECTURA INCEPTIONV3 [4] ROI NORMALIZADA EVALUADA CON LA ARQUITECTURA ResNet50 [5] ROI CON ECUALIZACIÓN DE HISTOGRAMA EVALUADA CON LA	

ARQUITECTURA INCEPTIONV3 [6] ROI CON ECUALIZACIÓN DE HISTOGRAMA EVALUADA CON LA ARQUITECTURA ResNet50 [7] ROI NORMALIZADA Y CON ECUALIZACIÓN DEL HISTOGRAMA EVALUADA CON LA ARQUITECTURA INCEPTIONV3 [8] ROI NORMALIZADA Y CON ECUALIZACIÓN DEL HISTOGRAMA EVALUADA CON LA ARQUITECTURA ResNet50.....	109
FIGURA 92. GRÁFICA CON LOS VALORES DE ESPECIFICIDAD OBTENIDOS PARA CADA PREPROCESAMIENTO CON DATOS DE RESERVA DEL CONJUNTO IDRID [54]. LA ABREVIACIÓN INC HACE REFERENCIA A LA ARQUITECTURA INCEPTION.....	110
FIGURA 93. GRÁFICA CON LOS VALORES DE ESPECIFICIDAD OBTENIDOS PARA CADA PREPROCESAMIENTO CON DATOS DE RESERVA DEL CONJUNTO IMO [23]. LA ABREVIACIÓN INC HACE REFERENCIA A LA ARQUITECTURA INCEPTION.....	111
FIGURA 94. GRÁFICA CON LOS VALORES DE ESPECIFICIDAD OBTENIDOS PARA CADA PREPROCESAMIENTO CON DATOS DE RESERVA DEL CONJUNTO MESSIDOR [55]. LA ABREVIACIÓN INC HACE REFERENCIA A LA ARQUITECTURA INCEPTION. ....	112
FIGURA 95. COMPARACIÓN DE PREDICCIONES. [1] IMAGEN DEL CONJUNTO E_OPHTHA [53] CLASIFICADA COMO ENFERMA. [2] MAPA DE CALOR GENERADO CON INFORMACIÓN DE LA IMAGEN [1]. [3] IMAGEN DEL CONJUNTO IMO [23] CLASIFICADA COMO ENFERMA. [4] MAPA DE CALOR GENERADO CON INFORMACIÓN DE LA IMAGEN [3]. ....	114
FIGURA 96. IMAGEN DE RETINA CLASIFICADA COMO LEVE. ....	115
FIGURA 97. IMAGEN DE RETINA CLASIFICADA COMO LEVE .....	115

# Índice de Tablas

TABLA 1. ESCALA CLÍNICA INTERNACIONAL EMITIDA POR LA ACADEMIA AMERICANA DE OFTALMOLOGÍA (AOO). .....	13
TABLA 2. COMPILACIÓN DE PUBLICACIONES SOBRE DETECCIÓN DE RETINOPATÍA DIABÉTICA DE MANERA AUTÓNOMA. ....	23
TABLA 3. REFERENCIA DE ABREVIATURAS EMPLEADAS EN LA TABLA 2. ....	24
TABLA 4. COMPARACIÓN DE LA FUNCIÓN DESCENDIENTE DE GRADIENTE. ....	43
TABLA 5. COMPARACIÓN DESCENDIENTE DE GRADIENTE Y ADAGRAD. ....	44
TABLA 6. COMPARACIÓN ADAGRAD Y RMSPROP. ....	45
TABLA 7. FUNCIONES DEL OPTIMIZADOR ADAM. ....	46
TABLA 8. ELEMENTO DE CORRECCIÓN EN OPTIMIZADOR ADAM. ....	46
TABLA 9. COLECCIONES DE IMÁGENES DE RETINOPATÍA DIABÉTICA. *RDNP RETINOPATÍA DIABÉTICA NO PROLIFERATIVA +N/D INFORMACIÓN NO DISPONIBLE. ....	75
TABLA 10. RESUMEN DE LOS CONJUNTOS DE DATOS DEFINIDOS. ....	77
TABLA 11. INFORMACIÓN DE LAS ARQUITECTURAS EMPLEADAS. ....	83
TABLA 12. AUMENTO DE DATOS. ....	83
TABLA 13. OPTIMIZADORES EMPLEADOS PARA EL ENTRENAMIENTO DE LA RED NEURONAL CONVOLUCIONAL.....	85
TABLA 14. HIPERPARÁMETROS INICIALES DE ENTRENAMIENTO. ....	85
TABLA 15. MODELOS UTILIZADOS PARA COMPARAR RESULTADOS DE OPTIMIZADORES. ....	87
TABLA 16. MODELOS ENTRENADOS PARA LA DETECCIÓN DE RETINOPATÍA DIABÉTICA.....	87
TABLA 17. TIPOS DE PREPROCESAMIENTO REALIZADO A LAS IMÁGENES.....	88
TABLA 18. CARACTERÍSTICAS DEL EQUIPO .....	90
TABLA 19. HIPERPARÁMETROS FINALES DE ENTRENAMIENTO.....	116
TABLA 20. TABLA CON TIEMPOS DE EJECUCIÓN DE LOS MODELOS UTILIZADOS. ....	119
TABLA 21. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES SIN PREPROCESAMIENTOS. ....	130
TABLA 22. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA. ....	130
TABLA 23. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES NORMALIZADAS. ....	130
TABLA 24. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES NORMALIZADAS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA.....	131
TABLA 25. RESULTADOS DE EVALUAR REGIONES DE INTERÉS SIN PREPROCESAMIENTOS. ....	131
TABLA 26. RESULTADOS DE EVALUAR REGIONES DE INTERÉS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA. ....	131
TABLA 27. RESULTADOS DE EVALUAR REGIONES DE INTERÉS NORMALIZADAS. ....	131
TABLA 28. RESULTADOS DE EVALUAR REGIONES DE INTERÉS NORMALIZADAS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA. ....	131
TABLA 29. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES SIN PREPROCESAMIENTO DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	132
TABLA 30. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES SIN PREPROCESAMIENTO DEL CONJUNTO DE RESERVA. ....	132
TABLA 31. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR IMÁGENES ORIGINALES SIN PREPROCESAMIENTO DE LOS DATOS DE RESERVA. ....	132

TABLA 32. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	132
TABLA 33. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DEL CONJUNTO DE RESERVA. ....	133
TABLA 34. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR IMÁGENES ORIGINALES CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS DATOS DE RESERVA.....	133
TABLA 35. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES NORMALIZADAS DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. .	133
TABLA 36. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES NORMALIZADAS DEL CONJUNTO DE RESERVA. ....	133
TABLA 37. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR IMÁGENES ORIGINALES NORMALIZADAS DE LOS DATOS DE RESERVA. ....	134
TABLA 38. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES NORMALIZADAS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	134
TABLA 39. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES NORMALIZADAS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DEL CONJUNTO DE RESERVA. ....	134
TABLA 40. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR IMÁGENES ORIGINALES NORMALIZADAS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS DATOS DE RESERVA.....	135
TABLA 41. RESULTADOS DE EVALUAR REGIONES DE INTERÉS SIN PREPROCESAMIENTOS DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	135
TABLA 42. RESULTADOS DE EVALUAR REGIONES DE INTERÉS SIN PREPROCESAMIENTOS DEL CONJUNTO DE RESERVA. ....	135
TABLA 43. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR REGIONES DE INTERÉS SIN PREPROCESAMIENTOS DE LOS DATOS DE RESERVA.	135
TABLA 44. RESULTADOS DE EVALUAR REGIONES DE INTERÉS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	136
TABLA 45. RESULTADOS DE EVALUAR REGIONES DE INTERÉS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DEL CONJUNTO DE RESERVA. ....	136
TABLA 46. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR REGIONES DE INTERÉS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS DATOS DE RESERVA. ....	136
TABLA 47. RESULTADOS DE EVALUAR REGIONES DE INTERÉS NORMALIZADAS DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN....	137
TABLA 48. RESULTADOS DE EVALUAR REGIONES DE INTERÉS NORMALIZADAS DEL CONJUNTO DE RESERVA. ....	137
TABLA 49. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR REGIONES DE INTERÉS NORMALIZADAS DE LOS DATOS DE RESERVA. ....	137
TABLA 50. RESULTADOS DE EVALUAR REGIONES DE INTERÉS NORMALIZADAS Y CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	137
TABLA 51. RESULTADOS DE EVALUAR REGIONES DE INTERÉS NORMALIZADAS Y CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DEL CONJUNTO DE RESERVA. ....	138
TABLA 52. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR REGIONES DE INTERÉS NORMALIZADAS Y CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS DATOS DE RESERVA.....	138

TABLA 53. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES SIN PREPROCESAMIENTOS DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	138
TABLA 54. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES SIN PREPROCESAMIENTOS DEL CONJUNTO DE RESERVA. ....	138
TABLA 55. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR IMÁGENES ORIGINALES SIN PREPROCESAMIENTO DE LOS DATOS DE RESERVA. ....	139
TABLA 56. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	139
TABLA 57. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DEL CONJUNTO DE RESERVA. ....	139
TABLA 58. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR IMÁGENES ORIGINALES CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS DATOS DE RESERVA. ....	140
TABLA 59. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES NORMALIZADAS DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	140
TABLA 60. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES NORMALIZADAS DEL CONJUNTO DE RESERVA. ....	140
TABLA 61. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR IMÁGENES ORIGINALES NORMALIZADAS DE LOS DATOS DE RESERVA. ....	140
TABLA 62. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES NORMALIZADAS Y CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	141
TABLA 63. RESULTADOS DE EVALUAR IMÁGENES ORIGINALES NORMALIZADAS Y CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DEL CONJUNTO DE RESERVA. ....	141
TABLA 64. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR IMÁGENES ORIGINALES NORMALIZADAS Y CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS DATOS DE RESERVA. ....	141
TABLA 65. RESULTADOS DE EVALUAR REGIONES DE INTERÉS SIN PREPROCESAMIENTO DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	141
TABLA 66. RESULTADOS DE EVALUAR REGIONES DE INTERÉS SIN PREPROCESAMIENTO DEL CONJUNTO DE RESERVA. ....	142
TABLA 67. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR REGIONES DE INTERÉS SIN PREPROCESAMIENTO DE LOS DATOS DE RESERVA. ....	142
TABLA 68. RESULTADOS DE EVALUAR REGIONES DE INTERÉS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	142
TABLA 69. RESULTADOS DE EVALUAR REGIONES DE INTERÉS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DEL CONJUNTO DE RESERVA. ....	142
TABLA 70. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR REGIONES DE INTERÉS CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS DATOS DE RESERVA. ....	143
TABLA 71. RESULTADOS DE EVALUAR REGIONES DE INTERÉS NORMALIZADAS DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	143
TABLA 72. RESULTADOS DE EVALUAR REGIONES DE INTERÉS NORMALIZADAS DEL CONJUNTO DE RESERVA. ....	143
TABLA 73. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR REGIONES DE INTERÉS NORMALIZADAS DE LOS DATOS DE RESERVA. ....	144
TABLA 74. RESULTADOS DE EVALUAR REGIONES DE INTERÉS NORMALIZADAS Y CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS CONJUNTOS DE ENTRENAMIENTO Y VALIDACIÓN. ....	144

TABLA 75. RESULTADOS DE EVALUAR REGIONES DE INTERÉS NORMALIZADAS Y CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DEL CONJUNTO DE RESERVA. ....	144
TABLA 76. MATRIZ DE CONFUSIÓN OBTENIDA AL EVALUAR REGIONES DE INTERÉS NORMALIZADAS Y CON ECUALIZACIÓN ADAPTATIVO DEL HISTOGRAMA DE LOS DATOS DE RESERVA.....	145



## Introducción

---

### 1.1 Imagen del fondo ocular

El primer intento en representar la retina fue completado por el médico francés Jean Mery, quien demostró que al sumergir un gato vivo en agua era posible observar sus vasos retinianos desde el exterior. La impracticidad de tal enfoque, para los humanos, llevó al científico checo Jan Purkyně a investigar métodos más viables [1]. Esos antecedentes permitieron la invención del oftalmoscopio en 1851 por Von Helmholtz [2]. De esta manera la inspección y evaluación de la retina se convertirían en actividades rutinarias para los oftalmólogos, produciendo las primeras imágenes de retina, Figura 1, publicadas por el oftalmólogo holandés Van Trigt en 1853 [3].

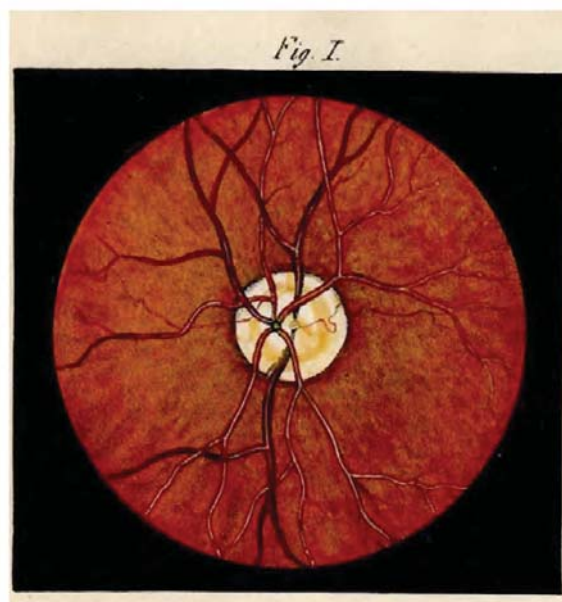


Figura 1. Primera imagen documentada de retina humana dibujada por Van Trigt en 1853.

Tres años después Eduard Jaeger utilizó el oftalmoscopio para la elaboración de un atlas de dibujos de fondo ocular. La realización del trabajo dio lugar a la identificación de cambios maculares diabéticos en forma de manchas amarillentas y extravasaciones que penetraban parte o el total del grosor de la retina. Con ello se propuso por primera vez la existencia de una relación entre la diabetes y la aparición de complicaciones oculares en la retina [4]. Estos descubrimientos requerían ser documentados; por lo que el uso de imágenes, para capturar el fondo ocular, resultaba ser la opción más apropiada tanto para evaluar pacientes como para documentar los hallazgos registrados. A principios de 1860, Henry Noyes de Nueva York y A.M. Rosebrugh de Toronto, construyeron cámaras de fondo ocular y lograron capturar imágenes de retina en animales. Hermann Helmholtz se mostró escéptico sobre la técnica declarando en la revista *The Ophthalmic Review*: *"Debo confesar que todavía no puedo creer que (fotografiar el fondo del ojo) tenga éxito en el sujeto humano; la dificultad principal es evidentemente el uso de luz solar suficientemente potente, que solo sería admisible con la vista completamente ciega"*. De esta manera dejaba claro las limitaciones existentes en la técnica [5]. Las investigaciones continuaron, buscando mejorar las implementaciones realizadas por Noyes y Rosebrugh. En el año de 1886 Howe, Jackman y Webster lograron fotografiar el fondo ocular de un ser humano [6]. Su método consistía en una exposición de dos minutos y medio que resultaba en una imagen con un reflejo corneal sobresaliente, pero un disco óptico apenas visible. Debido a que la impresión de imágenes era un proceso demasiado costoso para una publicación periódica, la reproducción de la imagen fue un grabado que simulaba la fotografía original. La técnica se publicó junto con la reproducción de la imagen en dos periódicos, Figura 2.

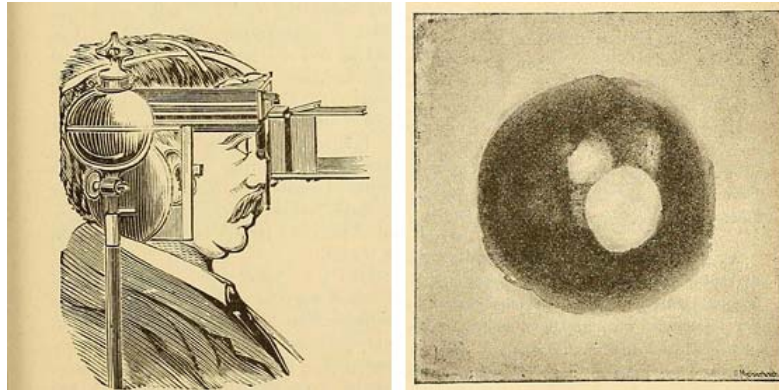


Figura 2. Del lado izquierdo se observa parte del proceso para la captura del fondo ocular. Del lado derecho un dibujo de la imagen obtenida.

Cuatro años después, la primera clasificación de retinopatía diabética fue realizada por Julius Hirschberg [7]. Partiendo de la información registrada y de sus observaciones estableció 4 tipos de clases.

- Retinitis centralis punctata
- Forma hemorrágica
- Infarto de retina
- Glaucoma hemorrágico

Junto con Julius Hirschberg otros personajes contribuyeron durante las últimas décadas del siglo XIX y los primeros años del siglo XX a sentar las bases necesarias para la oftalmología moderna, identificando necesidades del área e incrementando el número de integrantes de la comunidad oftalmológica. En 1926 se obtuvieron las primeras imágenes que capturaban 20 grados del fondo ocular convirtiéndose en el estándar de la época. Años después con el aumento en el ángulo de visión, las cámaras de 30 y 45 grados se convertirían en el nuevo estándar, Figura 3; esto debido a que se identificó que en la mayoría de los diagnósticos retinianos la información significativa se localizaba en la parte posterior del polo cubierta por ese ángulo de visión.

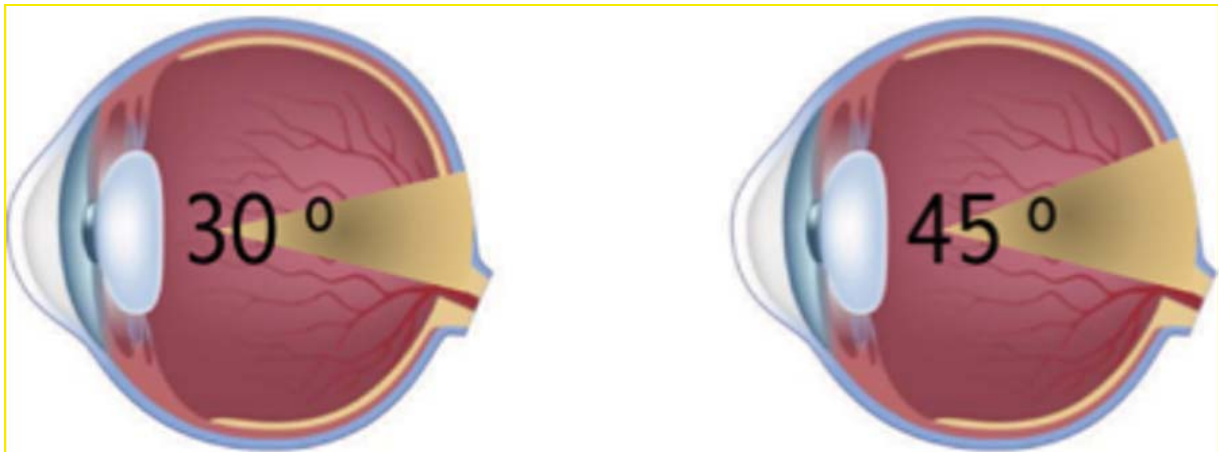


Figura 3. Grados de captura del fondo ocular.

En 1964 Allen utilizó la fotografía estereoscópica del fondo de ojo para representar la forma tridimensional de la retina. El oftalmólogo combinaba las imágenes de retina de varios ángulos en una forma tridimensional [8]. Anterior a esto, la fotografía del fondo de ojo era una representación bidimensional de los tejidos retinianos semitransparentes tridimensionales proyectados en el plano de la imagen; este método no proporcionaba toda la información necesaria para el especialista. La contribución de todos los oftalmólogos permitió una mayor comprensión de la retinopatía diabética, así como su tratamiento. En 1968 se llevó a cabo el Simposio Airlie House, Figura 4, donde por primera vez se estandarizó el sistema de clasificación de retinopatía diabética.



Figura 4. Foto grupal de los participantes del Simposio Airlie House en 1968.

## 1.2 Implementación de Inteligencia Artificial en la medicina

A la par de los avances médicos las primeras investigaciones en el área de Inteligencia Artificial comenzaban a surgir. En 1943 McCulloch, neurofisiólogo, y Pitts, matemático, redactaron un artículo sobre el proceso que, a partir de sus investigaciones, suponían que las neuronas realizaban durante su funcionamiento. Modelaron la primera neurona artificial, registrada, con el uso de circuitos eléctricos mediante el uso de operaciones binarias simples. Aunque la implementación de la primera neurona artificial causó expectativas entre la comunidad científica la última mitad del siglo XX se caracterizó por el registro de avances limitados y escepticismo; desacelerando el optimismo existente. En 1976 Gunn fue el primero en utilizar Inteligencia Artificial en un problema relacionado a la medicina. Su trabajo consistió en el uso de análisis computacional para diagnosticar dolor abdominal agudo [9]. Sus resultados reportaron un incremento del 10% de exactitud con respecto al método empleado en la época, sin embargo, su uso no se extendió en el diagnóstico de pacientes. Para 1984 la revolución médica producto de los alcances que se podían tener a través de sistemas computacionales era inevitable. Tomando ventaja de ello, Baudoin et al. [10] publicaron la realización del procesamiento de imágenes de retina para detectar microaneurismas; una lesión característica de retinopatía diabética. El inicio de este siglo se ha caracterizado por el resurgimiento de sistemas inteligentes para la realización de tareas médicas.

La línea del tiempo de la Figura 5 muestra los avances en oftalmología e Inteligencia Artificial registrados hasta el momento.

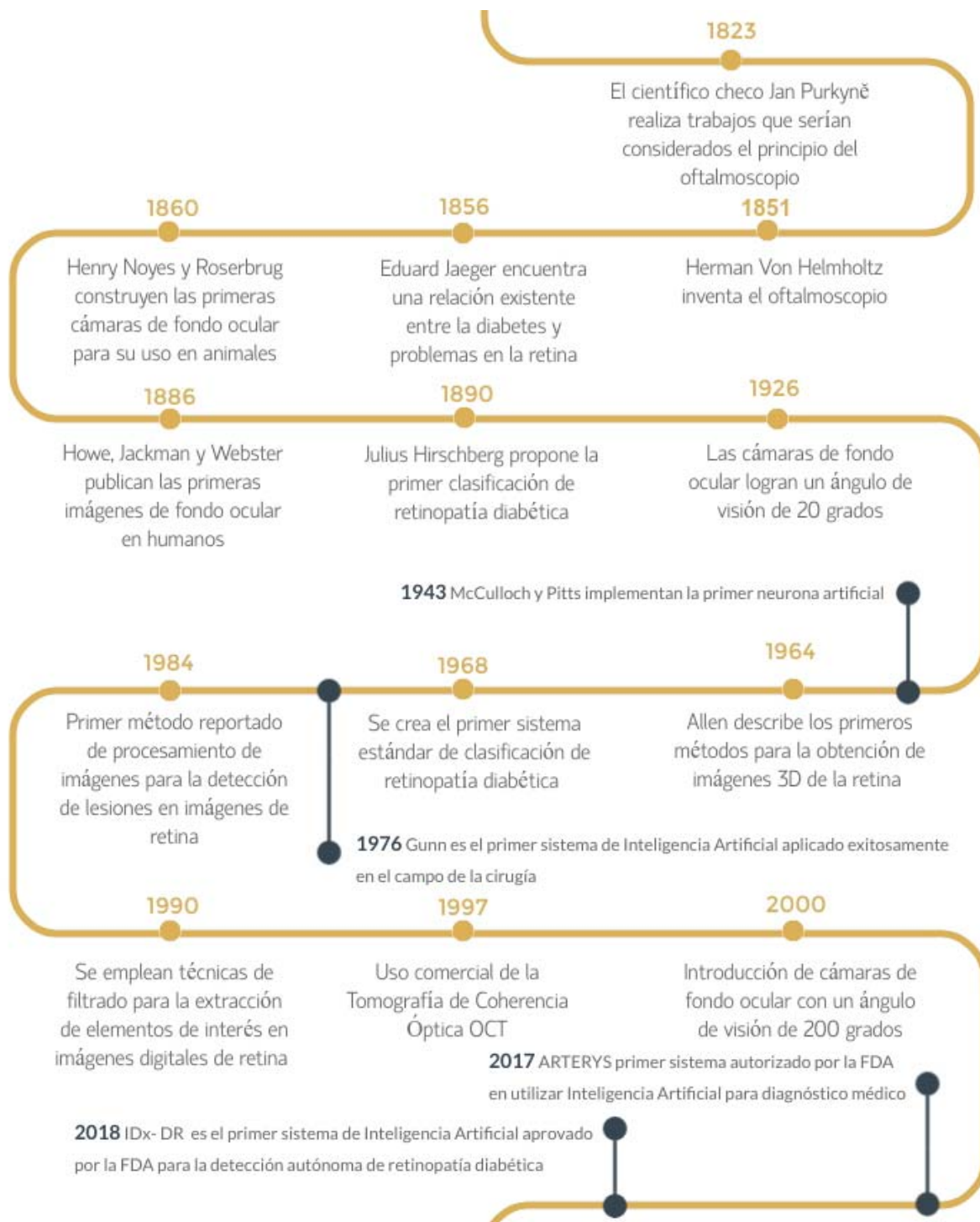


Figura 5. Evolución en el diagnóstico de retinopatía diabética.



## 1.3 Aprendizaje automático

Es una rama de la inteligencia artificial cuyo objetivo es comprender la estructura de los datos y ajustarlos en modelos para realizar predicciones. Aunque el aprendizaje automático es un campo de la informática, difiere de los enfoques computacionales tradicionales. En la informática tradicional, los algoritmos son conjuntos de instrucciones explícitamente programadas que se usan para calcular y/o resolver problemas. En cambio, los algoritmos de aprendizaje automático permiten que los sistemas de cómputo entrenen con un conjunto de datos de entrada y a partir de análisis estadístico generen los valores de salida deseados, facilitando así la construcción de modelos complejos. Con ello surge el aprendizaje supervisado y el aprendizaje no supervisado.

### 1.3.1 Aprendizaje supervisado

El algoritmo cuenta con datos de entrenamiento etiquetados con las salidas deseadas. El propósito de este método es que el algoritmo pueda generar un modelo al comparar la predicción del modelo con la salida esperada. De esta forma, tras varias iteraciones se logra ajustar el modelo al conjunto de datos evaluados. El aprendizaje supervisado, por lo tanto, utilizará los patrones localizados durante el entrenamiento para predecir valores de etiquetas en un nuevo conjunto de datos. Con el aprendizaje supervisado, un algoritmo podría utilizar imágenes separadas por clases A y B para entrenar un modelo, e identificar, después de la etapa de entrenamiento, si una nueva imagen pertenece a la clase A o B.

### 1.3.2 Aprendizaje no supervisado

En este tipo de aprendizaje el algoritmo se entrena con datos no etiquetados. Debido a que los datos no etiquetados son más abundantes, que los datos etiquetados, los métodos de aprendizaje automático que facilitan el aprendizaje no supervisado tienen un valor adicional. El objetivo del aprendizaje no supervisado radica en el descubrimiento de patrones, dentro de un conjunto de datos, conllevando al aprendizaje de características. Esto resulta particularmente útil en tareas como la de clasificación de datos.

### 1.3.3 Redes Neuronales Convolucionales

Las Redes Neuronales Convolucionales [ConvNets] son una clase de red neuronal profunda que a su vez forman parte de los métodos de aprendizaje automático supervisado y no supervisado basados en redes neuronales artificiales. Se considerarán supervisadas cuando se cuenta con la clasificación esperada para cada imagen. En el caso de retinopatía diabética se tendría que saber si la imagen es un caso que presenta la patología o un caso normal. Si sólo se cuenta con imágenes y se desconoce su clasificación se podría realizar un aprendizaje no supervisado y esperar que el modelo encuentre por sí mismo patrones que permitan su clasificación. Algunos investigadores que han trabajado bajo este esquema han logrado encontrar relaciones entre la imagen de retina, el sexo de la persona, problemas cardiovasculares y la edad del paciente [11]. Este tipo de red neuronal surge del estudio de procesos biológicos [12] donde el patrón de conectividad entre las neuronas se asemeja a la organización de la corteza visual animal. Las ConvNets utilizan relativamente poco preprocesamiento en comparación con otros algoritmos para el análisis de imágenes.

#### 1.3.3.1 *ImageNet*

El surgimiento de las ConvNets está relacionado con ImageNet, una colección de imágenes anotadas destinadas a la investigación de visión artificial, y el Desafío de Reconocimiento Visual de Gran Escala de ImageNet ILSVRC, competencia anual realizada entre los años 2010 y 2017. El objetivo de ImageNet es proporcionar un recurso para promover la investigación y el desarrollo de métodos mejorados para la visión artificial. El tipo de imágenes utilizadas se enlista a continuación:

- Información geológica
- Objetos naturales
- Deportes
- Artefactos
- Personas
- Animales
- Flora



Las tareas realizadas en el ILSVRC son:

- Clasificación. Se asigna una etiqueta de clase en función del objeto de interés en la imagen.
- Localización. Se indica una región de interés en función del objeto de interés en la imagen.
- Clasificación multiclase. Se indican una o más regiones de interés en función de los objetos de interés.
- Segmentación. Se delimita un área contenedora de los objetos de interés.

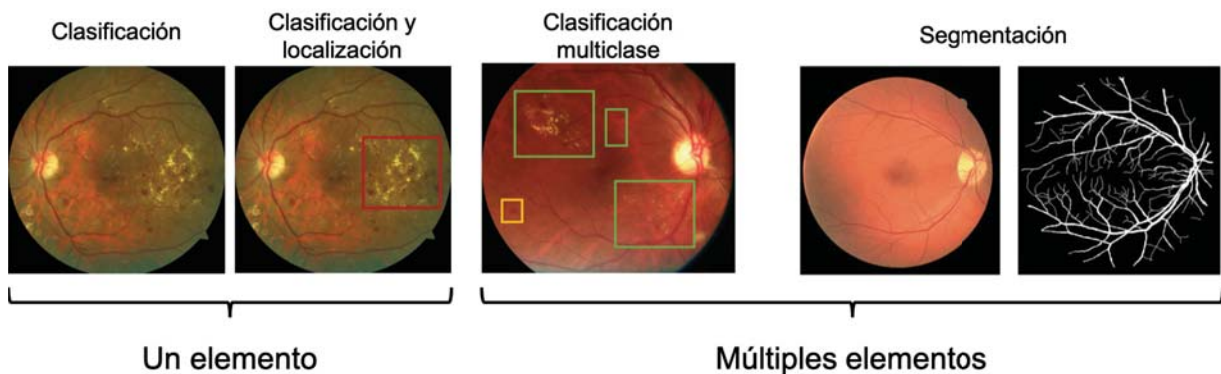


Figura 6. Tareas realizadas en el desafío ILSVRC.

Durante los primeros años de competencia, 2010 y 2011, las tasas de error 28% y 26%, respectivamente, se consideraban aceptables [13]. En 2012, el equipo SuperVision, integrado por Alex Krizhevsky, Ilya Sutskever y Geoffrey Hinton obtuvieron el 15% de error de clasificación, Figura 7, entrenando una ConvNet a la que nombraron AlexNet.

Desde 2012 las ConvNets dominaron la competencia del ILSVR logrando un error de clasificación del 3% en 2016; menor al error de clasificación humano estimado en 5.1% para el mismo conjunto de datos [14].

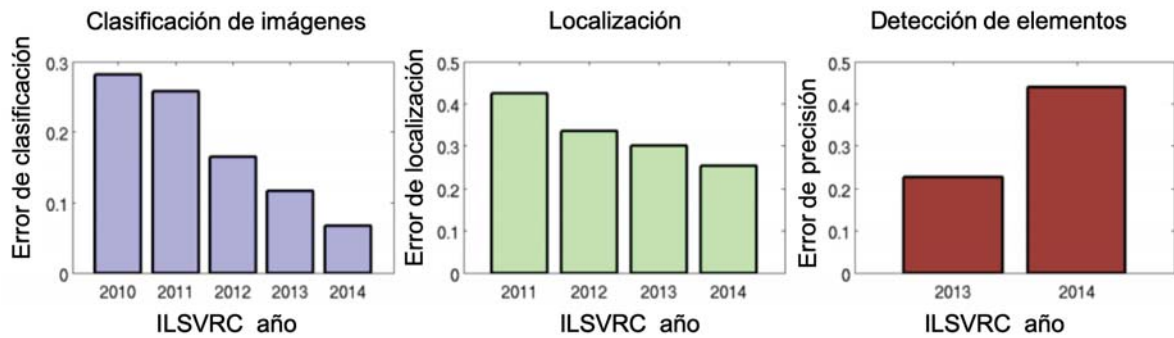


Figura 7. Resultados registrados en diferentes tareas de la competencia ILSVRC.

### 1.3.3.2 ConvNets en imágenes médicas

La validación de ConvNets en imágenes médicas resultaba ser una tarea que permitiría conocer la confiabilidad de este tipo de modelos. Radiografías, tomografías, resonancias magnéticas, medicina nuclear y ecografías, por mencionar algunas, son ejemplos de imágenes que emplean los médicos de manera cotidiana para la detección de lesiones y/o anomalías que permitan realizar un diagnóstico acertado de sus pacientes. El uso de ConvNets, como auxiliar médico, permitiría reducir errores incrementando así la exactitud en los diagnósticos realizados. Derivado de ello en 2018 W. David Pan et al. publicaron un trabajo donde reportan la detección automática de malaria en imágenes de muestras sanguíneas al microscopio con una exactitud mayor al 90% [15]. En el mismo año Heba Mohsen et al. trabajaron en la clasificación de tumores cerebrales a partir de imágenes de resonancia magnética logrando un AUC en la tarea de detección del 0.984 [16]. Okeke Stephen et al. utilizaron en 2019 ConvNets para la evaluación de radiografías de tórax y la detección de neumonía logrando una exactitud en la etapa de validación del 93.73% [17]. Ragab et al, en el mismo año, implementaron un sistema de detección asistida por computadora CAD para la detección de cáncer en mamografías consiguiendo una exactitud de detección cercana al 90% [18].

## 1.4 Diabetes

La diabetes es una enfermedad crónica que se origina porque el páncreas no sintetiza la cantidad de insulina que el cuerpo humano necesita, la elabora de una calidad inferior o no es capaz de utilizarla con eficacia [19]. En las personas con diabetes hay un exceso de glucosa en sangre (hiperglucemia), ya que no se distribuye de forma adecuada. La glucosa elevada puede ser perjudicial "para todo el organismo, pero principalmente para el corazón, el riñón y las arterias, por lo que las personas que tienen diabetes y no lo saben o no la tratan tienen más riesgo de problemas renales, infartos, pérdida de visión y amputaciones de miembros inferiores.

## 1.5 Retinopatía Diabética

Las personas con diabetes pueden tener una enfermedad ocular llamada retinopatía diabética. Esta enfermedad ocurre porque los niveles altos de azúcar en la sangre causan daño a los vasos sanguíneos en la retina. Estos vasos sanguíneos pueden hincharse y tener fugas de líquido. También pueden cerrarse e impedir que la sangre fluya. A veces, se generan nuevos vasos sanguíneos anormales en la retina. Todos estos cambios pueden hacerle perder la visión.

### 1.5.1 Etapas de la retinopatía diabética

La enfermedad ocular diabética presenta dos etapas principales.

- Retinopatía Diabética No Proliferativa [RDNP]

Es la etapa temprana de la enfermedad. Muchos vasos sanguíneos pequeños sufren pérdidas y hacen que la retina se hinche. A su vez esta etapa se puede clasificar en leve, moderada o severa dependiendo de las lesiones existentes en la retina. Esta es la razón más común por la que la gente con diabetes pierde la visión. Al tener retinopatía diabética no proliferativa la visión del paciente será borrosa.



Figura 8. Comparación de visión entre un paciente sano [izquierda] y un paciente con retinopatía diabética [derecha].

- Retinopatía Diabética Proliferativa [RDP]

Esta es la etapa más avanzada de la enfermedad ocular diabética. Se produce cuando la retina comienza a desarrollar nuevos vasos sanguíneos. Esto se denomina neovascularización. La retinopatía diabética proliferativa es muy grave y puede ocasionar una pérdida tanto de la visión central como la periférica.

La Tabla 1 muestra la clasificación de la enfermedad de acuerdo con los hallazgos observados en la retina.

<b>Nivel de severidad de la enfermedad</b>	<b>Hallazgos observados a la oftalmoscopia con dilatación pupilar</b>
Sin retinopatía	Sin anomalías
Retinopatía diabética no proliferativa leve	Presencia de al menos una microaneurisma.
Retinopatía diabética no proliferativa moderada	Presenta microaneurismas, exudados duros y hemorragias, superficiales y profundas, en al menos uno de los cuadrantes. Se aprecia también la presencia de manchas algodinosas, rosarios venosos y Anomalías Microvasculares Intrarretinianas AMIRs.

Nivel de severidad de la enfermedad	Hallazgos observados a la oftalmoscopia con dilatación pupilar
Retinopatía diabética no proliferativa severa	Presencia de hemorragias, superficiales y profundas, en cuatro cuadrantes o rosarios venosos en dos cuadrantes o AMIRs en al menos un cuadrante.
Retinopatía proliferativa temprana	Uno de los siguientes: <ul style="list-style-type: none"> <li>• Neovascularización</li> <li>• Hemorragia vítrea o prerretiniana</li> </ul>

Tabla 1. Escala clínica internacional emitida por la Academia Americana de Oftalmología (AOO).

En la Figura 9 es posible observar algunas lesiones descritas en la Tabla 1.

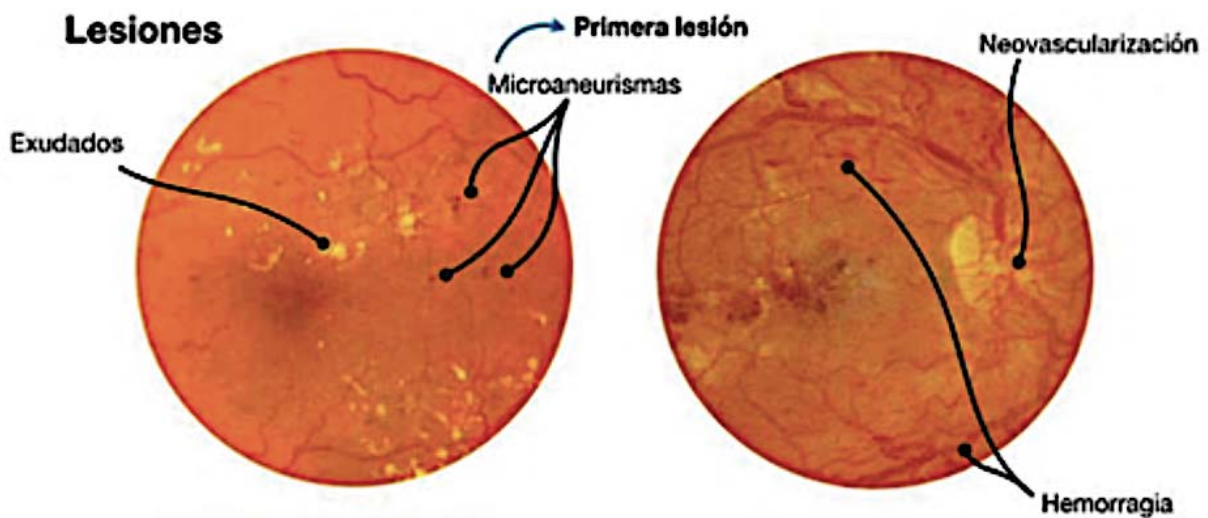


Figura 9. Lesiones características de retinopatía diabética

## 1.6 Objetivos

### 1.6.1 Objetivo principal

Implementar una metodología mediante el uso de una ConvNet para la detección de retinopatía diabética en imágenes de fondo ocular.

### 1.6.2 Objetivos secundarios

- Establecer un procedimiento para la normalización de las imágenes de retina.
- Implementar una ConvNet para la detección de retinopatía diabética.
- Ajustar hiperparámetros del modelo para lograr mejorar el desempeño del modelo.
- Validar el desempeño y la generalización del modelo.
- Visualizar las regiones de interés que tienen mayor peso durante la etapa de decisión.

## 1.7 Hipótesis

Implementando una ConvNet se puede tener un modelo confiable para la detección de retinopatía diabética en imágenes de fondo ocular.

## 1.8 Justificación

La detección de retinopatía diabética necesita de la evaluación de imágenes de retina para la detección de lesiones propias de la enfermedad. Una ConvNet emplea la información de las imágenes para construir una jerarquía de filtros no lineales que permitan distinguir las características relevantes de la imagen y lograr así su categorización. Este tipo de red es invariante a transformaciones espaciales por lo que las características relevantes, en este caso lesiones, pueden ser identificadas en cualquier región de la imagen. El uso de las ConvNets ha incrementado en los últimos

años debido al aumento de datos disponibles, el incremento de la capacidad de cómputo, el desarrollo de software para su uso y un creciente interés en las áreas de investigación y desarrollo de diferentes compañías. Países como Tailandia [20], India [21], España [22] y China [23] cuentan con programas pilotos, desarrollados en centros de investigación locales, para la detección de retinopatía diabética mediante inteligencia artificial. En México también existen algunas instituciones que se encuentran desarrollando sistemas con el mismo propósito, sin embargo, ninguna de ellas cuenta con licencia para su uso clínico y comercial. Querétaro cuenta con el Instituto Mexicano de Oftalmología [IMO] [24] que recibe pacientes de la zona metropolitana y zonas rurales aledañas a la ciudad. El uso de una ConvNet como auxiliar en el diagnóstico de la patología, corresponde a una herramienta que permitiría un diagnóstico en menor tiempo, el uso de un modelo estandarizado y la atención de un mayor número de pacientes. Al ser un modelo computacional, el diagnóstico se puede realizar en cualquier lugar donde se tenga acceso a un equipo de cómputo.

## 1.9 Planteamiento del problema

La Fundación del Cuidado de la Salud en California CHCF, por sus siglas en inglés, encontró recientemente que entre los pacientes con retinopatía diabética que fueron referidos a un especialista por un médico general, sólo el 23% consultó al especialista. Del resto, el 15% no estaba al tanto de la enfermedad, otro 15% no hizo una cita de detección, el 22% no asistió a su cita y el 25% optó por no recibir tratamiento [25]. En base a estos datos y con la proyección de crecimiento de la enfermedad resulta necesario el desarrollo de una herramienta que permita la detección de retinopatía diabética, empleando como entrada imágenes de retina y generando a su salida un diagnóstico del paciente; refiriéndolo a un especialista en caso de ser necesario. Esto permitiría que los médicos especialistas atiendan de manera oportuna sólo a aquellos pacientes que en realidad lo necesitan.

## Antecedentes

---

### 2.1 Publicaciones relevantes

Marco Albán y Tanner Gilligan [26] publicaron en 2016 el uso de Redes Neuronales Convolucionales y técnicas de eliminación de ruido para diagnosticar la presencia y la gravedad de retinopatía diabética en fotografías de angiografía con fluoresceína. Los datos empleados consistían en fotografías de retina con rangos variables de gravedad indicados por médicos. GoogLeNet, una red neuronal convolucional diseñada por Google fue utilizada para la clasificación de retinopatía diabética mostrando un AUC del 79% y una precisión del 45%. Estos resultados indican un desempeño satisfactorio al ser comparado con las métricas de evaluación humana. Los autores consideraron seguir explorando técnicas de eliminación de ruido y normalización de datos, así como el uso de un conjunto de clasificadores entrenados en los píxeles de las imágenes para la extracción de características útiles para la detección de retinopatía diabética.

En el mismo año Harry Pratta, Frans Coenenb, et al. [27] reportaron un trabajo similar para el diagnóstico de retinopatía diabética al desarrollar una Red Neuronal Convolucional e implementar un aumento en los datos. Los autores entrenaron la red utilizando unidades de procesamiento gráfico (GPUs) de alta gama en un conjunto de datos de 80 mil imágenes logrando una especificidad del 95%, una precisión del 75% y una sensibilidad del 30% en 5 mil imágenes de validación. La baja sensibilidad ocurrió en la clasificación de las clases leve y moderada. Los autores afirman que este tipo de modelos tienen el potencial de ser útiles para los médicos a medida que las redes y los conjuntos de datos continúen mejorando hasta lograr clasificaciones en tiempo real.

Darshit Doshi et al. [28] reportaron, también en 2016, el diseño y la implementación de Redes Neuronales Convolucionales profundas aceleradas por GPUs para diagnosticar y clasificar automáticamente imágenes de retina en 5 clases de



acuerdo con la gravedad de la enfermedad. La precisión alcanzada por un modelo único de Redes Neuronales Convolucionales fue de 0.386 en una métrica kappa ponderada cuadrática; un ensamblaje de tres modelos similares resultó en una precisión de 0.3996. Durante las tareas de clasificación, los autores trabajaron únicamente con el canal verde de las imágenes de retina.

Sagar Honnugar, Sanyam Mehra y Samuel Joseph [29] exploraron técnicas de aprendizaje automático para la identificación y clasificación de imágenes de retina con retinopatía diabética. El enfoque presentado se diferencia de los demás autores, hasta ahora citados, por no utilizar Redes Neuronales Convolucionales. Uno de los objetivos de su investigación fue la comparación de otras técnicas de aprendizaje automático con respecto a las Redes Neuronales Convolucionales. La metodología utilizada por los autores implica el preprocesamiento de la imagen, la extracción de características y el uso de un clasificador multiclase. Bajo este esquema de trabajo obtuvieron una precisión del 68%, 73% y 72% con los clasificadores regresión logística, bosque aleatorio y máquina de soporte de vectores respectivamente. Para mejorar la precisión de clasificación sin emplear redes neuronales los autores proponen combinar diferentes clasificadores a través de un enfoque de conjunto.

En 2017 Ratul Ghosh, Kuntal Ghosh y Sanjit Maitra [30] propusieron un enfoque de Redes Neuronales Convolucionales en conjunto con técnicas de eliminación de ruido para identificar características como microaneurismas y hemorragias en la retina. Utilizaron un conjunto de más de 30 mil imágenes y GPUs de alta gama durante la etapa de entrenamiento y validación. El modelo propuesto alcanzó el 95% de precisión para la clasificación de dos clases y 85% de precisión para la clasificación de cinco clases en 3 mil imágenes de validación. Teniendo en cuenta la heterogeneidad del conjunto de datos, los autores consideran que el rendimiento del modelo propuesto es satisfactorio, sin embargo, consideran implementar otras técnicas complejas de eliminación de ruido para incrementar la precisión de su modelo.

Suvajit Dutta, Bonthala CS Manideep, Shalva Rai y Vijayarajan V. [31] publicaron un trabajo comparativo de varios algoritmos de aprendizaje automático supervisado para detectar retinopatía diabética. Los autores utilizaron, inicialmente, el algoritmo de agrupación K-medias para la identificación de los niveles de gravedad en la imagen.

Después de determinar los niveles de gravedad de la imagen utilizaron Redes Neuronales para obtener el resultado de clasificación básica. Las Redes Neuronales se compararon con Redes Neuronales Profundas, las cuales mostraron un aumento en la precisión. Los autores también consideraron dos arquitecturas de Redes Neuronales Convolucionales debido a su buen desempeño al clasificar imágenes. Las Redes Neuronales Profundas mostraron el mejor desempeño, con una precisión del 75% sin embargo, las dos Redes Neuronales Convolucionales utilizadas mostraron una precisión del 68% y 72%. Aumentar el tamaño de la imagen de entrenamiento y establecer una nueva arquitectura de Red Neuronal Convolutiva son las estrategias que planean llevar a cabo los autores para incrementar la precisión del modelo.

Saboora Mohammadian, Ali Karsaz y Yaser M. Roshan [32] realizaron un análisis comparativo de nueve algoritmos clasificadores comunes para la detección de retinopatía diabética. Cada técnica de clasificación se optimizó en función de sus parámetros ajustables. Este análisis lo realizaron para contar con una comparativa y elegir el algoritmo de clasificación adecuado en futuras investigaciones. Los resultados del trabajo mostraron una tendencia en como los parámetros afectan la precisión de cada técnica de clasificación; así como las técnicas que presentan el mejor desempeño. De acuerdo con los resultados, el clasificador de procesos gaussianos mostró el mejor desempeño al detectar casos con retinopatía diabética. El algoritmo de Máquina de Soporte de Vectores y los enfoques de refuerzo adaptativo son clasificadores que también podrían ajustarse para lograr un rendimiento satisfactorio.

En 2018 Sheikh Saiful et al. [33] presentaron una red neuronal convolutiva profunda, que realiza la detección de microaneurismas; lesiones presentes en la etapa inicial de retinopatía diabética, junto con la asignación de una de las cinco clases establecidas para categorizar la gravedad de la enfermedad. La red propuesta alcanzó un 98% de sensibilidad y más del 94% de especificidad en la etapa inicial de detección. La puntuación kappa reportada fue de más de 0.85 en la clasificación de la gravedad. Los resultados experimentales mostraron la efectividad del algoritmo propuesto, así como un prometedor uso en aplicaciones clínicas.

Navoneel Chakrabarty [34] también empleó una red neuronal convolutiva para predecir si una imagen de retina presentaba lesiones de retinopatía diabética. El

procesamiento realizado por el autor implica la conversión de imágenes en el espacio RGB a escala de grises y un reajuste de dimensiones. La metodología propuesta por el autor se aplica en un conjunto de datos de 30 imágenes de retina de alta resolución. Los resultados reportados indican una precisión predictiva del 100% y una sensibilidad del 100%.

En 2019 Valentina Bellemo et al. [35] utilizaron un modelo de Inteligencia Artificial al combinar dos redes neuronales convolucionales para detectar retinopatía diabética en imágenes de retina. El modelo fue entrenado con 76,370 imágenes de retina de 13,099 pacientes con diabetes que habían participado en el Programa Integrado de Retinopatía Diabética de Singapur. Para la validación clínica, se utilizaron 4504 imágenes de fondo de retina de 3093 ojos de 1574 africanos con diabetes. El AUC del sistema para la detección de retinopatía diabética referible fue 0.973, con una sensibilidad del 92.25% y una especificidad del 89.04%. El modelo de Inteligencia Artificial identificó una mayor duración de la diabetes, un mayor nivel de hemoglobina glicosilada y un aumento de la presión arterial sistólica como factores de riesgo asociados con la retinopatía diabética referible. Esto muestra la posible aplicación y adopción de Inteligencia Artificial incluso cuando el modelo utilizado ha sido entrenado en una población diferente.

Finalmente, Misgina Tsighe Hagos y Shri Kant [36] implementaron técnicas de transferencia de aprendizaje en Redes Neuronales Convolucionales para clasificar dos clases de retinopatía diabética a partir de una cantidad limitada de datos. El modelo entrenado registró una precisión del 90.9% siendo un resultado alentador para su uso en aplicaciones clínicas. La Tabla 2 muestra una recopilación de información relevante de todos los trabajos citados en la automatización del diagnóstico de retinopatía diabética.

<b>Autor</b>	<b>Clasificador</b>	<b>Número de clases</b>	<b>Conjunto de datos</b>	<b>Espacio de color</b>	<b>Métrica</b>	<b>Entorno de trabajo</b>	<b>Hardware</b>
Marco Albán & Tanner Gilligan (2016) [26]	ConvNets	5	EyePacs 35k	RGB	AUC 79%	Caffe	GPU con 1536 núcleos CUDA
Darshit Doshi et al. (2016) [28]	ConvNets	5	EyePacs 35k	Canal verde	Kappa .3996	Theano	GPU GeForce serie GTX 780
Sagar Honnunar, Sanyam Mehra & Samuel Joseph (2016) [29]	LR RF SVM	3	EyePacs	Canal verde	Exactitud Entrenamiento LR .85 RF .90 SVM .83 Pruebas LR .68 RF .73 SVM .72	MATLAB	CPU

<b>Autor</b>	<b>Clasificador</b>	<b>Número de clases</b>	<b>Conjunto de datos</b>	<b>Espacio de color</b>	<b>Métrica</b>	<b>Entorno de trabajo</b>	<b>Hardware</b>
Harry Pratta, Frans Coenenb, et al. (2016) [27]	ConvNets	5	EyePacs 80k	RGB	Especificidad .95 exactitud .75 sensibilidad .30	Keras	NVIDIA K40c contains 2880 CUDA cores
Saboora Mohammadian, Ali Karsaz & Yaser M. Roshan (2017) [32]	AB DT NB G KNN QDA RF SVM MLP	No especificado	Messidor 1.1k	No especificado	Exactitud AB .8331 DT .7316 NB .7283 G .8707 KNN .7607 QDA .6247 RF .8028 SVM .8449 MLP .8125	Python Numpy	No especificado

<b>Autor</b>	<b>Clasificador</b>	<b>Número de clases</b>	<b>Conjunto de datos</b>	<b>Espacio de color</b>	<b>Métrica</b>	<b>Entorno de trabajo</b>	<b>Hardware</b>
Suvajit Dutta, Bonthala CS Manideep, Shalva Rai & Vijayarajan V (2017) [31]	NN DNN ConvNets	5	Sin especificar	Escala de grises	Exactitud Entrenamiento NN 62.7 DNN 89.6 CNN76.4 Pruebas NN 34.7 DNN75.7 CNN 72.4	Theano y Torch	GPU
Ratul Ghosh, Kuntal Ghosh & Sanjit Maitra (2017) [30]	ConvNets	2 & 5	EyePacs 30k	RGB	Exactitud 0.95 dos clases 0.85 cinco clases	Theano	GPU
Sheikh Saiful, Mahedi Hasan & Sohaib Abdullah (2018) [33]	ConvNets	2	EyePacs 35k	RGB	AUC .844	No especificado	No especificado

<b>Autor</b>	<b>Clasificador</b>	<b>Número de clases</b>	<b>Conjunto de datos</b>	<b>Espacio de color</b>	<b>Métrica</b>	<b>Entorno de trabajo</b>	<b>Hardware</b>
Navoneel Chakrabarty (2018) [34]	ConvNets	2	DataSet privado 30 de alta resolución	Escala de grises	AUC .94	Keras backend con Tensorflow	CPU
Valentina Bellemo et al. (2019) [35]	ConvNets	2	DataSet privado de población africana 4k	RGB	AUC .973	No especificado	No especificado
Misgina Tsighe Hagos & Shri Kant (2019) [36]	ConvNets	2	EyePacs 2.5k	RGB	Exactitud .909	Keras	Google Colaboratory GPU

Tabla 2. Compilación de publicaciones sobre detección de retinopatía diabética de manera autónoma.

<b>Técnica de procesamiento</b>	<b>Siglas empleadas</b>
AdaBoost	AB
Árbol de decisiones	DT
Naive-Bayes	NB
Proceso Gaussiano	G
Vecinos cercanos	KNN
Clasificador cuadrático	QDA
Bosque aleatorio	RF
Máquina de soporte de vectores	SVM
Perceptrón de varias capas	MLP
Red neuronal	NN
Red neuronal profunda	DNN
Regresión logística	LR
Bosque aleatorio	RF

Tabla 3. Referencia de abreviaturas empleadas en la Tabla 2.




## 2.2 Aplicaciones y casos de éxito

En 2018 la compañía IDx implementó el primer sistema de inteligencia artificial aprobado por la FDA para analizar imágenes de retina con el fin de diagnosticar retinopatía diabética [37]. El sistema se desarrolló en colaboración con el departamento de oftalmología y ciencias visuales de la Universidad de Iowa. Actualmente es utilizado en sistemas de salud de Iowa, Georgia, California, Maryland y Minnesota. El sistema se ha desempeñado de manera satisfactoria en los institutos afiliados obteniendo opiniones positivas por parte de los médicos.



Figura 10. Sistema IDX-DR para el diagnóstico de retinopatía diabética [38].

Eyenuk es otro sistema que combina inteligencia artificial con clasificación humana para diagnosticar retinopatía diabética. La herramienta permite canalizar a las clínicas de atención primaria con especialistas en atención ocular. De esta manera el sistema ha logrado incrementar el acceso a los exámenes de detección de retinopatía diabética, reducir los tiempos de espera y mejorar el cumplimiento del paciente al recibir los exámenes necesarios. Además de retinopatía diabética el sistema Eyenuk detecta otras afecciones oculares como lo son el glaucoma y la degeneración macular relacionada con la edad. La tecnología se está utilizando de manera comercial en Europa y como un dispositivo de investigación en prestigiosas instituciones clínicas en los Estados Unidos. También cuenta con una licencia de dispositivo médico para ser utilizado en Canadá.



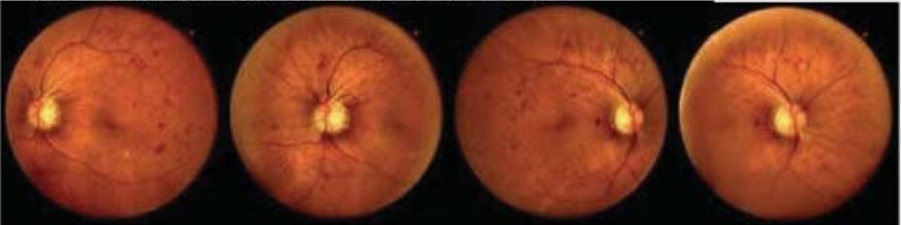
**EyeArt**  
 Diabetic Retinopathy Screening Report

Patient Information	General Information
<b>Patient ID:</b> <b>Patient Name:</b> <b>Date of Birth:</b> 01-01-1971 <b>Gender:</b> Male <b>Submission Date:</b> 2019/Feb/11 08:54 UTC	<b>Referring Location:</b> <b>Referring Provider:</b> <b>Encounter ID:</b> 61713 <b>EyeArt Control ID:</b> 61713 <b>Dilation Status:</b> Non Dilated

**Diabetic Retinopathy Screening Summary**

**Screening Result:** Positive for vision threatening diabetic retinopathy.

**Details:** Signs of Proliferative DR [4.2] with macular edema detected



61713\_i003
61713\_i004
61713\_i001
61713\_i002

\*Do not use the above thumbnail images for diagnostic purposes.

Figura 11. Reporte del diagnostico realizado por el sistema EYENUK. [39].

Pegasus es otro caso de éxito de un sistema que utiliza inteligencia artificial para detectar retinopatía diabética. El sistema analiza imágenes de la parte posterior del ojo, tomadas con cualquier cámara de retina o teléfono inteligente estándar, en busca de características indicativas de retinopatía diabética, glaucoma, degeneración macular relacionada con la edad, edema macular diabético y/o anomalías maculares y retinianas generales. La empresa desarrolladora tiene su sede en Londres y considera que Pegasus podría representar un ahorro significativo, en costos relacionados con la atención médica, para la economía del Reino Unido.

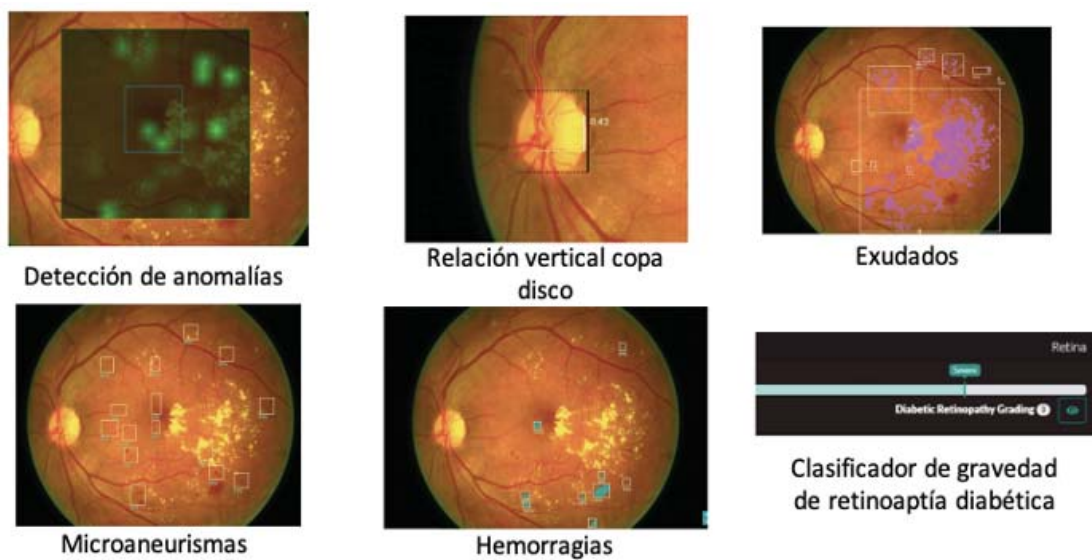


Figura 12. Anomalías relacionadas con RD detectadas automáticamente con el sistema Pegasus [40].

## Marco teórico

---

### 3.1 Procesamiento de imágenes

#### 3.1.1 Normalización

El proceso de normalización, también conocido como estiramiento del contraste, realiza un mapeo de las imágenes y asigna la intensidad mínima de cada imagen al valor mínimo del rango establecido. De la misma manera, asigna la intensidad máxima de cada imagen al valor máximo del rango establecido. La normalización de una imagen se define en la Ecuación 1.

$$I_N = (I - Min) \frac{nMax - nMin}{Max - Min} + nMin$$

Ecuación 1. Ecuación para la normalización de una imagen.

Donde:

$I_N$  = Nuevo valor del pixel evaluado

$I$  = Pixel evaluado

$Min$  = Valor mínimo del rango de intensidad

$Max$  = Valor máximo del rango de intensidad

$nMin$  = Nuevo valor mínimo del rango de intensidad

$nMax$  = Nuevo valor máximo del rango de intensidad

#### 3.1.2 Contraste de ecualización de histograma adaptativo limitado [CLAHE]

Es otra técnica utilizada para mejorar el contraste de las imágenes. Sin embargo, a diferencia de la normalización de imágenes, utiliza la distribución de probabilidad para asignar nuevos valores a los píxeles de las imágenes. Este tipo de ecualización se diferencia de una ecualización de histograma ordinaria debido a que se calculan varios

histogramas, cada uno correspondiente a una sección distinta de la imagen, para redistribuir los valores de luminosidad de la imagen. Por lo tanto, es adecuado para mejorar el contraste local y las definiciones de bordes en cada región de una imagen.

### 3.1.3 Anotación de imágenes

La anotación de imágenes es una técnica que permite crear datos de entrenamiento para algoritmos de Inteligencia Artificial. Implica la asignación de etiquetas con las cuales se proporciona a los modelos de clasificación la información de los objetos presentes en una imagen. Dependiendo de la necesidad de la tarea existen diferentes formas de anotación. En esta sección se mencionan aquellas relevantes para tareas médicas sin embargo J. Low [41] presenta un artículo más detallado al respecto.

#### 3.1.3.1 Cuadro delimitador

Es el tipo de anotación más utilizado y simple. Esta forma de anotación implica establecer un cuadro lo más cerca posible de los bordes de los objetos de interés en la imagen. El uso de los cuadros delimitadores se encuentra en la clasificación, localización y detección de objetos en diversas industrias incluida la atención médica.

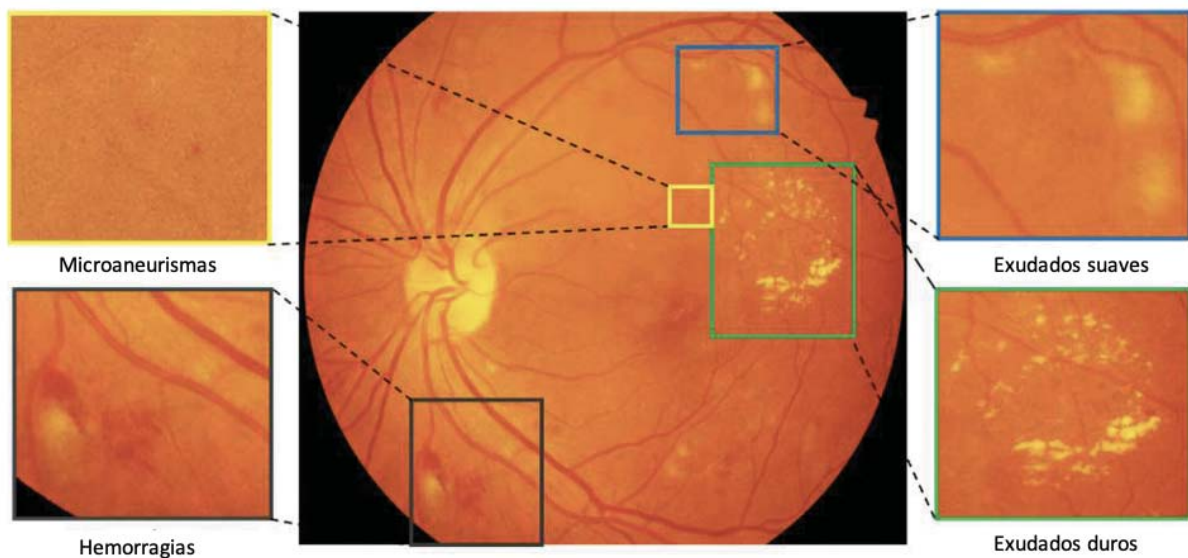


Figura 13. Anotaciones realizadas mediante cuadro delimitador.

### 3.1.3.2 Anotación poligonal

La anotación de polígono se usa para anotaciones más precisas en elementos de forma irregular; objetos no simétricos. La anotación de polígono generalmente requiere un alto nivel de precisión por parte del etiquetador.

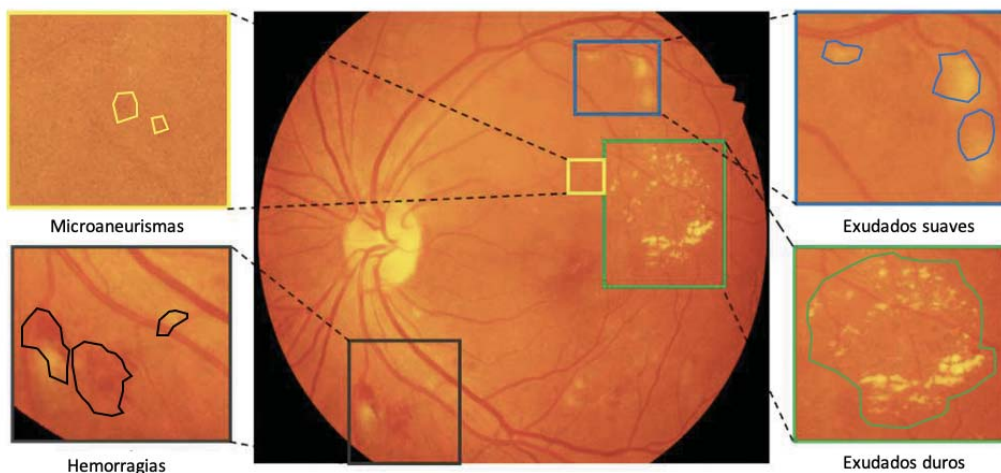


Figura 14. Anotaciones poligonales realizadas en una imagen de retina.

### 3.1.3.3 Segmentación Semántica

La segmentación semántica es la tarea de separar una imagen en múltiples secciones y asignar a cada píxel de interés una etiqueta con la clase que le corresponde. La segmentación semántica se usa comúnmente para la detección y localización de un objeto específico.

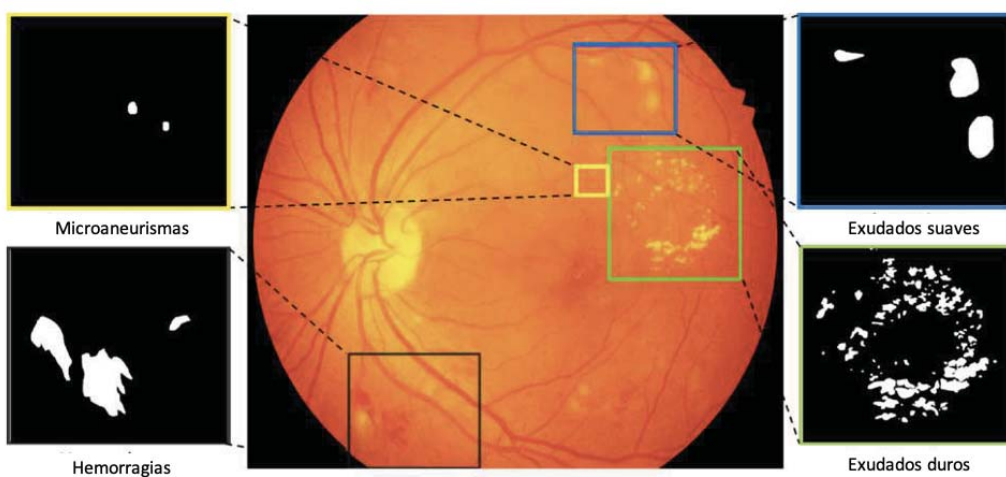


Figura 15. Segmentación semántica realizada en una imagen de retina.



## 3.2 Redes Neuronales

Una red neuronal también conocida como perceptrón multicapa es un modelo computacional inspirado en su homólogo biológico. Consiste en un conjunto de unidades llamadas neuronas artificiales conectadas entre sí y configuradas para lograr aproximar cualquier función mediante un algoritmo de aprendizaje. La información de entrada es evaluada en la red mediante diferentes operaciones para generar predicciones como valores de salida. La Figura 16 muestra el diagrama de una red neuronal simple de 3 entradas, una capa oculta y 2 salidas.

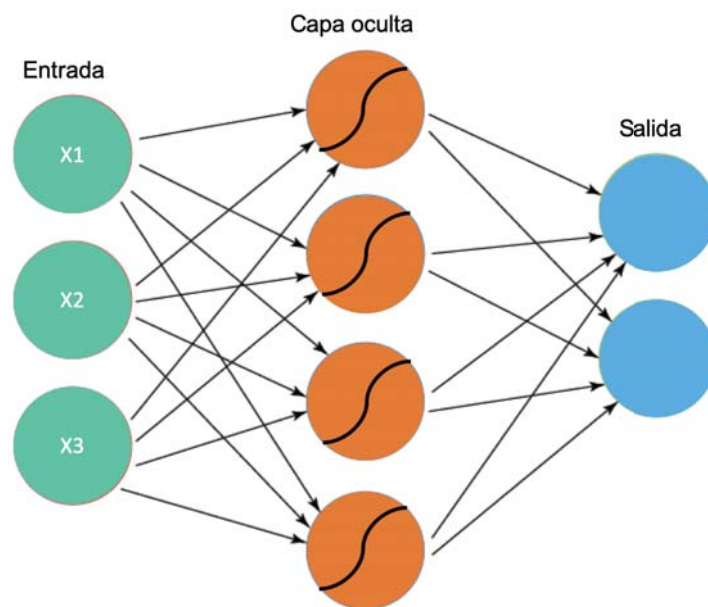


Figura 16. Diagrama de un perceptrón multicapa.

La salida de cada capa está dada por la Ecuación 2.

$$\hat{y}_i = \sigma(Wx^i + b)$$

Ecuación 2. Función para la predicción de valores.

- $\sigma$  Función de activación
- $W$  Conjunto de pesos utilizados por la red neuronal
- $x^i$  Conjunto de datos de entrada
- $b$  Valor de sesgo

Todos los valores obtenidos son evaluados en una función de activación. En este caso se considera la función Sigmoid que se encuentra definida por la Ecuación 3. Esta función tiene la característica de sólo generar valores entre 0 y 1, Figura 17, para cualquier valor de entrada.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Ecuación 3. Función Sigmoid.

La limitación de los valores de salida permite, en el caso de un clasificador binario, establecer el valor de 0.5 como umbral de decisión. En el caso de contar con las clases A y B se puede asignar a la clase A los valores obtenidos por debajo de 0.5 y a la clase B los valores superiores a 0.5.

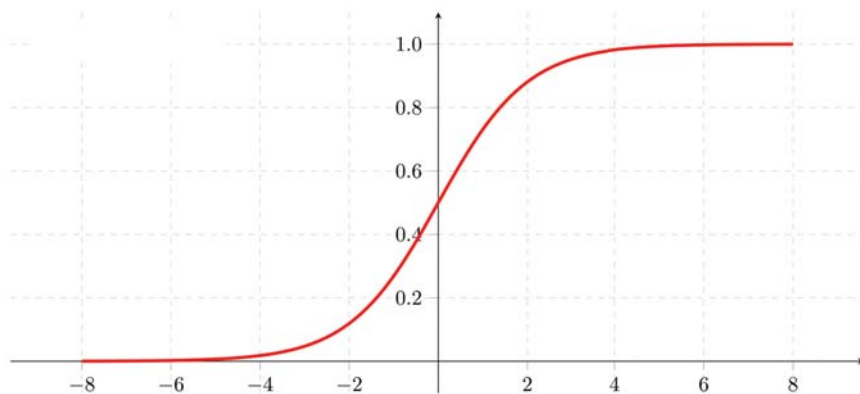


Figura 17. Gráfica de la función Sigmoid.

En el caso de clasificación multiclase el proceso es muy semejante. La única diferencia radica en el uso de la función Softmax, Ecuación 4, como función de activación. Con la función Softmax es posible obtener la probabilidad de cada clase considerando las probabilidades de las demás clases. De esta manera se obtiene un valor entre 0 y 1, para cada clase, y se garantiza que la suma de las probabilidades de todas las clases sea igual a 1.

$$\sigma(x_j) = \frac{e^{x_j}}{\sum_i e^{x_i}}$$

Ecuación 4. Función Softmax.



### 3.2.1 Entropía cruzada

El desempeño de las redes neuronales se puede evaluar a partir de las probabilidades obtenidas para cada valor de entrada. Estos valores se pueden multiplicar entre sí para obtener una única probabilidad de salida también conocida como probabilidad única del modelo. El principal objetivo de los algoritmos de aprendizaje automático es maximizar la probabilidad del modelo.

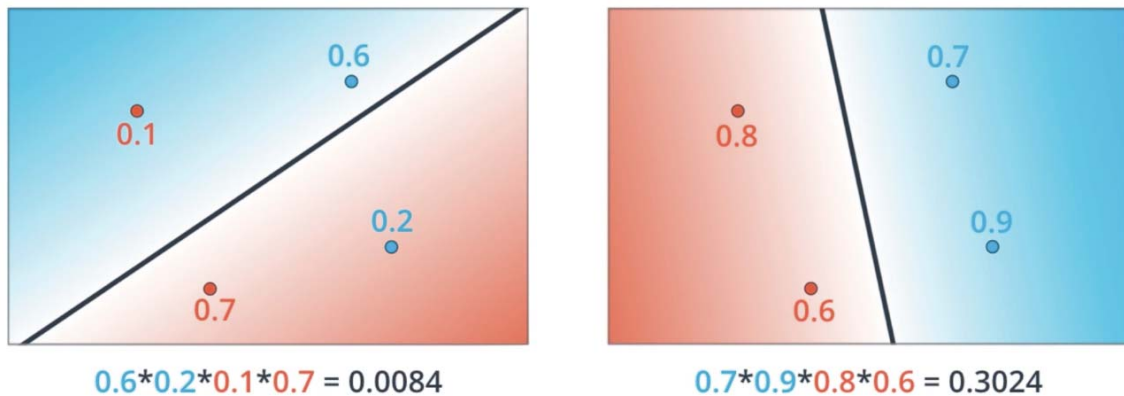


Figura 18. Probabilidades de un conjunto de datos.

En la Figura 18 se muestran dos modelos lineales que clasifican cada punto localizado en el plano. De acuerdo con el valor de probabilidad única del modelo, tiene un mejor desempeño de clasificación el modelo del lado derecho. Para mejorar el desempeño de cualquiera de los dos modelos es necesario realizar un ajuste que permita tener un mayor control sobre los cambios realizados a cada una de las probabilidades. Para ello es posible considerar la propiedad de los logaritmos de la Ecuación 5 y cambiar la multiplicación de probabilidades por una suma de probabilidades.

$$\log(ab) = \log(a) + \log(b)$$

Ecuación 5. Propiedad de los logaritmos.

Con ello se realizará la suma del logaritmo de cada probabilidad. Debido a que los valores de probabilidad se localizan entre 0 y 1 los resultados obtenidos serán negativos por lo que resulta apropiado considerar el logaritmo negativo de cada valor para obtener una suma de valores positivos. El proceso se realiza en la Figura 19. El resultado

obtenido se conoce como entropía cruzada. Entre menor sea su valor mejor será el desempeño del modelo.

$0.6 * 0.2 * 0.1 * 0.7 = 0.0084$ $\ln(0.6) + \ln(0.2) + \ln(0.1) + \ln(0.7)$ $-0.51 \quad -1.61 \quad -2.3 \quad -0.36$ $-\ln(0.6) - \ln(0.2) - \ln(0.1) - \ln(0.7) = 4.8$ $0.51 \quad 1.61 \quad 2.3 \quad 0.36$	$0.7 * 0.9 * 0.8 * 0.6 = 0.3024$ $\ln(0.7) + \ln(0.9) + \ln(0.8) + \ln(0.6)$ $-0.36 \quad -0.1 \quad -0.22 \quad -0.51$ $-\ln(0.7) - \ln(0.9) - \ln(0.8) - \ln(0.6) = 1.2$ $0.36 \quad 0.1 \quad .22 \quad 0.51$
--	---

Figura 19. Entropía cruzada de los modelos presentados en la Figura 18.

Formalmente la entropía cruzada se define por la Ecuación 6.  $y_i$  es el valor esperado a la salida; en el caso de un clasificador binario su valor es 0 o 1.  $p_i$  es la probabilidad obtenida de que la clase sea 0 o 1; su valor oscila entre 0 y 1. Si la clase  $y_i$  es 0, el primer elemento de la sumatoria es cero. Si la clase  $y_i$  es 1, el segundo elemento de la sumatoria es cero. De esta manera solo se considera la probabilidad de la clase esperada.

$$L = - \sum_{i=1}^m y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)$$

Ecuación 6. Definición matemática de la entropía cruzada.

### 3.2.2 Modelos no lineales

La capacidad de las redes neuronales para describir modelos no lineales permite que sean utilizadas en numerosas aplicaciones. La Figura 20 muestra la obtención de un modelo no lineal a partir de la combinación de dos modelos lineales. El valor de la probabilidad, 0.82, se obtiene evaluando en la Ecuación 3 la suma de las probabilidades de cada modelo lineal, 0.7 y 0.8.

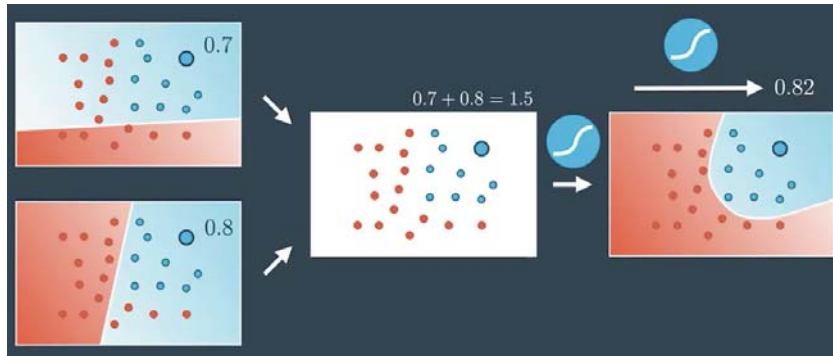


Figura 20. Modelo no lineal.

Gráficamente cada capa de la Figura 16 se representa en la Figura 21. Cada salida representa el modelo generado para las clases consideradas.

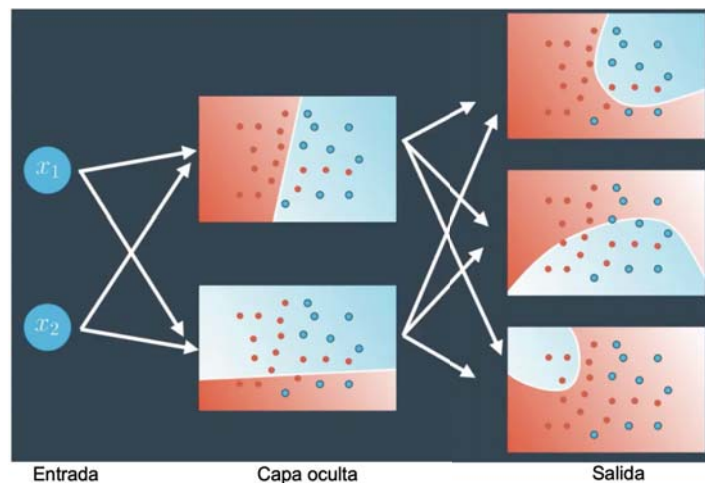


Figura 21. Modelos gráficos de red neuronal.

### 3.2.3 Tasa de aprendizaje

Durante el ajuste de los modelos de aprendizaje automático un parámetro importante es la tasa de aprendizaje. La Figura 22 ejemplifica la manera en es utilizado este parámetro. Primero se evalúan los datos de entrada para conocer la clasificación que se le asignará. En el ejemplo de la Figura 22 el valor de entrada es clasificado en la región azul lo cual es incorrecto, ya que se esperaría que sea clasificado en la región naranja. Para ajustar el modelo se obtiene la diferencia entre los coeficientes del clasificador actual y los valores del dato de entrada. A partir de ello se obtienen los coeficientes del nuevo clasificador.

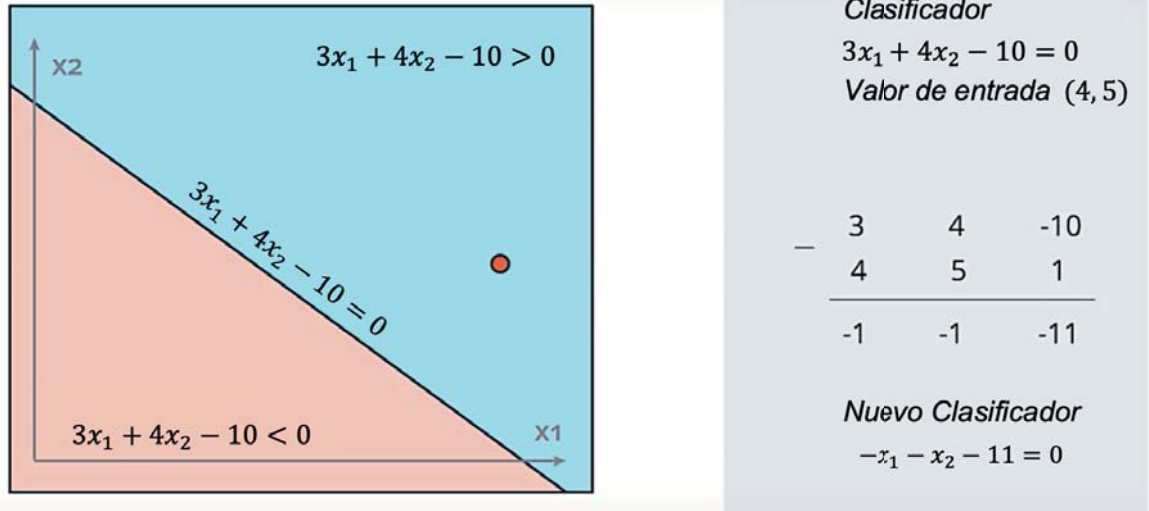


Figura 22. Clasificador lineal.

El nuevo clasificador asigna la etiqueta naranja al dato de entrada, lo cual es correcto. Sin embargo, considerando que los problemas de clasificación necesitan de modelos que se ajusten a una cantidad grande de datos, el método de la Figura 22 generaría muchas variaciones en la clasificación de todos los puntos. Un cambio en el punto 1 podría generar que el punto 2, clasificado correctamente con anterioridad, ahora sea clasificado incorrectamente. Por ello es necesario utilizar una tasa de aprendizaje con la cual se realizan cambios pequeños en el modelo y se evitan cambios drásticos que afecten la clasificación del resto de datos. En la Figura 23 se define una tasa de aprendizaje de 0.01. La tasa de aprendizaje multiplica a los valores del punto evaluado para encontrar el valor de coeficientes con una variación menor respecto a la Figura 22. Este paso se tiene que repetir tantas veces como sea necesario hasta lograr clasificar al punto considerado de manera correcta.

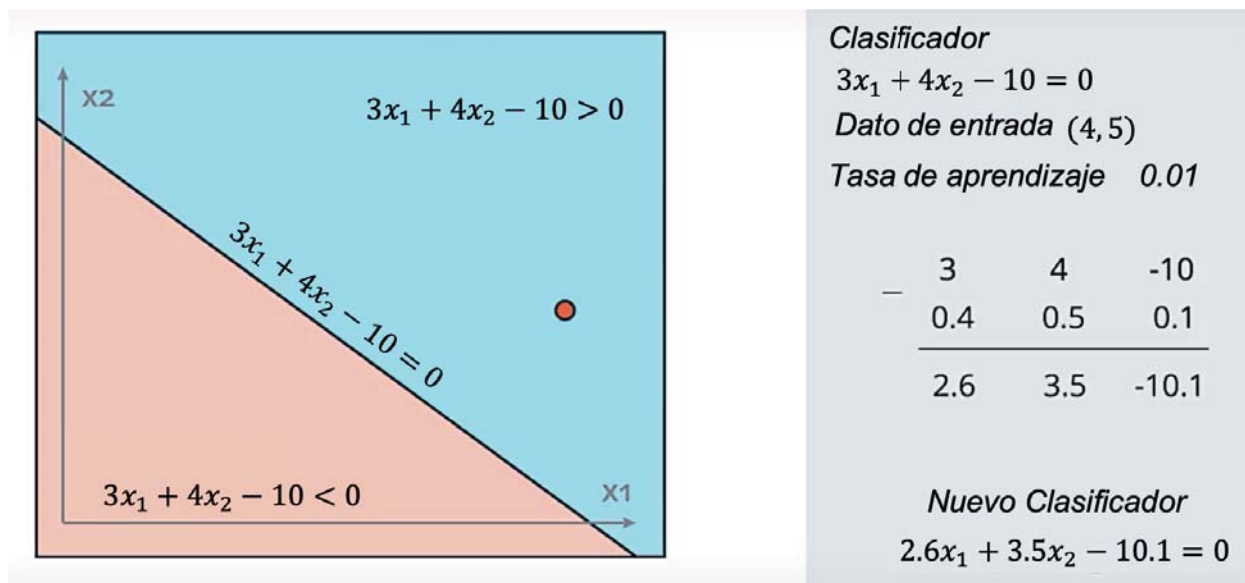


Figura 23. Clasificador lineal con tasa de aprendizaje.

### 3.2.4 Algoritmos de optimización

Dada una función  $f(x)$ , un algoritmo de optimización permite minimizar o maximizar el valor de  $f(x)$ . En el contexto de aprendizaje automático los algoritmos de optimización permiten encontrar el conjunto de parámetros de la red neuronal que permitan minimizar la función de pérdida Ecuación 6.

#### 3.2.4.1 Descenso de gradiente

El descenso de gradiente es un algoritmo que realiza la actualización de los parámetros, de una red neuronal, en el sentido opuesto al gradiente de la función objetivo Ecuación 6. El primer paso para obtener la función del descenso de gradiente es conocer la derivada de la función Sigmoid definida en la Ecuación 3.

$$\sigma'(x) = \sigma(x)(1 - \sigma(x))$$

Ecuación 7. Derivada de la función Sigmoid.

El error total en función de las predicciones y las salidas esperadas está dado por la Ecuación 8.

$$E = -\frac{1}{m} \sum_{i=1}^m (y_i \ln(\hat{y}_i) + (1 - y_i) \ln(1 - \hat{y}_i))$$

Ecuación 8. Fórmula del error.

Para minimizar el valor del error es necesario calcular el gradiente de la Ecuación 8 con respecto a todos los pesos utilizados en la red neuronal. El error total se considera como el promedio de los errores de cada peso.

$$\nabla E = \left( \frac{\partial}{\partial w_1} E, \dots, \frac{\partial}{\partial w_n} E, \frac{\partial}{\partial b} E \right)$$

Ecuación 9. Gradiente del error.

Ahora es necesario calcular la derivada del error con respecto a los pesos por lo que primero se calcula el cambio de los pesos con respecto a la Ecuación 2.

$$\frac{\partial}{\partial w_j} \hat{y} = \frac{\partial}{\partial w_j} \sigma(Wx + b)$$

Ecuación 10. Derivada de la predicción con respecto a pesos.

Desarrollando la Ecuación 10 se obtiene el resultado de la derivada de la predicción con respecto a los pesos.

$$\begin{aligned} \frac{\partial}{\partial w_j} \hat{y} &= \sigma(Wx + b)(1 - \sigma(Wx + b)) \frac{\partial}{\partial w_j} (Wx + b) \\ &= \hat{y}(1 - \hat{y}) \frac{\partial}{\partial w_j} (Wx + b) \\ &= \hat{y}(1 - \hat{y}) \frac{\partial}{\partial w_j} (w_1 x_1 + \dots + w_j x_j + \dots + w_n x_n + b) \\ &= \hat{y}(1 - \hat{y}) x_j \end{aligned}$$

Ecuación 11. Desarrollo de la Ecuación 10.

Al calcular la derivada del error con respecto a los pesos se obtiene una relación entre el error, los pesos, la predicción y la salida deseada.

$$\begin{aligned}
 \frac{\partial}{\partial w_j} E &= \frac{\partial}{\partial w_j} [-y \ln(\hat{y}) - (1 - y) \ln(1 - \hat{y})] \\
 &= -y \frac{\partial}{\partial w_j} \ln(\hat{y}) - (1 - y) \frac{\partial}{\partial w_j} \ln(1 - \hat{y}) \\
 &= -y \frac{1}{\hat{y}} \hat{y}(1 - \hat{y}) x_j - (1 - y) \frac{1}{(1 - \hat{y})} (-1) \hat{y}(1 - \hat{y}) x_j \\
 &= -y(1 - \hat{y}) x_j + (1 - y) \hat{y} x_j \\
 &= [-y + y \hat{y}] x_j + [\hat{y} - y \hat{y}] x_j \\
 &= x_j [-y + y \hat{y} + \hat{y} - y \hat{y}] \\
 &= (\hat{y} - y) x_j
 \end{aligned}$$

Ecuación 12. Desarrollo de la derivada del error con respecto a los pesos.

Un cálculo similar permite obtener la función que relaciona el cambio del error con el parámetro correspondiente al sesgo.

$$\frac{\partial}{\partial b} E = (\hat{y} - y)$$

Ecuación 13. Derivada del error con respecto al sesgo.

Ambos resultados permiten identificar que el cambio del error, también conocido como gradiente, está relacionado con la diferencia existente entre la predicción y el valor real. Un valor de gradiente pequeño indica que los pesos deben cambiar muy poco. Por el contrario, un valor de gradiente grande indica un mayor cambio en el valor de los pesos. Con los resultados obtenidos se define la Ecuación 14 y Ecuación 15 que permiten actualizar los parámetros de una red neuronal, a partir del gradiente obtenido, considerando también un valor  $\alpha$  correspondiente a la tasa de aprendizaje.

$$w'_i = w_i - \alpha \frac{\partial}{\partial w_j} E$$

Ecuación 14. Función para la actualización del valor de los pesos.

$$b = b - \alpha \frac{\partial}{\partial b} E$$

Ecuación 15. Función para la actualización del valor del sesgo.

Dentro del descenso de gradiente hay tres variantes que difieren en la cantidad de datos utilizados durante el cálculo de los parámetros del modelo. Dependiendo de la cantidad de datos, se realiza una compensación entre la precisión de la actualización de parámetros y el tiempo que lleva realizar una actualización.

### 3.2.4.1.1 *Descenso de gradiente por lotes*

Todos los datos del conjunto destinado para entrenamiento se toman en cuenta durante cada actualización. Se obtiene el promedio de los gradientes de todos los datos de entrenamiento para actualizar los parámetros.

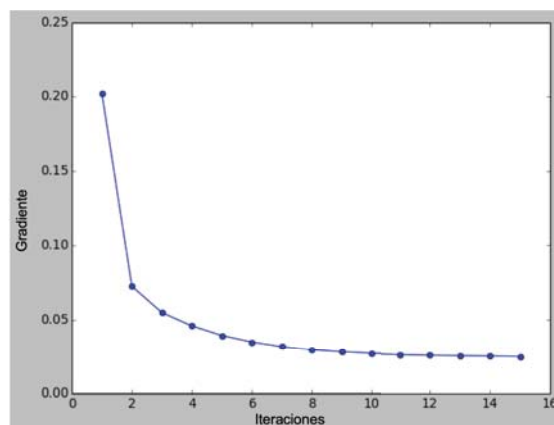


Figura 24. Descenso de gradiente por lote.

La gráfica resultante, Figura 24, es bastante uniforme debido a que se realiza un promedio de todos los gradientes del conjunto de entrenamiento para cada actualización de parámetros. El valor del gradiente disminuye tras cada iteración.



### 3.2.4.1.2 *Descenso de gradiente estocástico SGD*

Una desventaja del descenso de gradiente por lotes se da al trabajar con conjuntos de datos muy grandes donde para cada actualización de parámetros se consideran todos los ejemplos; algo que es costoso computacionalmente y además emplea una cantidad considerable de memoria. En SGD para cada actualización de parámetros se considera un solo ejemplo. Debido a ello se presenta una oscilación que permite saltar a mínimos locales nuevos y potencialmente mejores. Por otro lado, esto en última instancia complica la convergencia al mínimo exacto. Sin embargo, se ha demostrado que cuando se disminuye lentamente la tasa de aprendizaje, SGD muestra el mismo comportamiento de convergencia que el descenso de gradiente por lotes.

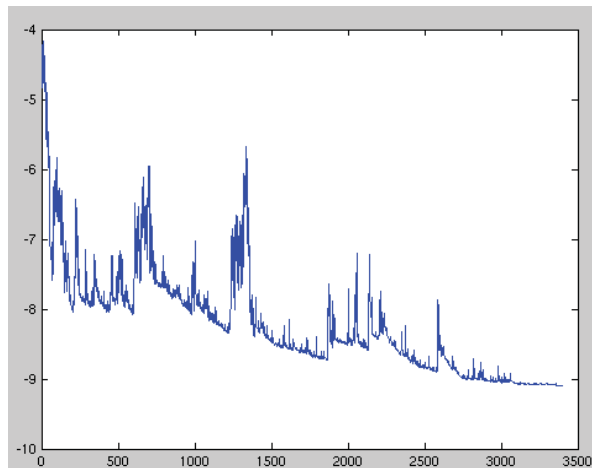


Figura 25. Descenso de gradiente estocástico.

### 3.2.4.1.3 *Descenso de gradiente por mini lotes*

Este método permite tener en un método las ventajas de las dos variantes descritas anteriormente. Utiliza un lote de datos de entrenamiento menor al conjunto de entrenamiento total conocido como mini lote. Al igual que en SGD, el gradiente promedio durante las iteraciones oscila ya que se realiza un promedio de una cantidad determinada de ejemplos a la vez. Al usar el descenso de gradiente por mini lotes, los parámetros se actualizan con frecuencia.

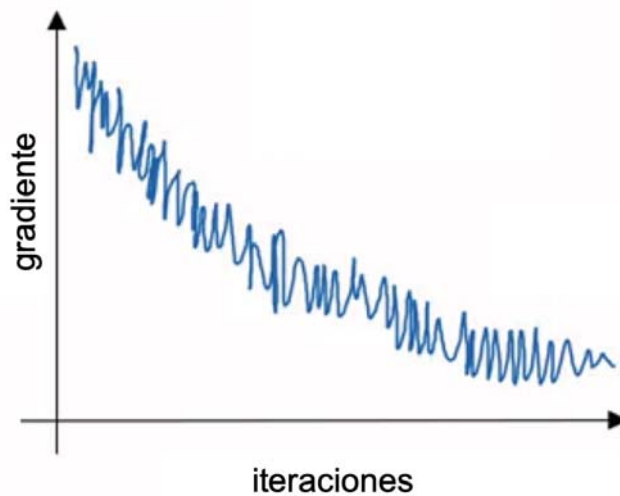


Figura 26. Descenso de gradiente por mini lotes.

### 3.2.4.2 Extensiones y variaciones

El gradiente descendiente es una técnica que en práctica puede presentar diferentes inconvenientes que repercuten en el tiempo de entrenamiento o en la localización de los parámetros óptimos. Estos problemas ocurren por la existencia de puntos de inflexión o mínimos locales donde un valor de gradiente cercano a cero no actualiza significativamente los parámetros de peso e impide continuar con el proceso de aprendizaje Figura 27.

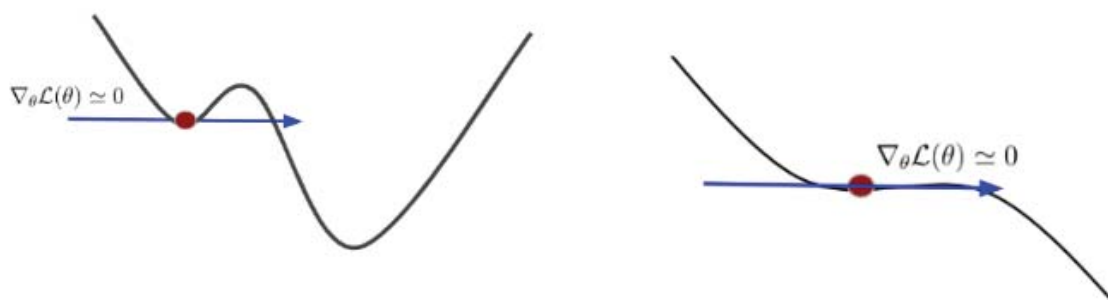


Figura 27. Mínimo local y punto de inflexión de una función.

Otro inconveniente es la presencia de oscilaciones debido al ruido y a las variaciones existentes en el conjunto de datos. Este fenómeno incrementa el tiempo de aprendizaje e inclusive puede imposibilitar la localización de un mínimo en la función objetivo. Por ello existen variaciones del descenso de gradiente que tienen como objetivo dar solución a los problemas descritos anteriormente.

### 3.2.4.2.1 *Descendiente de gradiente con momento*

Descendiente de gradiente	Descendiente de gradiente con momento
$w'_i = w_i - \alpha \frac{\partial}{\partial W} E$	$v_{t+1} = \rho v_t + \frac{\partial}{\partial W} E$ $w'_i = w_i - \alpha v_{t+1}$

Tabla 4. Comparación de la función descendiente de gradiente.

Es una variación del gradiente descendiente para evitar la oscilación que se puede presentar durante la localización del mínimo global. Matemáticamente el momento se representa por la multiplicación de los términos  $\rho$  y  $v$ . De forma intuitiva el momento acumula en el gradiente una velocidad  $v$  durante el entrenamiento. La variable  $\rho$  puede ser considerada una fuerza de fricción que disminuye la velocidad. En general, la velocidad aumenta con el tiempo, por lo que resulta posible salir de puntos de inflexión o mínimos locales debido a que el tamaño del paso depende del gradiente de la función en el punto actual y de la velocidad que se ha acumulado con el tiempo.



Descendiente de gradiente



Descendiente de gradiente con momento

Figura 28. Descendiente de gradiente y descendiente de gradiente con momento.

### 3.2.4.2.2 Algoritmo de Gradiente Adaptativo AdaGrad

En este método se realiza una suma de gradientes cuadrados durante la optimización. En lugar del término de impulso se cuenta con la variable  $g$  que es la suma de los gradientes al cuadrado.

Descendiente de gradiente con momento	AdaGrad
$v_{t+1} = \rho v_t + \frac{\partial}{\partial W} E$ $w'_i = w_i - \alpha v_{t+1}$	$g_0 = 0$ $g_{t+1} = g_t + \left[ \frac{\partial}{\partial W} E \right]^2$ $w'_i = w_i - \alpha \frac{\frac{\partial}{\partial W} E}{\sqrt{g_{t+1} + 1e^{-5}}}$

Tabla 5. Comparación descendiente de gradiente y AdaGrad.

Cuando se actualiza un parámetro de peso, se divide el gradiente actual por la raíz del término  $g$ . El optimizador AdaGrad se puede pensar como una función de pérdida de dos dimensiones en donde el gradiente de la función es pequeño en una dirección y grande en la otra. La suma de los gradientes pequeños hará que la suma cuadrada de los mismos gradientes sea aún más pequeña. Si durante la actualización de los pesos se divide el gradiente por un valor muy pequeño correspondiente a la suma de los gradientes cuadrados se generará un resultado muy grande. Por el contrario, en caso de que la suma de los gradientes sea considerablemente grande generará un resultado muy pequeño. De esta forma se acelera el proceso de actualización de parámetros en el eje con gradientes pequeños al incrementar el valor del cambio necesario. La actualización de valores donde el gradiente es mayor se realiza de manera más lenta. Como resultado el algoritmo realiza actualizaciones en cualquier dirección con la misma proporción.

### 3.2.4.2.3 Propagación cuadrática media RMSProp

AdaGrad	RMSProp
$g_0 = 0$ $g_{t+1} = g_t + \left[ \frac{\partial}{\partial W} E \right]^2$ $w'_i = w_i - \alpha \frac{\frac{\partial}{\partial w_j} E}{\sqrt{g_{t+1} + 1e^{-5}}}$	$g_0 = 0, \quad \beta \cong 0.9$ $g_{t+1} = \beta \cdot g_t + (1 - \beta) \left[ \frac{\partial}{\partial W} E \right]^2$ $w'_i = w_i - \alpha \frac{\frac{\partial}{\partial w_j} E}{\sqrt{g_{t+1} + 1e^{-5}}}$

Tabla 6. Comparación AdaGrad y RMSProp.

Cuando AdaGrad se utiliza en una cantidad significativa de valores, el término del gradiente es grande. Si el gradiente actual se divide por ese valor, el resultado obtenido es muy pequeño. Esto es semejante a utilizar una tasa de aprendizaje muy pequeña con lo cual incrementa drásticamente el tiempo de entrenamiento. En RMSProp también se realiza una suma de gradientes cuadrados sin embargo ese valor se multiplica por una tasa de decaimiento  $\beta$  que impide que el término continúe creciendo. La actualización de los parámetros ocurre de igual manera como sucede con Adagrad. RMSProp es menos propenso a detenerse en mínimos locales o puntos de inflexión ya que recorre un camino directo hacia el mínimo global de la función. Una visualización clara de ello se puede encontrar al final de [42], el cual es un recurso que profundiza en los algoritmos de optimización.

### 3.2.4.2.4 Estimación Adaptativa del Momento Adam

Hasta ahora el momento ha sido utilizado para construir la velocidad del gradiente y actualizar los parámetros del modelo hacia la dirección de esa velocidad. En el caso de AdaGrad y RMSProp, se utiliza la suma de los gradientes cuadrados para escalar el gradiente actual, por lo que es posible hacer actualizaciones de pesos con la misma relación en cada dimensión. Adam resulta de reunir en un algoritmo las ventajas de los optimizadores hasta ahora mencionados.

$m_0 = 0, v_0 = 0$	
$m_{t+1} = \beta_1 m_t + (1 - \beta_1) \frac{\partial}{\partial w_j} E$	Momento
$v_{t+1} = \beta_2 v_t + (1 - \beta_2) \left[ \frac{\partial}{\partial w_j} E \right]^2$	RMSProp
$w'_i = w_i - \frac{\epsilon}{\sqrt{v_{t+1}} + 1e^{-5}} m_{t+1}$	RMSProp + Momento

Tabla 7. Funciones del optimizador Adam.

La primera ecuación es semejante a la utilizada en el descenso de gradiente con momento. El término  $m$  es la velocidad y  $\beta_1$  el coeficiente de fricción. En el caso de Adam  $m$  se conoce como primer momento y  $\beta_1$  es un hiperparámetro. Además, se incluye el término  $1 - \beta_1$  que multiplica al gradiente. La segunda parte le confiere al algoritmo las ventajas de RMSProp al realizar la suma de los gradientes cuadrados. El factor  $1 - \beta_2$  multiplica al gradiente cuadrado. El término  $v$  es el segundo momento y  $\beta_2$  es un hiperparámetro. El primer y segundo momento se inicializan en cero. Después de la primera actualización, el segundo momento se vuelve muy cercano a cero. Cuando se actualizan los parámetros se hace una división por un segundo momento muy pequeño  $v$  resultando en una primera actualización de valores muy grande. La existencia de una primera actualización grande es resultado de inicializar el primer y segundo momento en cero.

$\hat{m}_{t+1} = \frac{m_{t+1}}{1 - \beta_1^t}$ $\hat{v}_{t+1} = \frac{v_{t+1}}{1 - \beta_2^t}$	Corrección del sesgo
$w'_i = w_i - \frac{\epsilon}{\sqrt{\hat{v}_{t+1}} + 1e^{-5}} \hat{m}_{t+1}$	RMSProp + Momento

Tabla 8. Elemento de corrección en optimizador Adam.

Para evitar una primera actualización grande un elemento de corrección se incluye en Adam. Después de la primera actualización del primer y segundo momento  $m$  y  $v$  se hace una estimación sin sesgar de ambos momentos tomando en cuenta el número de pasos  $t$  realizados. Como resultado, la primera actualización de los parámetros no es grande y no perturba el proceso de entrenamiento desde el inicio.

### 3.2.5 Propagación y retropropagación

El proceso de entrenamiento de una red neuronal implica encontrar los valores de pesos y sesgos que al ser operados en la red den lugar a la salida deseada. Debido a las interconexiones existentes. El proceso de propagación, Figura 29, permite conocer la respuesta de una red neuronal a partir de los datos de entrada proporcionados.

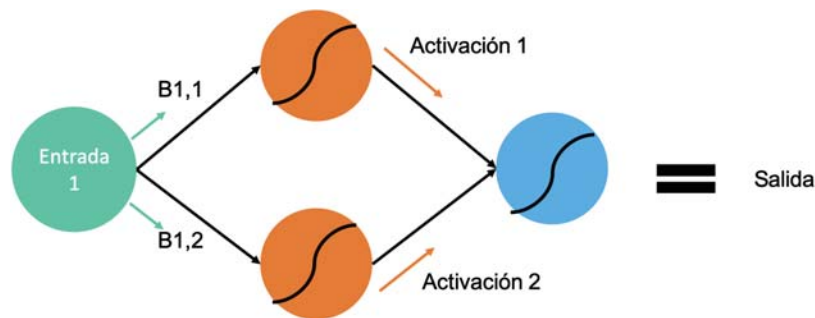


Figura 29. Propagación hacia delante.

El proceso de retropropagación, Figura 30, ocurre en el sentido contrario y permite calcular el error atribuible a cada neurona a través del descenso de gradiente.

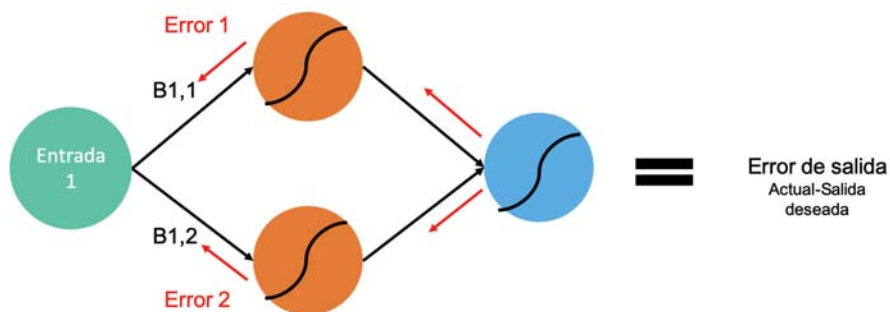


Figura 30. Retropropagación de red neuronal.

La Figura 31 muestra un conjunto de pesos asignados aleatoriamente a una red neuronal. Mediante el proceso de propagación y utilizando como activación la función Sigmoid se puede obtener la salida de la red en función de los valores de entrada. Si arbitrariamente se asigna que el valor 1 es la salida esperada se puede utilizar la Ecuación 12 para ajustar el valor de los pesos y mejorar el desempeño de la red.

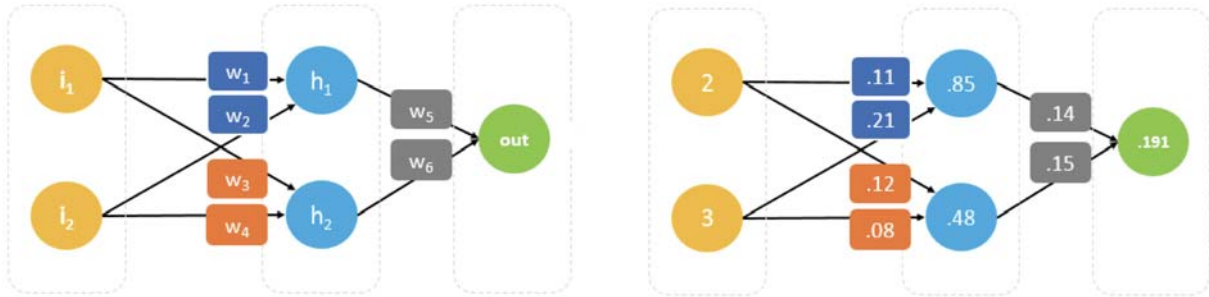


Figura 31. Red neuronal con valores arbitrarios.

La Ecuación 16 muestra el desarrollo de las operaciones involucradas en el proceso de propagación.

$$\begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} 0.11 & 0.12 \\ 0.21 & 0.08 \end{bmatrix} \begin{bmatrix} 0.14 \\ 0.15 \end{bmatrix} = \begin{bmatrix} 0.85 & 0.48 \end{bmatrix} \begin{bmatrix} 0.14 \\ 0.15 \end{bmatrix} = \begin{bmatrix} 0.191 \end{bmatrix}$$

Ecuación 16. Desarrollo de operaciones de la Figura 31.

Considerando la Ecuación 12 es posible calcular el error con respecto al peso w6.

$$\begin{aligned} \frac{\partial}{\partial w_6} E &= (\hat{y} - y)h_2 \\ &= (0.191 - 1) * 0.48 \\ &= -0.38832 \end{aligned}$$

Ecuación 17. Error con respecto al peso w6.

Considerando la Ecuación 12 también es posible calcular el error con respecto al peso w1.



$$\begin{aligned}\frac{\partial}{\partial w_1} E &= (\hat{y} - y)i_1 \\ &= (0.191 - 1) * 2 \\ &= -1.618\end{aligned}$$

Ecuación 18. Error con respecto al peso 1.

Considerando el resultado de la Ecuación 17 y la Ecuación 14 se puede conocer el nuevo valor del peso  $w_6$ .

$$\begin{aligned}w'_6 &= w_6 - \alpha \frac{\partial}{\partial w_j} E \\ &= 0.15 + \alpha[0.38832]\end{aligned}$$

Ecuación 19. Actualización del peso  $w_6$ .

Para el caso del peso  $w_1$  se puede utilizar la misma ecuación

$$\begin{aligned}w'_1 &= w_1 - \alpha \frac{\partial}{\partial w_1} E \\ &= 0.11 + \alpha[1.618]\end{aligned}$$

Ecuación 20. Actualización del peso  $w_1$ .

De acuerdo con la Ecuación 19 y la Ecuación 20 es posible generar un sistema de ecuaciones que permita realizar las operaciones que den lugar a la actualización de los pesos en función del error obtenido. Para este ejemplo se define la tasa de aprendizaje  $\alpha$  con un valor de 0.05.

$$\begin{aligned}\begin{bmatrix} w_5 \\ w_6 \end{bmatrix} &= \begin{bmatrix} w_5 \\ w_6 \end{bmatrix} - \alpha \begin{bmatrix} (\hat{y} - y) h_1 \\ (\hat{y} - y) h_2 \end{bmatrix} = \begin{bmatrix} w_5 \\ w_6 \end{bmatrix} - \begin{bmatrix} \alpha h_1 (\hat{y} - y) \\ \alpha h_2 (\hat{y} - y) \end{bmatrix} \\ \begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} &= \begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} - \alpha \begin{bmatrix} (\hat{y} - y) i_1 \\ (\hat{y} - y) i_2 \end{bmatrix} \begin{bmatrix} w_5 & w_6 \end{bmatrix} = \begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} - \begin{bmatrix} \alpha i_1 (\hat{y} - y) w_5 & \alpha i_1 (\hat{y} - y) w_6 \\ \alpha i_2 (\hat{y} - y) w_5 & \alpha i_2 (\hat{y} - y) w_6 \end{bmatrix}\end{aligned}$$

Ecuación 21. Actualización de los pesos de la Figura 31 en función del error.

$$\begin{bmatrix} w_5 \\ w_6 \end{bmatrix} = \begin{bmatrix} 0.14 \\ 0.15 \end{bmatrix} - 0.05 \left[ (-0.809) \begin{bmatrix} 0.85 \\ 0.48 \end{bmatrix} \right] = \begin{bmatrix} 0.14 \\ 0.15 \end{bmatrix} - \begin{bmatrix} -0.034 \\ -0.019 \end{bmatrix} = \begin{bmatrix} 0.17 \\ 0.17 \end{bmatrix}$$

$$\begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} = \begin{bmatrix} .11 & .12 \\ .21 & .08 \end{bmatrix} - 0.05 \left[ (-0.809) \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right] \begin{bmatrix} 0.14 & 0.15 \end{bmatrix} = \begin{bmatrix} .11 & .12 \\ .21 & .08 \end{bmatrix} - \begin{bmatrix} -0.011 & -0.012 \\ -0.017 & -0.018 \end{bmatrix} = \begin{bmatrix} .12 & .13 \\ .23 & .10 \end{bmatrix}$$

Ecuación 22. Sustitución numérica de la Ecuación 21.

Los resultados de la Ecuación 22 se utilizan para actualizar los pesos de la Figura 31. Con los nuevos valores, Figura 32, la predicción obtenida es de 0.26 la cual es más cercana a 1 en comparación con el 0.191 obtenido inicialmente. Este proceso se realiza durante varias iteraciones hasta lograr la construcción de un modelo que realice predicciones se ajuste a los datos de entrenamiento.

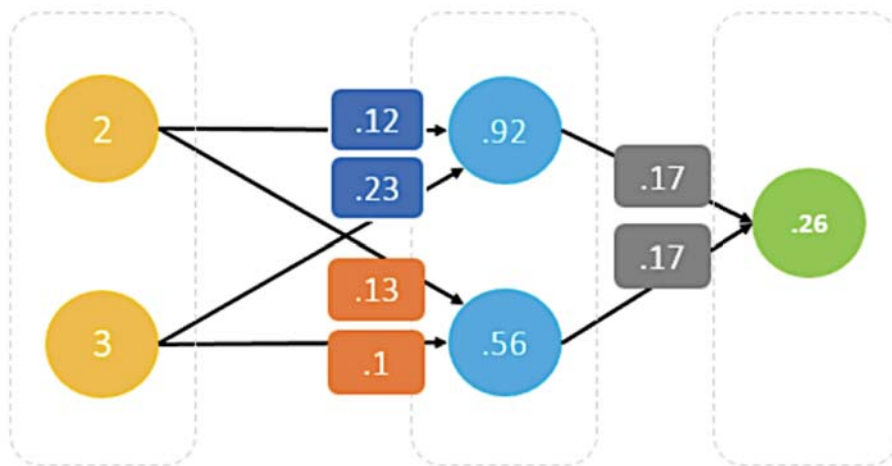


Figura 32. Actualización del valor de los pesos.

### 3.3 Redes Neuronales Convolucionales

#### 3.3.1 Imágenes digitales

Para los humanos la habilidad para el análisis de imágenes es innata. Identificar un entorno, así como los objetos que en él se localizan resulta ser una tarea sencilla. Debido a ello reconocemos patrones, generalizamos a partir de conocimiento previo y nos adaptamos a diferentes medios. Para el caso de los equipos de cómputo el registro de imágenes resulta ser diferente. Cada imagen guardada es representada por un arreglo numérico o matriz que se traduce en información visual para el usuario final. La

Figura 33 muestra una imagen en conjunto con una sección del arreglo numérico que la representa.



Figura 33. Del lado izquierdo se puede ver una imagen en el espacio de color RGB. Del lado derecho se representa una sección de los píxeles de un canal del espacio de color RGB pertenecientes a la misma imagen.

### 3.3.2 Conectividad local

Procesar una imagen con una red neuronal involucra su transformación a un vector de dimensión 1. En la Figura 34 la imagen de entrada con dimensiones 4 x 4 x 1 se redimensiona a un vector de 16 x 1 para ser procesada por la red. El vector de entrada, correspondientes a la imagen redimensionada, se conecta totalmente a las 4 neuronas de la capa oculta generando una salida única.

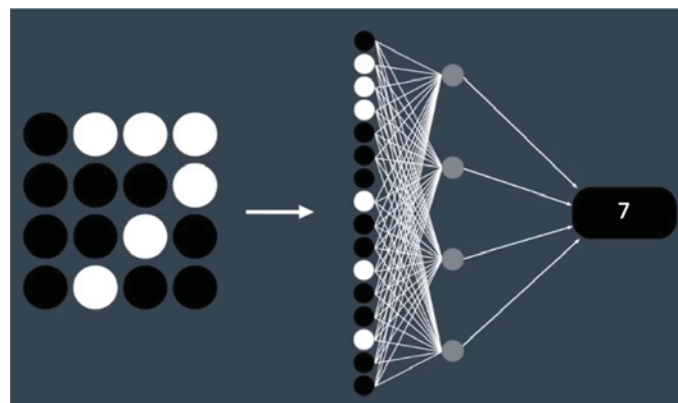


Figura 34. Entrada totalmente conectada a la capa oculta.

Realizar la conexión de todos los valores de la imagen con cada neurona de la capa oculta puede resultar innecesario y redundante. En el caso de la Figura 34 cada neurona de la capa oculta retiene información de todos los píxeles que conforman la imagen original. En lugar de ello se pueden asignar regiones locales, de la imagen original, a cada neurona de la capa oculta. La Figura 35 utiliza la conectividad local a partir de la división en 4 secciones de la imagen original. Con la conectividad local se utiliza un menor número de parámetros; lo cual reduce el tiempo de cómputo, se aprende de las características importantes y se logra una mejor generalización. La operación que permite una conectividad local se conoce como convolución y es la operación básica en las Redes Neuronales Convolucionales.

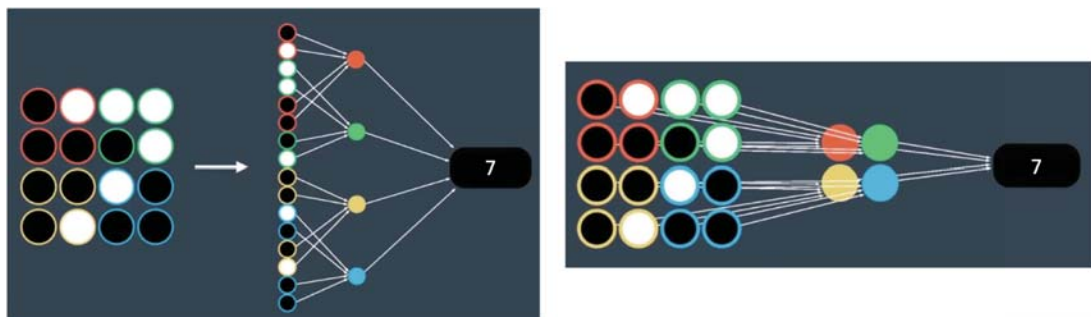


Figura 35. Conectividad local en redes neuronales.

### 3.3.3 Operadores

#### 3.3.3.1 Convolución

La primera capa de una ConvNet es la capa convolucional. El tamaño de la conectividad local dependerá del operador o filtro utilizado. En la Figura 36 el filtro utilizado tiene dimensiones de 3 x 3 generando una conectividad local de 9 píxeles para cada región de la imagen. En el caso de la Figura 35 el filtro utilizado cuenta con dimensiones de 2 x 2 realizando una conexión local a 4 píxeles.

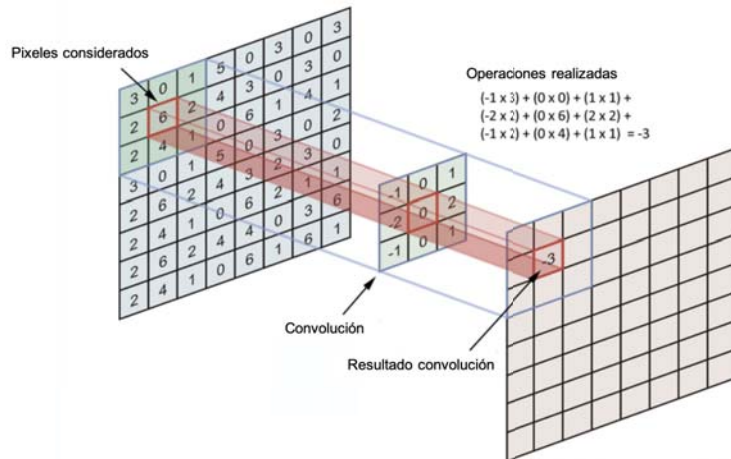


Figura 36. Operación de convolución en un arreglo matricial.

El número de salidas de la operación convolución dependerá del número de filtros utilizados. A estas salidas se les conoce como mapas de activación. La Figura 37 muestra el valor numérico de algunos filtros, así como su resultado al aplicarlo en la imagen original. En el caso de una ConvNet se utilizan cientos de filtros, en cada capa de convolución, para extraer las características relevantes de la imagen original.

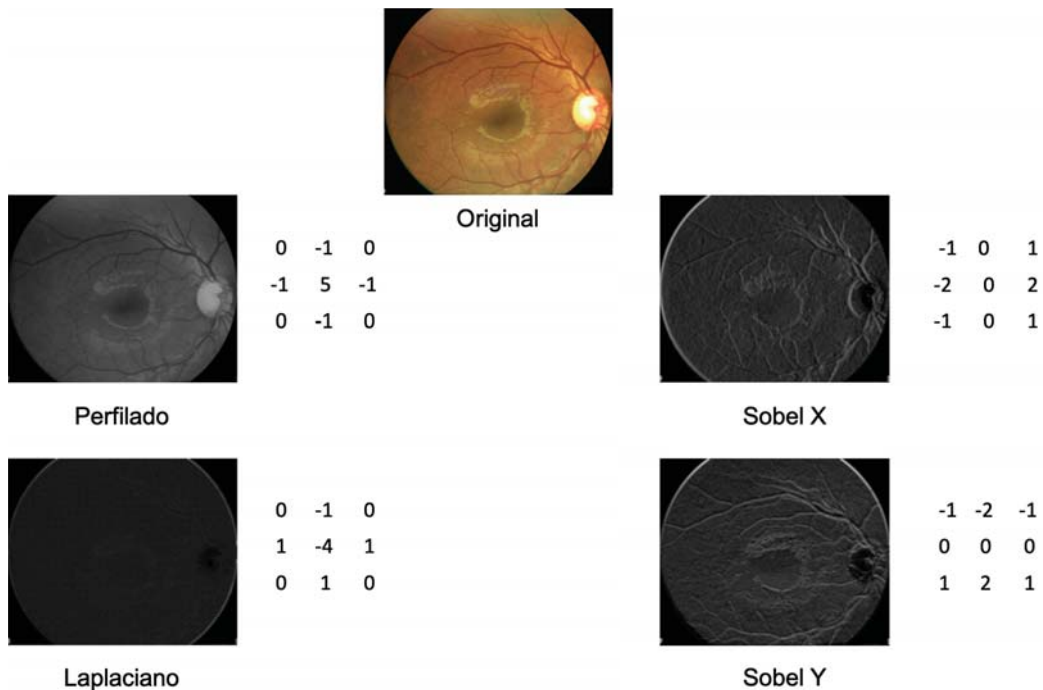


Figura 37. Filtros empleados en el procesamiento de imágenes.

En la Figura 38 se muestra un filtro de 7x7 empleado para la detección de curvas. En el arreglo numérico, los valores que coinciden con la posición de la curva son mayores en comparación con el resto.

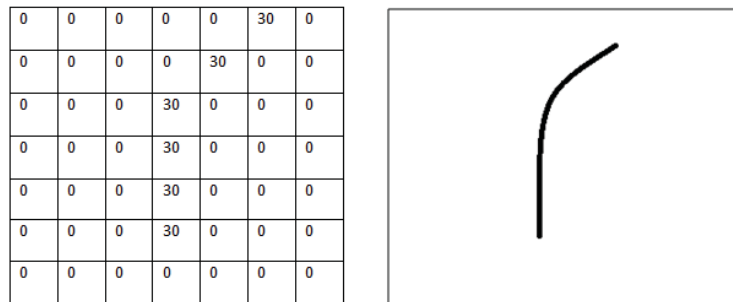
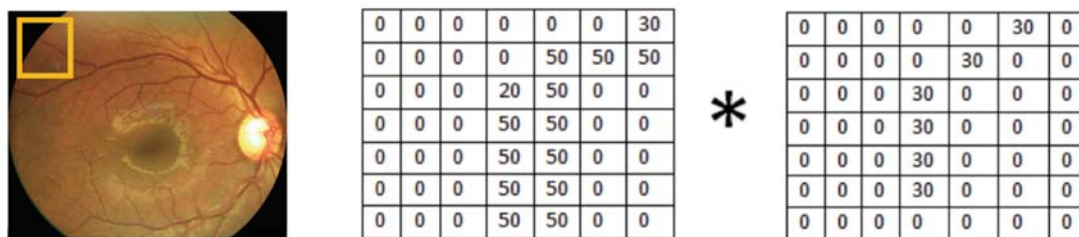


Figura 38. Del lado izquierdo representación numérica de un filtro. Del lado derecho representación gráfica del mismo filtro.

Al emplear el filtro de la Figura 38 y realizar la operación de convolución sobre la región marcada en la imagen de la Figura 39 se obtiene un valor alto.



$$\text{Convolución} = (50 \times 30) + (50 \times 30) + (50 \times 30) + (20 \times 30) + (50 \times 30) = 6600$$

Figura 39. De izquierda a derecha. Imagen con región de interés señalada. Matriz de un canal de la región de interés. Filtro para la detección de curvas.

Para la Figura 40 la misma operación de convolución, evaluada en una sección diferente produce un valor igual a cero. Por lo tanto, el resultado de la convolución depende tanto del filtro como de la sección considerada. El filtro de la Figura 38 permite obtener valores altos en regiones donde se localizan curvas semejantes a la descrita por el filtro.

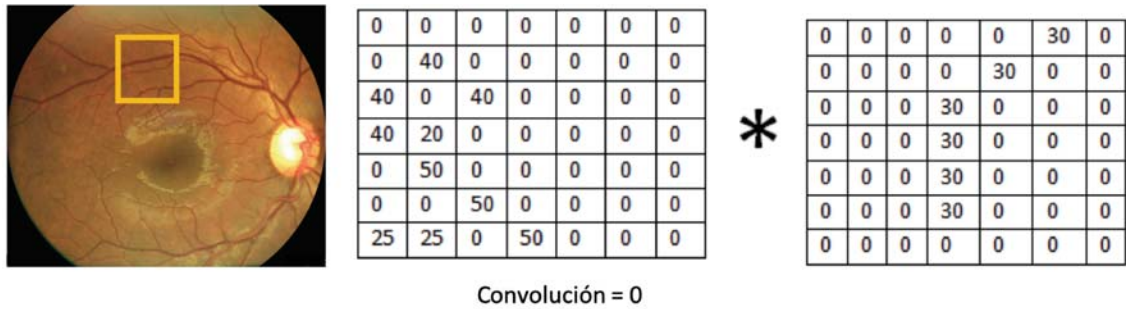


Figura 40. De izquierda a derecha. Imagen con región de interés señalada. Matriz de un canal de la región de interés. Filtro para la detección de curvas.

### 3.3.3.2 Paso [Stride]

Es un parámetro dentro de las ConvNets que indica el número de píxeles que recorre el filtro de convolución cada vez que opera una nueva región. Cuando el paso tiene un valor de una unidad, el filtro se mueve un píxel a la vez. En caso de ser igual a dos, el filtro se mueve dos píxeles después de cada operación de convolución. El valor del paso determina la cantidad de información retenida, así como las dimensiones del mapa de activación.

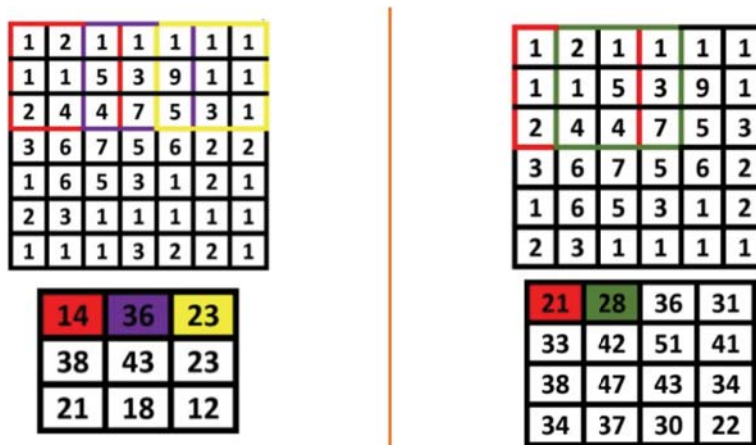


Figura 41. Del lado izquierdo convolución con paso igual a dos. Del lado derecho operación de convolución con paso igual a uno.



### 3.3.3.3 Relleno [Padding]

Es un parámetro no entrenable en las ConvNets. En ocasiones resulta conveniente rellenar la matriz, correspondiente a la imagen de entrada, con ceros alrededor del borde Figura 42. Esto permite realizar la operación de convolución en elementos que se localizan al borde de la imagen. El relleno con ceros también permite ajustar el tamaño final de los mapas de activación.

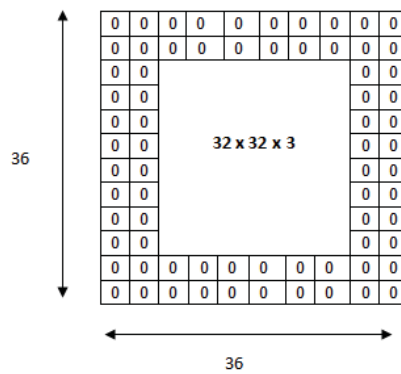


Figura 42. Relleno con ceros a una imagen de 32 X 32 X 3.

### 3.3.3.4 Unidad Lineal Rectificada [ReLU]

La Unidad Lineal Rectificada ReLU es la función de activación comúnmente empleada después de la operación de convolución. Se comporta de manera lineal para valores positivos y regresa cero para todos los valores negativos Figura 43.

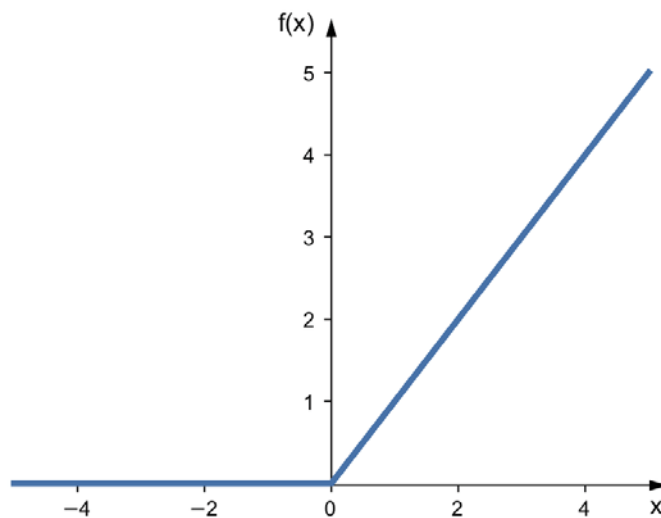


Figura 43. Función de activación ReLU.



Debido a la naturaleza de la función el costo computacional es bajo. Por lo tanto, el modelo toma menos tiempo para su entrenamiento. También converge más rápido; la linealidad implica que la pendiente no se estabiliza o satura, aunque  $x$  incremente. Otro punto que considerar es la existencia de una escasez en la activación dado que ReLU es cero para todas las entradas negativas. Debido a ello existe la probabilidad de que algunas unidades no se activen en absoluto. Esto puede ser deseable considerando que este tipo de modelos buscan emular las funciones neuronales realizadas biológicamente. Si bien contamos con miles de millones de neuronas, no todas ellas se activan todo el tiempo. Cumplen con diferentes roles y son activadas por diferentes señales.

### 3.3.3.5 Agrupamiento [Pooling]

La agrupación espacial es un parámetro no entrenable que reduce las dimensiones de cada mapa de activación. De esta manera se disminuye el número de parámetros y cálculos en la red, se evita el sobreajuste y se impacta el costo computacional de procesamiento. La agrupación espacial le confiere la propiedad de invariabilidad al modelo. La presencia de transformaciones, distorsiones y/o traslaciones de la información no cambia la salida de agrupación debido a que se toma el valor máximo, promedio o la suma de los valores de entrada Figura 44.

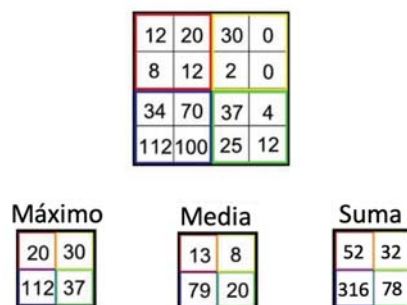


Figura 44. Salidas de la operación de agrupación.

### 3.3.4 Técnicas de regularización

Durante la fase de entrenamiento los modelos de aprendizaje automático buscan ajustar sus valores a un conjunto de datos de entrenamiento. Esto implica la identificación de patrones, así como el manejo adecuado del ruido presente. Estos factores en conjunto pueden provocar la generación de modelos sobre ajustados. Un modelo sobre ajustado presenta un excelente desempeño con los datos de entrenamiento, pero una generalización pobre al ser utilizado en un conjunto de datos diferentes. Utilizando técnicas de regularización es posible resolver este problema.

#### 3.3.4.1 $L1$ & $L2$

Es la forma de regularización más común. El objetivo es actualizar la función del error Ecuación 8 añadiendo otro término conocido como término de regularización  $\lambda$ . Este método busca limitar la complejidad del modelo a través de la reducción del valor de los pesos [43]. El uso de modelos simples permite una mejor generalización y evita la generación de un sobre ajuste Figura 45.

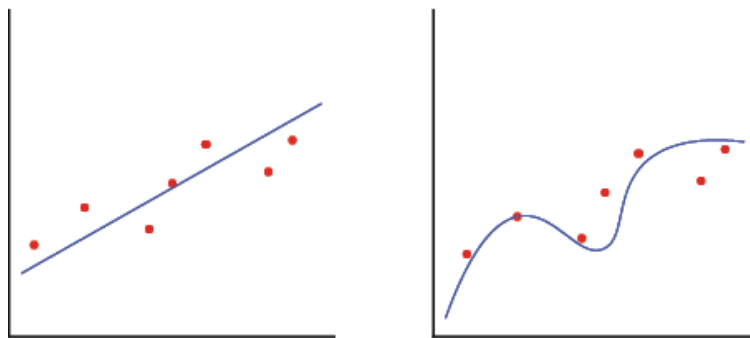


Figura 45. Modelo simple y modelo complejo.

Considerando el uso de los términos de regularización se puede redefinir la Ecuación 8.

$$E' = E + \lambda \sum_{j=1}^N |w_j| \quad \text{L1}$$

$$E' = E + \lambda \sum_{j=1}^N w_j^2 \quad \text{L2}$$

Ecuación 23. Regularización L1 y L2.

El valor de L1 elimina algunos parámetros ya que durante la reducción del valor de los pesos algunos coeficientes se vuelven cero. L2 también reduce el valor de los pesos sin embargo evita asignarle un coeficiente igual a cero para mantener la influencia de todos los parámetros durante las predicciones del modelo.

#### 3.3.4.2 Deserción [Dropout]

Esta técnica de regularización consiste en desactivar aleatoriamente neuronas en el modelo afectando las conexiones de entrada y salida en ese punto. En cada ciclo de entrenamiento se selecciona un conjunto diferente de neuronas lo que resulta en un conjunto diferente de conexiones disponibles. De esta manera se evita una adaptación entre neuronas que perjudique el desempeño del modelo. En el caso de las ConvNets esta técnica sucede comúnmente en la capa previa a la capa de decisión y no en todo el modelo.

#### 3.3.4.3 Aumento de datos

Otra técnica de regularización consiste en incrementar el número de datos de entrenamiento. Una forma sencilla de lograrlo es realizando transformaciones a los datos de entrada; Figura 46. En el caso de las imágenes se pueden rotar, trasladar, escalar o acercar.

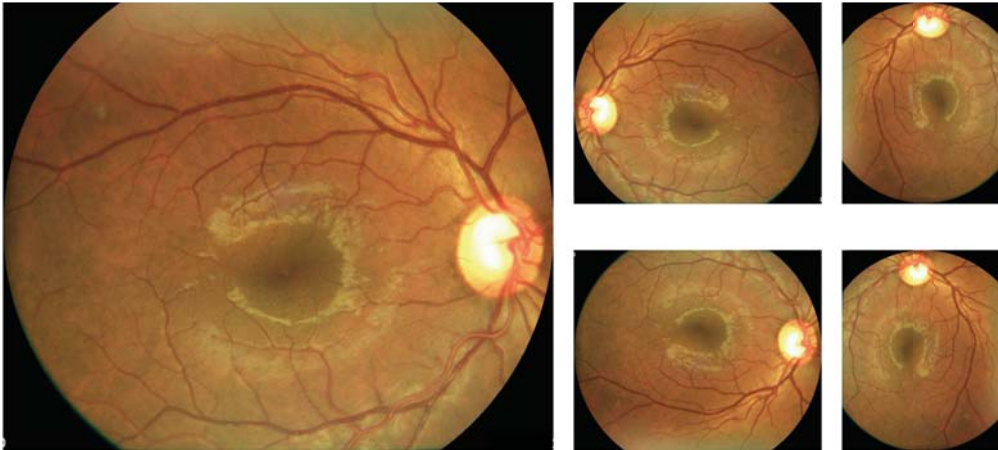


Figura 46. Aumento de datos en imágenes.

#### 3.3.4.4 *Detención temprana*

Cuando el desempeño de un modelo, en el conjunto de datos de validación, no mejora o por el contrario incrementa su error, es viable realizar una detención temprana del entrenamiento.

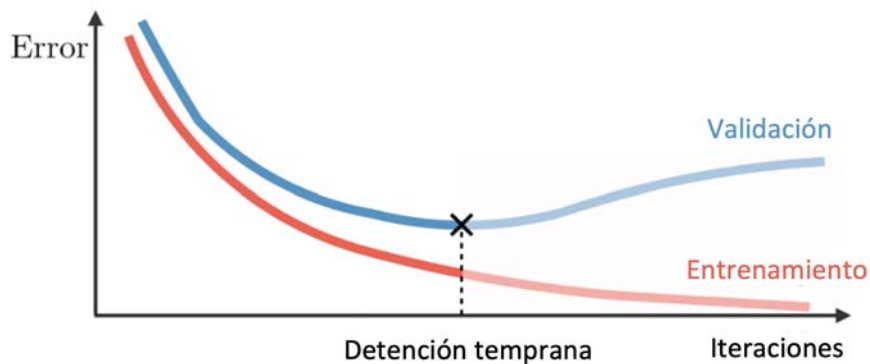


Figura 47. Gráfica detención temprana.

En la Figura 47 el punto marcado con x es el óptimo para realizar una detención temprana. Después de esa iteración el error de validación comienza a incrementar y el error de entrenamiento continúa descendiendo. Este comportamiento es un indicador de un sobre ajuste.

### 3.3.5 Capas de Clasificación

#### 3.3.5.1 Capa Totalmente Conectada [FC] y de Agrupación Promedio Global [GAP]

El uso de una capa Totalmente Conectada, FC, en una ConvNet tiene la finalidad de clasificar una imagen a partir de los resultados de la convolución. El resultado de la convolución se redimensiona a un vector de dimensión  $n \times 1$ , donde cada valor representa la probabilidad de que ciertas características correspondan a una clase. Por ejemplo, si se cuenta con una imagen de retinopatía diabética, las características que representan lesiones propias de la enfermedad deben tener altas probabilidades para etiquetar la imagen como enferma. La cantidad de parámetros en la capa totalmente conectada es proporcional a los parámetros de la ConvNet. Por consecuencia, en caso de tener una arquitectura compleja, se necesitan mayores recursos computacionales además de surgir la posibilidad de un sobre ajuste. Esta situación podría no ser controlada mediante el uso de técnicas de regularización.

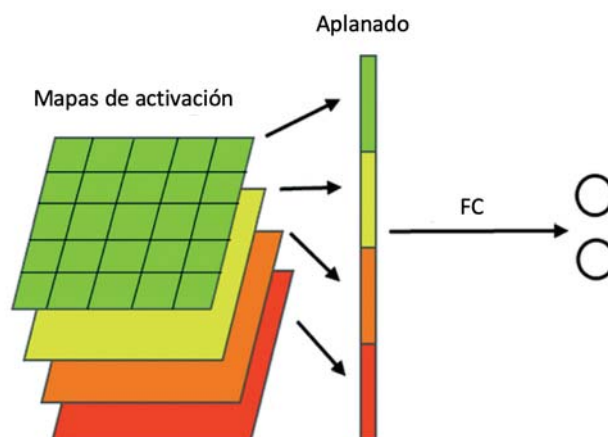


Figura 48. Capa totalmente conectada.

El uso de una capa de Agrupación Promedio Global GAP reduce la posibilidad de sobre ajuste al reducir el número total de parámetros en el modelo. Las GAP, Figura 49, reducen cualquier tensor con dimensiones  $h \times w \times d$  a dimensiones  $1 \times 1 \times d$  a partir de la media de todos los valores.

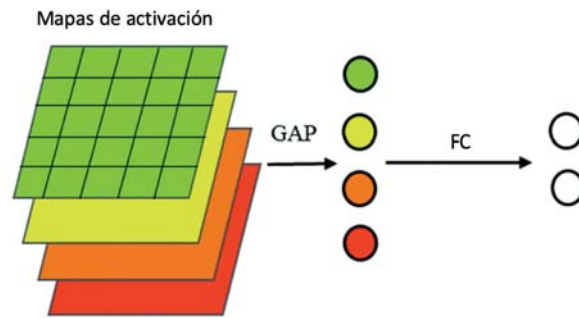


Figura 49. Capa de Agrupación Promedio Global.

### 3.3.6 Transferencia de aprendizaje

Durante la etapa de entrenamiento resultaría útil contar con valores iniciales de pesos que se aproximen a los óptimos para la extracción de características relevantes en los datos de entrada. En general los filtros presentes en las primeras capas de una ConvNet se encargan de la extracción de características de bajo nivel como los son colores o bordes Figura 50. La siguiente capa reconoce formas simples de la imagen de entrada, a partir de la información proporcionada por la primera capa Figura 50. La capa 3, Figura 51, aprende a reconocer texturas y elementos que constituyen al objeto de interés. Finalmente, los filtros en las últimas capas, Figura 51, se activan por características de alto nivel como lo serían hemorragias o microaneurismas. Con la transferencia de aprendizaje es posible utilizar los pesos entrenados en otros modelos y extraer características de bajo y alto nivel.

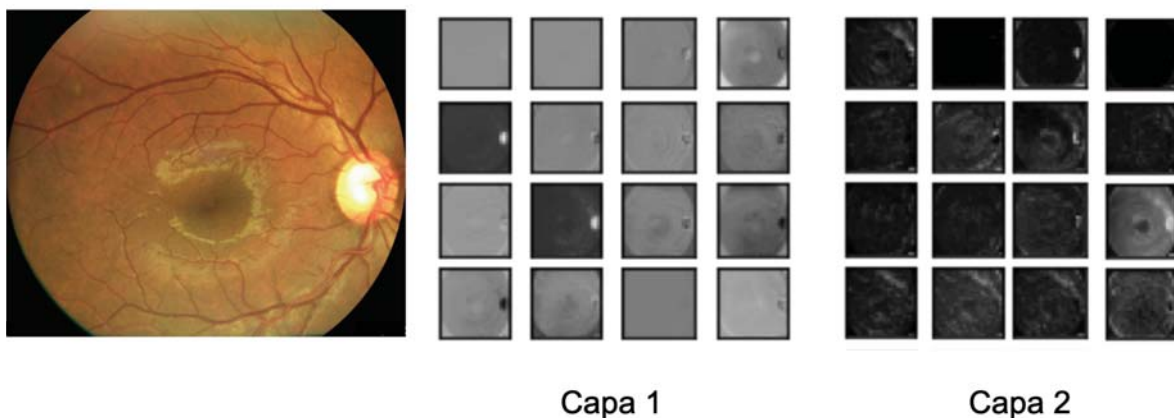


Figura 50. Primeras capas de una ConvNet.

Utilizando transferencia de aprendizaje no es necesario contar con un conjunto de datos significativamente grande, no se requiere de una gran capacidad computacional y se logra una convergencia en menor tiempo.

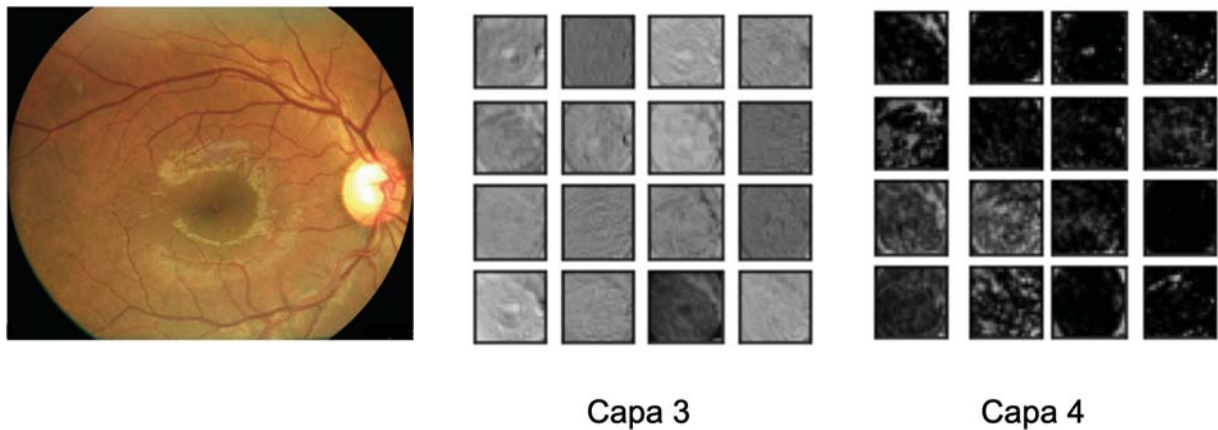


Figura 51. Últimas capas de una ConvNet.

La Figura 52 muestra una tabla de decisión que relaciona el tamaño de los datos de entrenamiento y su semejanza con los datos del modelo previamente entrenado. A partir de ello es posible conocer las capas del modelo que deben ser entrenadas.

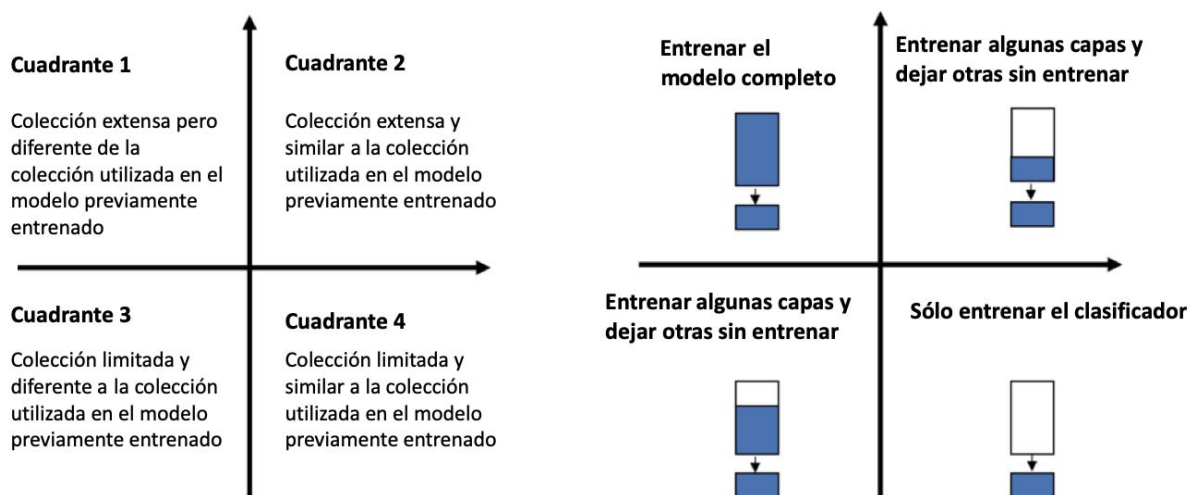


Figura 52. Tabla de decisión para transferencia de aprendizaje.

### 3.3.7 Arquitectura de una ConvNet

Con los elementos hasta ahora descritos se define la arquitectura de una ConvNet. En ella varían el número y la organización de bloques que la conforman. La arquitectura determina la cantidad de parámetros, el tiempo necesario de entrenamiento y el tamaño del modelo.

A continuación, se enlistan algunas arquitecturas reportadas en el estado del arte que han sido referentes durante la evolución de las ConvNets.

- LeNet (1990) Yann LeCun, Leon Bottou, Yosuha Bengio y Patrick Haffner propusieron una arquitectura de red neuronal para el reconocimiento de caracteres escritos a mano e impresos en máquina, que llamaron LeNet-5 [44].
- AlexNet (2012) Alex Krizhevsky et al. [45] publicaron AlexNet, una versión más profunda y amplia de LeNet que permitió la reducción del error reportado en ILSVRC. Se puede atribuir a este trabajo un avance significativo con respecto a los enfoques anteriores y la aplicación generalizada actual de las ConvNets.
- ZF Net (2013) El ganador del ILSVRC 2013 fue una ConvNet de Matthew Zeiler y Rob Fergus. Se realizó a partir del ajuste de los hiperparámetros de la arquitectura AlexNet [46].
- GoogLeNet (2014) El ganador del ILSVRC 2014 fue la ConvNet de Szegedy et al. Su principal contribución fue el desarrollo de un módulo de inicio que disminuyó drásticamente el número de parámetros en la red. Cuatro millones en comparación con los sesenta millones de AlexNet [47]. Inception V3 es la variante más reciente de GoogLeNet.



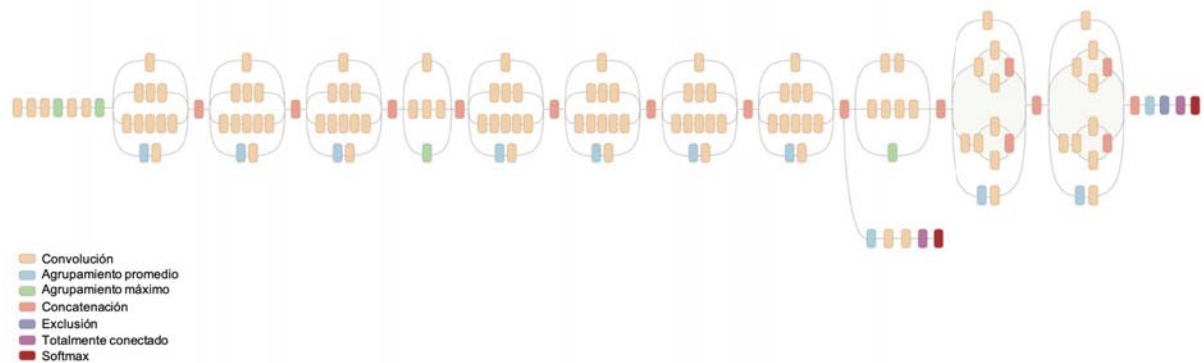


Figura 53. Arquitectura Inception V3.

- VGGNet (2014) Fue el segundo lugar en ILSVRC 2014. Demostró que la profundidad de la red, número de capas, es un componente crítico para un buen desempeño [48].
- ResNet (2015) Las redes residuales ganaron el ILSVRC de 2015. Actualmente son la opción predeterminada para usar ConvNets en la práctica [49].

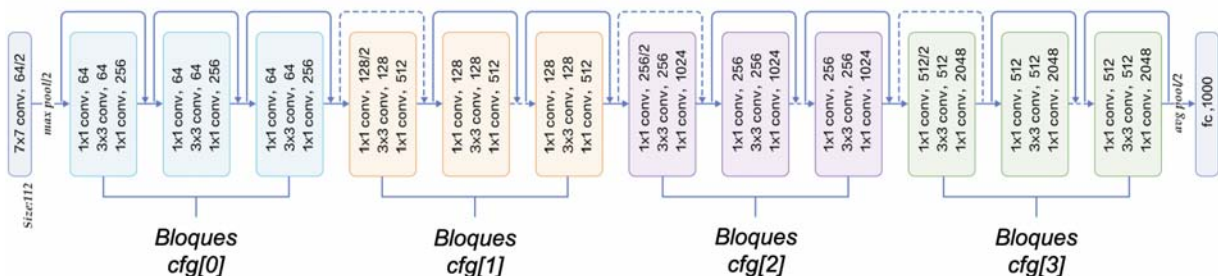


Figura 54. Arquitectura ResNet.

- DenseNet (2016) Esta arquitectura tiene cada capa directamente conectada a todas las demás de forma anticipada. Se ha demostrado que tiene un mejor desempeño sobre arquitecturas anteriores en tareas de reconocimiento de objetos [50].

La Figura 55 muestra la relación entre la cantidad de operaciones y la exactitud que reporta cada arquitectura con el conjunto de imágenes ImageNet. El tamaño del círculo correspondiente a cada arquitectura es proporcional a la memoria empleada por el modelo.

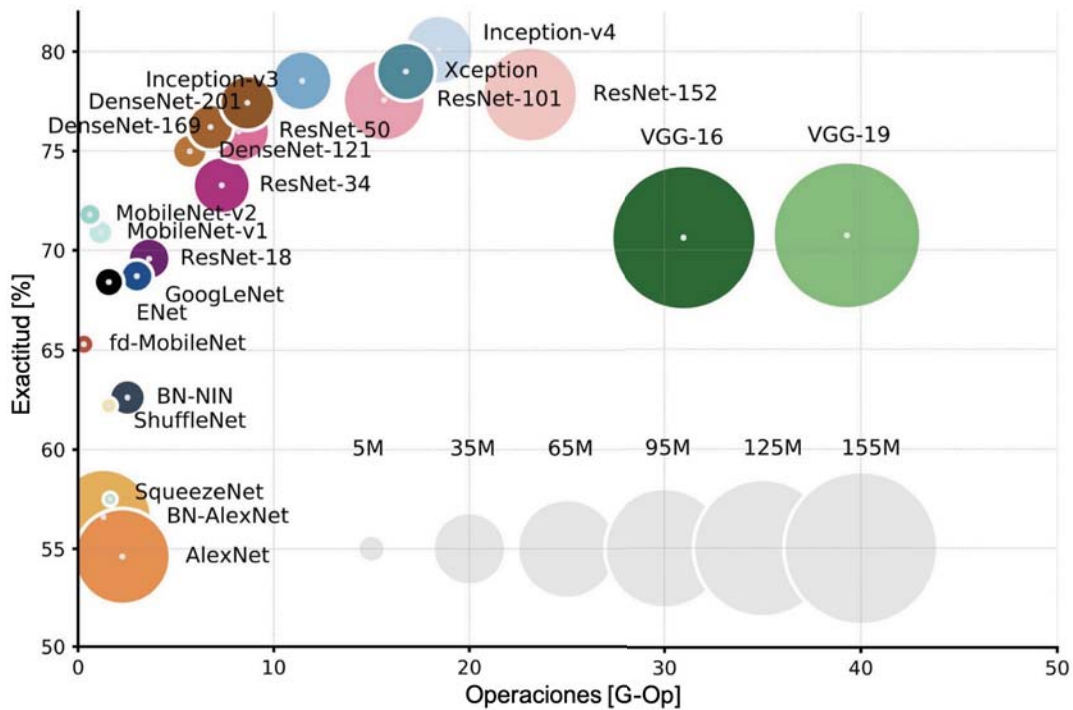


Figura 55. Relación entre exactitud y número de operaciones de arquitecturas estándar.

### 3.3.8 Entrenamiento de ConvNets

Una vez considerados todos los elementos, hasta ahora descritos, es posible iniciar el proceso de entrenamiento.

1. Se inicializan los parámetros del modelo. Esto puede ser de manera aleatoria o mediante transferencia de aprendizaje.
2. La red toma un conjunto de imágenes de entrenamiento como entrada, se realiza la propagación hacia adelante (convolución, ReLU, agrupación, capa totalmente conectada o Agrupación Promedio Global) y se obtienen las probabilidades de salida para cada clase Figura 56.
3. Se calcula el error de la predicción.

4. Se usa retropropagación para calcular los gradientes del error con respecto a todos los pesos en la red y así actualizar los valores de los pesos que minimicen el error de salida. Otros parámetros como el número de filtros, el tamaño del filtro y la arquitectura de la red se establecen desde el inicio y no cambian durante el proceso de entrenamiento.
5. Se repiten los pasos 2 a 4 con todas las imágenes del conjunto de entrenamiento.

Dentro de una ConvNet existen una serie de parámetros conocidos como hiperparámetros. Los hiperparámetros controlan directamente el comportamiento del algoritmo de entrenamiento causando un impacto significativo en el desempeño del modelo. Los hiperparámetros utilizados en las ConvNets son tasa de aprendizaje, tamaño del lote, número de iteraciones, cantidad de capas ocultas, unidades ocultas y funciones de activación.

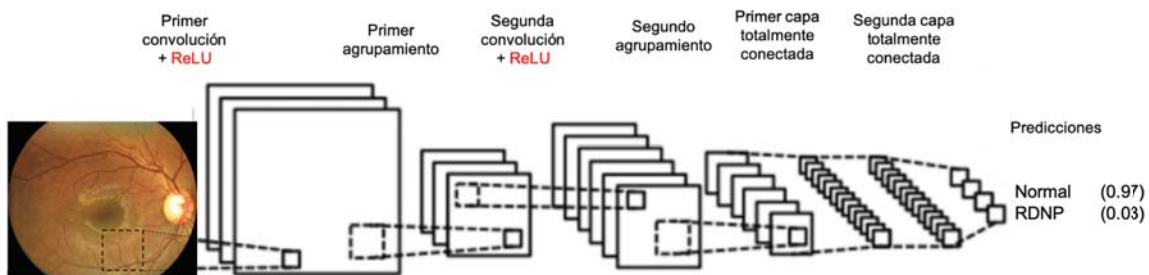


Figura 56. Proceso de propagación en una ConvNet.

## 3.4 Datos

### 3.4.1 Conjunto de datos

El proceso de entrenamiento, validación y pruebas es necesario para la construcción y evaluación de una ConvNet. Por ello, contar con una cantidad significativa de datos es indispensable al momento de implementar un modelo de aprendizaje automático.

### 3.4.2 División de datos

Los datos utilizados durante el desarrollo de modelos de aprendizaje automático comúnmente se dividen en tres grupos; entrenamiento, validación y pruebas. Los datos consisten en un vector de entrada y otro de salida. En el caso de imágenes, como ya se mencionó, el vector de entrada corresponde al valor de los píxeles que describen la imagen original, mientras que el vector de salida será la predicción asignada a esa imagen.

- Datos de entrenamiento

Son el conjunto de datos utilizados para ajustar los parámetros del modelo. En el caso del aprendizaje supervisado, el modelo se entrena con este conjunto de datos comparando la predicción con la salida deseada para cada elemento del conjunto de entrenamiento.

- Datos de validación

El modelo entrenado se usa para realizar predicciones en un segundo conjunto de datos denominado datos de validación. Los datos de validación proporcionan una evaluación imparcial del modelo en el conjunto de datos de entrenamiento mientras se ajustan los hiperparámetros del modelo. Los datos de validación se pueden utilizar para realizar una detención temprana en caso de sobre ajuste.

- Datos de prueba

El conjunto de datos de prueba se utiliza para proporcionar una evaluación imparcial del ajuste y la generalización final del modelo una vez terminado el entrenamiento. Estos datos también se denominan datos de reserva.

Al momento de dividir los datos, resulta común que el mayor porcentaje se asigne a los datos de entrenamiento. El porcentaje restante se divide entre los conjuntos de validación y pruebas. La Figura 57 muestra algunos porcentajes utilizados en la literatura al momento de dividir el conjunto de datos iniciales.

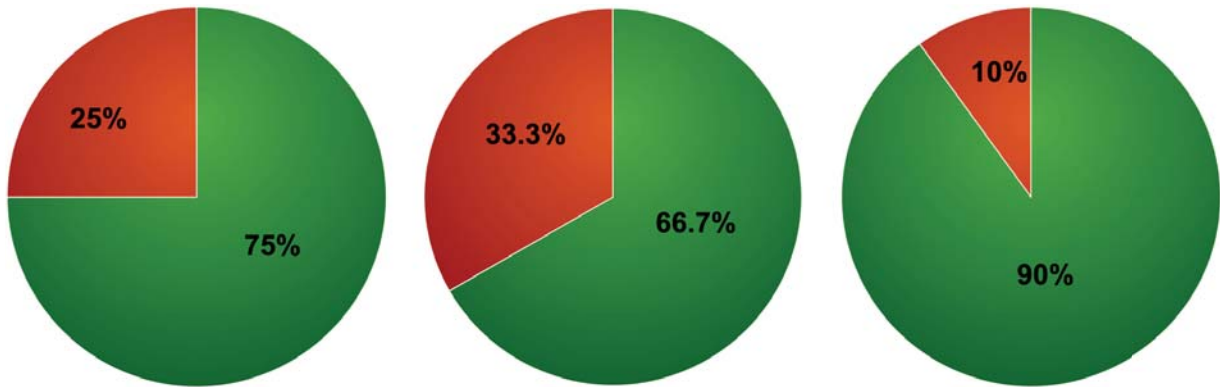


Figura 57. Porcentajes comúnmente utilizados para la división de datos. La región verde corresponde a los datos de entrenamiento, la región roja a los datos de validación y reserva.

Todas las imágenes del conjunto de entrenamiento se sometieron a transformaciones para evitar un sobreajuste y mejorar la generalización del modelo; como ya se mencionó en la sección de regularización, Figura 46. Dado que el modelo está siendo constantemente entrenado con nuevas versiones ligeramente modificadas de los datos de entrada, se construye un modelo más robusto.

### 3.4.3 Métricas de evaluación

Durante la fase de entrenamiento es necesario establecer métricas que permitan monitorear el desempeño del modelo. Al tratarse de un problema de clasificación, en el que se involucran predicciones, resulta útil el uso de la exactitud y la pérdida. La pérdida se encuentra definida en la Ecuación 6. La exactitud está definida por la Ecuación 24.

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

Ecuación 24. Exactitud.

TP corresponde a los verdaderos positivos, TN a los verdaderos negativos, FP a los falsos positivos y FN a los falsos negativos. Al ser un problema de clasificación médica también resulta apropiado evaluar el modelo con las métricas de sensibilidad, Ecuación 25, y especificidad, Ecuación 26.

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

Ecuación 25. Sensibilidad.

$$E = \frac{TN}{TN + FP}$$

Ecuación 26. Especificidad.

Además de las métricas ya mencionadas resulta conveniente considerar la precisión, Ecuación 27, que permite comparar el desempeño de modelos cuando el costo de detectar falsos positivos es alto. En contraste, el alcance, Ecuación 28, permite comparar modelos cuando hay un costo grande asociado con la detección de falsos negativos. Finalmente, la métrica F1 resulta de la evaluación del balance entre la precisión y el alcance.

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

Ecuación 27. Precisión.

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

Ecuación 28. Alcance.

$$F1 = \frac{2 \times p \times r}{p + r}$$

Ecuación 29. Valor F1.

## Metodología

La Figura 58 muestra el diagrama de la metodología utilizada para la implementación de una ConvNet entrenada para la detección de retinopatía diabética.

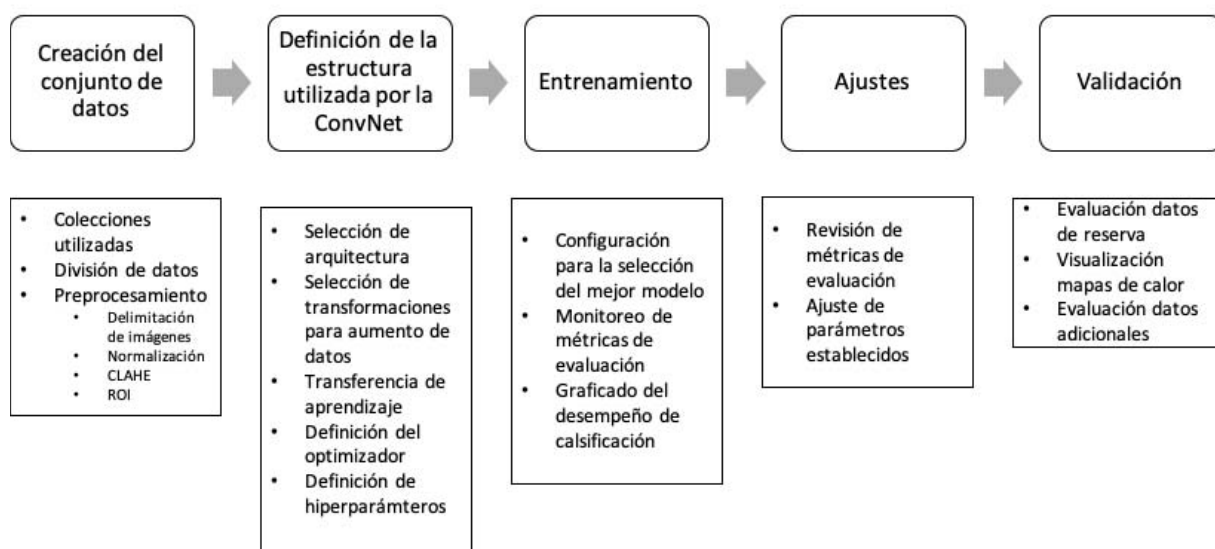


Figura 58. Metodología utilizada para la detección de retinopatía diabética.

### 4.1 Creación del conjunto de datos

Para entrenar una red neuronal es necesario contar con una cantidad de datos grande y representativos de la tarea que se quiere automatizar. Pete Warden propone, en el caso de tareas de clasificación, contar con al menos mil ejemplos por clase [51].

#### 4.1.1 Colecciones utilizadas

La Tabla 9 resume la información de las colecciones de imágenes, de retinopatía diabética, a las que se tuvo acceso. La colección Eye Pacs es la única que provee la relación de imágenes del ojo derecho e izquierdo por paciente Figura 59.

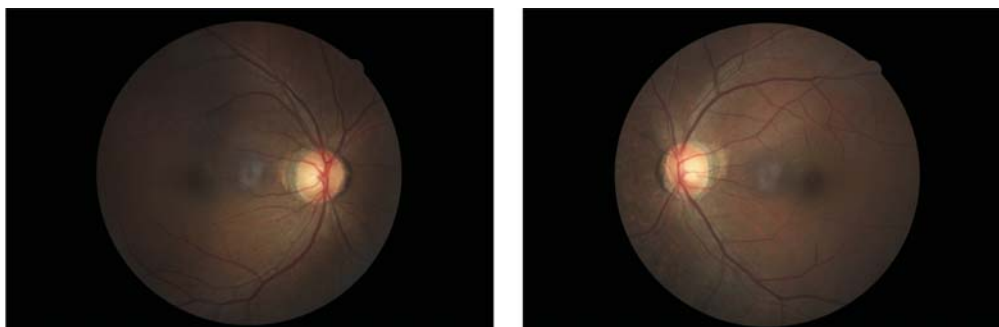


Figura 59. Imagen del ojo izquierdo y derecho de un mismo paciente.

La detección de retinopatía diabética se realiza en cada imagen de manera independiente por lo que no es indispensable contar con la relación de ambos ojos para realizar la evaluación de las imágenes. Las colecciones DIARETDB [52], E\_ophtha [53] y ROC [54] son las únicas que cuentan con una clasificación realizada a partir de anotaciones. Este tipo de etiquetado indica la región dentro de la imagen donde se localizan las lesiones de la enfermedad. La Figura 60 muestra la imagen original, así como las anotaciones realizadas para diferentes lesiones. El resto de las colecciones son clasificadas en las clases Leve, Moderada, Severa, Proliferativa o Normal. Esta clasificación resulta ser general en comparación con la realizada a partir de anotaciones.

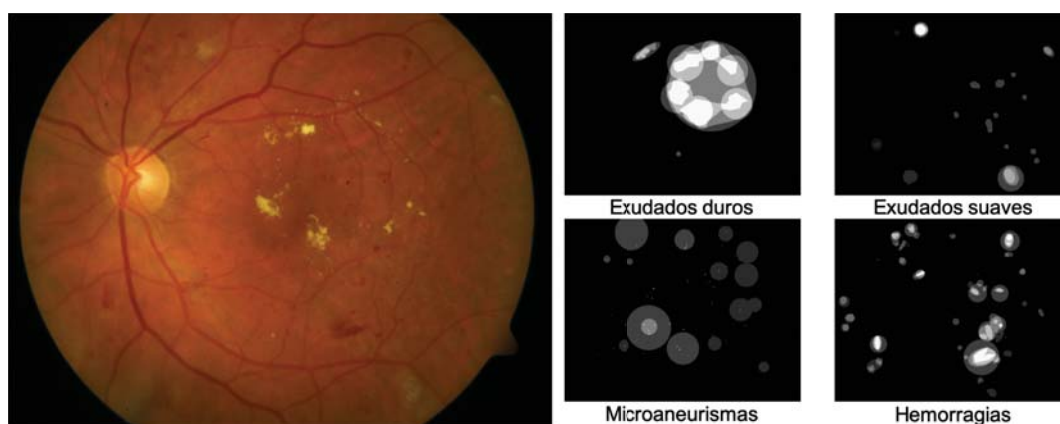




Figura 60. Anotaciones realizadas en una imagen de la colección DIARETDB.

Nombre de la colección	Clases	Anotaciones	Cantidad de imágenes	Dimensiones	Ángulo de captura	Proveedor
DIARETDB [público] [52]	2	Normal Exudados duros Hemorragias Puntos rojos Exudados suaves	5 84 *RDNP	1152x1500x3	50	Laboratorio de procesamiento de información de la Universidad tecnológica de Lappeenranta
E_ophtha [público] [53]	2	Normal Microaneurismas Exudados	100 90 *RDNP	960x1440x3	45	Asistencia pública de los hospitales de París
Eye Pacs [público] [55]	5	Normal Leve Moderada Severa Proliferativa	20K 20K 20K 20K 8.7K	2592x1944x3 4752x3168x3	+N/d	Eye Pacs
IMO [privado] [24]	5	Normal Leve Moderada Severa Proliferativa	669 62 148 70 79	1536x1152x3 2124x2056x3	+N/d	Instituto Mexicano de Oftalmología

Nombre de la colección	Clases	Anotaciones	Cantidad de imágenes	Dimensiones	Ángulo de captura	Proveedor
IDRID [público] [56]	5	Normal Leve Moderada Severa Proliferativa	168 25 157 86 58	2847x3423x3	+N/d	Indian Diabetic Retinopathy Image Dataset
Messidor [público] [57]	4	Normal Leve Moderada Severa	547 149 239 252	1488x2240x3	45	Consortio Messidor
ROC [público] [54]	2	Microaneurismas	52 *RDNP	576x768x3	+N/d	Universidad de Iowa

Tabla 9. Colecciones de imágenes de retinopatía diabética. \*RDNP Retinopatía diabética no proliferativa +N/d Información no disponible.

#### 4.1.2 División de los datos

Considerando el tipo de etiquetado existente, se realizaron dos conjuntos de datos, los cuales se muestran en la Tabla 10. El primero se conforma por las imágenes sin anotaciones considerando todas las imágenes divididas en las clases Leve, Moderada, Severa y Normal. El segundo conjunto son únicamente las imágenes con anotaciones específicas. En el caso de las imágenes sin anotaciones las clases consideradas son No Referible y Referible. La primera incluye a todas las imágenes clasificadas como Normal y Leve. El resto de las clases corresponden al grupo Referible. En el caso de las imágenes con anotaciones se realizó la división en las clases Normal y RDNP; siendo la clase RDNP todas aquellas secciones de imágenes con anotaciones de lesiones.

	<b>Imágenes sin anotaciones</b>	<b>Imágenes con anotaciones</b>
<b>Colecciones consideradas</b>	IMO [24] Messidor [57] IDRID [56]	DIARETDB [52] E_ophtha [53] ROC [54]
<b>Clases existentes</b>	Normal Leve Moderada Severa Proliferativa	Normal Exudados duros Hemorragias Puntos rojos Exudados suaves Microaneurismas
<b>Clases consideradas</b>	No Referible Referible	Normal Retinopatía diabética no proliferativa RDNP

	<b>Imágenes sin anotaciones</b>	<b>Imágenes con anotaciones</b>
<b>Cantidad de datos de entrenamiento</b>	NR 1134 R 658	N 2173 RDNP 2173
<b>Cantidad de datos de validación</b>	NR 162 R 94	N 271 RDNP 271
<b>Cantidad de datos de reserva</b>	NR 324 R 188	N 271 RDNP 271

Tabla 10. Resumen de los conjuntos de datos definidos.

#### 4.1.3 Preprocesamiento

Las imágenes utilizadas presentan una variedad de condiciones, Figura 61, debido a factores como la cámara utilizada, la variación en la iluminación, la existencia de artefactos y la presencia de cambios de orientación. La suma de todos estos factores contribuye a la generación de ruido siendo necesario el uso de técnicas de preprocesamiento que permitan contrarrestar esta situación. Lograr una normalización de las imágenes permitiría contar con un proceso robusto para el diagnóstico de retinopatía diabética. En la Figura 62 se indican los pasos del preprocesamiento realizado.

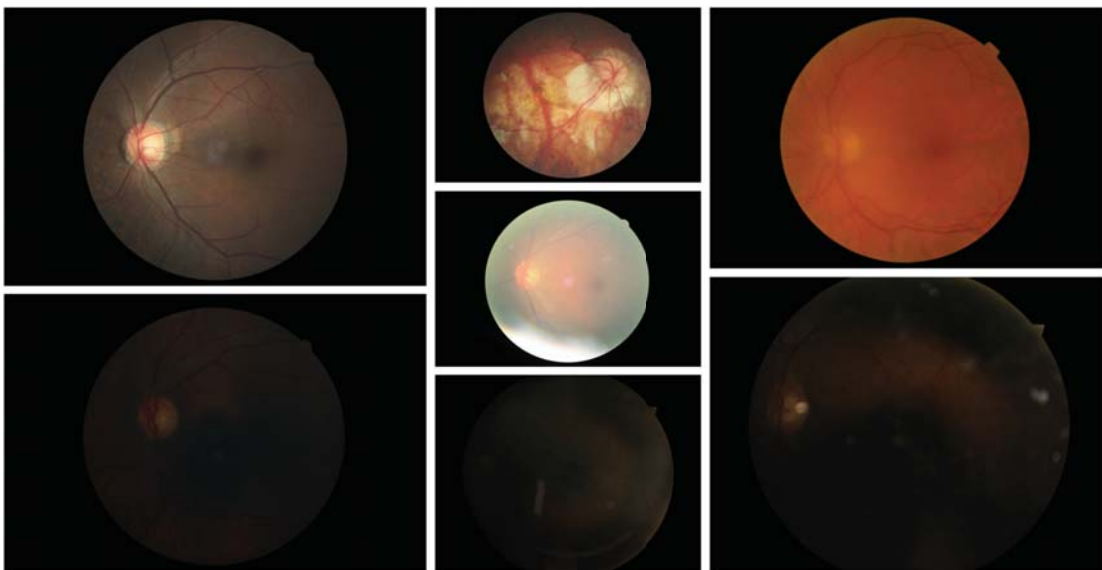


Figura 61. Ejemplo de variaciones presentes en las imágenes de retina.



Figura 62. Pasos del preprocesamiento.

El último paso del preprocesamiento solo se realiza en el conjunto de datos con anotaciones. Debido a que la ConvNet es entrenada con el preprocesamiento de la Figura 62, las imágenes que se evalúen, una vez entrenada la red, tendrán que ser sometidas al mismo preprocesamiento antes de ser evaluadas por el modelo.

#### 4.1.3.1 *Delimitación de imágenes*

El primer paso consiste en delimitar las imágenes para remover los bordes en color negro. Esas regiones no proporcionan información relevante para la detección de retinopatía diabética por lo que se elimina información innecesaria y por lo tanto se reduce el tiempo de procesamiento Figura 63.

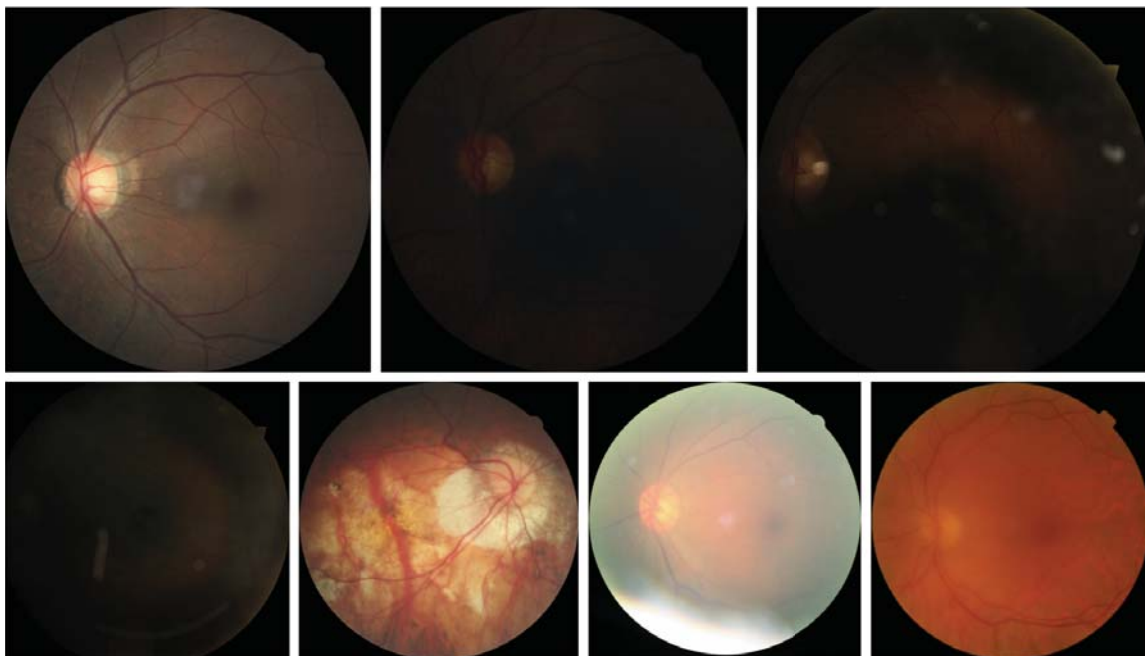


Figura 63. Imágenes de retina recortadas.

#### 4.1.3.2 Normalización

Dado que las imágenes de fondo ocular se toman iluminando la retina, una fuente de iluminación mal alineada crea un gradiente de iluminación a través de la imagen. Con un promedio local es posible eliminar ese gradiente homogenizando las condiciones de iluminación. En la Figura 64 se muestra el resultado de dos imágenes normalizadas.

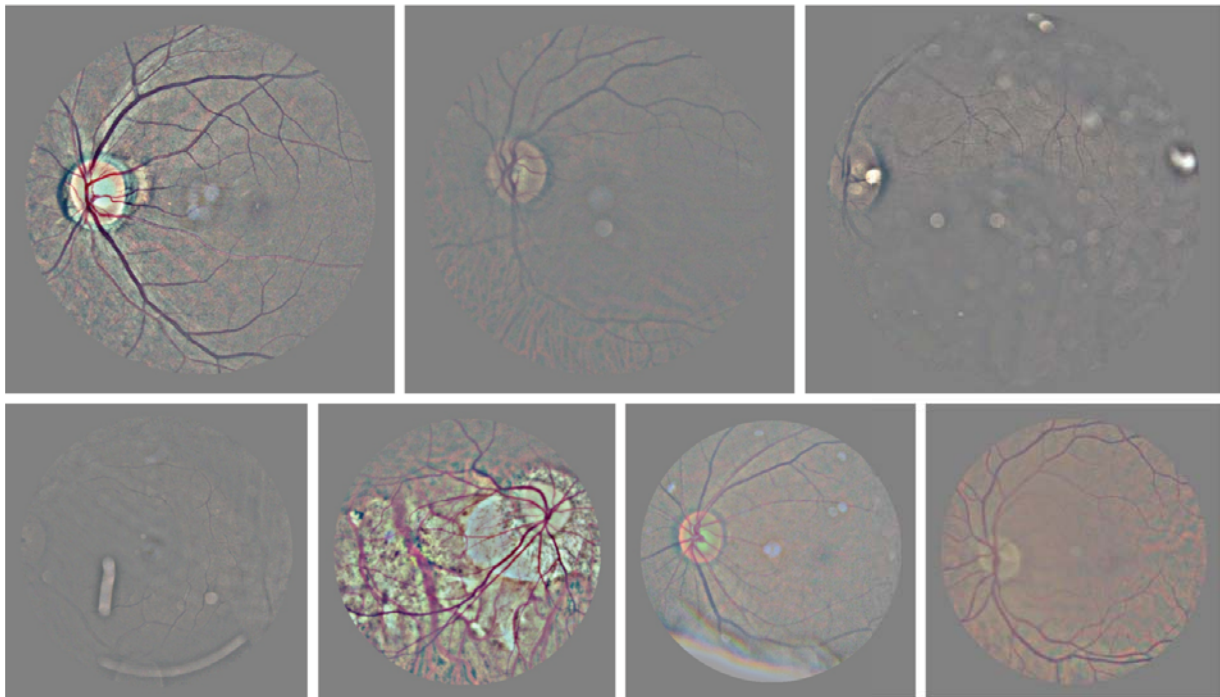


Figura 64. Imágenes de retina normalizadas.

#### 4.1.3.3 Contraste de ecualización de histograma adaptativo limitado [CLAHE]

La ecualización del histograma de la imagen permite maximizar su contraste. De esta manera, se incrementa la diferencia relativa de la intensidad entre los píxeles, realizando los elementos que la conforman.

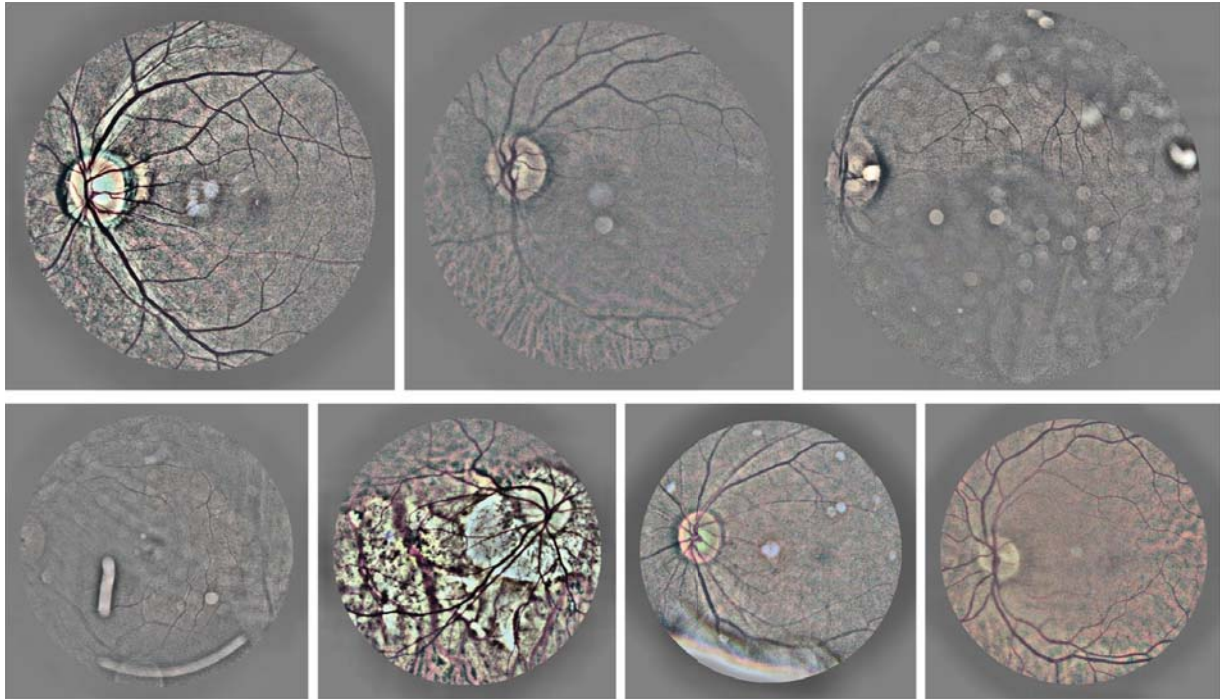


Figura 65. Imágenes de retina ecualizadas.

#### 4.1.3.4 *Regiones de interés*

Las regiones de interés sólo se obtienen del conjunto de datos imágenes con anotaciones. Para realizar este paso es necesario considerar los dos formatos en los que se presentan las anotaciones. El primero se muestra en la Figura 60 y consiste en el uso de una imagen en escala de grises que indica las regiones donde se localizan las lesiones anotadas. Bajo este formato inicialmente es necesario redimensionar la imagen original al menor valor próximo que permita lograr una división exacta de la imagen en regiones de 299 x 299 píxeles Figura 66.

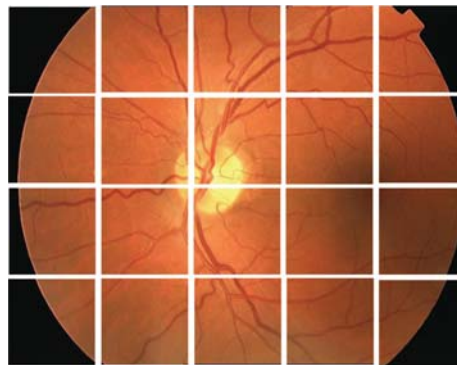


Figura 66. División de imagen original.



En caso de tener una imagen de 2592 x 1944 x 3 se redimensionaría a 2392 x 1794 x 3 con lo cual es posible dividir la imagen en regiones exactas de 299 x 299 píxeles. Después de ello se evalúa cada región para verificar si en ella se localizan píxeles diferentes de cero; esto indicaría la presencia de lesiones. En caso de ser cierto, la región de la imagen original se guarda. En caso de ser falso se realiza la evaluación de una nueva región Figura 67.

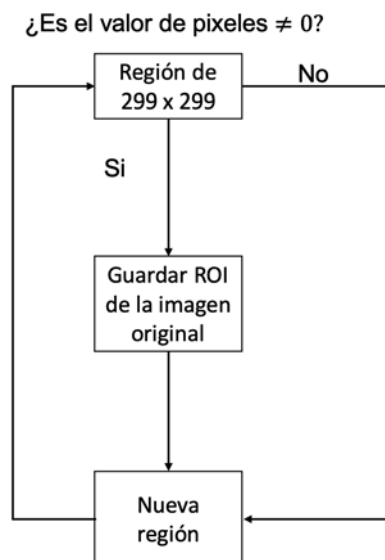


Figura 67. Árbol de decisión para la localización de anotaciones.

El segundo formato es a partir de un archivo de texto que contiene las coordenadas de las anotaciones realizadas. Estas coordenadas pueden indicar el centro de la anotación, en el caso de anotaciones circulares, o los valores de dos vértices, en el caso de anotaciones rectangulares. Bajo este formato se realiza una iteración a través de cada valor de coordenadas y se hace un complemento a la anotación que permita contar con regiones de 299 x 299. En la Figura 68 se muestra en color verde la sección o coordenada señalada por un experto. En color negro se muestra la zona complementaria que permite obtener una región de 299 x 299. En el caso de coordenadas el punto dado se toma como centro y se realiza un recorrido del mismo número de unidades que permita formar el cuadrado esperado. En el caso del rectángulo, se realiza una diferencia entre el valor de las coordenadas y las coordenadas necesarias para lograr una región de 299 x 299 píxeles.

Para las imágenes normales se realizó una división como la de la Figura 66. Debido a que no hay regiones que tengan mayor peso al momento de la decisión, se eligieron aleatoriamente regiones de interés hasta lograr que las clases RDNP y Normal se igualaran.



Figura 68. Anotaciones bajo formato de coordenadas.

## 4.2 Definición de la Red Neuronal Convolutiva

La siguiente sección define los elementos que conforman la ConvNet para la detección de retinopatía diabética.

### 4.2.1 Selección de la arquitectura

En la literatura se pueden encontrar diferentes arquitecturas de ConvNets utilizadas para la clasificación de imágenes. Estas arquitecturas pueden estar definidas para la realización de tareas específicas; nombradas arquitecturas definidas por el autor, o para su uso en competencias de clasificación de imágenes; conocidas como arquitecturas estándar. En trabajos relacionados a la detección de retinopatía diabética se han reportado resultados satisfactorios con ambos tipos de arquitecturas. Por ello, y con la finalidad de encontrar la arquitectura que genere las mejores predicciones se

consideraron un total de 4 arquitecturas para entrenar con el mismo conjunto de datos. Dos arquitecturas son diseñadas para la detección de retinopatía diabética y las dos restantes están enfocadas al procesamiento de imágenes de ImageNet. La Tabla 11 desglosa información respecto a las arquitecturas empleadas.

<b>Arquitectura</b>	<b>Dimensiones entrada</b>	<b>Exactitud</b>	<b>Parámetros entrenables</b>	<b>Parámetros no entrenables</b>
Darshit Doshi et al. [28]	512 x 512	0.3996 [Kappa]	886,849	0
Kele Xu et al. [58]	224 x 224	0.945 [Porcentaje]	2,811,169	0
ResNet50	224 x 224	0.749 [Porcentaje]	22,059,650	53,120
InceptionV3	299 x 299	0.779 [Porcentaje]	22,293,410	34,432

Tabla 11. Información de las arquitecturas empleadas.

#### 4.2.2 Selección de transformaciones para aumento de datos

Para mejorar la generalización del modelo y evitar un sobreajuste es necesario realizar un aumento de datos como se mencionó en la sección de regularización Figura 46. Las transformaciones de la Tabla 12 son comunes durante el entrenamiento de ConvNets por lo que son consideradas como punto de inicio. Sin embargo, en caso de obtener resultados no satisfactorios, durante la etapa de entrenamiento, se considerará la posibilidad de incrementar el número de transformaciones.

<b>Transformación</b>	<b>Valor</b>
Reescalado	1/255
Rango corte de imagen	0.2
Rango de enfoque	0.2
Rango de rotación	20
Rango de desplazamiento horizontal	0.2
Rango de desplazamiento vertical	0.2
Giro horizontal	Verdadero

Tabla 12. Aumento de datos.

Las imágenes resultado de las transformaciones realizadas son las que se presentan a la ConvNet durante el entrenamiento como lo menciona Adrian Rosebrock [58]. En la fase de predicciones la única transformación que debe aplicarse a la imagen es la del reescalado.

#### 4.2.3 Transferencia de aprendizaje

Las arquitecturas propuestas por autores no reportan ningún archivo del que se puedan obtener los pesos resultado de su entrenamiento. Por lo tanto, la única opción, para el proceso de entrenamiento, es iniciar los parámetros de la red con valores aleatorios. Por otro lado, en el caso de las arquitecturas estándar, es posible iniciar el entrenamiento con los pesos utilizados durante la evaluación del conjunto de imágenes ImageNet; estos valores se encuentran disponibles en un archivo de descarga pública. Se tiene que considerar que el conjunto de imágenes de ImageNet es significativamente diferente de un conjunto de imágenes médicas. En base a ello y de acuerdo con las recomendaciones realizadas en la Figura 52 será necesario entrenar las últimas capas; correspondientes a características de alto nivel. Las primeras capas conservarán los valores transferidos.

#### 4.2.4 Definición del optimizador

Los optimizadores utilizados durante la fase de entrenamiento, así como sus hiperparámetros se resumen en la Tabla 13. La selección de estos optimizadores radica en que son utilizados recurrentemente en modelos para la clasificación de imágenes. Cada optimizador se evalúa de manera independiente por lo que es necesario realizar un nuevo entrenamiento para identificar cual se desempeña mejor. En base a los primeros resultados se elegirá un optimizador y se ajustarán los valores de sus hiperparámetros.

Optimizador	Tasa de aprendizaje	Parámetros	
RMSProp	0.01	Rho=0.9	
SGD	0.01	Momento=0	
Adam	0.01	Beta_1=0.9	Beta_2=0.999

Tabla 13. Optimizadores empleados para el entrenamiento de la Red Neuronal Convolutiva.

#### 4.2.5 Hiperparámetros

Además de los hiperparámetros existentes en el optimizador, también es necesario definir el tamaño del lote, número de iteraciones y funciones de activación. La Tabla 14 resume los valores que servirán como punto de inicio y que han sido tomados de recomendaciones realizadas en la literatura.

Hiperparámetro	Valor
Tasa de aprendizaje	0.01
Tamaño de lote	8
Número de iteraciones	50
Funciones de activación	ReLU y Sigmoid en la capa final

Tabla 14. Hiperparámetros iniciales de entrenamiento.

### 4.3 Entrenamiento

Trabajar con una gran cantidad de datos, durante el entrenamiento de una red neuronal, requiere el uso de tarjetas de procesamiento gráfico que permitan reducir el tiempo de cómputo. Además de ello es necesario considerar la selección de un entorno de trabajo que permita la implementación de los elementos considerados durante la definición del modelo. La información referente a ambas consideraciones se localiza en la sección 4.6 de la metodología.

### 4.3.1 Configuración para la selección del mejor modelo

Durante la etapa de entrenamiento es común que las métricas de evaluación oscilen Figura 69. Esto implica que el modelo con el mejor desempeño podría presentarse antes de la última iteración realizada. En la Figura 69 se muestra en color azul la gráfica correspondiente a la pérdida evaluada en los datos de validación. La gráfica muestra resultados volátiles y se puede observar que antes de la iteración 10 se localiza el menor valor de pérdida. Por ello resulta útil configurar el entorno de trabajo de tal manera que guarde el mejor modelo obtenido a partir de la métrica considerada.

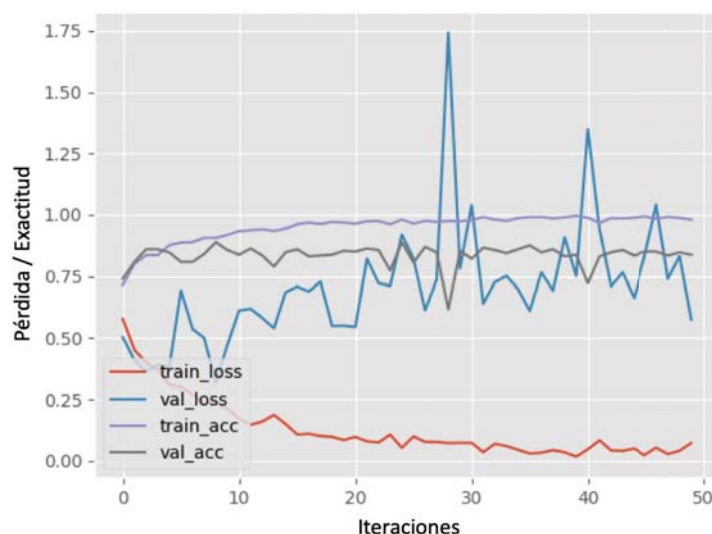


Figura 69. Gráfica de las métricas de exactitud y pérdida de un modelo clasificador.

### 4.3.2 Gráficas de desempeño del modelo de clasificación

Al finalizar el entrenamiento es conveniente graficar la evolución de las métricas de evaluación durante el entrenamiento. Estas gráficas proporcionan información respecto al desempeño del modelo, tiempo de convergencia y la presencia de sobreajuste.

### 4.3.3 Definición de los modelos empleados

La primera parte del entrenamiento se realizó con la arquitectura definida por Chollet [60]. Se trata de una arquitectura simple utilizada para la clasificación de animales. El objetivo de emplear esta arquitectura es identificar al optimizador que

mejores resultados genere, a partir de un costo computacional bajo, para utilizarlo con las arquitecturas definidas en la sección previa.

Modelo	Imágenes utilizadas	Optimizador
Chollet [59]	Original	RMSProp
Chollet [59]	Original	SGD
Chollet [59]	Original	Adam

Tabla 15. Modelos utilizados para comparar resultados de optimizadores.

Una vez identificado el optimizador con el mejor desempeño, se realizará el entrenamiento de los modelos definidos en la Tabla 16. Cada modelo es entrenado con las variaciones, al conjunto de datos, registradas en la Tabla 17. Considerando eso se tendrá un total de 48 modelos para la detección de retinopatía diabética.

<b>Modelo</b>	<b>Imágenes utilizadas</b>	<b>Transferencia de aprendizaje</b>
Darshit [28]	ROI	No
Darshit [28]	Original	No
Kele [58]	ROI	No
Kele [58]	Original	No
Inception V3	ROI	No
Inception V3	Original	No
Inception V3	ROI	Si
Inception V3	Original	Si
ResNet 50	ROI	No
ResNet 50	Original	No
ResNet 50	ROI	Si
ResNet 50	Original	Si

Tabla 16. Modelos entrenados para la detección de retinopatía diabética.

La sección correspondiente a Original de la Tabla 17 consiste en no realizar ninguna transformación a la imagen original.

Original	CLAHE
Normalizada	Normalizada y CLAHE

Tabla 17. Tipos de preprocesamiento realizado a las imágenes.

## 4.4 Ajustes

La sección de ajustes corresponde a la identificación de cambios que se pueden realizar, en el modelo construido, para mejorar su desempeño en el conjunto de datos utilizado.

### 4.4.1 Revisión de métricas de evaluación y ajuste de hiperparámetros

Las gráficas de desempeño proporcionan información respecto a la exactitud del entrenamiento, los valores de pérdida, la presencia de sobreajustes y los puntos de convergencia. Si después de la fase de entrenamiento se cuenta con una exactitud y pérdida no satisfactorios se pueden realizar algunos cambios.

- Revisar datos de entrada. Factores como ruido o la falta de representatividad de una clase generan resultados no satisfactorios.
- Evaluar el desempeño de otros optimizadores. El optimizador resulta ser un elemento importante para que el conjunto de parámetros del modelo dé lugar a un desempeño satisfactorio.
- En caso de utilizar transferencia de aprendizaje, modificar la configuración establecida. Bajo el escenario de trabajar con imágenes médicas, se consideraría incrementar el número de capas a entrenar.
- Incrementar el número de iteraciones. Esto solo es útil si las gráficas de desempeño del modelo de clasificación no han logrado converger en las iteraciones establecidas.
- Suspender el entrenamiento en caso de identificar sobreajuste.



Slav Ivanov [61] proporciona una lista aún más extensa de ajustes que se pueden realizar en caso de obtener resultados poco satisfactorios después del entrenamiento de una red neuronal.

## 4.5 Validación

El último paso, correspondiente a la validación, consiste en utilizar los datos que no fueron empleados durante la fase de entrenamiento. Esto permite evaluar la generalización del modelo al conocer su comportamiento frente a nuevos conjuntos de datos.

### 4.5.1 Evaluación de datos de reserva

Durante la división de datos se asignó un porcentaje a los datos de reserva. Esos datos están etiquetados de la misma forma que los datos de entrenamiento y validación; por lo tanto, se conoce la clase a la que pertenecen. Para evaluar este conjunto de datos es necesario realizar el reescalado descrito en la Tabla 12 así como el preprocesamiento utilizado durante la fase de entrenamiento.

### 4.5.2 Evaluación de datos adicionales

Además de los datos de reserva resulta conveniente la evaluación del modelo en un conjunto de datos ajeno a los empleados durante el entrenamiento y la validación. Esto debido a que la división de datos inicial se realiza a partir de un conjunto que proviene de la misma fuente; lo cual puede generar un sesgo en la toma de decisiones. En este trabajo sólo es posible realizar una evaluación de datos adicionales con los modelos entrenados a partir de regiones de interés. El conjunto de datos del IMO [24], IDRID [56] y Messidor [57] pueden ser utilizados para realizar una validación adicional de los modelos.

### 4.5.3 Visualización del modelo

Además de la evaluación con datos de reserva y datos adicionales, resulta necesaria la interpretación de los resultados obtenidos. En problemas de carácter médico es necesario comprender lo que sucede detrás de los modelos que diagnostican enfermedades en pacientes. Para el caso de imágenes, la visualización de modelos permite mostrar, de manera gráfica, aquellos elementos que tienen mayor peso durante la toma de decisiones.

## 4.6 Hardware y Software empleado

### 4.6.1 Configuración del equipo

Los algoritmos de aprendizaje profundo son computacionalmente costosos, por lo que resulta favorable contar con un equipo que permita el entrenamiento en un periodo de tiempo razonable. La Tabla 18 contiene información sobre el equipo utilizado. Adicionalmente se emplearon dos tarjetas gráficas con memoria de 4GB cada una. Para el uso de las tarjetas, durante el entrenamiento de la ConvNet, se instaló la plataforma de computación en paralelo CUDA 9.0. También se instaló cuDNN que es una biblioteca con rutinas estándar para redes neuronales profundas como propagación hacia adelante y hacia atrás, agrupación, normalización y capas de activación.

<b>Sistema operativo</b>	GNU/Linux
<b>Distribución</b>	Ubuntu 16.04.6 LT
<b>Procesador</b>	Intel(R) Xeon(R) CPU E5-2620 v2 2.
<b>Arquitectura</b>	64 bits
<b>CPUs</b>	24
<b>Tarjetas gráficas</b>	NVIDIA Quadro K5000

Tabla 18. Características del equipo.

#### 4.6.2 Entorno de trabajo

El entorno de trabajo utilizado es Keras desarrollado por el ingeniero François Chollet [62]. Keras es una biblioteca de redes neuronales de código abierto escrita en Python. Se puede ejecutar sobre TensorFlow, Microsoft Cognitive Toolkit o Theano. Se eligió debido a que es amigable, modular y extensible, permitiendo la implementación de arquitecturas en poco tiempo. La versión 2.2.4 de Keras es utilizada a través de Python 2.7.12 con el soporte de la biblioteca 1.8.0 de TensorFlow.

## Resultados y discusión

---

En este trabajo se realizó la implementación de diferentes modelos para diagnosticar retinopatía diabética no proliferativa. En primera instancia se evaluó el desempeño de los optimizadores Adam, RMSProp y SGD en un mismo conjunto de datos. El optimizador Adam mostró el mejor desempeño para la clasificación de imágenes debido a su adaptabilidad en el valor de la tasa de aprendizaje, por lo que fue utilizado para el entrenamiento de todos los modelos reportados en esta investigación.

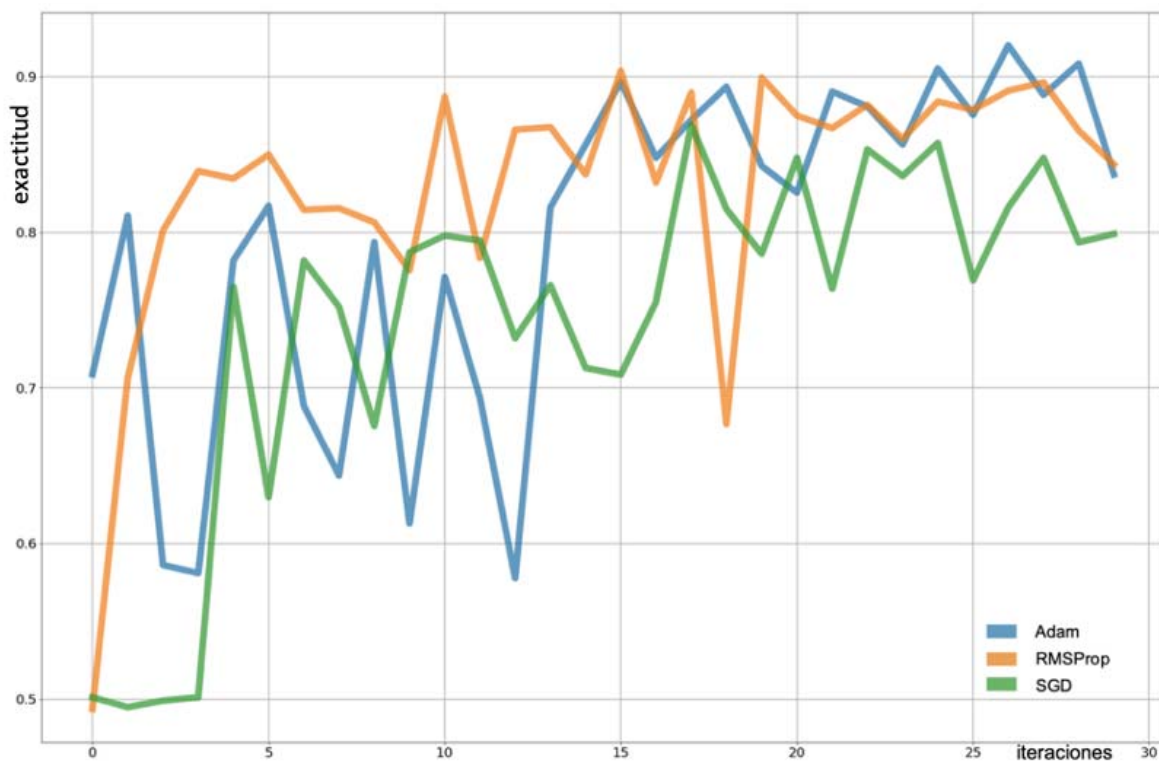


Figura 70. Evaluación de optimizadores en un conjunto limitado de imágenes de retina.

A partir de los diferentes preprocesamientos realizados se puede observar en la Figura 71 que el tipo de preprocesamiento no afecta significativamente la exactitud del modelo entrenado. En contraste el tipo de arquitectura utilizada es un factor que influye considerablemente la exactitud del modelo entrenado. La implementación de transferencia de aprendizaje es un factor que incrementó significativamente la exactitud de los modelos independientemente de su tipo de arquitectura. También se validó que el método de regiones de interés permite localizar lesiones microscópicas, como lo son las microaneurismas, ya que al trabajar con ese tipo de datos se obtuvieron los mayores valores de exactitud.

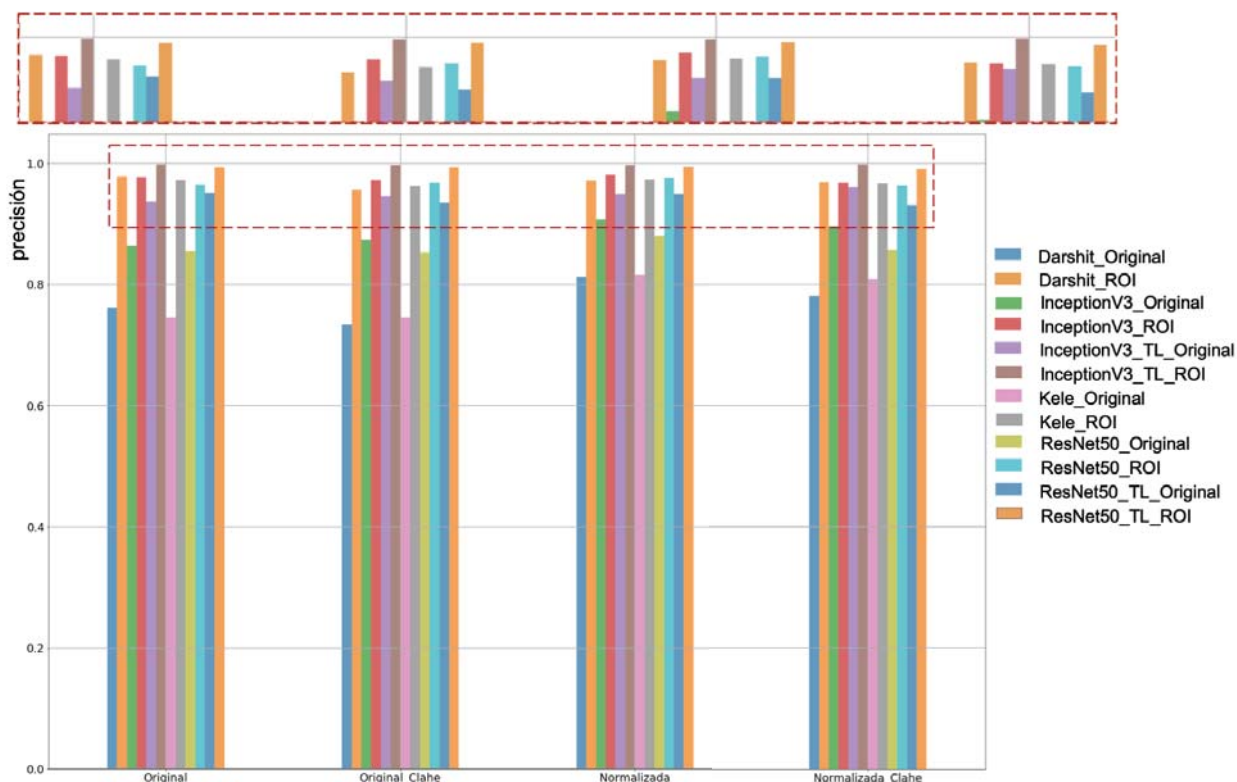


Figura 71. Exactitud de los modelos utilizando diferentes procesamientos y arquitecturas. La terminación \_ROI corresponde a los resultados de modelos entrenados con regiones de interés. La terminación \_Original corresponde a los modelos entrenados con las imágenes originales. La abreviación TL se utiliza en los modelos que emplearon transferencia de aprendizaje.

En la Figura 72 se reportan los resultados de la pérdida de cada modelo, los cuales son consistentes con los mostrados en la Figura 71. Los modelos que presentan menor pérdida son aquellos con mayores valores de exactitud. A partir de la gráfica de la Figura 72 es posible identificar aquellos modelos con mayor confiabilidad para la detección de retinopatía diabética. Los modelos con arquitectura Inception y ResNet entrenados con Regiones de Interés y transferencia de aprendizaje muestran los menores valores de pérdida y por lo tanto resultan ser los más confiables.

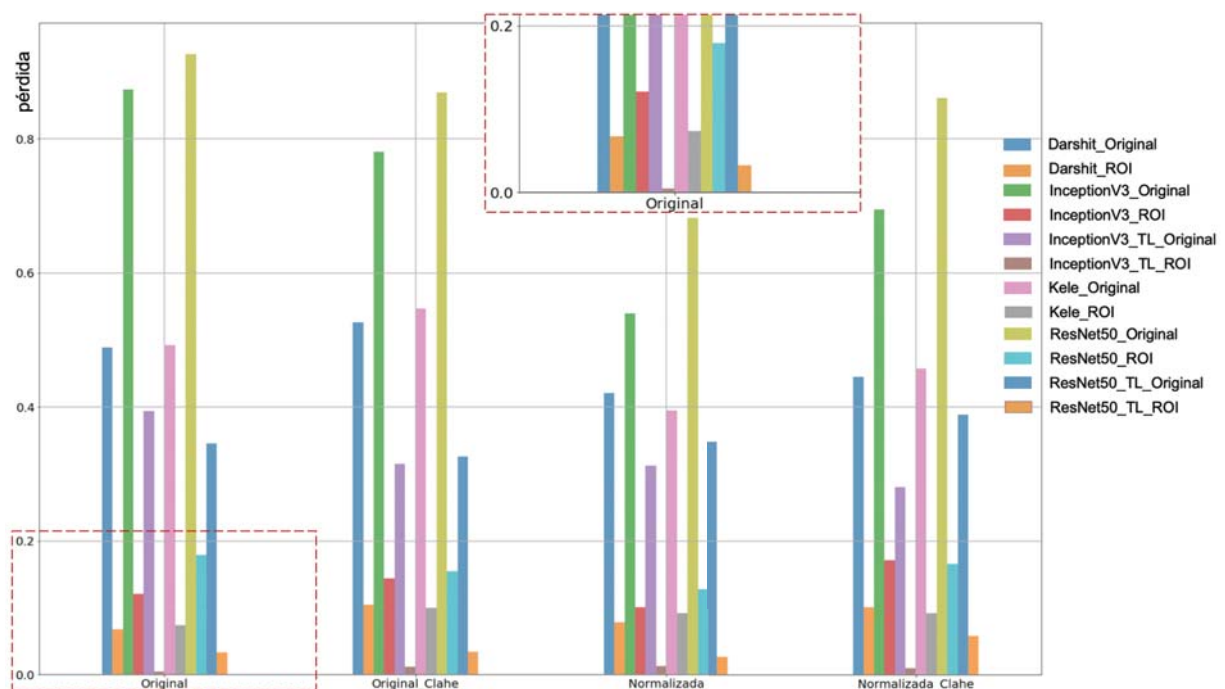


Figura 72. Resultados de las pérdidas de los modelos empleados. La terminación \_ROI corresponde a los resultados de modelos entrenados con regiones de interés. La terminación \_Original corresponde a los modelos entrenados con las imágenes originales. La abreviación TL se utiliza en los modelos que emplearon transferencia de aprendizaje.

En la Figura 73 se muestran los resultados de evaluar los modelos con mayor exactitud y menor pérdida en la métrica F1. Estos resultados permiten corroborar que los modelos con arquitecturas Inception y ResNet, entrenados con transferencia de aprendizaje y bajo el formato de regiones de interés tienen un mayor equilibrio entre los

valores de precisión y alcance. Por lo tanto, se esperaría que estos modelos muestren el mejor desempeño al momento de tener un uso clínico.

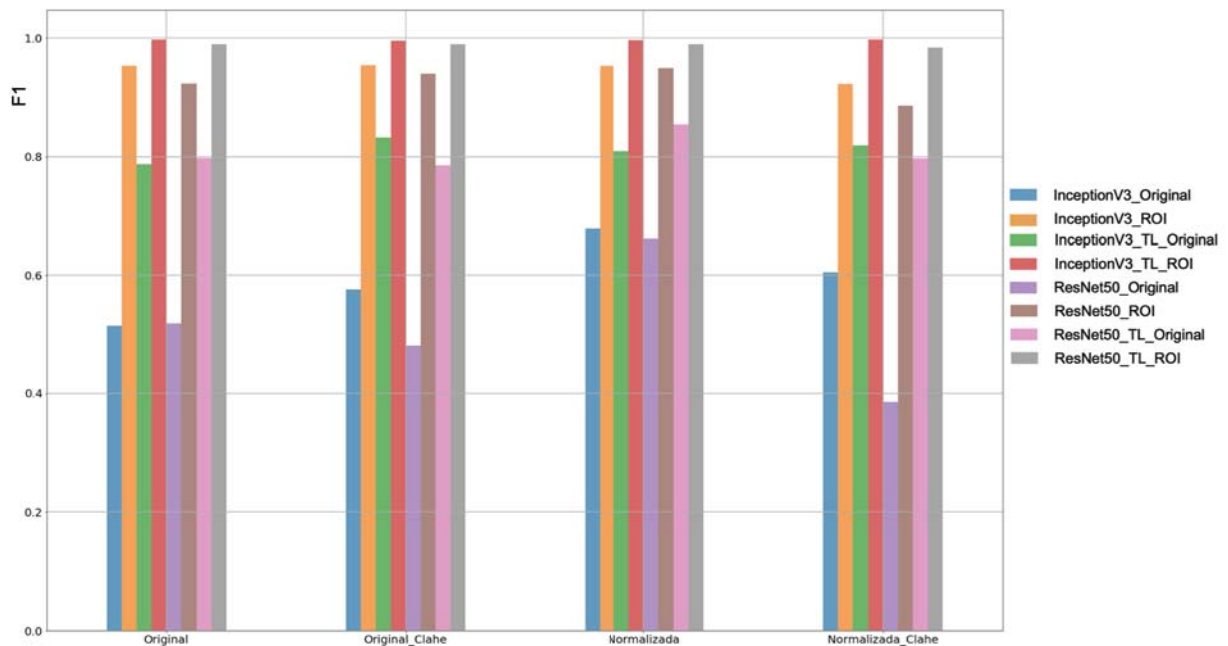


Figura 73. Evaluación de los modelos con mayor exactitud y menor pérdida en la métrica F1. La terminación `_ROI` corresponde a los resultados de modelos entrenados con regiones de interés. La terminación `_Original` corresponde a los modelos entrenados con las imágenes originales. La abreviación TL se utiliza en los modelos que emplearon transferencia de aprendizaje.

Se realizó una segunda evaluación utilizando datos de un conjunto de imágenes de reserva para corroborar la generalización de los modelos. Las primeras evaluaciones corresponden únicamente a los modelos que presentaron el mejor desempeño al ser entrenados con imágenes originales. La Figura 74 muestra la exactitud de dos arquitecturas evaluadas en los diferentes preprocesamientos establecidos. Los resultados obtenidos indican un comportamiento satisfactorio de todos los modelos independientemente del preprocesamiento utilizado.

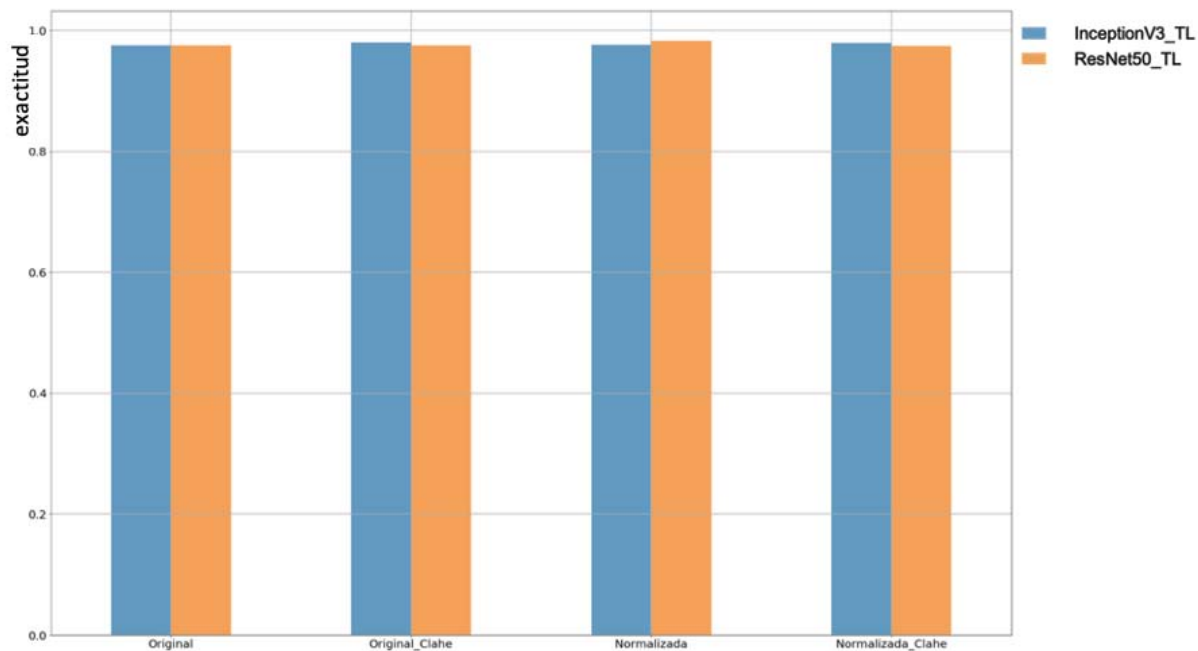


Figura 74. Exactitud de cada preprocesamiento evaluado con los datos de reserva. La abreviatura TL indica que los modelos fueron entrenados con transferencia de aprendizaje.

La Figura 75 y Figura 76 muestran los resultados de sensibilidad y especificidad, respectivamente, de los modelos evaluados en la Figura 74 con el conjunto de imágenes de reserva. Los resultados obtenidos muestran una clasificación satisfactoria entre casos sanos y casos con retinopatía diabética, ya que las variaciones derivadas del tipo de preprocesamiento y de la arquitectura utilizada son mínimas. Se tiene la hipótesis de que una validación con un mayor número de imágenes podría generar resultados significativamente diferentes entre los modelos considerados.



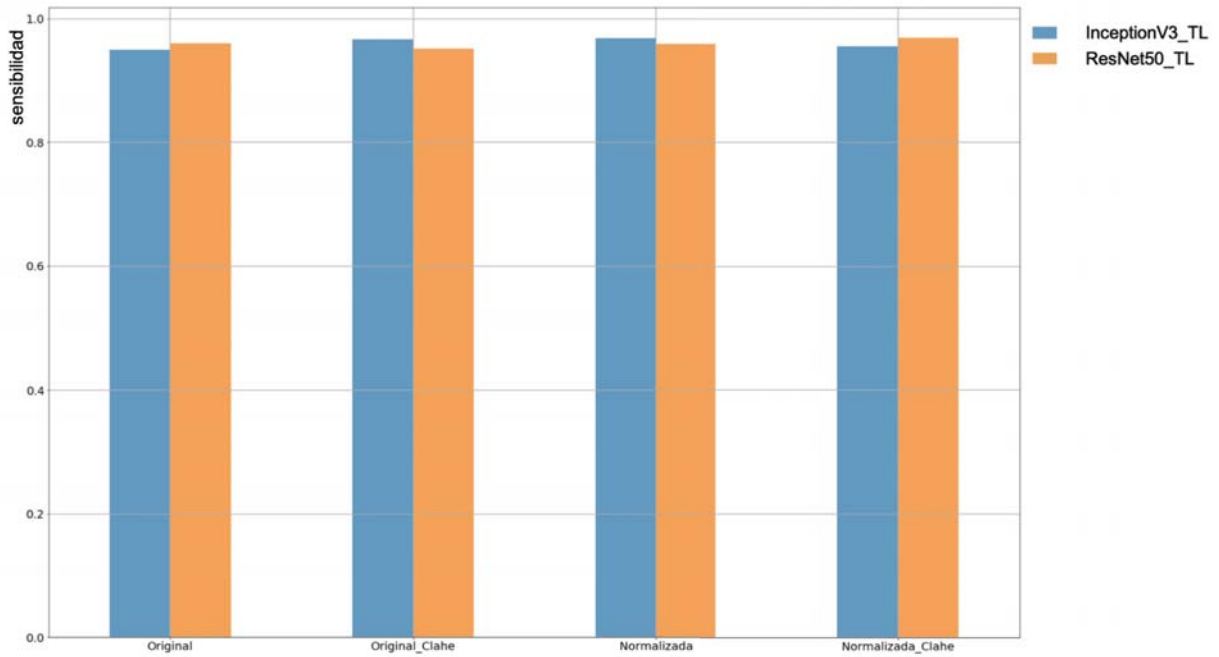


Figura 75. Gráfica con los valores de sensibilidad obtenidos para cada preprocesamiento con los datos de reserva. La abreviatura TL indica que los modelos fueron entrenados con transferencia de aprendizaje.

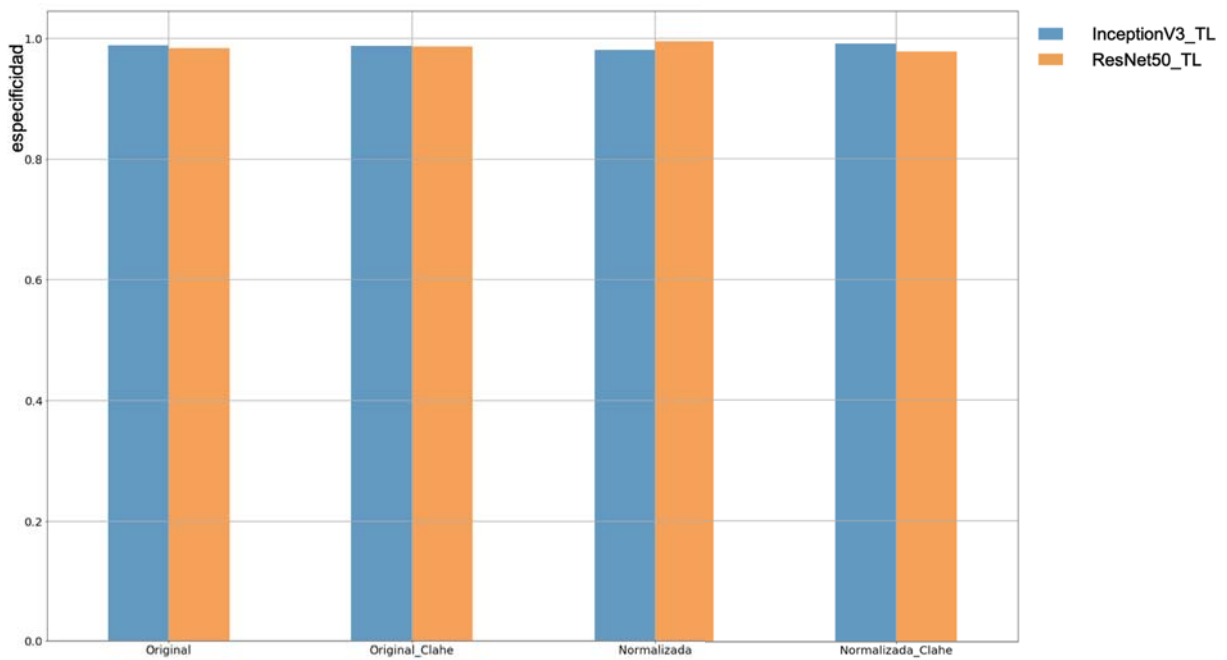


Figura 76. Gráfica con los valores de especificidad obtenidos para cada preprocesamiento con los datos de reserva. La abreviatura TL indica que los modelos fueron entrenados con transferencia de aprendizaje.

La última evaluación realizada a los modelos entrenados con imágenes originales corresponde a la visualización de mapas de activación, que son útiles para tener una interpretación gráfica de las regiones dentro de la imagen que tienen mayor peso durante la realización de predicciones. Las regiones en color rojo maximizan la activación del modelo. La Figura 77 muestra, en la sección 1, una imagen de retina con lesiones propias de retinopatía diabética. Los mapas de activación obtenidos en la sección 2 y 3 corresponden a modelos con arquitecturas Inception y ResNet, respectivamente, entrenados con transferencia de aprendizaje. Aunque ambas arquitecturas tienen zonas de activación en común hay zonas que la arquitectura ResNet, en la sección 3, no considera. Las dos arquitecturas omiten la zona inferior de la imagen en la cual hay una posible lesión correspondiente a una microaneurisma. El problema de la detección de lesiones microscópicas ya se había mencionado en la sección de la metodología correspondiente a regiones de interés. Se tiene la hipótesis de que con ese método se prevengan situaciones como la actual.

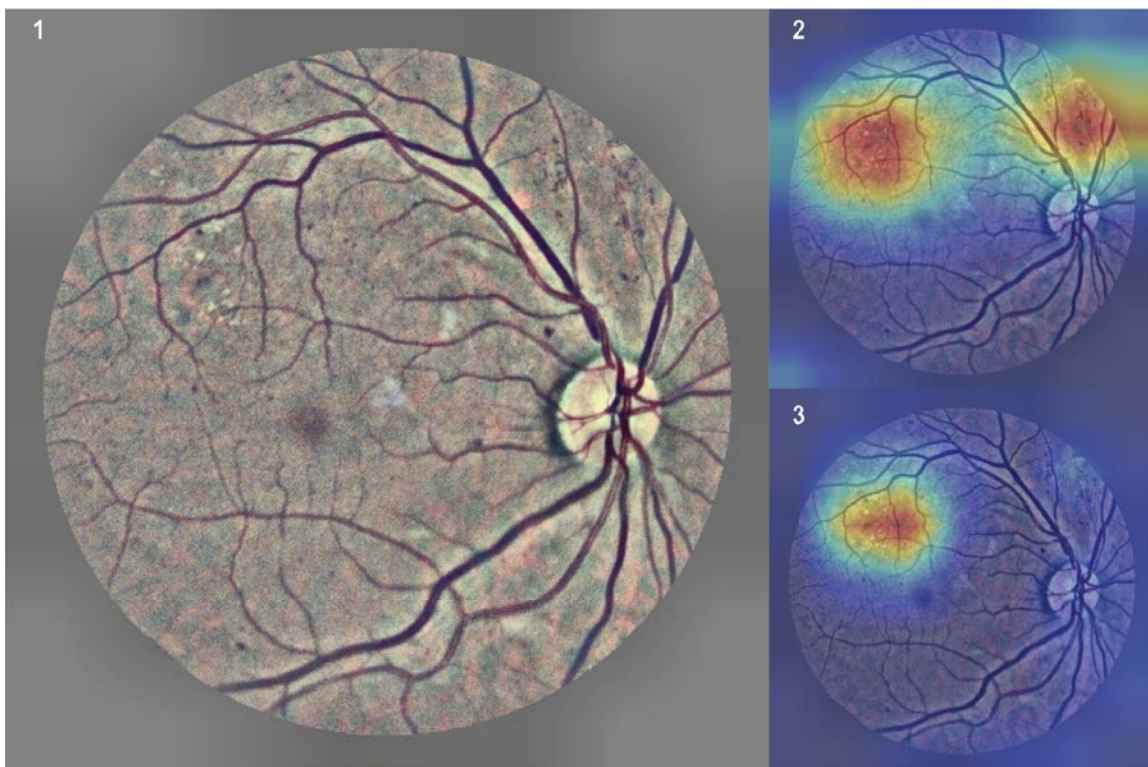


Figura 77. Visualización de imagen de retina. El área en color rojo es la que mayor peso tiene al momento de realizar la decisión. [1] Imagen original de retina [2] Visualización del modelo de decisión InceptionV3 [3] Visualización del modelo de decisión ResNet50.

La Figura 78 muestra el mapa de activación de otra imagen de retina diagnosticada con retinopatía diabética. De la misma manera, que en la Figura 77, las arquitecturas Inception y ResNet tienen regiones en común sin embargo los mapas de activación no son idénticos entre sí. De igual manera el peso de la activación por regiones de interés es considerablemente diferente. Aún con ello ambos modelos calificaron la imagen con retinopatía diabética.

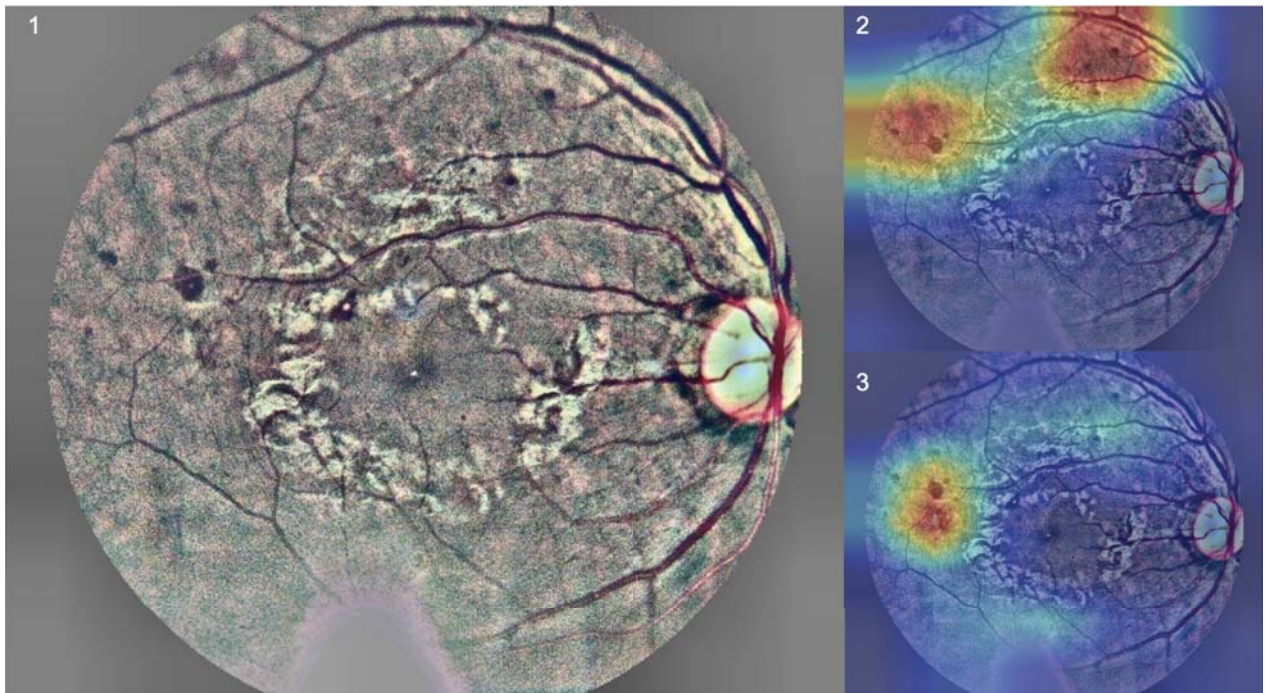


Figura 78. Visualización de imagen de retina. El área en color rojo es la que mayor peso tiene al momento de realizar la decisión. [1] Imagen original de retina [2] Visualización del modelo de decisión InceptionV3 [3] Visualización del modelo de decisión ResNet50.

En la Figura 79 se puede ver el mismo fenómeno ocurrido en la Figura 77 y Figura 78. Debido en todas las imágenes la arquitectura Inception ha tenido un mayor número de zonas de alta activación en regiones con lesiones propias de retinopatía diabética se puede considerar que es un modelo más robusto para su detección. Sin embargo, la arquitectura ResNet ha tenido un mejor desempeño de segmentación de las regiones



con lesiones de retinopatía diabética. Aunque este no es un trabajo enfocado a la segmentación resulta importante considerar esa característica para futuras aplicaciones.

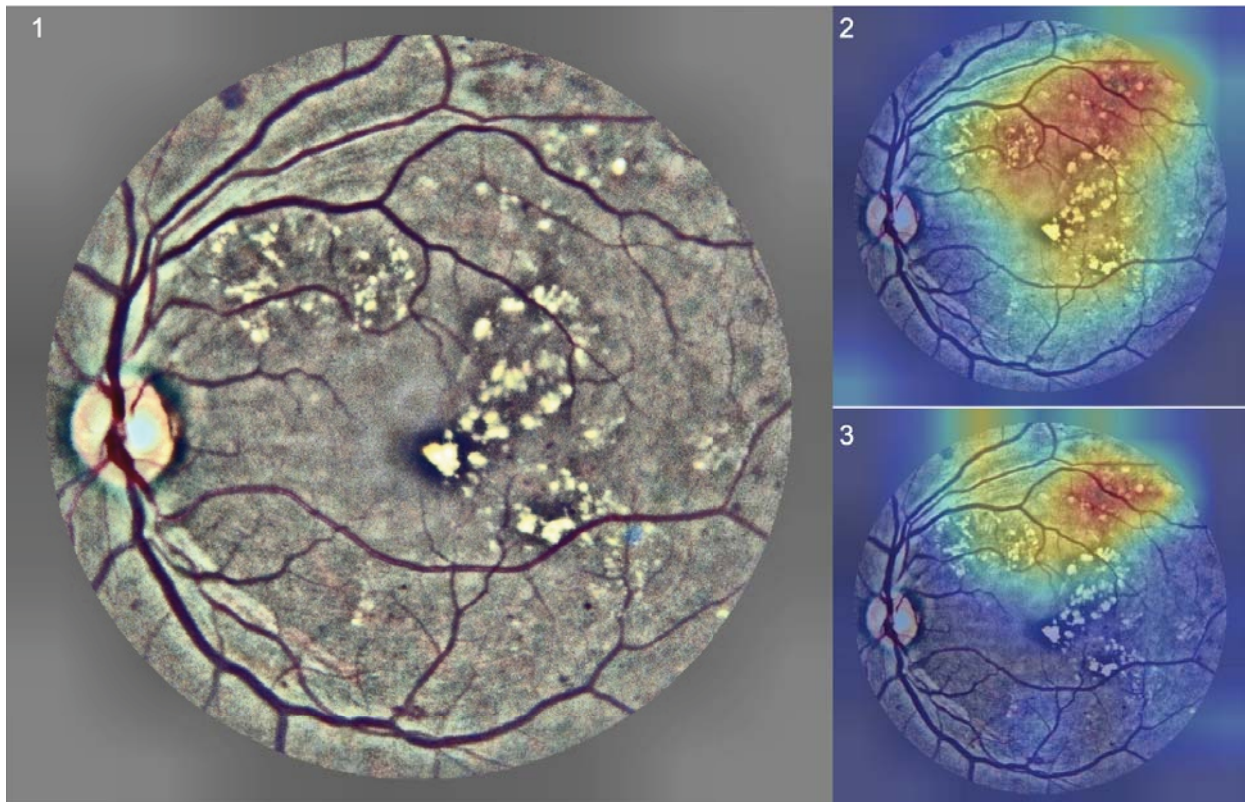


Figura 79. Visualización de imagen de retina. El área en color rojo es la que mayor peso tiene al momento de realizar la decisión. [1] Imagen original de retina [2] Visualización del modelo de decisión InceptionV3 [3] Visualización del modelo de decisión ResNet50.

La Figura 80 muestra zonas de activación más semejantes entre las arquitecturas consideradas. Esto permite visualizar de una manera clara como los parámetros establecidos en el modelo afectan en las regiones que se consideran.

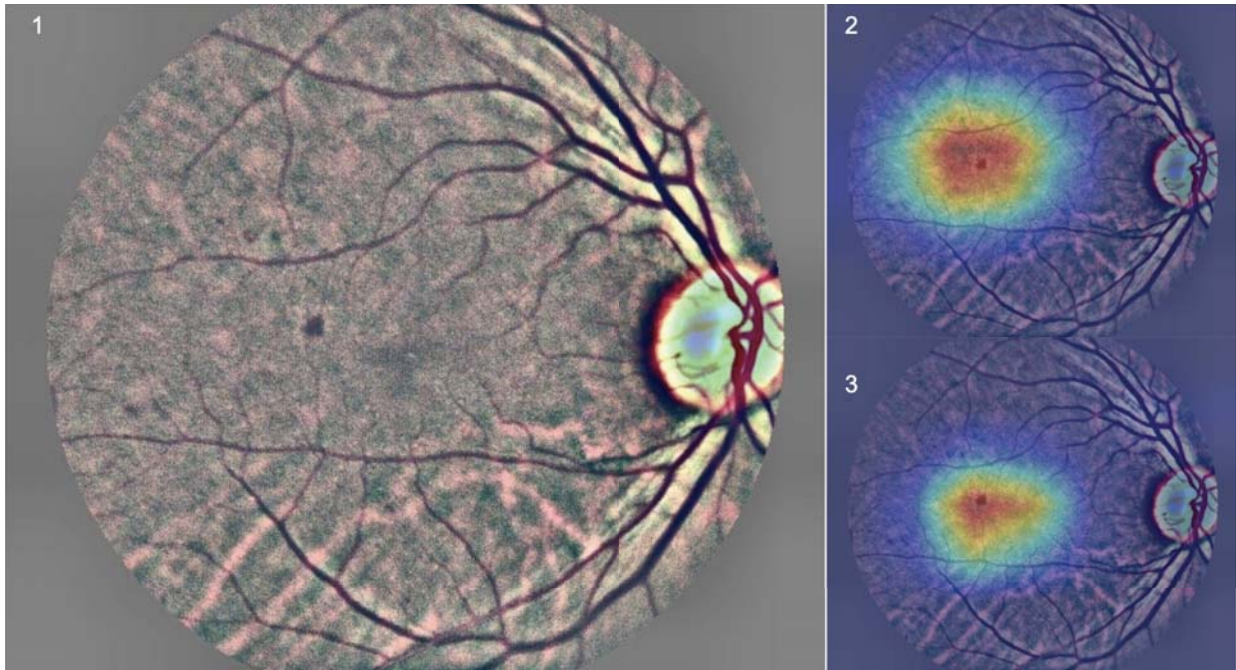


Figura 80. Visualización de imagen de retina. El área en color rojo es la que mayor peso tiene al momento de realizar la decisión. [1] Imagen original de retina [2] Visualización del modelo de decisión InceptionV3 [3] Visualización del modelo de decisión ResNet50.

Finalmente, la Figura 81 fue sometida a los preprocesamientos establecidos para después realizar una comparación de los mapas de activación generados. En la Figura 82 se puede identificar que dependiendo del preprocesamiento se asignan diferentes zonas de activación. Todas las zonas indicadas presentan lesiones de retinopatía diabética por lo que a partir de ello se podría realizar un ensamble de modelos y así contar con un modelo más robusto de clasificación.



Figura 81. Imagen de retina original.

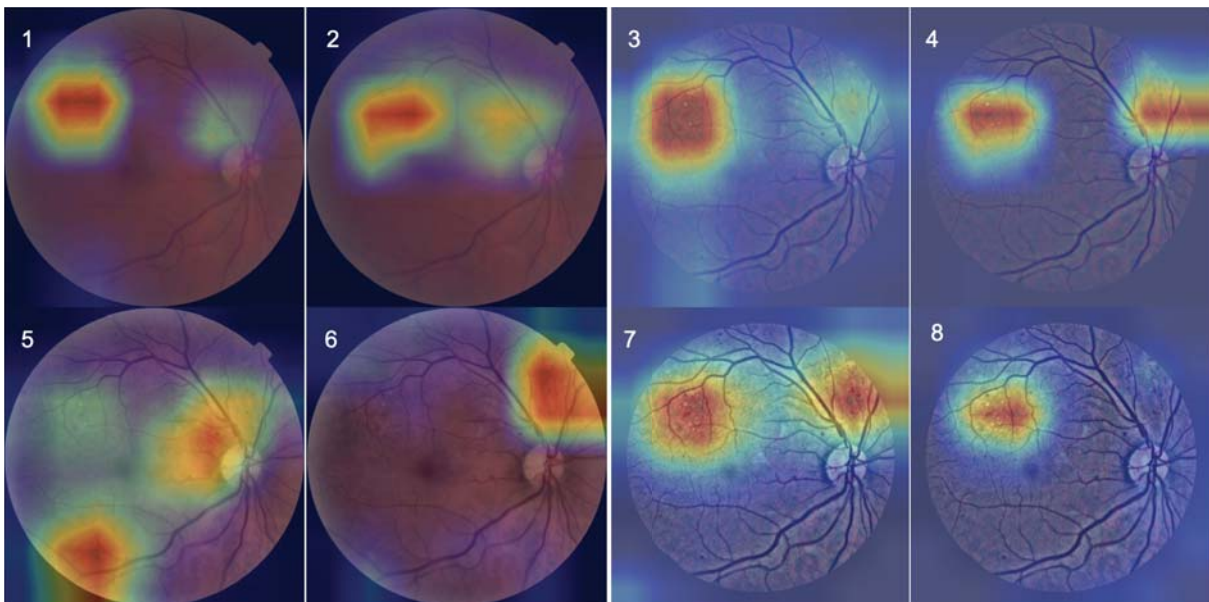


Figura 82. Comparación de la respuesta de los modelos de decisión entrenados al utilizar diferentes preprocesamientos. Se utiliza la imagen de la Figura 81 para su comparación. [1] Imagen original evaluada con la arquitectura InceptionV3 [2] Imagen original evaluada con la arquitectura ResNet50 [3] Imagen normalizada evaluada con la arquitectura InceptionV3 [4] Imagen normalizada evaluada con la arquitectura ResNet50 [5] Imagen con ecualización de histograma evaluada con la arquitectura InceptionV3 [6] Imagen con ecualización de histograma evaluada con la arquitectura ResNet50 [7] Imagen normalizada y con ecualización del histograma evaluada con la arquitectura InceptionV3 [8] Imagen normalizada y con ecualización del histograma evaluada con la arquitectura ResNet50.

Los resultados reportados permiten clasificar casos de pacientes con retinopatía moderada y severa, sin embargo, se cuenta con la limitación de no distinguir entre un paciente sano y uno con retinopatía diabética no proliferativa leve. El problema que se presenta es el tamaño que tienen las lesiones existentes en casos de retinopatía diabética no proliferativa leve. El uso de regiones de interés permitiría conservar una mayor cantidad de información en cada capa e incrementar así la probabilidad de localizar lesiones microscópicas en las imágenes de retina. La Figura 83 muestra los resultados de exactitud, en una nueva validación, de los modelos que presentaron el mejor desempeño al ser entrenados con regiones de interés. Para esta evaluación se utilizaron datos de los conjuntos E\_ophtha [55] y DIARETDB [52]. La exactitud registrada es superior al reportado por los modelos que trabajan con imágenes de retina original. Estos resultados indican que al utilizar el método de regiones de interés se resuelve la limitación existente en el método original y se incrementa la exactitud de los modelos.

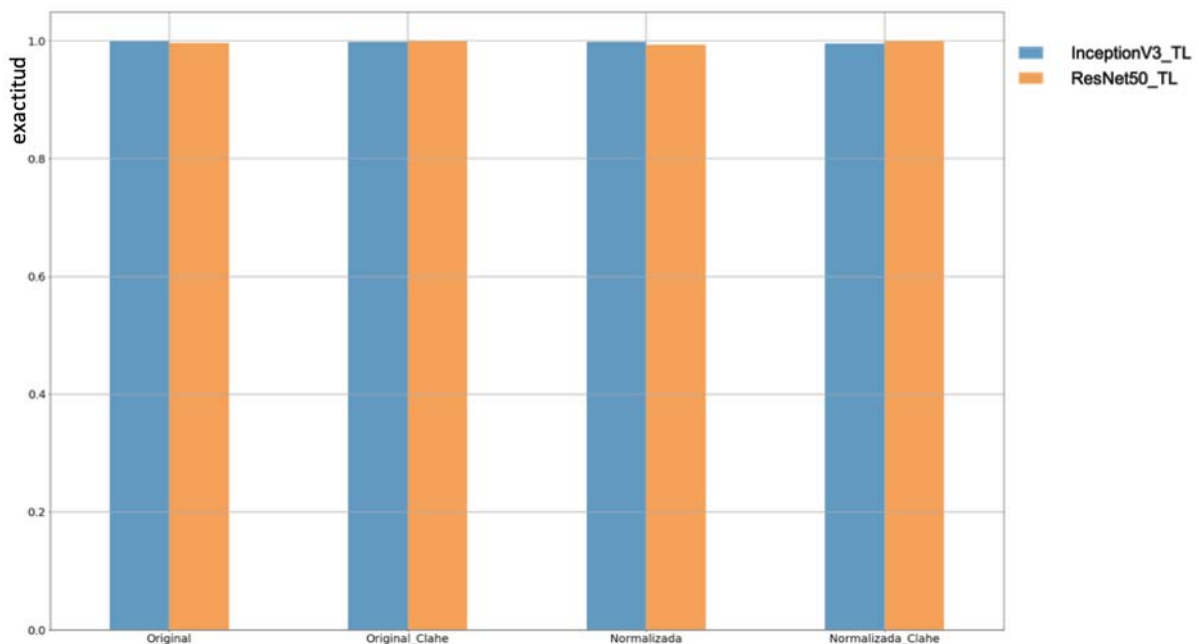


Figura 83. Exactitud con relación a los preprocesamientos empleados para las arquitecturas Inception y ResNet en datos de reserva. La abreviatura TL indica que los modelos fueron entrenados con transferencia de aprendizaje.



En la Figura 84 y Figura 85 se presentan los resultados de sensibilidad y especificidad, respectivamente, de los modelos evaluados en la Figura 83 con el conjunto de imágenes mencionado. Los resultados obtenidos son superiores a los reportados con imágenes originales y al igual que en los demás casos reportados, las variaciones derivadas del tipo de preprocesamiento y de la arquitectura seleccionada son mínimas. Al igual que con las imágenes originales, se tiene la hipótesis de que una validación con un mayor número de imágenes podría generar resultados significativamente diferentes entre los modelos considerados.

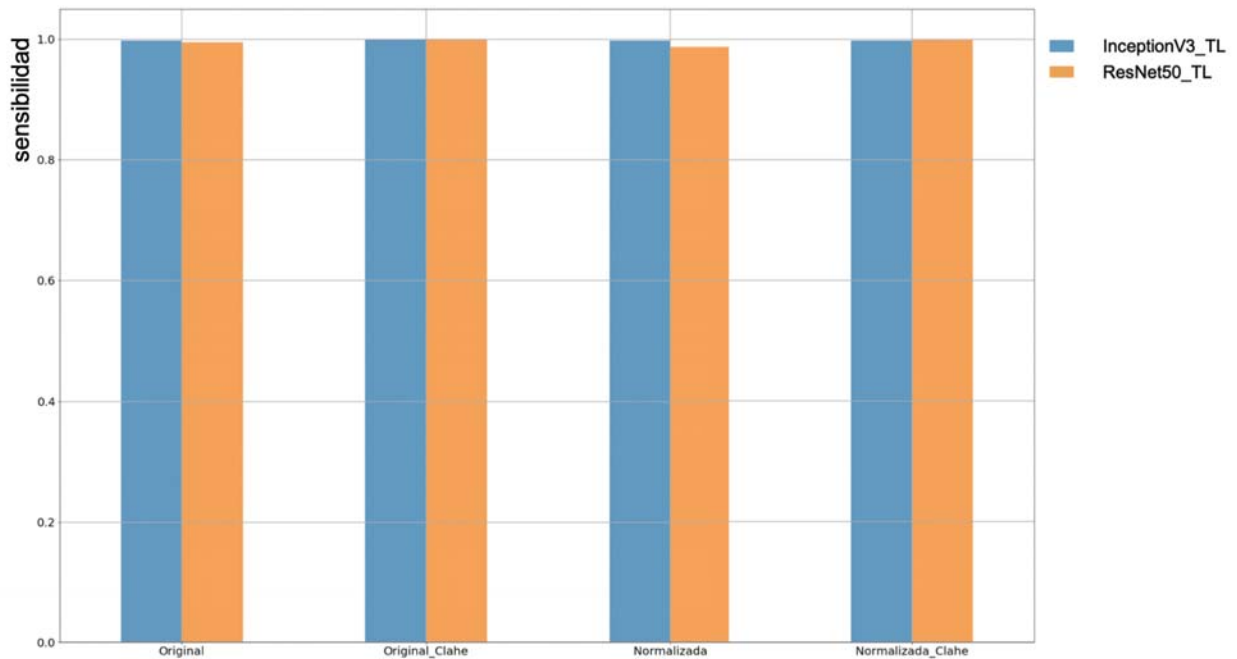


Figura 84. Gráfica con los valores de sensibilidad obtenidos para cada preprocesamiento con los datos de reserva. La abreviatura TL indica que los modelos fueron entrenados con transferencia de aprendizaje.



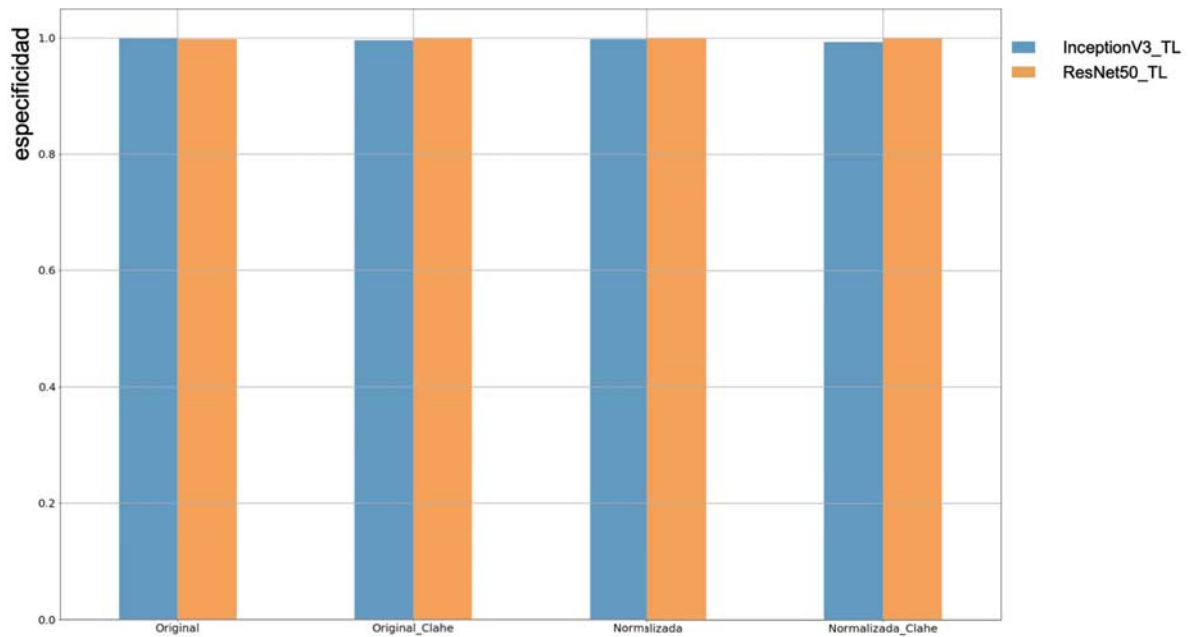


Figura 85. Gráfica con los valores de especificidad obtenidos para cada preprocesamiento con los datos de reserva. La abreviatura TL indica que los modelos fueron entrenados con transferencia de aprendizaje.

Para el método de regiones de interés también se obtuvieron los mapas de activación de algunas imágenes evaluadas. La Figura 86 muestra en la sección 1 a la imagen original; la sección 2 corresponde a la misma imagen, pero normalizada. Las imágenes normalizadas se utilizaron para la visualización de los modelos debido a que bajo este tipo de preprocesamiento se obtuvieron resultados, en las métricas de evaluación, ligeramente superiores a los demás. La sección 3 y sección 4 correspondientes a la arquitectura Inception y ResNet respectivamente comparten zonas de activación en común. Sin embargo, como ya se indicó en las imágenes originales estas zonas no resultan ser idénticas entre sí.

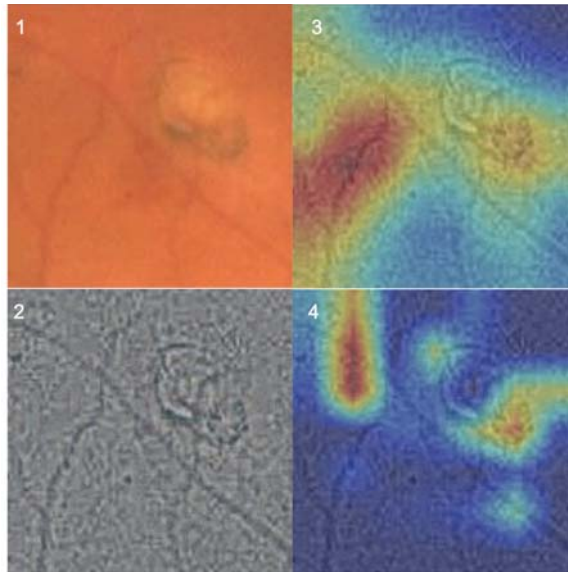


Figura 86. Visualización de región de interés [ROI] obtenida a partir de una imagen original de retina. [1] ROI original [2] ROI preprocesada [3] Visualización del modelo de decisión InceptionV3 [4] Visualización del modelo de decisión ResNet50.

La Figura 18 ejemplifica cómo la arquitectura ResNet, sección 4, realiza una segmentación más detallada de las lesiones presentes en comparación con la arquitectura Inception de la sección 3.

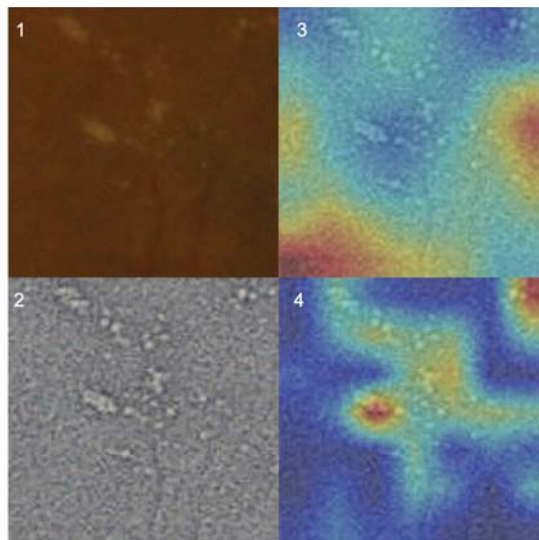


Figura 87. Visualización de región de interés [ROI] obtenida a partir de una imagen original de retina. [1] ROI original [2] ROI preprocesada [3] Visualización del modelo de decisión InceptionV3 [4] Visualización del modelo de decisión ResNet50.

En la Figura 88 se puede observar el mismo fenómeno que el registrado en la Figura 87. Esta característica, presente en la arquitectura ResNet, podría ser utilizada para realizar tareas de segmentación de las lesiones presentes en la imagen.

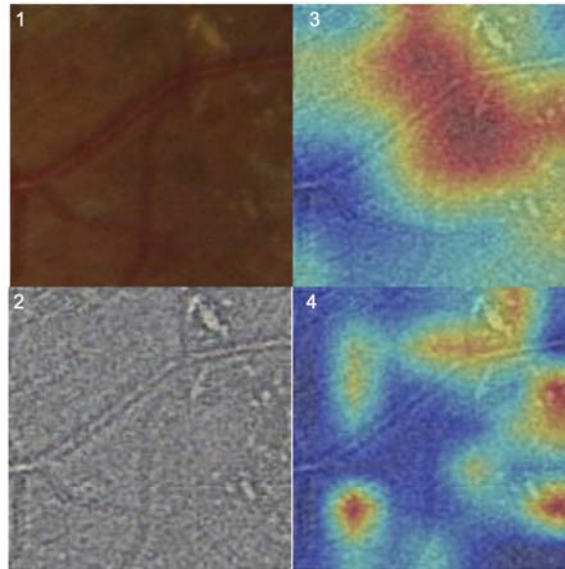


Figura 88. Visualización de región de interés [ROI] obtenida a partir de una imagen original de retina. [1] ROI original [2] ROI preprocesada [3] Visualización del modelo de decisión InceptionV3 [4] Visualización del modelo de decisión ResNet50.

El mapa de activación de la Figura 89 muestra que el modelo entrenado con la arquitectura Inception, sección 3, podría estar considerando secciones que no contienen lesiones propias de retinopatía diabética. Aunque en principio el modelo considera la sección donde se ubica una microaneurisma, esa no es la única sección que maximiza la activación del modelo. En contraste el modelo ResNet considera únicamente la zona donde se ubica la lesión mencionada.

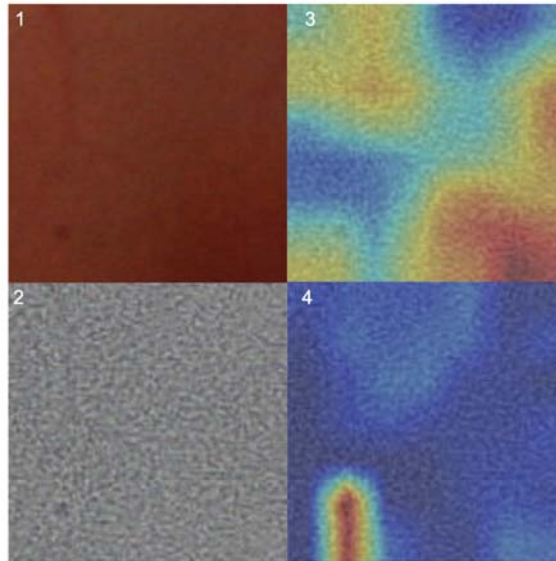


Figura 89. Visualización de región de interés [ROI] obtenida a partir de una imagen original de retina. [1] ROI original [2] ROI preprocesada [3] Visualización del modelo de decisión InceptionV3 [4] Visualización del modelo de decisión ResNet50.

Como sucedió con las imágenes originales la Figura 90 fue sometida a los preprocesamientos establecidos para después realizar una comparación de los mapas de activación generados. En la Figura 91 se puede identificar que dependiendo del preprocesamiento las zonas de activación presentan variaciones. La sección 7 es la única que considera, además de las regiones esperadas, zonas donde no parecen localizarse lesiones propias de retinopatía diabética.



Figura 90. Región de interés extraída de una imagen de retina.

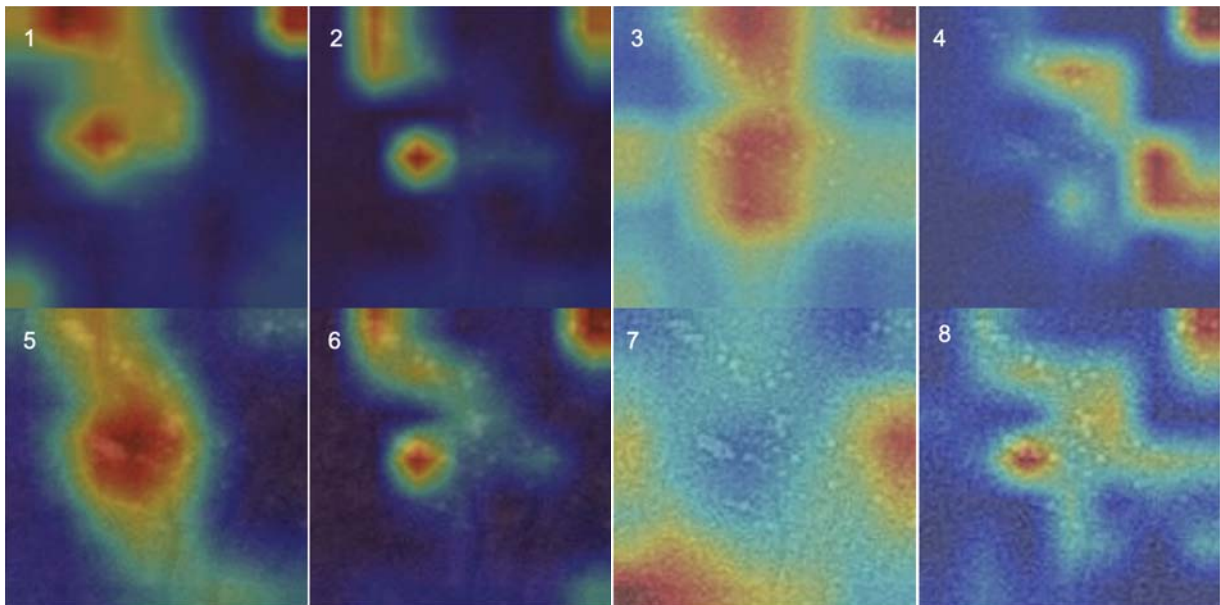


Figura 91. Comparación de la respuesta de los modelos de decisión entrenados al utilizar diferentes preprocesamientos. Se utiliza la imagen de la Figura 92 para su comparación. [1] ROI evaluada con la arquitectura InceptionV3 [2] ROI evaluada con la arquitectura ResNet50 [3] ROI normalizada evaluada con la arquitectura InceptionV3 [4] ROI normalizada evaluada con la arquitectura ResNet50 [5] ROI con ecualización de histograma evaluada con la arquitectura InceptionV3 [6] ROI con ecualización de histograma evaluada con la arquitectura ResNet50 [7] ROI normalizada y con ecualización del histograma evaluada con la arquitectura InceptionV3 [8] ROI normalizada y con ecualización del histograma evaluada con la arquitectura ResNet50.

Los resultados obtenidos con el método de regiones de interés demuestran ser superiores a los obtenidos con el método de imágenes originales. Con los datos disponibles fue posible realizar una validación adicional de los modelos entrenados. La validación consistió en tomar regiones de interés de las imágenes de los conjuntos IMO [24], IDRID [56] y Messidor [57]. Estos conjuntos no cuentan con anotaciones por lo que se esperaba que en las imágenes clasificadas con retinopatía diabética no proliferativa al menos una región de interés sea clasificada como enferma. En el caso de las imágenes normales ninguna región de interés se debería clasificar como enferma. La Figura 92 muestra los resultados de exactitud obtenidos de la evaluación de los modelos con el conjunto de datos IDRID [56]. Se puede observar que la clase Normal es clasificada de manera satisfactoria en la mayoría de los modelos. En el caso de las clases diagnosticada con retinopatía diabética no se registra un desempeño satisfactorio en ninguno de los modelos considerados.

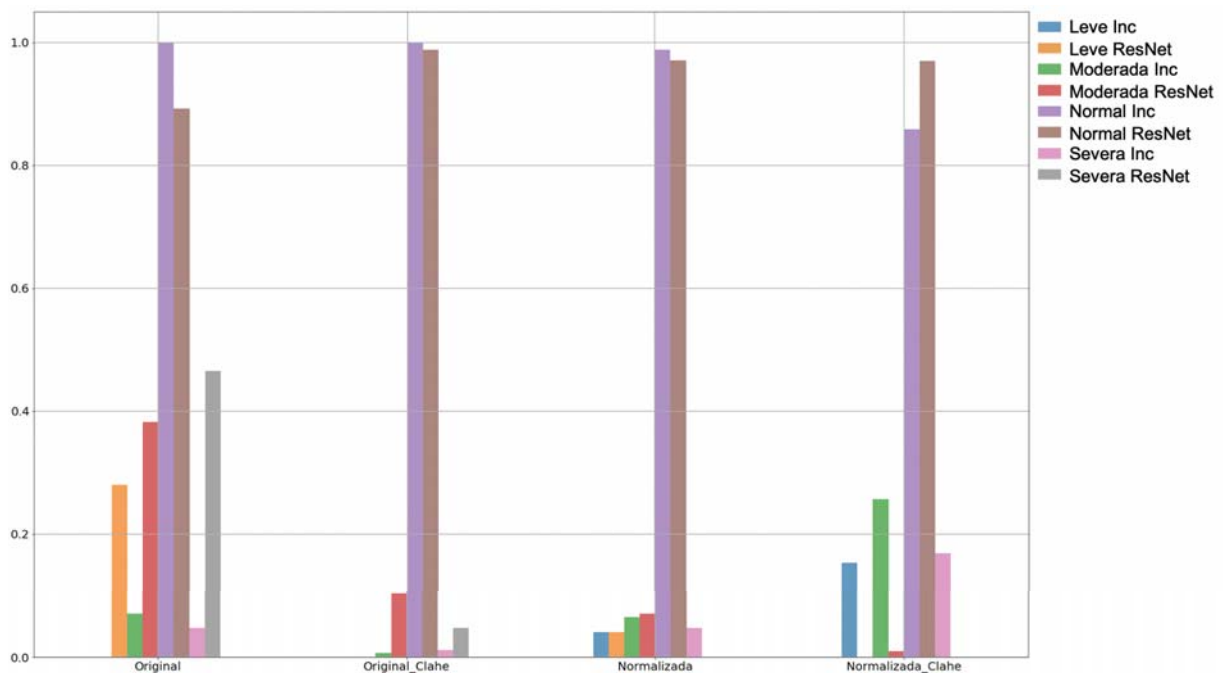


Figura 92. Gráfica con los valores de especificidad obtenidos para cada preprocesamiento con datos de reserva del conjunto IDRID [56]. La abreviación Inc hace referencia a la arquitectura Inception.

La Figura 93 muestra los resultados de exactitud al evaluar las imágenes del conjunto IMO [24]. Los resultados son similares a los obtenidos en la Figura 92, siendo satisfactoria la clasificación de la clase Normal pero no la de las clases con retinopatía diabética no proliferativa.

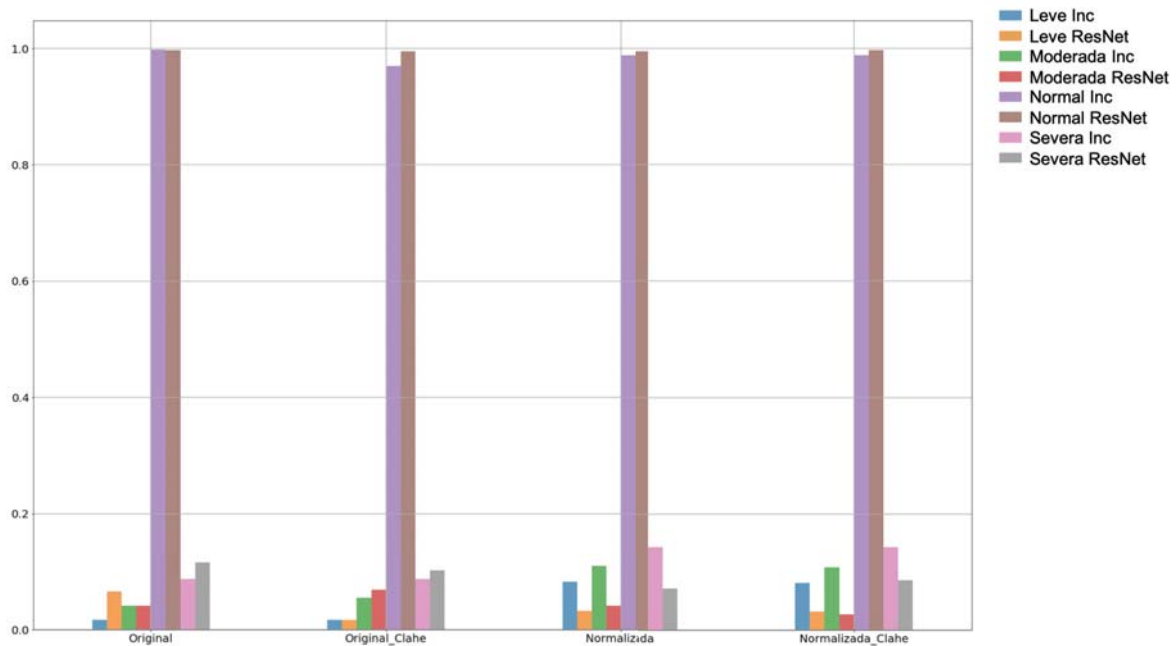


Figura 93. Gráfica con los valores de especificidad obtenidos para cada preprocesamiento con datos de reserva del conjunto IMO [24]. La abreviación Inc hace referencia a la arquitectura Inception.

Finalmente, la Figura 94 presenta los resultados de la evaluación del conjunto Messidor [51]. El desempeño de los modelos es similar al registrado en la Figura 92 y Figura 93. Estos resultados son significativamente diferentes a los obtenidos en la Figura 83, Figura 84 y Figura 85 en donde también se utilizan datos de reserva pero de los conjuntos E\_ophtha [55] y DIARETDB [52].

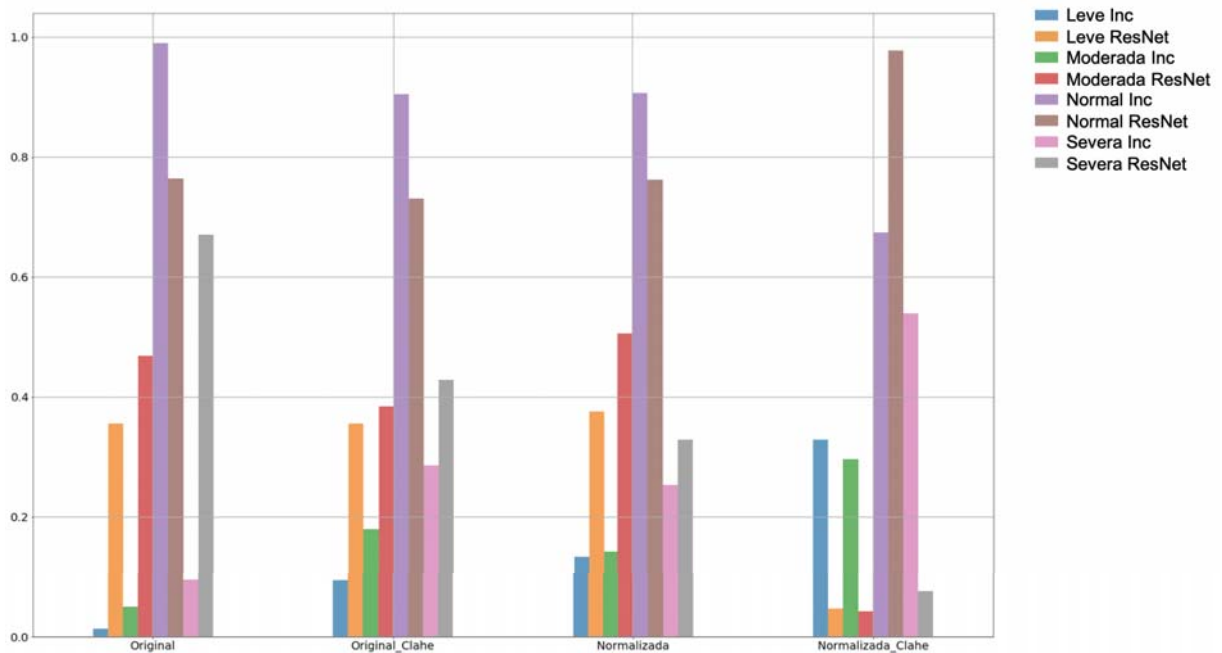


Figura 94. Gráfica con los valores de especificidad obtenidos para cada preprocesamiento con datos de reserva del conjunto Messidor [57]. La abreviación Inc hace referencia a la arquitectura Inception.

En el apéndice 2 sección se localizan tablas de referencia con los valores numéricos de todas las gráficas mostradas anteriormente.

Después de un análisis exhaustivo y de la realización de diferentes ajustes a los modelos de decisión, se determinaron los factores que conllevan a los resultados obtenidos en la Figura 92, Figura 93 y Figura 94.

- Conjunto de datos pequeño

El número de imágenes con el que se cuenta es pequeño, comparado con otros conjuntos de datos reportados en la literatura para la realización de la misma tarea. Alban et al. [26] reportan el uso de más de nueve mil imágenes enfermas para el entrenamiento de una ConvNet. La limitación del número de imágenes impide una generalización satisfactoria al momento de evaluar los modelos en un nuevo conjunto de datos. Aunque fueron utilizadas técnicas para el aumento de imágenes, en la Figura 95 se puede visualizar la diferencia entre las regiones de activación consideradas. La sección 1 de la Figura 95 corresponde a la región de una imagen original, del conjunto E\_ophtha [55], con lesiones propias de retinopatía diabética. El mapa de activación de la sección 2,



obtenido a partir de la sección 1, corresponde con lo esperado ya que las lesiones son las que maximizan la activación del modelo. Imágenes de ese conjunto de datos también fueron utilizadas durante la fase de entrenamiento y validación por lo que se asume que comparten características similares. La sección 3 de la Figura 95 es una región de una imagen original, del conjunto IMO [24]. Al igual que en la sección 1, la imagen contiene lesiones propias de retinopatía diabética. El mapa de activación obtenido se encuentra en la sección 4 y en él se puede observar que las regiones de activación no corresponden con las lesiones presentes. En los modelos entrenados con regiones de interés ninguna imagen del conjunto IMO [24], IDRID [56] y Messidor [57] fue utilizada. Esto explicaría la falta de generalización frente a datos extraídos de esos conjuntos. Además de ello la Figura 96 y Figura 97 muestran dos ejemplos en donde resulta complicado identificar las lesiones que clasificarían a las imágenes en la clase retinopatía diabética no proliferativa. La suma de ambos factores contribuye al bajo desempeño de clasificación mostrado en la Figura 92, Figura 93 y Figura 94.

- No se tiene una muestra de todas las posibles lesiones

Además de los factores mencionados es importante considerar que las colecciones utilizadas solamente proporcionan información sobre exudados, hemorragias, puntos rojos y microaneurismas. Cualquier lesión diferente de las mencionadas no contribuye a la clasificación de las imágenes dentro de la clase correspondiente a retinopatía diabética no proliferativa.

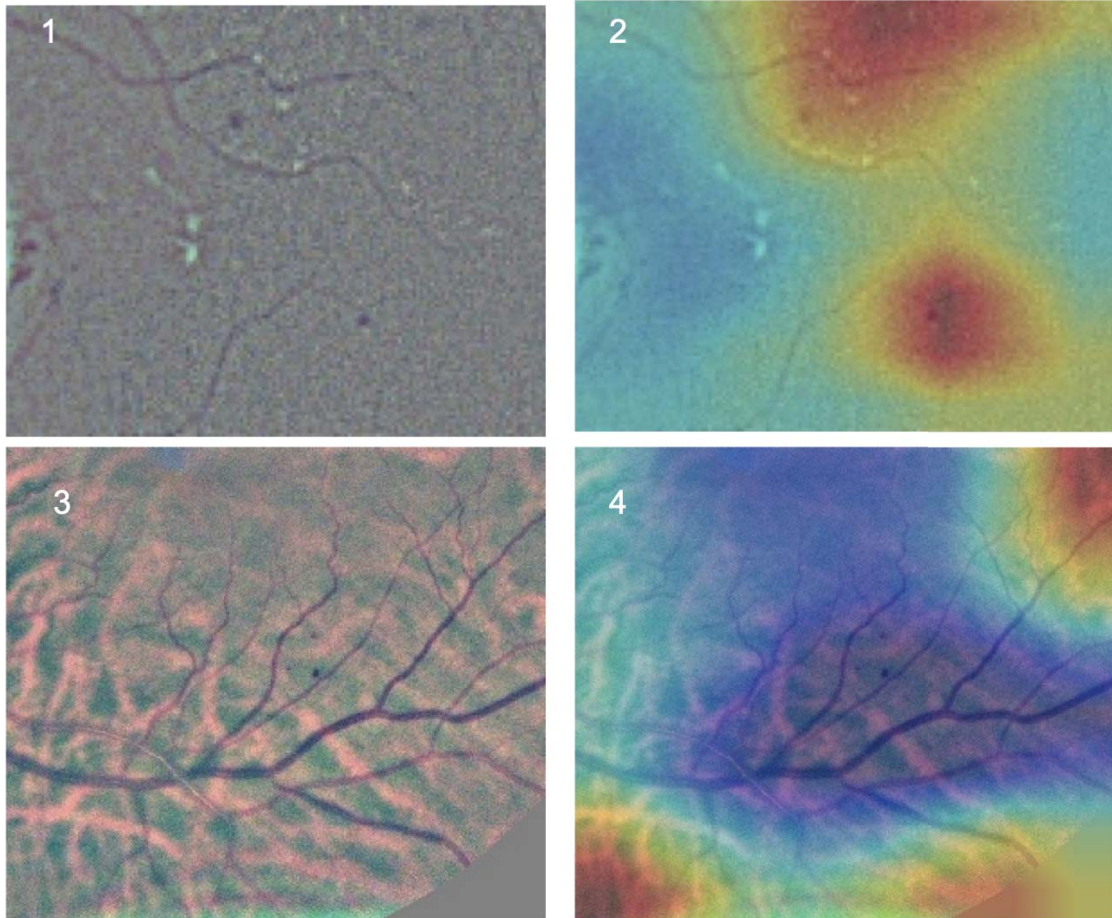


Figura 95. Comparación de predicciones. [1] Imagen del conjunto E\_ophtha [55] clasificada como enferma. [2] Mapa de calor generado con información de la imagen [1]. [3] Imagen del conjunto IMO [24] clasificada como enferma. [4] Mapa de calor generado con información de la imagen [3].

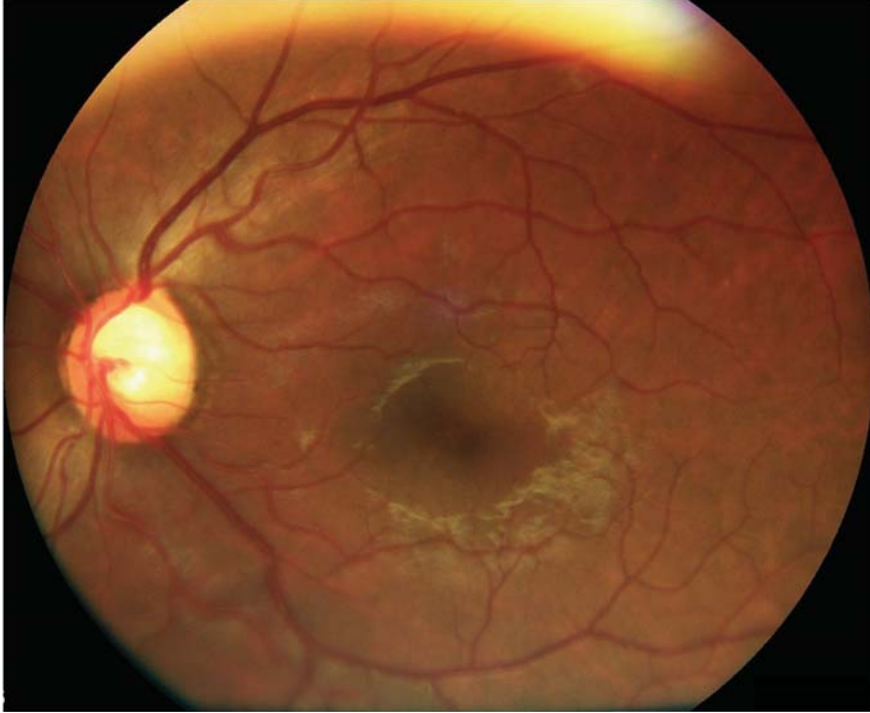


Figura 96. Imagen de retina clasificada como leve.

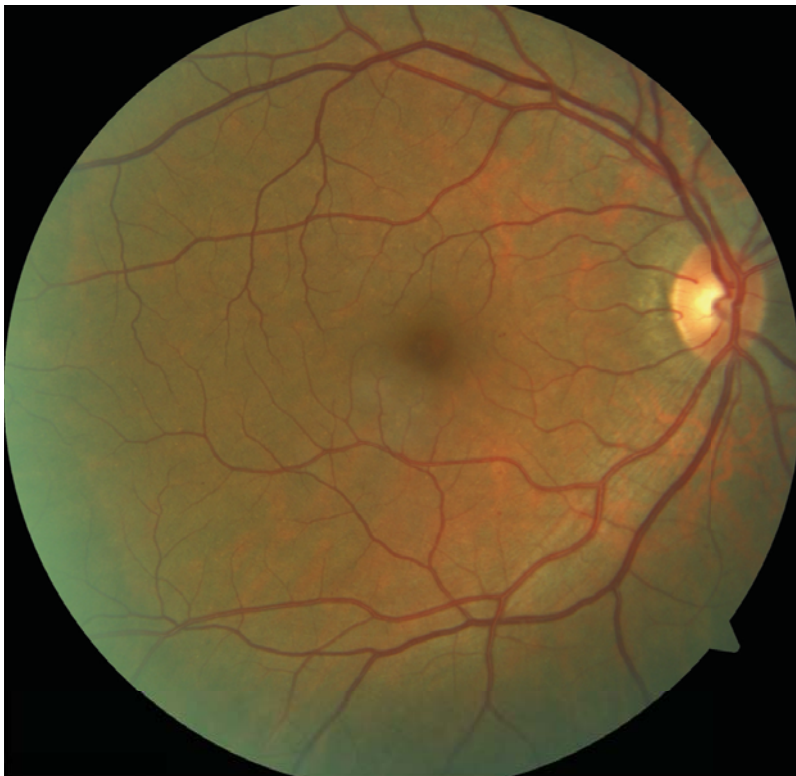


Figura 97. Imagen de retina clasificada como leve

La Tabla 19 muestra los valores utilizados, en los hiperparámetros, que mostraron el mejor comportamiento después de la realización de varios ajustes a los modelos.

Hiperparámetro	Valor
Tasa de aprendizaje	0.001
Tamaño de lote	16
Número de iteraciones	30
Funciones de activación	ReLU y Sigmoid en la capa final

Tabla 19. Hiperparámetros finales de entrenamiento.

La Tabla 20 muestra el tiempo de entrenamiento de los modelos evaluados para la realización de esta investigación. El valor de los hiperparámetros utilizados se muestra en la Tabla 19. Se puede observar que el uso de transferencia de aprendizaje no afecta significativamente el tiempo total de entrenamiento en comparación con los modelos que no lo emplean. Esto es algo esperado debido a que, al utilizar transferencia de aprendizaje, únicamente se asignan los pesos de un modelo previamente entrenado a un nuevo modelo por entrenar. Por lo mismo únicamente se afecta el número de iteraciones necesarias para lograr una convergencia, pero no así el tiempo de entrenamiento para realizar cada iteración con el conjunto de imágenes dado. El aumento en el tiempo de entrenamiento al no utilizar GPUs es notorio. En los resultados reportados se tiene que considera que se trabajó con un equipo con procesador de 24 núcleos; esta característica es poco común en un equipo de cómputo promedio. Por ello al realizar el entrenamiento en un equipo de menor gama el tiempo de entrenamiento aumentaría considerablemente. El número de GPUs utilizados durante el entrenamiento modificó ligeramente el tiempo de entrenamiento total. Esto tiene una mayor trascendencia a medida que se aumenta el número de imágenes empleadas, así como la cantidad de iteraciones necesarias para lograr la convergencia del modelo.

Arquitectura	Transferencia de aprendizaje	Imágenes empleadas	Número total de imágenes de entrenamiento	Número de GPUs	Tiempo de entrenamiento
Darshit Doshi [28]	No	Originales [299]	1792	0	162.5 min
Kele Xu [58]	No	Originales [299]	1792	0	219.5 min
Darshit Doshi [28]	No	Originales [299]	1792	1	132 min
Kele Xu [58]	No	Originales [299]	1792	1	130.5 min
Darshit Doshi [28]	No	Originales [299]	1792	2	126 min
Kele Xu [58]	No	Originales [299]	1792	2	126.5 min
Inception V3	No	Originales [299]	1792	0	360 min
ResNet 50	No	Originales [224]	1792	0	356 min
Inception V3	No	Originales [299]	1792	1	142 min
ResNet 50	No	Originales [224]	1792	1	127 min
Inception V3	No	Originales [299]	1792	2	135 min

Arquitectura	Transferencia de aprendizaje	Imágenes empleadas	Número total de imágenes de entrenamiento	Número de GPUs	Tiempo de entrenamiento
ResNet 50	No	Originales [224]	1792	2	123 min
Inception V3	Si	Originales [299]	1792	0	359 min
ResNet 50	Si	Originales [224]	1792	0	354 min
Inception V3	Si	Originales [299]	1792	1	144 min
ResNet 50	Si	Originales [224]	1792	1	129.5 min
Inception V3	Si	Originales [299]	1792	2	134 min
ResNet 50	Si	Originales [224]	1792	2	122 min
Inception V3	No	Regiones de interés [299]	4346	0	834 min
ResNet 50	No	Regiones de interés [224]	4346	0	839 min
Inception V3	No	Regiones de interés [299]	4346	1	209.5 min
ResNet 50	No	Regiones de interés [224]	4346	1	141.5 min
Inception V3	No	Regiones de interés [299]	4346	2	151.5 min

Arquitectura	Transferencia de aprendizaje	Imágenes empleadas	Número total de imágenes de entrenamiento	Número de GPUs	Tiempo de entrenamiento
ResNet 50	No	Regiones de interés [224]	4346	2	112.4 min
Inception V3	Si	Regiones de interés [299]	4346	0	832 min
ResNet 50	Si	Regiones de interés [224]	4346	0	834 min
Inception V3	Si	Regiones de interés [299]	4346	1	205.5 min
ResNet 50	Si	Regiones de interés [224]	4346	1	141 min
Inception V3	Si	Regiones de interés [299]	4346	2	153 min
ResNet 50	Si	Regiones de interés [224]	4346	2	112.4 min

Tabla 20. Tabla con tiempos de ejecución de los modelos utilizados.

## Conclusiones

---

Las Redes Neuronales Convolucionales tienen un desempeño satisfactorio de clasificación al momento de trabajar con imágenes médicas. La principal limitación de este tipo de modelos es la necesidad de contar con un gran conjunto de datos. Ese conjunto de datos debe ser representativo de la patología que se quiere detectar para asegurar una generalización adecuada del modelo.

En base a los resultados obtenidos, los modelos entrenados con imágenes originales tienen una mayor probabilidad de registrar resultados satisfactorios al ser evaluados en un nuevo conjunto de datos. Esto debido a que fueron entrenados con datos de tres centros médicos diferentes además de considerar la información del total de la imagen. Esto último permite la consideración del total de lesiones presentes en las imágenes. En el caso de las regiones de interés sólo se utilizaron imágenes de dos centros diferentes y el tipo de lesiones etiquetadas fue limitado. Sin embargo, los modelos entrenados con regiones de interés pueden ser utilizados como punto de inicio, mediante transferencia de aprendizaje, al momento de realizar un nuevo entrenamiento con un nuevo conjunto de imágenes que permita la consideración de un mayor número de características. Para mejorar los resultados obtenidos con regiones de interés es necesario tener al menos uno de los siguientes escenarios.

- Recopilar una mayor cantidad de imágenes anotadas. Esta información tendría que ser proporcionada por un grupo de especialistas encargados de identificar, en imágenes de retina, las zonas donde existen lesiones propias de retinopatía diabética.



- Incorporar nuevos operadores a las arquitecturas existentes que permitan retener una mayor cantidad de información a partir del uso de una menor cantidad de datos.
- El uso de aprendizaje supervisado débil, en el cual sólo es necesario una clasificación general de la imagen y a partir de ello se inspeccionan los elementos que la conforman para extraer las características que le confieren la clase asignada. En el caso de imágenes de retina si se cuenta con una imagen clasificada como enferma, se dividiría el total de la imagen en regiones pequeñas, de igual forma como sucedió con el método de regiones de interés. Al estar clasificada como enferma se asume que al menos una de las regiones en las que fue dividida tendrá características semejantes al resto de las clasificadas como enfermas. Por lo tanto, se guarda la información de esas regiones y se descarta la información del resto de las regiones. De esta manera en nuevas imágenes, se buscará localizar por regiones las características aprendidas durante el entrenamiento.
- Utilizar técnicas de aprendizaje no supervisado. Bajo este esquema no sería necesario asignar clases al conjunto de datos utilizado. Sólo sería necesario entrenar al modelo con un conjunto de datos representativo de la patología para extraer características relevantes y lograr la generación del modelo de clasificación.

El uso de Inteligencia Artificial en la industria del cuidado de la salud continuará desarrollándose y recopilando información que permita tener modelos más exactos y precisos para la atención de los pacientes. La gran ventaja de este tipo de sistemas es su capacidad para atender con la misma calidad y bajo las mismas normas a cualquier persona independientemente de su condición social, su ubicación geográfica o su situación económica.

## Bibliografía

- [1] Keeler, «150 years since Babbage's ophthalmoscope,» *Arch Ophthalmol*, pp. 115(11):1456-7, Nov 1997.
- [2] H. v. Helmholtz, Descripción de un espejo ocular para el examen de la retina en el ojo vivo, Alemania: Springer-Verlag Berlín Heidelberg, 1851.
- [3] A. C. v. Trigt, Dissertatio ophthalmologica inauguralis De speculo oculi ejusque usu observation, Utrecht: Trajecti ad Rhenum : P.W. van de Weijer, 1853.
- [4] J. E, «Beitr zur Pathol des Auges,» Wien, 1856, p. vol p. 33 Fig 12.
- [5] H. H., «A Quartely Journal of Ophtalmic Surgery & Science,» *The Ophthalmic Review*, vol. 1, p. 312, 1856.
- [6] W. J. Jackman WT, «On photographing the retina of the living eye,» *Philadelphia Photographer*, Philadelphia, 1886; 23:340-341.
- [7] H. J, «Über diabetische Netzhautentzündung,» *Dtsch Med Wochenschr*, 1890;13:1181.
- [8] A. L, «Ocular fundus photography: Suggestion for achieving consistently good pictures and instructions for stereoscopic photography,» *Am J Ophthalmol*, 1964 Jan, pp. 57():13-28.
- [9] G. AA., «The dignosis of acute abdominal pain with computer analysis,» *J R Coll Surg Edinb*, 1976;21:170-172.
- [10] L. B. K. J. Budoin CE, «Automatic detection of microaneurysms in diabetic fluorescein angiography,» *Rev Epidemiol Sante Publique*, pp. 32(3-4):254-61, 1984.
- [11] K. Fukushima, «Neocognitron,» 2007. [En línea]. Available: 1717. doi:10.4249/scholarpedia.1717. .
- [12] J. Brownlee, «Machine Learning Mastery,» *Deep Learning for Computer Vision*, 1 Mayo 2019. [En línea]. Available: <https://machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/>. [Último acceso: 5 Septiembre 2019].

- [13] O. Shalev, «LinkedIn,» Blog, 5 Abril 2017. [En línea]. Available: <https://www.linkedin.com/pulse/deep-learning-overhyped-ofir-shalev>. [Último acceso: 5 Septiembre 2019].
- [14] W. D. D. Y. & W. D. Pan, «Classification of Malaria-Infected Cells Using Deep Convolutional Neural Networks.,» *Machine Learning: Advanced Techniques and Emerging Applications*, p. 159, 2018.
- [15] H. E.-D. E. S. A. E.-H. E. S. M. & S. A. B. M. Mohsen, «Classification using deep learning neural networks for brain tumors,» *Future Computing and Informatics Journal*, pp. 3(1), 68-71., 2018.
- [16] O. S. M. M. U. J. & J. D. U. Stephen, «An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare,» *Journal of healthcare engineering*, vol. 2019, 2019.
- [17] D. A. S. M. M. S. & R. J. Ragab, «Breast cancer detection using deep convolutional neural networks and support vector machines,» *PeerJ*, 7, e6201., 2019.
- [18] C. Plus, «Cuidate Plus,» Blog, 12 11 2018. [En línea]. Available: <https://cuidateplus.marca.com/enfermedades/digestivas/diabetes.html>. [Último acceso: 8 Junio 2020].
- [19] Reuters, «Milenio,» Blog, 13 Diciembre 2018. [En línea]. Available: <https://www.milenio.com/tecnologia/google-lanza-sistema-inteligencia-artificial-evaluar-enfermedad-ocular-diabetica-tailandia>. [Último acceso: 2 Mayo 2020].
- [20] C. Metz, «New York Times,» Blog, 14 Marzo 2019. [En línea]. Available: <https://www.nytimes.com/es/2019/03/14/espanol/india-usa-inteligencia-artificial-para-combatir-la-ceguera-diabetica.html>. [Último acceso: 2 Mayo 2020].
- [21] E. PRESS, «20 minutos,» Blog, 30 Mayo 2019. [En línea]. Available: <https://www.20minutos.es/noticia/3656298/0/hospital-candelaria-aplica-inteligencia-artificial-para-detectar-retinopatia-diabetica/>. [Último acceso: 2 Mayo 2020].

- [22] Desconocido, «XINHUA español,» Blog, 2 Junio 2018. [En línea]. Available: [http://spanish.xinhuanet.com/2018-06/02/c\\_137225544.htm](http://spanish.xinhuanet.com/2018-06/02/c_137225544.htm). [Último acceso: 2 Mayo 2020].
- [23] «Instituto Mexicano de Oftalmología,» [En línea]. Available: <https://imoiap.com.mx>. [Último acceso: 3 Mayo 2020].
- [24] J. Cuadros, «California Health Care Foundation,» Blog, 02 03 2016. [En línea]. Available: <https://www.chcf.org/blog/artificial-intelligence-eases-burden-of-blindness-in-diabetes/>. [Último acceso: 08 06 2020].
- [25] M. & G. T. Alban, «Automated detection of diabetic retinopathy using fluoresceon angiography photographs,» *Report of standard education*, 2016.
- [26] F. C. D. M. B. S. P. H. Y. Z. Harry Pratta, «Convolutional Neural Networks for Diabetic Retinopathy,» *International Conference On Medical Imaging Understanding and Analysis*, 2016.
- [27] D. Doshi, A. Shenoy, D. Sidhpura y P. Gharpure, «Diabetic Retinopathy detection using deep convolutional neural networks,» *International Conference on Computing, Analytics and Security Trends* , 19-21 Dec 2016.
- [28] S. M. S. J. Sagar Honnugar, «Diabetic Retinopathy Identification and Severity Classification,» *Report of standford education.*, 2016.
- [29] R. G. K. & M. S. Ghosh, «Automatic detection and classification of diabetic retinopathy stages using CNN,» *4th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 550-554, 2017.
- [30] B. C. M. S. R. V. V. Suvajit Dutta, «A comparative study of deep learning models for medical image classification,» *Materials Science and Engineering Conference Series*, vol. 263, p. 042097, 2017.
- [31] S. Mohammadian, A. Karsaz y Y. M. Roshan, «A comparative analysis of classification algorithms in diabetic retinopathy screening,» *7th International Conference on Computer and Knowledge Engineering (ICCKE)*, 2017.

- [32] S. M. S. H. M. M. & A. S. Islam, «Deep Learning based Early Detection and Grading of Diabetic Retinopathy Using Retinal Fundus Images.,» *arXiv preprint arXiv:1812.10595.*, 2018.
- [33] N. Chakrabarty, «A Deep Learning Method for the detection of Diabetic Retinopathy.,» *5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, pp. 1-5, 2018.
- [34] V. L. Z. W. L. G. N. Q. D. X. Y. Y. M. Y. .. & L. M. L. Bellemo, «Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study.,» *The Lancet Digital Health.*, pp. e35-e44., 2019.
- [35] M. T. & K. S. Hagos, «Transfer Learning based Detection of Diabetic Retinopathy from Small Dataset.,» *arXiv preprint arXiv:1905.07203.*, 2019.
- [36] I. Technologies., «FDA permits marketing of IDx-DR for automated detection of diabetic retinopathy in primary care [press release].,» [www.eyediagnosis.net/single-post/2018/04/12/FDA-permits-marketing-of-IDx-DR-for-automated-detection-of-diabetic-retinopathy-in-primary-care.](http://www.eyediagnosis.net/single-post/2018/04/12/FDA-permits-marketing-of-IDx-DR-for-automated-detection-of-diabetic-retinopathy-in-primary-care.), April 12, 2018. .
- [37] IDx, «Eye Diagnosis,» [En línea]. Available: <https://www.eyediagnosis.co/idx-dr-eu-1>. [Último acceso: 09 06 2020].
- [38] L. B. V. S. F. Bandello, «Next Sight,» Publication, 2019. [En línea]. Available: [https://www.nextsight.info/blog/press/41-eye\\_doctor](https://www.nextsight.info/blog/press/41-eye_doctor). [Último acceso: 09 06 2020].
- [39] Visulytix, Digital brochure, 2019. [En línea]. Available: <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE3claH>. [Último acceso: 09 06 2020].
- [40] J. Low, «Medium,» Blog, [En línea]. Available: <https://medium.com/supahands-techblog/what-is-image-annotation-caf4107601b7>. [Último acceso: 09 06 2020].
- [41] S. Ruder, «An overview of gradient descent optimization algorithms,» Blog, 19 01 2016. [En línea]. Available: <https://ruder.io/optimizing-gradient-descent/index.html#gradientdescentvariants>. [Último acceso: 16 02 2020].

- [42] J. M. Valverde, «Lipman's Artificial Intelligence Directory,» Blog, 19 January 2018. [En línea]. Available: <http://laid.delanover.com/difference-between-l1-and-l2-regularization-implementation-and-visualization-in-tensorflow/>. [Último acceso: 31 01 2020].
- [43] Y. B. L. B. Y. & H. P. LeCun, «Gradient-based learning applied to document recognition,» *Proceedings of the IEEE*, vol. 86(11), pp. 2278-2324, 30 Septiembre 1998.
- [44] A. S. I. & H. G. E. Krizhevsky, «Imagenet classification with deep convolutional neural networks.,» *In Advances in neural information processing systems*, pp. 1097-1105, 2012.
- [45] M. D. & F. R. Zeiler, «Visualizing and understanding convolutional networks.,» *European conference on computer vision. Springer, Cham.*, 818-833.
- [46] C. L. W. J. Y. S. P. R. S. A. D. ... & R. A. Szegedy, «Going deeper with convolutions,» *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9, 2015.
- [47] K. & Z. A. Simonyan, «Very deep convolutional networks for large-scale image recognition,» *arXiv preprint arXiv:1409.1556*, 2014.
- [48] K. Z. X. R. S. & S. J. He, «Deep residual learning for image recognition,» *Proceedings of the IEEE conference on computer vision and pattern recognition* , pp. 770-778, 2016.
- [49] G. L. Z. V. D. M. L. & W. K. Q. Huang, «Densely connected convolutional networks,» *Proceedings of the IEEE conference on computer vision and pattern recognition* , pp. 4700-4708, 2017.
- [50] P. Warden, «Pete Warden's blog,» Blog, 14 Diciembre 2017. [En línea]. Available: <https://petewarden.com/2017/12/14/how-many-images-do-you-need-to-train-a-neural-network/>.
- [51] V. K. J.-K. K. e. a. Tomi Kauppi, «IMAGE DATABASE,» 19 06 2007. [En línea]. Available: <http://www.it.lut.fi/project/imageret/diaretdb1/>.

- [52] ADCIS, «E\_ophtha,» 2013. [En línea]. Available: <http://www.adcis.net/en/third-party/e-ophtha/>.
- [53] E. & C. H. Foundation, «Kaggle,» 2015. [En línea]. Available: <https://www.kaggle.com/c/diabetic-retinopathy-detection>.
- [54] IDRiD, «Diabetic Retinopathy segmentation and grading challenge,» Data, [En línea]. Available: <https://idrid.grand-challenge.org/Data/>. [Último acceso: 21 02 2020].
- [55] ADCIS, «ADCIS,» 2014. [En línea]. Available: <http://www.adcis.net/es/descargas-de-software-de-terceros/messidor-es/>.
- [56] U. o. Iowa, «Retinopathy Online Challenge,» Data, 2007. [En línea]. Available: <http://webeye.ophth.uiowa.edu/ROC/>. [Último acceso: 21 02 2020].
- [57] K. F. D. & M. H. Xu, «Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image,» *Molecules*, pp. 22(12), 2054., 2017.
- [58] A. Rosebrock, «PyImageSearch,» Tutorial, 8 Julio 2019. [En línea]. Available: <https://www.pyimagesearch.com/2019/07/08/keras-imagedatagenerator-and-data-augmentation/>.
- [59] F. Chollet, «Building powerful image classification models using very little data,» Tutorials, 05 06 2016. [En línea]. Available: <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html>. [Último acceso: 20 02 2020].
- [60] S. Ivanov, «37 Reasons why your Neural Network is not working,» Blog, 25 07 2017. [En línea]. Available: <https://blog.slavv.com/37-reasons-why-your-neural-network-is-not-working-4020854bd607>. [Último acceso: 20 02 2020].
- [61] «Wikipedia,» 18 Junio 2019. [En línea]. Available: <https://es.wikipedia.org/wiki/Keras>.
- [62] «Arterys receives 510(k) clearance for Arterys software for cloud-based medical image visualization and quantification [press release].,» [www.prnewswire.com/news-releases/arterys-receives-510k-clearance-for-arterys-software-for-cloud-based-medical-image-visualization-and-quantification-](http://www.prnewswire.com/news-releases/arterys-receives-510k-clearance-for-arterys-software-for-cloud-based-medical-image-visualization-and-quantification-)

300355923.html?tc=eml\_clearime. Accessed November 20, 2018.>, November 2, 2016. .

- [63] «Kernel (image processing),» 26 Agosto 2019. [En línea]. Available: [https://en.wikipedia.org/wiki/Kernel\\_\(image\\_processing\)](https://en.wikipedia.org/wiki/Kernel_(image_processing)). [Último acceso: 10 Septiembre 2019].
- [64] Stanford, «Convolutional Neural Networks for Visual Recognition,» [En línea]. Available: <http://cs231n.github.io/convolutional-networks/>. [Último acceso: 10 Septiembre 2019].
- [65] D. BRITZ, «WILDML Artificial Intelligence, Deep Learning, and NLP,» 7 Noviembre 2015. [En línea]. Available: <http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/>.
- [66] «Quora,» 25 Diciembre 2017. [En línea]. Available: <https://www.quora.com/How-is-a-convolutional-neural-network-able-to-learn-invariant-features>.
- [67] «GitHub,» 14 Octubre 2017. [En línea]. Available: <https://github.com/rasbt/python-machine-learning-book/blob/master/faq/difference-deep-and-normal-learning.md>.
- [68] «StackExchange,» 18 Julio 2017. [En línea]. Available: <https://stats.stackexchange.com/questions/182102/what-do-the-fully-connected-layers-do-in-cnns/182122#182122>.
- [69] Y. Jia, «Caffe,» Blog, [En línea]. Available: <https://caffe.berkeleyvision.org>.
- [70] V. N. & G. H. Alex Krizhevsky, «The CIFAR-10 dataset,» [En línea]. Available: <https://www.cs.toronto.edu/~kriz/cifar.html>.
- [71] S. a. S. C. Ioffe, «Batch normalization: Accelerating deep network training by reducing internal covariate shift.,» *Proceedings of The 32nd International Conference on Machine Learning*, p. 448–456, 2015.
- [72] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens y Z. and Wojna, «Rethinking the inception architecture for computer vision,» *arXiv preprint arXiv:1512.00567..*
- [73] V. Cheplygina, «GroundAI,» Blog, 12 Octubre 2018. [En línea]. Available: <https://www.groundai.com/project/cats-or-cat-scans-transfer-learning-from-natural-or-medical-image-source-datasets/1>. [Último acceso: 09 01 2020].



[74] SHAOANLU, «SHAOANLU,» Blog, 29 05 2017. [En línea]. Available: <https://shaoanlu.wordpress.com/2017/05/29/sgd-all-which-one-is-the-best-optimizer-dogs-vs-cats-toy-experiment/>. [Último acceso: 09 01 2020].

[75] J. Cuadros, «California Health Care Foundation,» Blog, 2 Marzo 2016. [En línea]. Available: <https://www.chcf.org/blog/artificial-intelligence-eases-burden-of-blindness-in-diabetes/>. [Último acceso: 2 Mayo 2020].

---

## Apéndice

---

### 1. Tablas con valores utilizados en resultados

#### Inicialización aleatoria

Autor	Entrenamiento		Validación	
	Pérdida	Exactitud	Pérdida	Exactitud
<b>Darshit Doshi [28]</b>	0.4121	0.7985	0.4885	0.7617
<b>Kele Xu [58]</b>	0.4391	0.7907	0.4925	0.7461

Tabla 21. Resultados de evaluar imágenes originales sin preprocesamientos.

Autor	Entrenamiento		Validación	
	Pérdida	Exactitud	Pérdida	Exactitud
<b>Darshit Doshi [28]</b>	0.5159	0.7478	0.5266	0.7344
<b>Kele Xu [58]</b>	0.4914	0.7645	0.5466	0.7461

Tabla 22. Resultados de evaluar imágenes originales con equalización adaptativo del histograma.

Autor	Entrenamiento		Validación	
	Pérdida	Exactitud	Pérdida	Exactitud
<b>Darshit Doshi [28]</b>	0.2927	0.8738	0.4209	0.8125
<b>Kele Xu [58]</b>	0.3661	0.8437	0.3948	0.8164

Tabla 23. Resultados de evaluar imágenes originales normalizadas.

Autor	Entrenamiento		Validación	
	Pérdida	Exactitud	Pérdida	Exactitud
<b>Darshit Doshi [28]</b>	0.3507	0.8421	0.4451	0.7812
<b>Kele Xu [58]</b>	0.2717	0.8694	0.4571	0.8086

Tabla 24. Resultados de evaluar imágenes originales normalizadas con ecualización adaptativo del histograma.

Autor	Entrenamiento		Validación	
	Pérdida	Exactitud	Pérdida	Exactitud
<b>Darshit Doshi [28]</b>	0.07032	0.9721	0.0671	0.9778
<b>Kele Xu [58]</b>	0.0780	0.9688	0.0737	0.9725

Tabla 25. Resultados de evaluar regiones de interés sin preprocesamientos.

Autor	Entrenamiento		Validación	
	Pérdida	Exactitud	Pérdida	Exactitud
<b>Darshit Doshi [28]</b>	0.1080	0.9576	0.1051	0.9556
<b>Kele Xu [58]</b>	0.0669	0.9744	0.1004	0.9630

Tabla 26. Resultados de evaluar regiones de interés con ecualización adaptativo del histograma.

Autor	Entrenamiento		Validación	
	Pérdida	Exactitud	Pérdida	Exactitud
<b>Darshit Doshi [28]</b>	0.0579	0.9790	0.0777	0.9713
<b>Kele Xu [58]</b>	0.1146	0.9548	0.0914	0.9734

Tabla 27. Resultados de evaluar regiones de interés normalizadas.

Autor	Entrenamiento		Validación	
	Pérdida	Exactitud	Pérdida	Exactitud
<b>Darshit Doshi [28]</b>	0.0764	0.9688	0.1015	0.9681
<b>Kele Xu [58]</b>	0.0508	0.9800	0.0918	0.9670

Tabla 28. Resultados de evaluar regiones de interés normalizadas con ecualización adaptativo del histograma.

## Inicialización aleatoria

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.8654	0.9410	0.7001	0.8549	0.9438	0.6493
InceptionV3	0.8673	0.9516	0.6731	0.8647	0.9461	0.6815

Tabla 29. Resultados de evaluar imágenes originales sin preprocesamiento de los conjuntos de entrenamiento y validación.

Arquitectura	Datos de reserva		
	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.8493	0.9301	0.6718
InceptionV3	0.8522	0.9395	0.6527

Tabla 30. Resultados de evaluar imágenes originales sin preprocesamiento del conjunto de reserva.

ResNet 50		InceptionV3	
83	105	80	108
49	274	43	280
Alcance	Precisión	Alcance	Precisión
0.6287	0.4414	0.6504	0.4255
F1	0.5186	F1	0.5144

Tabla 31. Matriz de confusión obtenida al evaluar imágenes originales sin preprocesamiento de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.8651	0.9727	0.5774	0.8529	0.9591	0.5804
InceptionV3	0.8844	0.9579	0.7207	0.8745	0.9517	0.7025

Tabla 32. Resultados de evaluar imágenes originales con ecualización adaptativo del histograma de los conjuntos de entrenamiento y validación.

<b>Datos de reserva</b>			
Arquitectura	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.8561	0.9634	0.5759
InceptionV3	0.8688	0.9475	0.6930

Tabla 33. Resultados de evaluar imágenes originales con ecualización adaptativo del histograma del conjunto de reserva.

<b>ResNet 50</b>		<b>InceptionV3</b>	
68	120	91	97
27	296	37	286
Alcance	Precisión	Alcance	Precisión
0.7157	0.3617	0.7109	0.4840
F1	0.4805	F1	0.5759

Tabla 34. Matriz de confusión obtenida al evaluar imágenes originales con ecualización adaptativo del histograma de los datos de reserva.

<b>Entrenamiento</b>				<b>Validación</b>		
Arquitectura	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.8807	0.9306	0.7838	0.8803	0.9283	0.7885
InceptionV3	0.9170	0.9767	0.7881	0.9078	0.9713	0.7701

Tabla 35. Resultados de evaluar imágenes originales normalizadas de los conjuntos de entrenamiento y validación.

<b>Datos de reserva</b>			
Arquitectura	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.8825	0.9266	0.7994
InceptionV3	0.8962	0.9566	0.7696

Tabla 36. Resultados de evaluar imágenes originales normalizadas del conjunto de reserva.

ResNet 50		InceptionV3	
117	71	112	76
49	274	30	293
Alcance	Precisión	Alcance	Precisión
0.7048	0.6223	0.7887	0.5957
F1	0.6609	F1	0.6787

Tabla 37. Matriz de confusión obtenida al evaluar imágenes originales normalizadas de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.8495	0.9815	0.4424	0.8568	0.9788	0.5038
InceptionV3	0.8796	0.9625	0.6859	0.8960	0.9718	0.7225

Tabla 38. Resultados de evaluar imágenes originales normalizadas con ecualización adaptativo del histograma de los conjuntos de entrenamiento y validación.

Arquitectura	Datos de reserva		
	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.8502	0.9831	0.4377
InceptionV3	0.8816	0.9636	0.6905

Tabla 39. Resultados de evaluar imágenes originales normalizadas con ecualización adaptativo del histograma del conjunto de reserva.

ResNet 50		InceptionV3	
48	140	93	95
13	310	26	297
Alcance	Precisión	Alcance	Precisión
0.7868	0.2553	0.7815	0.4946
F1	0.3855	F1	0.6057

Tabla 40. Matriz de confusión obtenida al evaluar imágenes originales normalizadas con ecualización adaptativo del histograma de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9687	0.9618	0.9755	0.9651	0.9602	0.9700
InceptionV3	0.9784	0.9815	0.97530	0.9771	0.9878	0.9659

Tabla 41. Resultados de evaluar regiones de interés sin preprocesamientos de los conjuntos de entrenamiento y validación.

Arquitectura	Datos de reserva		
	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9615	0.9536	0.9691
InceptionV3	0.9766	0.9750	0.9781

Tabla 42. Resultados de evaluar regiones de interés sin preprocesamientos del conjunto de reserva.

ResNet 50		InceptionV3	
450	30	459	21
44	438	24	458
Alcance	Precisión	Alcance	Precisión
0.9109	0.9375	0.9503	0.9562
F1	0.9240	F1	0.9532

Tabla 43. Matriz de confusión obtenida al evaluar regiones de interés sin preprocesamientos de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9759	0.9699	0.98179	0.9677	0.9678	0.9677
InceptionV3	0.9777	0.9802	0.9752	0.9724	0.9763	0.9685

Tabla 44. Resultados de evaluar regiones de interés con ecualización adaptativo del histograma de los conjuntos de entrenamiento y validación.

Arquitectura	Datos de reserva		
	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9693	0.9575	0.9806
InceptionV3	0.9771	0.9782	0.9759

Tabla 45. Resultados de evaluar regiones de interés con ecualización adaptativo del histograma del conjunto de reserva.

ResNet 50		InceptionV3	
461	19	457	23
40	442	21	461
Alcance	Precisión	Alcance	Precisión
0.9201	0.9604	0.9560	0.9520
F1	0.9398	F1	0.9539

Tabla 46. Matriz de confusión obtenida al evaluar regiones de interés con ecualización adaptativo del histograma de los datos de reserva.



Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9762	0.9777	0.9748	0.9764	0.9790	0.9738
InceptionV3	0.9806	0.9925	0.9681	0.9816	0.9917	0.9713

Tabla 47. Resultados de evaluar regiones de interés normalizadas de los conjuntos de entrenamiento y validación.

Arquitectura	Datos de reserva		
	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9749	0.9803	0.9694
Inception V3	0.9770	0.9898	0.9636

Tabla 48. Resultados de evaluar regiones de interés normalizadas del conjunto de reserva.

ResNet 50		Inception V3	
451	29	446	34
19	460	10	469
Alcance	Precisión	Alcance	Precisión
0.9595	0.9395	0.9780	0.9291
F1	0.9493	F1	0.9529

Tabla 49. Matriz de confusión obtenida al evaluar regiones de interés normalizadas de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9532	0.9926	0.9074	0.9633	0.9970	0.9252
InceptionV3	0.9668	0.9406	.9907	0.9680	0.9451	0.9889

Tabla 50. Resultados de evaluar regiones de interés normalizadas y con ecualización adaptativo del histograma de los conjuntos de entrenamiento y validación.

Datos de reserva			
Arquitectura	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9489	0.9971	0.8905
InceptionV3	0.9593	0.9244	0.9901

Tabla 51. Resultados de evaluar regiones de interés normalizadas y con ecualización adaptativo del histograma del conjunto de reserva.

ResNet 50		InceptionV3	
385	95	470	10
3	476	68	411
Alcance	Precisión	Alcance	Precisión
0.9922	0.8020	0.8736	0.9791
F1	0.8870	F1	0.9233

Tabla 52. Matriz de confusión obtenida al evaluar regiones de interés normalizadas y con ecualización adaptativo del histograma de los datos de reserva.

## Inicialización con transferencia de aprendizaje

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9701	0.9874	0.9384	0.9509	0.9726	0.9116
InceptionV3	0.9662	0.9910	0.9191	0.9372	0.9636	0.8888

Tabla 53. Resultados de evaluar imágenes originales sin preprocesamientos de los conjuntos de entrenamiento y validación.

Datos de reserva			
Arquitectura	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9275	0.9512	0.8852
InceptionV3	0.9275	0.9658	0.8534

Tabla 54. Resultados de evaluar imágenes originales sin preprocesamientos del conjunto de reserva.

ResNet 50		InceptionV3	
146	42	137	51
32	291	23	300
Alcance	Precisión	Alcance	Precisión
0.8202	0.7765	0.8562	0.7287
F1	0.7977	F1	0.7873

Tabla 55. Matriz de confusión obtenida al evaluar imágenes originales sin preprocesamiento de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9469	0.9765	0.8907	0.9352	0.9571	0.8961
InceptionV3	0.9812	0.9934	0.9593	0.9450	0.9631	0.9130

Tabla 56. Resultados de evaluar imágenes originales con ecualización adaptativo del histograma de los conjuntos de entrenamiento y validación.

Arquitectura	Datos de reserva		
	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9256	0.9595	0.8615
InceptionV3	0.9403	0.9619	0.9013

Tabla 57. Resultados de evaluar imágenes originales con ecualización adaptativo del histograma del conjunto de reserva.

ResNet 50		InceptionV3	
139	49	152	36
27	296	25	298
Alcance	Precisión	Alcance	Precisión
0.8373	0.7393	0.8587	0.8085
F1	0.7852	F1	0.8328

Tabla 58. Matriz de confusión obtenida al evaluar imágenes originales con ecualización adaptativo del histograma de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9762	0.9970	0.9373	0.9490	0.9790	0.8920
InceptionV3	0.9589	0.9738	0.9325	0.9490	0.9629	0.9247

Tabla 59. Resultados de evaluar imágenes originales normalizadas de los conjuntos de entrenamiento y validación.

Arquitectura	Datos de reserva		
	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9510	0.9867	0.8808
InceptionV3	0.9295	0.9425	0.9074

Tabla 60. Resultados de evaluar imágenes originales normalizadas del conjunto de reserva.

ResNet 50		InceptionV3	
147	41	153	35
9	314	37	286
Alcance	Precisión	Alcance	Precisión
0.9423	0.7819	0.8052	0.8138
F1	0.8546	F1	0.8094

Tabla 61. Matriz de confusión obtenida al evaluar imágenes originales normalizadas de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9709	0.9709	0.9709	0.9313	0.9404	0.9162
InceptionV3	0.9648	0.9897	0.9176	0.9607	0.9818	0.9222

Tabla 62. Resultados de evaluar imágenes originales normalizadas y con ecualización adaptativo del histograma de los conjuntos de entrenamiento y validación.

Arquitectura	Datos de reserva		
	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9236	0.9326	0.9088
InceptionV3	0.9383	0.9732	0.8710

Tabla 63. Resultados de evaluar imágenes originales normalizadas y con ecualización adaptativo del histograma del conjunto de reserva.

ResNet 50		InceptionV3	
153	35	143	45
43	280	18	305
Alcance	Precisión	Alcance	Precisión
0.7806	0.8138	0.8881	0.7606
F1	0.7968	F1	0.8194

Tabla 64. Matriz de confusión obtenida al evaluar imágenes originales normalizadas y con ecualización adaptativo del histograma de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9975	0.9990	0.9959	0.9932	0.9958	0.9905
InceptionV3	0.9994	0.9997	0.9990	0.9989	1.0	0.9979

Tabla 65. Resultados de evaluar regiones de interés sin preprocesamiento de los conjuntos de entrenamiento y validación.

**Datos de reserva**

Arquitectura	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9948	0.9989	0.9905
InceptionV3	0.9989	0.9989	0.9989

Tabla 66. Resultados de evaluar regiones de interés sin preprocesamiento del conjunto de reserva.

**ResNet 50**

**InceptionV3**

471	9	479	1
1	481	1	481
Alcance	Precisión	Alcance	Precisión
0.9978	0.9812	0.9979	0.9979
F1	0.9894	F1	0.9979

Tabla 67. Matriz de confusión obtenida al evaluar regiones de interés sin preprocesamiento de los datos de reserva.

**Entrenamiento**

**Validación**

Arquitectura	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9971	0.9958	0.9985	0.9932	0.9927	0.9937
InceptionV3	0.9998	0.9997	0.9998	0.9979	0.9979	0.9979

Tabla 68. Resultados de evaluar regiones de interés con ecualización adaptativo del histograma de los conjuntos de entrenamiento y validación.

**Datos de reserva**

Arquitectura	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9948	0.9927	0.9968
InceptionV3	0.9979	0.9979	0.9979

Tabla 69. Resultados de evaluar regiones de interés con ecualización adaptativo del histograma del conjunto de reserva.

ResNet 50		InceptionV3	
477	3	478	2
7	475	2	480
Alcance	Precisión	Alcance	Precisión
0.9855	0.9937	0.9958	0.9958
F1	0.9895	F1	0.9958

Tabla 70. Matriz de confusión obtenida al evaluar regiones de interés con ecualización adaptativo del histograma de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9967	0.9951	0.9984	0.9947	0.9915	0.9979
InceptionV3	0.9988	0.9994	0.9981	0.9973	0.9968	0.9979

Tabla 71. Resultados de evaluar regiones de interés normalizadas de los conjuntos de entrenamiento y validación.

Arquitectura	Datos de reserva		
	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9947	0.9937	0.9958
InceptionV3	0.9984	0.9989	0.9979

Tabla 72. Resultados de evaluar regiones de interés normalizadas del conjunto de reserva.

ResNet 50		InceptionV3	
476	4	478	2
6	473	1	478
Alcance	Precisión	Alcance	Precisión
0.9875	0.9916	0.9979	0.9958
F1	0.9895	F1	0.9968

Tabla 73. Matriz de confusión obtenida al evaluar regiones de interés normalizadas de los datos de reserva.

Arquitectura	Entrenamiento			Validación		
	Exactitud	Especificidad	Sensibilidad	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9941	0.9930	0.9951	0.9910	0.9850	0.9969
InceptionV3	0.9990	0.9992	0.9989	0.9984	0.9978	0.9989

Tabla 74. Resultados de evaluar regiones de interés normalizadas y con ecualización adaptativo del histograma de los conjuntos de entrenamiento y validación.

Arquitectura	Datos de reserva		
	Exactitud	Especificidad	Sensibilidad
ResNet 50	0.9921	0.9895	0.9948
InceptionV3	0.9989	0.9989	0.9989

Tabla 75. Resultados de evaluar regiones de interés normalizadas y con ecualización adaptativo del histograma del conjunto de reserva.



<b>ResNet 50</b>		<b>InceptionV3</b>	
475	5	479	1
10	469	1	478
Alcance	Precisión	Alcance	Precisión
0.9793	0.9895	0.9979	0.9979
F1	0.9843	F1	0.9979

Tabla 76. Matriz de confusión obtenida al evaluar regiones de interés normalizadas y con ecualización adaptativo del histograma de los datos de reserva.