TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE QUERÉTARO

Segmentación semántica de imágenes de melanomas mediante capa convolucional Chebyshev.

TESIS

Que para obtener el título de

Maestro en Ciencias de Datos

P R E S E N T A

Fernando García Ramírez

DIRECTOR DE TESIS

Mtro. Miguel Ángel Guzmán Rivera

CODIRECTOR DE TESIS

Dr. Efrén Gorrostieta Hurtado

ASESOR

Dr. Víctor Alejandro González Huitrón

Santiago de Querétaro, Querétaro, 25 de mayo de 2026

Educación
Secretaría de Educación Pública



Instituto Tecnológico de Querétaro
División de Estudios de Posgrado e Investigación

Santiago de Querétaro, 25/mayo/2026
Oficio No. DEPI/069/2026
Asunto: ***Autorización de impresión de tesis***

FERNANDO GARCÍA RAMÍREZ
ESTUDIANTE DE MAESTRÍA EN CIENCIA DE DATOS
PRESENTE

De acuerdo con el Reglamento para Exámenes Profesionales de la Dirección General de Educación Superior Tecnológica, se le autoriza la impresión de la Tesis, para obtener el Grado de MAESTRÍA EN CIENCIA DE DATOS, titulada:

"Segmentación semántica de imágenes de melanomas mediante capa convolucional Chebyshev"

Para el correspondiente Examen de Grado.

ATENTAMENTE
Excelencia en Educación Tecnológica.
La tierra será como sean los hombres.

EDUCACIÓN
INSTITUTO TECNOLÓGICO DE QUERÉTARO
DIVISIÓN DE POSGRADO E INVESTIGACIÓN

GABRIELA PINEDA CHACÓN
JEFA DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

ccp. Coordinación de Maestría en Ciencia de Datos
Archivo

ERA*



2026 año de Margarita Maza

Educación
Secretaría de Educación Pública



Instituto Tecnológico de Querétaro
División de Estudios de Posgrado e Investigación

Santiago de Querétaro, 25/mayo/2026
Oficio No. DEPI/068/2026
Asunto: ***Constancia de No Plagio***

A QUIEN CORRESPONDA:

Por medio de la presente se hace constar que el trabajo de tesis con título: **"SEGMENTACIÓN SEMÁNTICA DE IMÁGENES DE MELANOMAS MEDIANTE CAPA CONVOLUCIONAL CHEBYSHEV"** ha sido revisado por medio de la herramienta de software TURNITIN, cuyo resultado se anexa a la presente y **no se ha encontrado evidencias de plagio en su realización.** El autor de dicho trabajo, estudiante de **Maestría en Ciencia de Datos, Fernando García Ramírez**, es el responsable de la autenticidad y originalidad del mismo y manifiesta que para su desarrollo ha utilizado diversas citas para su soporte, mismas que han sido marcadas a lo largo del mismo y listadas al final como REFERENCIAS bibliográficas.

Se extiende la presente para la continuación del proceso de obtención del grado de Maestría en Ciencia de Datos, y a petición del interesado.

ATENTAMENTE
Excelencia en Educación Tecnológica.
La tierra será como sean los hombres.

EDUCACIÓN
INSTITUTO TECNOLÓGICO DE QUERÉTARO
DIVISIÓN DE POSGRADO E INVESTIGACIÓN

GABRIELA PINEDA CHACÓN
JEFA DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

ccp. Coordinacion de Maestría en Ciencia de Datos
Archivo

ERA*



2026 año de Margarita Maza

CONSENTIMIENTO DE DIFUSIÓN Y CESIÓN DE DERECHOS.

Santiago de Querétaro, Qro. 21 de Mayo de 2026.

El que suscribe, egresado de Maestría en Ciencia de Datos; de manera libre y voluntaria autorizo al Centro de Información del Tecnológico Nacional de México Instituto Tecnológico de Querétaro a difundir la **TESIS** de mi autoría con el título Segmentación semántica de imágenes de melanomas mediante capa convolucional Chebyshev (en lo sucesivo la "TESIS"), misma que adjunto en archivo electrónico para fines académicos, científicos y tecnológicos, mediante formato USB o digital, desde Internet, Intranet, Repositorio Institucional y Nacional y en general cualquier formato conocido o por conocer.

Lo anterior en términos de los antecedentes y consideraciones siguientes:

I. Que la presente cesión de derechos de la **TESIS** se transfiere en virtud de los estipulado en el numeral 3.9 Examen de grado, 4.5 De los productos académicos y de productividad y 6.1 De los productos académicos y su propiedad de los Lineamientos para la Operación de los Estudios de Posgrado del Tecnológico Nacional de México.

II. Manifiesto bajo propuesta de decir la verdad, que la **TESIS** es original, inédita y propia, no existiendo impedimento de ninguna naturaleza para la cesión de derechos que se está haciendo, respondiendo además por cualquier acción de reinvindicación, plagio u otra clase de reclamación que al respecto pudiera sobrevivir.

III. Que la titularidad de los derechos de autor de la **TESIS** en términos del artículo 27 de la Ley Federal de Derecho de Autor (reproducción en todas sus modalidades, transformación o adaptación, comunicación pública, distribución y en general cualquier tipo de explotación) se pueda realizar por cualquier medio conocido o por conocer, son propiedad única y exclusiva del Tecnológico Nacional de México. La cesión de derechos se realiza con carácter permanente e irrevocable.

IV. Que no he otorgado ni otorgaré por ningún motivo, consentimiento alguno para la utilización de la **TESIS** por cualquier medio y su materialización en productos o servicios de cualquier naturaleza a favor de ninguna persona física o moral distinta al Tecnológico Nacional de México.

V. Estoy enterada(o) de la presente cesión de derechos, firmando al final del presente documento para su certificación

Dicha **TESIS** estará disponible al estudiantado de esta Institución a partir del Mayo, fecha en la cual se puede difundir la **TESIS**.

Estudiante: Fernando García Ramírez

No. de Control: M23140011 Correo electrónico: M23140011@queretaro.tecnm.mx Carrera: Maestría en Ciencia de Datos

CONSENTIMIENTO DE DIFUSIÓN Y CESIÓN DE DERECHOS.

Titulo de la TESIS: Segmentación semántica de imágenes de melanomas mediante capa convolucional Chebyshev

Área del conocimiento: Ciencia de Datos

Palabras clave de la obra: Segmentación semántica, melanoma, redes neuronales convolucionales, polinomios de Chebyshev Procesamiento de imágenes médicas.



FERNANDO GARCÍA RAMÍREZ

DEDICATORIA

Dedico este trabajo de tesis a mi madre, Ma. del Carmen Ramírez, a mis hermanos, María Clara García Ramírez y Daniel Julián Gárcía Ramírez, así como a mis sobrinas Clara Elizabeth Torres García y Martha Regina Torres García, por su apoyo inquebrantable a lo largo de mi trayectoria académica. Su aliento constante y enseñanzas me han inspirado a perseguir mis sueños y a no rendirme ante los desafíos y por su presencia y amor incondicional, que siempre me han brindado fuerza. A toda mi familia, cuya confianza y respaldo han sido fundamentales en cada paso que he dado.

A mis compañeros del ITQ, quienes han sido un pilar fundamental en este viaje, su apoyo incondicional, sus valiosas ideas y su camaradería han hecho que cada desafío sea más llevadero. Gracias por compartir su tiempo, su sabiduría y su motivación.

Y un agradecimiento especial a mis amigos, quienes han estado a mi lado en los momentos más difíciles, brindándome su apoyo y motivación. Su amistad ha sido un refugio y un impulso en este camino.

¡Gracias a todos por ser parte de mi vida!

AGRADECIMIENTOS

Agradezco al Tecnológico Nacional de México Campus Querétaro por otorgarme las condonaciones económicas y apoyo administrativo para que pudiera realizar mis estudios, y por permitirme estudiar un posgrado de calidad y proporcionarme el espacio para cursar la maestría y realizar mi trabajo de investigación.

A mi asesor de tesis el Mtro. Miguel Ángel Guzmán Rivera por sus recomendaciones y apoyo para llevar a cabo este trabajo de tesis. También, al resto de mis profesores Dr. Víctor Alejandro González Huitrón y Dr. Efren Gorrostieta Hurtado por compartir sus conocimientos y su experiencia.

Al departamento del posgrado por todo su apoyo en el proceso administrativo.

DECLARATORIA DE AUTENTICIDAD

Por la presente declaro que, salvo cuando se haga referencia específica al trabajo de otras personas, el contenido de esta tesis es original y no se ha presentado total o parcialmente para su consideración para cualquier otro título o grado en esta o cualquier otra Universidad. Esta tesis es resultado de mi propio trabajo y no incluye nada que sea resultado de algún trabajo realizado en colaboración, salvo que se indique específicamente en el texto.

Fernando García Ramírez

Querétaro, Querétaro, México, 2026

RESUMEN

El cáncer de piel es el segundo más frecuente en México, donde el melanoma maligno presenta una mortalidad del 30% por detección tardía. La escasez de especialistas y la complejidad en la diferenciación visual de lesiones dificultan el diagnóstico oportuno. La segmentación semántica precisa de imágenes dermatoscópicas es determinante para evaluar criterios clínicos (asimetría, bordes, color) que condicionan el pronóstico del paciente.

La presente investigación desarrolla un modelo de segmentación semántica basado en la arquitectura U-Net, integrando funciones de preprocesamiento fundamentadas en polinomios de Chebyshev en los bloques convolucionales. Los polinomios de Chebyshev poseen propiedades de ortogonalidad y minimización del error máximo (propiedad minimax) que facilitan la captura de variaciones sutiles en gradientes de bordes y texturas complejas. La hipótesis plantea que esta integración incrementará la estabilidad en la generalización del modelo, medida mediante el coeficiente de Intersección sobre Unión (IoU) y el coeficiente de Dice.

La metodología comprende: (1) preprocesamiento del conjunto ISIC 2017 con 2,600 imágenes dermatoscópicas reescaladas a 512×512 píxeles; (2) implementación de la arquitectura Xception-Cheb incorporando polinomios de Chebyshev de grado K=3 con normalización al intervalo [-1,1]; (3) entrenamiento supervisado mediante optimizador Adam (500 épocas, lote de 8, tasa de aprendizaje 0.001); y (4) evaluación comparativa frente a arquitecturas de referencia (Xception, EfficientNet-b3, ResNet-18, MobileOne-s0).

Los resultados experimentales evidencian que Xception-Cheb alcanzó un IoU de 0.7836 en prueba y 0.8777 en entrenamiento, con una diferencia de 0.0941 que representa la menor brecha observada entre las arquitecturas evaluadas, inferior a ResNet-18 (0.1126) y MobileOne-s0 (0.1048). Aunque el IoU absoluto resulta moderado frente a EfficientNet-b3 (0.9137), la estabilidad en la generalización demostrada constituye una ventaja estratégica en contextos clínicos donde la robustez y reproducibilidad son determinantes.

Las conclusiones establecen que la incorporación de polinomios de Chebyshev como función de preprocesamiento mejora la estabilidad en la generalización y reduce el sobreajuste mediante una base funcional matemáticamente rigurosa. La arquitectura propuesta ofrece un equilibrio favorable entre precisión, estabilidad y fundamentación teórica, justificando su aplicación en sistemas de apoyo al diagnóstico clínico. Se identifican como líneas futuras la optimización mediante regularización avanzada, validación en bases independientes e integración con sistemas de diagnóstico asistido por computadora. La contribución principal radica en demostrar empíricamente que bases funcionales ortogonales pueden mejorar la estabilidad de arquitecturas convolucionales profundas, estableciendo un marco metodológico reproducible para segmentación biomédica que contribuya a reducir la mortalidad por melanoma en el sistema de salud mexicano.

PALABRAS CLAVES

Cáncer de piel

Coeficiente de Dice

Imágenes dermatoscópicas

IoU (Intersección sobre Unión)

ISIC (International Skin Imaging Collaboration)

Lesiones cutáneas

Melanoma

Ortogonalidad

Polinomios de Chebyshev

Propiedad minimax

Redes neuronales

Segmentación semántica.

Sesgo y varianza

Sobreajuste (overfitting)

U-Net

INDICE GENERAL

Capítulo 1. Introducción. 01
1.1. Fundamentos. 02
1.2. Descripción del problema. 03
1.3. Justificación. 04
1.4. Hipótesis. 05
1.5. Objetivos. 05
1.5.1. Objetivo general. 05
1.5.2. Objetivos particulares. 05

Capítulo 2. Estado del arte. 07
2.1. Antecedentes. 07
2.1.1 Cáncer en la piel. 07
2.1.2 Regla ABCD. 08
2.2. Modelos de aprendizaje automático en la detección y segmentación semántica de cáncer de piel. 10

Capítulo 3. Marco teórico. 16
3.1. Aprendizaje profundo y la imagenología médica. 16
3.1.1. Aprendizaje supervisado. 16
3.1.2. La imagenología en las redes neuronales artificiales. 19
3.2. Redes neuronales. 21
3.2.1. Perceptrón multicapa. 21
3.2.2. Redes neuronales convolucionales. 24
3.2.3. Redes neuronales recurrentes. 28
3.3. Arquitecturas de redes neuronales convolucionales. 35
3.3.1. Xception. 36
3.3.2. EfficientNet. 38
3.3.3. ResNet. 39
3.3.4. MobileOne. 41
3.4. Modelos o arquitecturas para segmentación semántica. 43
3.4.1. U-Net. 43
3.4.2. Redes piramidales de características FPN. 45
3.4.3. Red de análisis de escenas piramidal PSPNet. 47
3.4.4. SegFormer. 49
3.5. Regularización. 50
3.5.1. Compensación entre sesgo y varianza. 51
3.5.2. Tipos de regularización con modelos lineales. 53
3.6. Medidas de calidad. 55
3.6.1. Exactitud. 56
3.6.2. $\boldsymbol{F_1}$-Score. 57
3.6.3. Intersección sobre Unión IoU. 57
3.7. Polinomios de Chebyshev. 58

3.7.1. Definición. ……... 59
3.7.2. Propiedades. …… 61

Capítulo 4. Implementación del modelo. …….. 64
4.1. Metodología. …….. 64
4.2. Adquisición del conjunto de datos. …… 65
4.3. Análisis y selección del modelo de segmentación semántica. ……... 65
4.3.1. Arquitectura empleada. ……... 66
4.3.1.1. Entrada U-Net. ……. 67
4.3.1.2. Parte contractiva (codificador). …… 68
4.3.1.3. Cuello de botella (Bottleneck). …… 73
4.3.1.4. Parte expansiva (decodificador). …….. 74
4.3.1.5. Salida U-Net. …… 78
4.4. Discusión de resultados. …… 79
4.4.1. Entrenamiento. …… 79
4.4.2. Evaluación y predicción. ……. 81
4.5. Métricas del modelo. …….. 81
4.6. Análisis de los resultados de cada modelo. …… 84

Capítulo 5. Resultados. ……... 87
5.1. Conclusiones. ……. 87
5.2. Trabajos futuros. …… 89
5.2.1. Optimización del modelo y exploración de bases funcionales ortogonales alternativas. ……..... 89
5.2.2. Estrategias avanzadas de regularización y aumento de datos. … 90
5.2.3. Integración con sistemas de apoyo al diagnóstico clínico (CAD). ……... 90
5.2.4. Exploración de arquitecturas emergentes y técnicas de aprendizaje por transferencia. ……..... 90

Referencias. ……. 89

Anexos. ……. 01
A. Articulo CITCA. …….. 01
B. Articulo Eureka ENC…… 09
C. Trabajo en el Priemer Foro Nacional de la Agenda Estratégica de la Salud del Tecnológico Nacional de México…… 14

INDICE DE FIGURAS

Figura 2.1. Tipos de cáncer en población mexicana. ……..... 07
Figura 2.2. Imágenes cutáneas con características para la regla ABCD. …….. 09
Figura 2.3. Estructura general por capas del método propuesto. Arquitectura de DilatedSkinNet. …… 10
Figura 2.4. Marco y diagrama de flujo del sistema propuesto. …… 11
Figura 2.5. Arquitectura propuesta (E-CNN) de detección y clasificación de lesiones cutáneas. ……... 13
Figura 3.1. Gráfica de una red neuronal con tres capas. ……... 18
Figura 3.2. El espectro electromagnético. …… 19
Figura 3.3. Representación esquemática de una neurona artificial. ……. 22
Figura 3.4. Representación esquemática de un perceptrón multicapa. ……. 23
Figura 3.5. Representación esquemática de una red neuronal convolucional. ……. 28
Figura 3.6. Estructura de la red neuronal recurrente simple de Elman. ……..... 29
Figura 3.7. Gráfico computacional utilizado para calcular la pérdida de entrenamiento en una red neuronal recurrente. …… 31
Figura 3.8. Red neuronal recurrente. ……..... 32
Figura 3.9. Red neuronal recurrente desplegada en el tiempo. ……. 33
Figura 3.10. Representación de la arquitectura Xception. ……..... 37
Figura 3.11. Representación de la arquitectura EfficientNet. ……... 39
Figura 3.12. Representación de la arquitectura ResNet. ……... 40
Figura 3.13. Representación del bloque MobileOne. ……... 42
Figura. 3.14. Arquitectura U-net. …….. 44
Figura 3.15. Arquitectura de redes piramidales de características. …….. 45
Figura 3.16. Esquema de arquitectura de red de escenas piramidal PSPNet. ……... 47
Figura 3.17. Arquitectura SegFormer. …….. 49
Figura 3.18. Grafica del punto de equilibrio de sesgo – varianza. ……... 52
Figura 3.19. Representación gráfica de la función de Chebyshev. ……... 61
Figura. 4.1. Metodología de desarrollo del proyecto. ……... 64
Figura. 4.2. Segmentación semántica. ……... 66
Figura. 4.3. Arquitectura U-net. ……. 67
Figura. 4.4. Entrada a la arquitectura U-Net. …… 67
Figura. 4.5. Convolución. ……. 70
Figura. 4.6. Función ReLU. …….. 71
Figura. 4.7. Muestreo máximo. ……. 71
Figura. 4.8. Conexión de salto. …… 72
Figura. 4.9. Cuello de botella. …….. 73
Figura. 4.10. Parte expansiva de la U-Net. ……... 74
Figura. 4.11. Convolución transpuesta con un kernel de 2×2. ……... 76
Figura. 4.12. Conexión de salto (líneas grises). …… 77
Figura. 4.13. Obtención de una máscara de segmentación semántica. ……. 78
Figura. 4.14. Comparativa visual de IoU en entrenamiento, prueba y brecha de generalización por arquitectura. ……... 82

Figura. 4.15. Imagen representativa de la obtención de una máscara de segmentación semántica. …….. 84
Figura. 4.16. Relación entre el IoU obtenido en el conjunto de prueba y la estabilidad de generalización. ……... 85

ÍNDICE DE TABLAS

Tabla 01. Resumen comparativo de diferentes métricas. 14
Tabla 02. Imágenes de prueba y entrenamiento SIIM-ISIC. 65
Tabla 03. Hiperparámetros de entrenamiento. 80
Tabla 04. Métricas del modelo. 82
Tabla 05. Métricas de eficiencia computacional y rendimiento por parámetro. ……... 83

Capítulo 1. Introducción.

El cáncer de piel constituye un desafío creciente para la salud pública a escala global, situación que también se observa en México, donde en la última década se ha documentado un aumento sostenido en la incidencia de esta enfermedad (Secretaría de Salud México, 2024). Este incremento epidemiológico ha llevado a que el cáncer cutáneo se ubique actualmente como el segundo tipo de cáncer más frecuente en la población mexicana, lo que obliga a reconsiderar las prioridades en prevención, diagnóstico y atención sanitaria. Entre las diversas entidades clínicas que conforman el espectro del cáncer de piel, el melanoma maligno merece especial atención por su carácter agresivo y su elevada capacidad metastásica; su pronóstico depende de manera crítica del momento del diagnóstico, de modo que la detección y el tratamiento en fases tempranas resultan determinantes para reducir la mortalidad asociada (Gaceta de la Facultad de Medicina UNAM, 2025). En consecuencia, la creciente carga de enfermedad impone la necesidad de fortalecer estrategias de vigilancia epidemiológica, programas de educación en foto protección y acceso oportuno a servicios dermatológicos, así como de promover investigaciones que permitan caracterizar factores de riesgo específicos en el contexto mexicano.

A pesar de la gravedad intrínseca del melanoma, en México el diagnóstico con frecuencia se realiza en planos avanzados (Instituto Nacional de Cancerología INCan, 2023). Diversos factores entre ellos el cambio climático, el incremento de la radiación ultravioleta y la insuficiente cultura de prevención, han contribuido al aumento de los casos anuales. Según la Secretaría de Salud (2024), cuando la detección es tardía este padecimiento presenta una tasa de mortalidad cercana al 30%. La principal barrera para un diagnóstico oportuno radica en la insuficiencia de especialistas dermatólogos frente a la demanda poblacional y en la complejidad inherente a la diferenciación visual entre un melanoma y un nevo benigno. En este contexto, la precisión en la delimitación de la lesión —proceso conocido como segmentación— resulta esencial para el análisis subsecuente de la asimetría, los bordes y el color, criterios ABCDE que condicionan de manera determinante el pronóstico del paciente.

Ante la necesidad de herramientas de apoyo clínico más rápidas y precisas, el Aprendizaje Automático (Machine Learning) y, en particular, el Aprendizaje Profundo (Deep Learning) han emergido como soluciones inivadoras en el ámbito de la dermatología. Las técnicas de segmentación semántica basadas en redes neuronales convolucionales posibilitan el procesamiento de imágenes dermatoscópicas con un nivel de detalle que, en numerosos casos, iguala o complementa la capacidad perceptiva del observador clínico. Esta capacidad técnica plantea una reconfiguración de los flujos diagnósticos al ofrecer información cuantitativa y reproducible sobre características morfológicas de las lesiones cutáneas (Salinas, M. P., 2024).

La integración de algoritmos de visión artificial no pretende suplantar la labor del médico, sino proporcionarle una segunda opinión matemática que contribuya a disminuir la tasa de falsos negativos y a optimizar la priorización de casos en sistemas de salud con recursos limitados. Desde una perspectiva argumentativa, este enfoque refuerza la complementariedad entre juicio clínico y herramientas computacionales: el algoritmo aporta consistencia y velocidad en el análisis de grandes volúmenes de imágenes, mientras que el clínico conserva la responsabilidad del diagnóstico final y de las decisiones terapéuticas.

Técnicamente, el empleo de capas convolucionales permite que los modelos aprendan de forma automática a reconocer texturas y patrones morfológicos complejos presentes en las imágenes dermatoscópicas. La segmentación precisa del área afectada facilita la extracción de métricas críticas como asimetría, irregularidad de bordes, variabilidad cromática y otras características cuantificables que son determinantes para evaluar la necesidad de una biopsia urgente. En consecuencia, la combinación de segmentación avanzada y análisis cuantitativo constituye una herramienta prometedora para mejorar la sensibilidad diagnóstica y reducir retrasos en la atención, siempre que su implementación vaya acompañada de validación clínica rigurosa y de protocolos que garanticen su uso ético y seguro.

1.1 Fundamentos.

El melanoma constituye un tumor maligno derivado de los melanocitos, cuya relevancia clínica radica en su potencial metastásico y en la marcada dependencia del pronóstico respecto al momento del diagnóstico. La detección temprana se apoya fundamentalmente en la dermatoscopia, técnica de imagen que elimina la reflexión de la capa córnea y permite la visualización de estructuras cutáneas más profundas y sutiles que escapan al examen clínico directo. La interpretación de las imágenes dermatoscópicas se fundamenta, de manera tradicional y aún vigente, en criterios morfológicos entre los que destacan la asimetría y la irregularidad de los bordes; estos parámetros constituyen indicadores diagnósticos que, combinados con la evaluación de color y patrón, orientan la decisión clínica sobre la necesidad de biopsia y tratamiento. En consecuencia, la dermatoscopia no solo mejora la sensibilidad diagnóstica frente a lesiones pigmentadas, sino que también articula un marco objetivo para la estratificación del riesgo y la priorización de intervenciones en contextos asistenciales.

El procesamiento digital de imágenes se entiende como la aplicación sistemática de algoritmos sobre una matriz de datos para extraer información relevante y cuantificable. En este ámbito, la segmentación semántica constituye la tarea de asignar a cada píxel de una imagen una categoría específica, transformando representaciones visuales en mapas interpretables que permiten análisis posteriores.

El presente proyecto se orienta a la segmentación semántica de imágenes dermatoscópicas por medio de la arquitectura U-Net. Para abordar esta problemática se integran conceptos de aproximación funcional mediante polinomios de Chebyshev, cuya propiedad de minimizar

el error máximo de aproximación resulta especialmente útil para modelar variaciones de intensidad y texturas complejas propias del melanoma. La incorporación de dichos polinomios facilita que la arquitectura de la red capture cambios sutiles en el gradiente de los bordes, aspectos que con frecuencia escapan a los métodos convencionales. En consecuencia, la combinación de segmentación semántica y aproximación por polinomios de Chebyshev constituye una estrategia metodológica prometedora para mejorar la precisión en la delimitación de lesiones y, por ende, la fiabilidad de las métricas diagnósticas derivadas.

1.2 Descripción del problema.

A nivel global, el melanoma constituye aproximadamente el 1% de los casos de cáncer cutáneo, si bien es responsable de la mayor parte de las muertes asociadas a esta patología (American Cancer Society, 2025). En el contexto mexicano, la situación reviste especial gravedad, dado que con frecuencia el diagnóstico se realiza en estudios avanzados, circunstancia que incrementa la tasa de mortalidad hasta alrededor del 30% (Secretaría de Salud, 2024). Esta realidad epidemiológica se ve potenciada por determinantes geográficos y sociales propios del país —entre ellos, elevados niveles de radiación solar y el acceso limitado a especialistas en dermatología en áreas rurales y suburbanas—, factores que dificultan la detección temprana y la atención oportuna de los pacientes (Gaceta de la Facultad de Medicina, 2023). En consecuencia, la combinación de una alta letalidad relativa del melanoma y las barreras estructurales del sistema de salud subraya la necesidad de estrategias integradas de prevención, diagnóstico y distribución de recursos clínicos.

A pesar de los avances en visión artificial, los métodos tradicionales de procesamiento de imágenes muestran limitaciones significativas al enfrentarse a la amplia variabilidad de las lesiones cutáneas, que se manifiesta en bordes difusos, presencia de vello, diversidad de foto tipos y fluctuaciones en la iluminación de las tomas dermatoscópicas. Cuando la segmentación inicial resulta imprecisa, se compromete la validez de todos los análisis subsecuentes —por ejemplo, la evaluación de la asimetría y la caracterización de los bordes—, lo que puede traducirse en **falsos negativos** (cáncer no detectado) o en **biopsias innecesarias**. En consecuencia, la robustez y la exactitud de los algoritmos de segmentación constituyen un requisito metodológico indispensable para garantizar decisiones clínicas fiables y minimizar tanto el subdiagnóstico como la intervención.

Ante este escenario, la integración de funciones ortogonales se propone como una solución robusta para la aproximación de lesiones en las fronteras complejas en imágenes dermatoscópicas. Las funciones ortogonales, y en particular los polinomios de Chebyshev, ofrecen propiedades numéricas favorables —minimización del error máximo, buena condición numérica y capacidad para representar variaciones de alta frecuencia— que pueden complementar la extracción jerárquica de características realizada por capas convolucionales. Al incorporar estas bases en la arquitectura y el proceso de optimización, se espera mejorar la regularización del modelo y reducir la propensión a errores sistemáticos derivados de ruido, variaciones de iluminación o artefactos como el vello.

El problema central de esta investigación radica en: ¿Cómo mejorar la regularización en la segmentación semántica en imágenes de melanomas mediante el uso de capas convolucionales optimizadas con polinomios de Chebyshev, con el fin de facilitar una detección temprana, objetiva y matemáticamente eficiente en el contexto de la salud pública?

1.3 Justificación.

La presente investigación se fundamenta en la necesidad imperante de optimizar los procesos de diagnóstico oncológico en México a través de herramientas tecnológicas de alta precisión. La relevancia de este trabajo se articula bajo tres ejes principales:

- **Justificación Social y de Salud Pública.** En México, el melanoma maligno representa un reto crítico debido a que el 30% de los casos se detectan en etapas donde la supervivencia es reducida (**Secretaría de Salud, 2024**). La implementación de un sistema de segmentación automatizado permitirá universalizar el acceso a diagnósticos preliminares precisos, especialmente en unidades de salud de primer nivel que carecen de dermatólogos especialistas. Al facilitar una detección temprana, este proyecto contribuye directamente a la reducción de la mortalidad y de los costos asociados a tratamientos oncológicos avanzados.
- **Justificación Técnica y Científica.** Desde una perspectiva técnica, las Redes Neuronales Convolucionales (CNN) convencionales presentan limitaciones al delimitar bordes de lesiones con bajo contraste o texturas irregulares. La innovación de esta tesis radica en la optimización de las capas convolucionales mediante **polinomios de Chebyshev**. A diferencia de los filtros estándar, los polinomios de Chebyshev ofrecen una propiedad de **aproximación uniforme**, lo que permite capturar los gradientes de la imagen con un error mínimo distribuido (propiedad minimax). Esto resulta en una segmentación semántica más robusta y matemáticamente superior, capaz de modelar la morfología del melanoma con una fidelidad que las arquitecturas genéricas no alcanzan.
- **Justificación Metodológica.** Este estudio propone un puente entre el análisis numérico clásico y el aprendizaje profundo contemporáneo. El uso de funciones ortogonales para la extracción de características no solo mejora la precisión del modelo (IoU y Coeficiente Dice), sino que también aporta a la interpretabilidad del modelo, un factor clave en la medicina computacional. Proporcionar un sistema con sustento matemático riguroso aumenta la confiabilidad de la herramienta ante la comunidad médica.

1.4 Hipótesis.

La integración de polinomios de Chebyshev como función de preprocesamiento dentro de la arquitectura U-Net para la segmentación semántica permitirá obtener una delimitación de las lesiones de melanoma con métricas de rendimiento competitivas, medidas mediante el coeficiente de similitud de Dice (DSC) y la Intersección sobre Unión (IoU), logrando una mayor estabilidad en la generalización del modelo, facilitando así una herramienta de diagnóstico más objetiva para el entorno clínico.

1.5 Objetivos.

En esta sección se presenta el objetivo general y los objetivos específicos que se definieron para el desarrollo de este trabajo de investigación.

1.5.1 Objetivo general.

Desarrollar un modelo de segmentación semántica para imágenes dermatoscópicas de melanomas mediante el diseño de una arquitectura de red neuronal tipo U-Net que integre funciones de preprocesamiento basadas en polinomios de Chebyshev en los bloques convolucionales, con el fin de incrementar la estabilidad en la generalización, para mejorar el apoyo diagnóstico en el contexto de la salud.

1.5.2 Objetivos particulares.

Para llevar a cabo el objetivo general de esta investigación, se definieron 5 objetivos específicos que se listan a continuación:

1. **Analizar y preprocesar** el conjunto de datos de imágenes dermatoscópicas utilizando bases de datos como ISIC 2017, aplicando técnicas de normalización y eliminación de artefactos para asegurar la calidad de los datos de entrenamiento.
2. **Diseñar e implementar** la formulación matemática de los polinomios de Chebyshev dentro de los bloques convolucionales para facilitar la extracción de características espectrales y de textura en la lesión.
3. **Entrenar y validar** la arquitectura de segmentación propuesta, comparando su desempeño contra modelos de convolución estándar (como U-Net tradicional) bajo las mismas condiciones de entrenamiento.
4. **Evaluar el rendimiento** del modelo utilizando métricas estadísticas de visión artificial, como el coeficiente de similitud de **Dice**, el **IoU** (Intersection over Union) y la regularización, para verificar el cumplimiento de la hipótesis planteada.

5. **Documentar y analizar** la relevancia de los resultados obtenidos en función de su aplicabilidad en el entorno clínico, considerando la variabilidad de imágenes dermatoscópicas presentes en la población.

Capítulo 2. Estado del arte.

2.1 Antecedentes.

El capítulo de antecedentes expone los conceptos y fundamentos teóricos que sustentan el desarrollo de la presente investigación. Se mencionan los principales tipos de cáncer de piel, así como los métodos de evaluación empleados por especialistas en dermatología. Además, se describen los modelos existentes en el campo de la detección y clasificación de lesiones cutáneas mediante inteligencia artificial. Finalmente, se abordan nociones clave relacionadas con el aprendizaje automático, incluyendo sus distintas modalidades, con el propósito de contextualizar el enfoque metodológico adoptado.

2.1.1 Cáncer de piel.

Uno de los tipos de cáncer más frecuentes en la población mexicana es el cáncer de piel, el cual se clasifica en cuatro variantes principales: carcinoma basocelular, carcinoma epidermoide, carcinoma de anexos y melanoma, siendo este último uno de los más letales. La Figura 2.1 presenta ejemplos visuales representativos de cada uno de estos tipos de cáncer cutáneo.



Figura 2.1. Tipos de cáncer en población mexicana.

En México, la mortalidad por cáncer ha registrado un aumento significativo, con un incremento del 78% en los últimos 19 años (Carlos-Ortega, B., 2024). El cáncer de piel, particularmente el melanoma, presenta una tendencia creciente en su incidencia, atribuida a factores como la exposición excesiva a radiación ultravioleta y modificaciones en los patrones de comportamiento social. La detección temprana de estos tipos de cáncer resulta fundamental para mejorar las tasas de supervivencia; sin embargo, el diagnóstico enfrenta limitaciones derivadas de la variabilidad en la interpretación de imágenes dermatoscópicas y la escasez de especialistas en diversas regiones del país.

En la actualidad, se dispone de pocos enfoques que, de manera simultánea, sean no invasivos y económicamente viables para contribuir de forma generalizada al diagnóstico y caracterización objetiva de las lesiones cutáneas. Esta limitación constituye un desafío relevante en el ámbito clínico, especialmente en contextos con recursos limitados o acceso restringido a tecnologías especializadas. Por ello, se vuelve imperativo promover el desarrollo de metodologías accesibles y precisas que mejoren la evaluación dermatológica, optimicen los procesos de atención médica y disminuyan la dependencia de procedimientos costosos o invasivos.

Ante la presencia de casos sospechosos, los especialistas en dermatología suelen realizar exámenes visuales de la piel mediante técnicas de epiluminiscencia. El cáncer de piel comprende un conjunto de enfermedades caracterizadas por la proliferación descontrolada de células cutáneas anormales, lo que conduce a la formación de tumores. La mayoría de estos tumores se originan por la exposición prolongada a radiación ultravioleta (UV) y el daño dérmico sin protección, factores que pueden alterar el ADN celular y desencadenar mutaciones responsables de la transformación maligna de los tejidos. A pesar de los avances en las técnicas diagnósticas, la detección temprana continúa siendo un desafío, especialmente en etapas iniciales donde los signos clínicos son sutiles o fácilmente confundibles con lesiones benignas. En este escenario, el desarrollo de herramientas tecnológicas basadas en inteligencia artificial y procesamiento de imágenes representa una alternativa prometedora para incrementar la precisión diagnóstica, disminuir la subjetividad en la evaluación clínica y facilitar el acceso a métodos de detección eficaces en contextos con recursos limitados [].

En México, la frecuencia del cáncer de piel ha experimentado un incremento significativo, con un registro anual de entre 11,000 y 16,000 nuevos casos. Particularmente, el melanoma representa una preocupación crítica, al ser responsable del 80% de las muertes asociadas a esta enfermedad cutánea (Gallegos, 2024).

2.1.2 Regla ABCD

Los algoritmos de detección médica constituyen una de las primeras herramientas empleadas para determinar si una lesión cutánea presenta características malignas o benignas (Nachbar, 1994; Zalaudek, 2006; Henning, 2007). En 1994, Nachbar desarrolló un método subjetivo basado en la percepción visual de la lesión, conocido como la regla ABCD, la cual considera criterios como asimetría, bordes, color, forma y diámetro. Por su simplicidad y aplicabilidad, este algoritmo se ha convertido en uno de los más utilizados para la evaluación inicial de lesiones cutáneas, ya sea mediante inspección directa o con el uso de un dermatoscopio.

El algoritmo médico ABCD se compone de cuatro criterios fundamentales para la evaluación de lesiones cutáneas:

- **Asimetría (A):** La lesión se divide en dos ejes perpendiculares entre sí, formando ángulos de 90°, con el objetivo de obtener la menor puntuación de asimetría posible. Se determina si la lesión es simétrica, y por cada eje en el que se detecta asimetría, se asigna un punto.
- **Bordes (B):** La lesión se segmenta mediante ocho cortes radiales que permiten identificar la presencia de bordes abruptos. Cada segmento que presenta un borde irregular recibe un punto adicional.
- **Colores (C):** Las lesiones pueden contener uno o varios de los siguientes colores: blanco, parduzco, marrón oscuro, negro, azul y rojo. Estos colores son producto de la vascularización y la concentración de melanina. Se suma un punto por cada color identificado.
- **Estructuras dermatoscópicas (D):** Se evalúa la presencia de estructuras como puntos, manchas, redes pigmentadas y zonas no estructuradas. Por cada tipo de estructura observada, se añade un punto.

Este sistema permite una clasificación inicial de las lesiones cutáneas, facilitando la identificación de aquellas que requieren atención especializada.

Las características previamente descritas se ponderan de la siguiente manera:

$$TDS = (1.3 \cdot A) + (0.1 \cdot B) + (0.5 \cdot C) + (0.5 \cdot D), \tag{2.1}$$

la Puntuación Dermatoscópica Total (TDS) permite clasificar las lesiones cutáneas según su grado de sospecha. Si el valor es inferior a 4.75, la lesión se considera benigna; si se encuentra entre 4.75 y 5.45, se clasifica como sospechosa; y si supera los 5.45 puntos, se interpreta como maligna. Los valores de ponderación utilizados en la regla ABCD han sido establecidos de manera empírica mediante estudios clínicos y análisis estadísticos (Nachbar, 1994). La Figura 2.2 presenta un ejemplo de imágenes en formato RGB en las que se aplican los criterios de dicha regla.



Figura 2.2. Imágenes cutáneas con características para la regla ABCD.

2.2. Modelos de aprendizaje automático en la detección y segmentación semántica de cáncer de piel.

(Kaur, R., 2022) aplicaron una red neuronal convolucional (CNN) para la segmentación de lesiones en imágenes dermatoscópicas de cáncer de piel, utilizando conjuntos de datos independientes correspondientes a los años 2016, 2017 y 2018 del repositorio ISIC (International Skin Imaging Collaboration). El modelo propuesto, basado en convoluciones dilatadas (atrous), funciones de activación leakyReLU y softmax, cuya arquitectura general se presenta en la Figura 2.3, fue entrenado con 4,446 imágenes, validado con 520 y evaluado con 1,525, alcanzando métricas destacadas de rendimiento como precisión (ACC), índice de Jaccard (JAC) y coeficiente de Dice (DICE), además de requerir menor tiempo de entrenamiento.

Para evaluar la capacidad de generalización del modelo, se entrenó la red exclusivamente con el conjunto ISIC 2018 y se probó en los conjuntos PH2, ISIC 2016 e ISIC 2017. Los resultados mostraron un rendimiento superior, con un margen de mejora de aproximadamente ±1% en comparación con las redes entrenadas individualmente en cada conjunto. Por ejemplo, la precisión (ACC) aumentó de 94.0% a 95.0% en ISIC 2016 y de 87.9% a 88.8% en ISIC 2017.

La red logró extraer con éxito las características relevantes de diversas imágenes dermatoscópicas de cáncer de piel, generando mapas segmentados de alta precisión. Los resultados experimentales evidenciaron que el modelo propuesto segmentó de manera eficaz las áreas correspondientes a las lesiones, lo cual representa un avance significativo para el desarrollo futuro de sistemas CAD altamente eficientes orientados a la clasificación de melanoma y no melanoma.

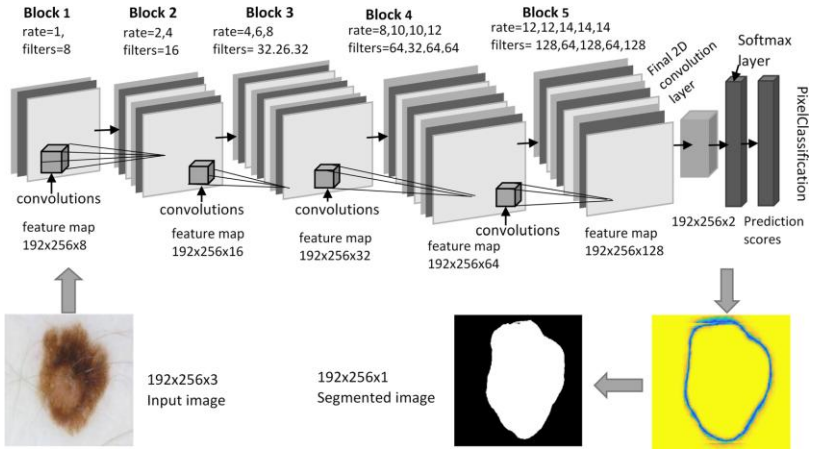


Figura 2.3. Estructura general por capas del método propuesto. Arquitectura de DilatedSkinNet.

(Shihab, A., 2021) aplicaron el clúster K-vecinos para la segmentación de imágenes de tumores cutáneos y una red neuronal convolucional (CNN) para la clasificación del tumor como benigno o maligno. Usan un total de 22,080 imágenes médicas basadas en el conjunto de datos digitales DermQuest, la cual dividen en 80% de imágenes para entrenamiento y 20% para prueba.

Ahmed Shihab et al. [02] implementaron una técnica de segmentación automática de imágenes dermatoscópicas de tumores cutáneos utilizando el algoritmo de clúster K-vecinos, complementado con una red neuronal convolucional (CNN) para la clasificación de las lesiones como benignas o malignas.

El estudio se basó en un conjunto de 22,080 imágenes médicas provenientes de la base de datos digital DermQuest, dividido en un 80% para entrenamiento y un 20% para prueba. El modelo propuesto permitió delimitar con precisión los bordes de las lesiones cutáneas y fue evaluado mediante métricas como sensibilidad, especificidad y precisión. Los resultados obtenidos evidenciaron una alta exactitud diagnóstica, alcanzando un 99.7% en la identificación de cáncer de piel.

(Adegun, A. A., 2020) propusieron un enfoque basado en aprendizaje profundo para la detección y segmentación automática de lesiones de melanoma. El método implementado se fundamenta en una arquitectura codificadora-decodificadora, cuya arquitectura general se presenta en la Figura 2.4, en la cual las subredes de codificación y decodificación están interconectadas mediante una serie de vías de salto. Estas conexiones permiten alinear los niveles semánticos entre los mapas de características generados por el codificador y los del decodificador, lo que favorece un aprendizaje más eficiente y una extracción de características más precisa.

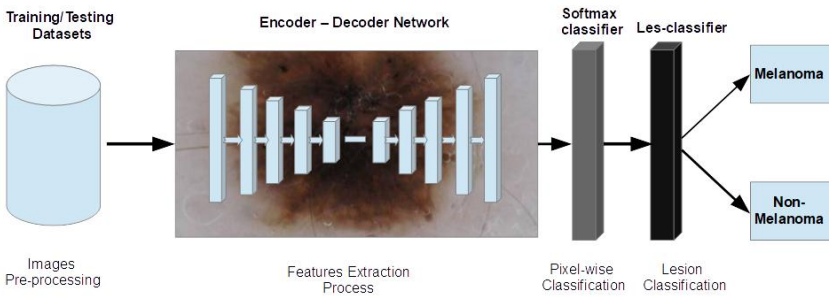


Figura 2.4. Marco y diagrama de flujo de arquitectura codificador – decodificador.

El modelo propuesto emplea un enfoque multietapa y multiescala, incorporando un clasificador softmax para realizar la clasificación por píxeles de las lesiones cutáneas. Asimismo, se implementa un método denominado *Lesion-classifier*, el cual permite distinguir entre melanoma y no melanoma a partir de los resultados obtenidos en la clasificación por píxeles. Para la fase experimental, se utilizaron dos bases de datos: ISIC 2017, con un total de 2,000 imágenes, y la del Hospital Pedro Hispano (PH2), que contiene 200 imágenes de lesiones cutáneas.

Los resultados obtenidos evidencian una precisión del 95% y un coeficiente de Dice del 92% en el conjunto ISIC 2017, así como una precisión del 95% y un coeficiente de Dice del 93% en el conjunto PH2. En los resultados obtenidos, la red fue estructurada en etapas, cada una encargada de una fase específica del proceso de aprendizaje y de la extracción de características. Esta organización permitió que el método clasificara las lesiones cutáneas en melanoma y no melanoma con base en los resultados generados por el clasificador softmax. Dicha función presenta la ventaja de consumir menos recursos computacionales, al no requerir una reponderación de las muestras, a diferencia de otras funciones de pérdida comúnmente utilizadas.

(Sara, U., 2025) desarrollaron un sistema basado en redes neuronales convolucionales para el análisis de lesiones cutáneas cancerosas. En su estudio, incorporaron el modelo SLC-CNN como extractor de características, el cual se emplea para la clasificación de las lesiones mediante un clasificador SVM, y para la segmentación utilizando el algoritmo XGBoost. El sistema propuesto procesa una imagen de prueba de cáncer de piel a través de etapas de preprocesamiento orientadas a la clasificación y segmentación. Una vez extraídas las características con SLC-CNN, estas se utilizan para determinar si la lesión es benigna o maligna, y posteriormente, con base en el resultado de la clasificación, se aplica el modelo XGBoost previamente entrenado para segmentar con precisión la región afectada.

A partir de la base de datos ISIC (International Skin Imaging Collaboration), se recopilaron 3,755 imágenes dermatoscópicas de cáncer de piel, clasificadas en dos categorías: 1,955 correspondientes a lesiones benignas y 1,800 a malignas. Estas imágenes fueron etiquetadas y almacenadas en formato JSON, a partir del cual se generaron los datos de máscara necesarios para la segmentación. En el estudio, se empleó el modelo extractor de características SLC-CNN tanto para la clasificación como para la segmentación de las lesiones, lo que permitió reducir significativamente el tiempo de entrenamiento y mejorar la precisión en los resultados. El modelo alcanzó una precisión del 99.6% en la clasificación y del 95.25% en la segmentación de las imágenes analizadas.

(Banasode, P., 2021) emplearon técnicas de aprendizaje automático para la detección de cáncer de piel, utilizando como algoritmo de clasificación la máquina de vectores de soporte (SVM). El procedimiento consistió en analizar las imágenes cutáneas y transformarlas a los formatos BGR-Gray y BGR-HSV, con el fin de facilitar su interpretación por parte del sistema en códigos binarios. Para su estudio, utilizaron la base de datos ISIC 2018, que contiene un total de 5,341 imágenes correspondientes a casos de melanoma y no melanoma, de las cuales seleccionaron 1,751 imágenes específicas de melanoma para el entrenamiento del modelo.

El conjunto de imágenes de entrenamiento fue utilizado para realizar pruebas con el clasificador SVM, el cual permite identificar las áreas de piel afectadas como melanoma. Para ello, se extrajeron características relevantes de las imágenes, tales como el comportamiento asimétrico, la distribución del color y la irregularidad de los bordes. Estas características fueron integradas en un vector que se introdujo en el modelo SVM para la

clasificación de las lesiones cutáneas. Los resultados obtenidos mediante el método propuesto evidenciaron una precisión del 96.9 % en la detección de melanoma.

(Kandhro, I. A., 2024) emplearon en su estudio el modelo preentrenado VGG19 (*Visual Geometry Group 19*), complementado con la técnica de *max pooling*, para la predicción de cáncer de piel. Este modelo se encargó de extraer características relevantes de las imágenes, clasificándolas en casos malignos o benignos. Posteriormente, dichas características fueron introducidas en un método de aprendizaje automático para su procesamiento. Los autores demostraron que la combinación con el modelo mejorado E-VGG19 (*enhanced Visual Geometry Group 19*) incrementa la precisión en la detección y clasificación de lesiones cutáneas cancerosas.

El experimento fue realizado utilizando el conjunto de datos ISIC 2020, el cual contiene 33,126 imágenes dermatoscópicas que representan diversas lesiones cutáneas, incluyendo casos benignos y malignos, cuya arquitectura general se presenta en la Figura 2.5. Para el desarrollo del modelo, se asignó el 70 % de las imágenes al proceso de entrenamiento y el 30% restante a las fases de validación y prueba. El modelo E-VGG16 demostró una capacidad efectiva para identificar correctamente las lesiones cutáneas. La implementación de aprendizaje por transferencia, junto con técnicas de aumento de datos, permitió alcanzar una exactitud del 91.01% en el entrenamiento y del 84.22% en la validación.

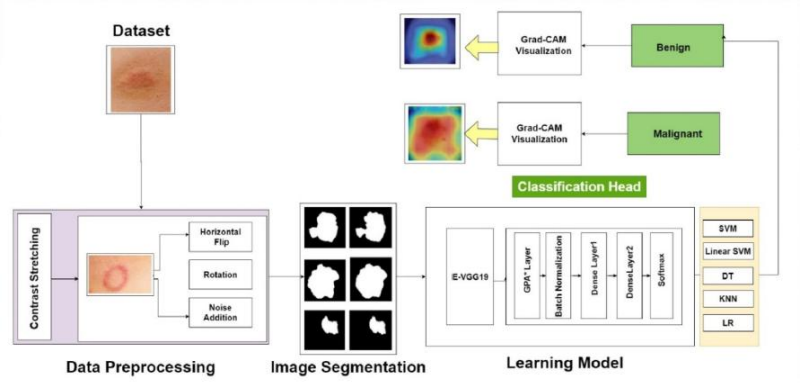


Figura 2.5. Arquitectura propuesta (E-CNN) de detección y clasificación de lesiones cutáneas.

La Tabla 01 presenta un resumen comparativo de las métricas utilizadas por los modelos previamente descritos. Estas medidas permiten evaluar el desempeño de cada enfoque en tareas de segmentación y clasificación de lesiones cutáneas, proporcionando una base cuantitativa para analizar su eficacia diagnóstica. Entre las métricas consideradas se incluyen la exactitud, sensibilidad, especificidad y otros indicadores relevantes para el diagnóstico de lesiones cutáneas, lo que permite observar tendencias y comportamientos particulares frente a distintos tipos de lesiones. Posteriormente, en el capítulo de resultados, se interpretan estos hallazgos a la luz de estudios previos y se reflexiona sobre su aplicabilidad en entornos reales, destacando la importancia de seleccionar métricas adecuadas según el contexto clínico y tecnológico.

Tabla 01. Resumen comparativo de diferentes modelos de aprendizaje automático.

REFERENCIA	BACKBONE	TIPO	ARQUITECTURA	CONJUNTO DE DATOS	EXACTITUD % (ACCURACY)	SENSIBILIDAD % (SENSITIVITY)	ESPECIFICIDAD % (SPECIFICITY)	JAC %	DICE %
KAUR, R., 2021	CNN	Segmentación	DilatedSkinNet	ISIC 2018	94.2	---	---	89.1	94.2
SHIHAB, A., 2021	K-Vecinos	Segmentación	ReLU	DermQuest	99.7	99.8	99.4	---	---
ADEGUN, A. A., 2020	CNN	Segmentación	Softmax	PH2	95.0	97.0	96.0	---	92.0
SARA, U., 2025	SLC-CNN	Clasificación	XGBoost	ISIC	99.6	---	---	---	---
BANASODE, P., 2021	SVM	Clasificación		ISIC 2018	96.9	95.7	90.2	---	---
KANDHRO, I. A., 2024	SVM & RF	Clasificación	E-VGG 19	ISIC 2020	88.0	85.0	---	---	---

En México, el cáncer de piel constituye una enfermedad de alta prevalencia, con un incremento notable en la incidencia y mortalidad, especialmente en el caso del melanoma, considerado el subtipo más letal. La detección temprana resulta esencial para mejorar el pronóstico; no obstante, enfrenta limitaciones significativas debido a la variabilidad en el diagnóstico visual y a la insuficiencia de especialistas en dermatología. Ante este panorama, se ha impulsado el desarrollo de métodos no invasivos y accesibles, respaldados por inteligencia artificial, con el propósito de optimizar la evaluación dermatológica y contribuir a una identificación más precisa y oportuna de lesiones cutáneas malignas.

En México, el melanoma es el tipo de cáncer de piel más letal, con una incidencia creciente atribuida a la radiación ultravioleta. La evaluación visual mediante la regla ABCD es limitada, lo que ha impulsado el uso de inteligencia artificial. Modelos como CNN, SVM y VGG19, entrenados con bases de datos ISIC y PH2, han logrado precisiones superiores al 95% en segmentación y clasificación de lesiones cutáneas, demostrando la eficacia del aprendizaje profundo en el diagnóstico automatizado.

No obstante, la mayoría de estos modelos no incorporan mecanismos de regularización con fundamento matemático explícito, lo que limita su estabilidad en la generalización ante la variabilidad propia de las imágenes dermatoscópicas. Esta limitación motiva la exploración de enfoques que integren bases funcionales ortogonales, como los polinomios de Chebyshev, dentro de arquitecturas de segmentación semántica profunda.

Capítulo 3. Marco teórico

En este capítulo se presentan los fundamentos teóricos que sustentan el desarrollo del modelo propuesto. Se abordan los principios del aprendizaje profundo supervisado aplicado a la imagenología médica, los tipos de redes neuronales relevantes para la segmentación semántica, las funciones de activación empleadas, las arquitecturas de referencia en segmentación semántica, los conceptos de regularización y las métricas de evaluación utilizadas. Finalmente, se introducen los polinomios de Chebyshev y sus propiedades matemáticas, cuya incorporación como función de preprocesamiento en los bloques convolucionales de la arquitectura U-Net constituye la contribución central de esta investigación.

El enfoque adoptado a lo largo del capítulo es el del aprendizaje supervisado, paradigma en el que el modelo aprende a partir de pares de imágenes y máscaras de segmentación anotadas, lo que resulta adecuado para la naturaleza del problema clínico abordado y para el conjunto de datos utilizado (Goodfellow, 2017).

3.1. Aprendizaje profundo y la imagenología médica.

3.1.1 Aprendizaje supervisado.

La tarea básica del aprendizaje supervisado; es predecir una salida aleatoria $\boldsymbol{Y}$ a partir de una entrada aleatoria $\boldsymbol{X}$, utilizando una función de predicción $\boldsymbol{g}: \boldsymbol{x} \mapsto \boldsymbol{y}$ que pertenece a una clase de funciones de aproximación $\boldsymbol{G}$. De manera más general, es posible que queramos predecir una salida con valor vectorial y utilizando una función de predicción $\boldsymbol{g}: \boldsymbol{x} \mapsto \boldsymbol{y}$ de la clase $\boldsymbol{G}$.

En el contexto del aprendizaje automático, la clase $\boldsymbol{G}$ se denomina espacio de hipótesis o universo de modelos posibles, y la capacidad de representación de un espacio de hipótesis $\boldsymbol{G}$ es simplemente su complejidad. Supongamos que tenemos una clase de funciones $\boldsymbol{G}_L$, indexada por un parámetro L que controla la complejidad de la clase, de modo que $\boldsymbol{G}_L \subset \boldsymbol{G}_{L+1} \subset \boldsymbol{G}_{L+2} \subset \cdots$. Al seleccionar una clase de funciones adecuada, debemos tener en cuenta la compensación entre la aproximación y la estimación. Por un lado, la clase $\boldsymbol{G}_L$ debe ser lo suficientemente compleja (rica) como para representar con precisión la función de predicción óptima desconocida g^*, lo que puede requerir un L muy grande. Por otro lado, los subconjuntos de la clase $\boldsymbol{G}_L$ deben ser lo suficientemente simples como para entrenarse con un pequeño error de estimación y con unas exigencias mínimas de memoria del ordenador, lo que puede requerir un L pequeño.

Para balancear estos objetivos contrapuestos, es útil que la clase $\boldsymbol{G_{L+1}}$, más compleja, se pueda construir fácilmente a partir de una $\boldsymbol{G_L}$ existente y más simple. La clase de funciones $\boldsymbol{G_L}$, más simple, puede construirse modificando una clase $\boldsymbol{G_{L-1}}$ aún más simple, y así sucesivamente.

Una clase de funciones que permite esta construcción jerárquica natural es la clase de *redes neuronales*. Conceptualmente, una red neuronal con L capas es un modelo de regresión paramétrica no lineal cuya capacidad de representación puede controlarse fácilmente mediante L.

La salida de una red neuronal se asocia como la composición repetida de funciones lineales y no lineales. Esta representación de la salida proporcionará una clase flexible de funciones no lineales que se pueden diferenciar fácilmente. Como resultado, el entrenamiento de la red mediante métodos de optimización implica principalmente operaciones matriciales estándar que se pueden realizar de manera muy eficiente.

Históricamente, las redes neuronales se diseñaron originalmente para imitar el funcionamiento del cerebro humano, con los nodos de la red modelando las neuronas y los enlaces de la red modelando los axones que conectan las neuronas.

Sin embargo, hay que tomar en cuenta que crear algoritmos eficientes de aprendizaje automático imitando el funcionamiento del cerebro humano han sido tan infructuosos como los intentos de construir aviones imitando el aleteo de las aves. En cambio, muchos algoritmos automáticos eficaces se han inspirado en ideas matemáticas para la aproximación de funciones. Una de estas ideas es el siguiente resultado fundamental.

Teorema 3.1: Kolmogorov (1957). *Toda función continua* $g^*: [0,1]^p \longmapsto \mathbb{R}$ *con* $p \geq 2$ *puede escribirse como,*

$$g^*(\boldsymbol{x}) = \sum_{j=1}^{2p+1} h_j \left(\sum_{i=1}^{p} h_{ij}(x_i) \right), \tag{3.1}$$

donde $\{h_j, h_{ij}\}$ *es un conjunto de funciones continuas univariantes que dependen de* g^*.

Este resultado establece que toda función continua definida en un espacio de alta dimensión puede representarse como una composición de funciones continuas unidimensionales más simples. La composición de los mapeos necesarios para calcular la salida $g^*(\boldsymbol{x})$ para una entrada $\boldsymbol{x} \in \mathbb{R}^p$ se muestra en la Figura 3.1, que representa una gráfica de una *red neuronal* con tres capas, denotadas como $l = 0,1,2$.

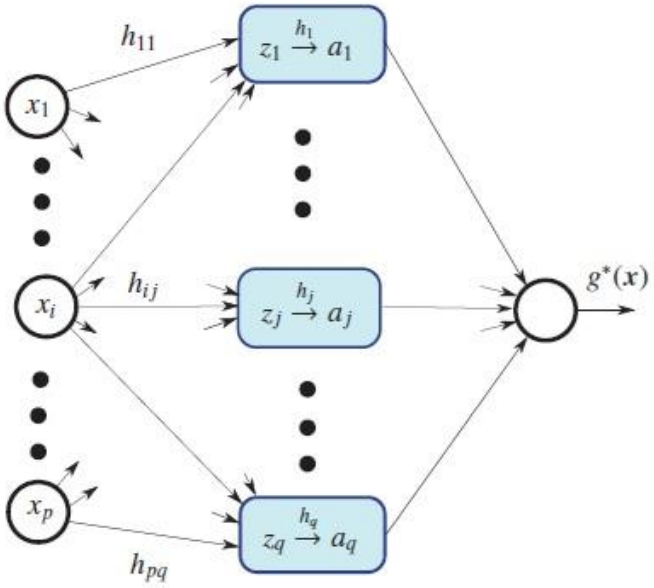


Figura 3.1. Toda función continua $g^*: [0,1]^p \longmapsto \mathbb{R}$ puede representarse mediante una red neuronal con una capa oculta $(l = 1)$, una capa de entrada $(l = 0)$ y una capa de salida $(l = 2)$. Adaptado de *Data science and machine learning: Mathematical and statistical methods* (p. 324), por Kroese, Botev, Taimre y Vaisman, 2019, Chapman & Hall/CRC.

.

En particular, cada componente p de la entrada $\boldsymbol{x}$ se representa como un nodo en *la capa de entrada* $(l = 0)$. En *la capa oculta* $(l = 1)$ hay $q \coloneqq 2p + 1$ nodos, cada uno de los cuales está asociado a un par de variables (z, a) con valores:

$$z_j \coloneqq \sum_{i=1}^{p} h_{ij}(\boldsymbol{x}) \quad y \quad a_j \coloneqq h_j(z_j). \tag{3.2}$$

Un enlace entre los nodos (z_j, a_j) y x_i con peso h_{ij} significa que el valor de z_j depende del valor de x_i a través de la función h_{ij}. Así, la capa de salida $(l = 2)$ representa el valor $g^*(\boldsymbol{x}) = \sum_{j=1}^{q} a_j$. Las flechas del gráfico recuerdan que la secuencia de los cálculos se ejecuta de izquierda a derecha, o desde la capa de entrada $l = 0$ hasta la capa de salida $l = 2$. En la práctica, no se conoce el conjunto de funciones $\{h_j, h_{ij}\}$, por que dependen de la incógnita g^*. En el caso de que g^* sea lineal, entonces todas las funciones unidimensionales $(2p + 1)(p + 1)$ también serán lineales. Sin embargo, en general, cada una de las funciones en $\{h_j, h_{ij}\}$ son no lineales.

Desafortunadamente, el Teorema 3.1 solo afirma la existencia de $\{h_j, h_{ij}\}$, y no indica cómo construir estas funciones no lineales. Una forma de salir de esta situación es sustituir estas $(2p + 1)(p + 1)$ funciones desconocidas por un número mucho mayor de funciones no lineales conocidas llamadas *funciones de activación*.

3.1.2. La imagenología en las redes neuronales artificiales.

El espectro electromagnético representa el rango completo de todos los tipos de radiación electromagnética. Desde las ondas de radio hasta los rayos gamma, dicho espectro resulta esencial para comprender cómo se propagan e interactúan con la materia las distintas formas de energía.

El espectro electromagnético constituye un concepto esencial en la física, ya que comprende todas las longitudes de onda de la radiación electromagnética. Dichas ondas se propagan por el espacio a la velocidad de la luz y están formadas por campos eléctricos y magnéticos que oscilan. Este espectro se segmenta en regiones de acuerdo con la longitud de onda y la frecuencia, cada una con características y aplicaciones particulares. La comprensión del espectro electromagnético resulta clave en disciplinas que incluyen la astronomía, las comunicaciones, la medicina y las ciencias ambientales.

El espectro electromagnético se clasifica en distintas regiones, cada una caracterizada por propiedades específicas en función de la frecuencia, la longitud de onda y la energía de la radiación. Dicho espectro se representa en la Figura 3.2 (Cox Figueroa, E. J., 2023).

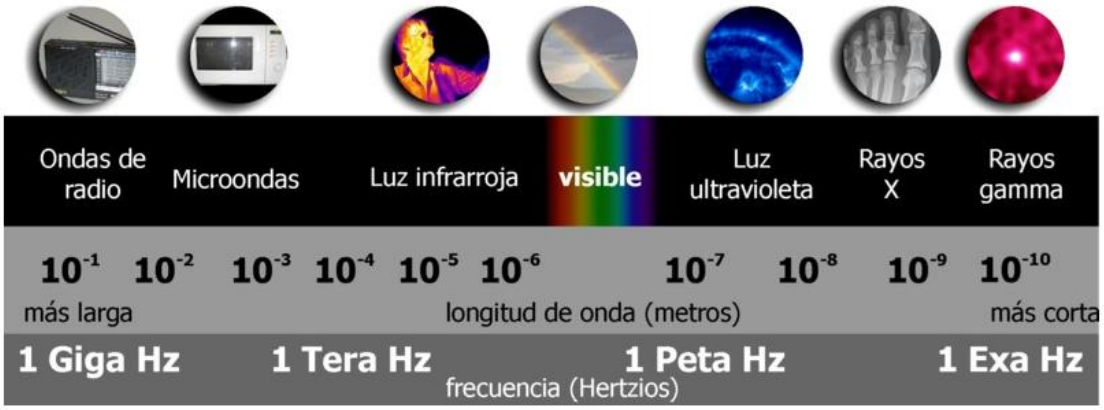


Figura 3.2. El espectro electromagnético. La frecuencia (y con ella la energía) aumenta de izquierda a derecha en este esquema. André Oliva / Public domain.

La figura 3.2 muestra que las ondas de baja frecuencia presentan una longitud de onda considerablemente mayor en comparación con aquellas de alta frecuencia, como los rayos gamma.

Existen dos tipos de radiación: ionizante y no ionizante. La radiación ionizante posee una cantidad de energía suficiente para desprender electrones de los átomos, fenómeno conocido como ionización. Este tipo de radiación puede alterar los átomos presentes en organismos vivos, lo que representa un riesgo para la salud al provocar daños en los tejidos y en el ADN. Las fuentes de radiación ionizante incluyen equipos de rayos X, partículas cósmicas provenientes del espacio exterior y elementos radiactivos, estos últimos emiten radiación ionizante como resultado de la desintegración de sus átomos (Cox Figueroa, E. J., 2023).

(Lifider, 2020) El espectro electromagnético, ordenado de mayor a menor energía transportada por el fotón, se divide en siete regiones principales de radiación electromagnética:

- **Gamma**: corresponde a las radiaciones de mayor energía, emitidas por núcleos atómicos.
- **Rayos X**: generados por electrones de los átomos, comúnmente utilizados en procedimientos de diagnóstico por imagen, como las radiografías.
- **Ultravioleta**: radiaciones altamente energéticas, capaces de inducir daños en la piel, incluyendo la generación de cáncer cutáneo.
- **Visible**: de energía intermedia, estas radiaciones estimulan el sistema visual humano, permitiendo la percepción de luz y color.
- **Infrarrojo**: responsables de la sensación térmica y del bronceado cutáneo, asociadas con la emisión de calor.
- **Microondas**: empleadas en sistemas de radar, telecomunicaciones y en dispositivos para el calentamiento de alimentos.
- **Radio**: representan las radiaciones de menor energía, utilizadas en la transmisión de señales de radio y televisión.

La radiación ultravioleta (UV) constituye una forma de radiación no ionizante emitida tanto por el Sol como por fuentes artificiales, entre ellas las camas bronceadoras. Si bien puede aportar ciertos beneficios al organismo humano, como la síntesis de vitamina D, también representa riesgos para la salud (Centro Nacional de Salud Ambiental, 2023). La radiación UV se clasifica en tres tipos principales según su longitud de onda: ultravioleta A (UVA), ultravioleta B (UVB) y ultravioleta C (UVC). La mayor parte de la radiación UV que alcanza la superficie terrestre corresponde al tipo UVA, aunque una fracción de la UVB también logra penetrar la atmósfera. Tanto la radiación UVA como la UVB pueden generar efectos adversos en la salud; sin embargo, la radiación UVA posee una mayor capacidad de penetración en las capas profundas de la piel y se mantiene relativamente constante a lo largo del año (Cox Figueroa, E. J., 2023).

Diversos estudios han demostrado que la exposición prolongada y sin protección a la radiación UV, especialmente a los rayos UVA y UVB, puede provocar alteraciones en el ADN de las células cutáneas, lo que incrementa el riesgo de desarrollar cáncer de piel. Los rayos UVB son los principales responsables de las quemaduras solares y se consideran el factor más relevante en la aparición de los tres tipos principales de cáncer de piel (American Cancer Society, s.f.).

3.2. Redes neuronales.

Las redes neuronales son soluciones informáticas basadas en algoritmos de aprendizaje automático que buscan imitar el funcionamiento del cerebro humano. Procesan datos con mayor eficiencia que los sistemas tradicionales, destacándose en el reconocimiento de patrones y la resolución de problemas. Se conocen también como redes neuronales artificiales (RNA) o simuladas (SNN). Forman parte del aprendizaje automático y son esenciales en el desarrollo del aprendizaje profundo. Su arquitectura se inspira en la estructura cerebral, con conexiones altamente interrelacionadas. Esta configuración permite replicar los procesos de indicación de las neuronas biológicas.

Las computadoras tradicionales funcionan mediante reglas lógicas y cálculos definidos. En cambio, las redes neuronales pueden procesar tanto funciones lógicas como datos no estructurados, como imágenes, vídeos y voz. A diferencia de los sistemas convencionales, las redes neuronales requieren entrenamiento para mejorar su precisión. Este proceso de optimización resulta rentable y útil en aplicaciones de inteligencia artificial. Las redes neuronales ofrecen una herramienta poderosa en ciencias de la computación. Son capaces de clasificar y agrupar datos con gran velocidad. Esto permite realizar tareas como reconocimiento de voz e imagen en minutos. Superan en eficiencia el tiempo que requeriría un experto humano para las mismas tareas.

3.2.1. Perceptrón multicapa.

El Perceptrón Multicapa, también conocido como MLP (Multi Layer Perceptron), constituye un modelo de aprendizaje automático basado en redes neuronales artificiales, caracterizado por la presencia de tres o más capas de neuronas densamente conectadas. Esta configuración implica que cada neurona en una capa establece comunicación con todas las neuronas de la capa anterior para recibir información, y transmite dicha información a todas las neuronas de la capa siguiente (Torres, 2020).

En este tipo de arquitectura se distinguen tres clases de capas: la capa de entrada, cuyas neuronas reciben los datos iniciales (inputs); la capa de salida, encargada de generar las salidas del modelo (outputs) o las predicciones; y una o más capas ocultas, situadas entre las anteriores. Las neuronas pertenecientes tanto a las capas ocultas como a la capa de salida transforman la información mediante una suma ponderada de las entradas recibidas, a la cual se añade una constante denominada sesgo. Posteriormente, se aplica una función de activación o transferencia. Los valores generados en las capas ocultas se transmiten a las neuronas de la capa siguiente, mientras que los valores obtenidos en la capa de salida corresponden a las predicciones del modelo. Los coeficientes que ponderan las entradas recibidas por cada neurona se denominan pesos sinápticos (García, R. A., 2024).

El concepto de aprendizaje en redes neuronales se refiere al proceso de optimización de los parámetros del modelo, específicamente los pesos sinápticos y los sesgos, con el objetivo de generar una salida lo más cercana posible a la deseada. La profundidad del aprendizaje, conocida como *deep learning*, está asociada con el número de capas ocultas que conforman la arquitectura de la red neuronal. En la Figura 3.3 se presenta el modelo gráfico de una neurona artificial junto con sus componentes fundamentales.

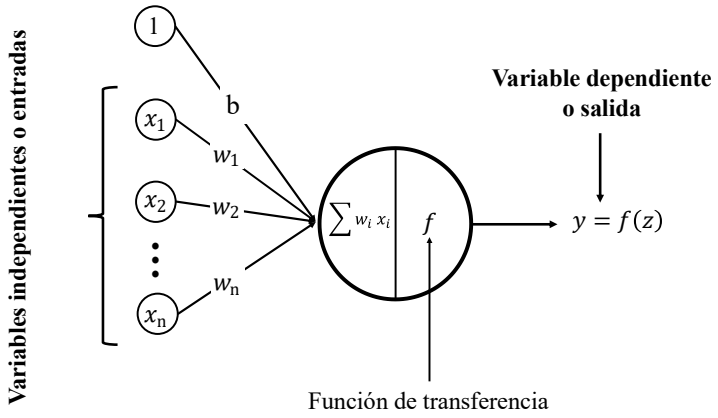


Figura 3.3. Representación esquemática de una neurona artificial. Elaboración propia.

El círculo representa una neurona artificial, a la cual ingresan los datos expresados mediante las variables $x_1, x_2, \cdots, x_n$, denominadas *entradas* (inputs*)*. Cada dato x_i se encuentra ponderado por un peso correspondiente w_i. Además, se incluye una entrada adicional de valor $x_0 = 1$, ponderada por un coeficiente b, el cual actúa como *sesgo* (bias) de la neurona. El procesamiento realizado por la neurona puede describirse en dos etapas principales, iniciando con el cálculo de la entrada ponderada o *preactivación*, simbolizada por z, según se muestra a continuación:

$$z = \sum_{i=1}^{n} w_i x_i + b, \tag{3.3}$$

a continuación, se procede con la determinación de la activación y, definida como:

$$y = f(z), \tag{3.4}$$

donde f representa una función de activación o de transferencia, e y corresponde a la activación de la neurona, la cual determina el valor de salida.

En el año 1958, Frank Rosenblatt presentó el Perceptrón, un modelo inspirado en investigaciones previas realizadas por Warren McCulloch y Walter Pitts. A diferencia del enfoque utilizado en el modelo de McCulloch-Pitts, en el Perceptrón cada entrada binaria (con valores de 0 o 1) se multiplicaba por un peso específico. Si la suma ponderada de estas entradas superaba un valor mínimo θ, denominado umbral, la neurona respondía con un valor de 1; en caso contrario, la respuesta era 0.

El Perceptrón simple propuesto por Rosenblatt puede configurarse para reproducir determinadas funciones lógicas, como la función OR y AND. Por ejemplo, al ajustar adecuadamente los pesos y el sesgo, es posible obtener una salida que represente la función OR. No obstante, no existe una configuración de parámetros (pesos y sesgo) que permita reproducir la función XOR, ya que los patrones $\{(x_1, x_2), y\}$ que la definen donde y representa la salida deseada para la entrada $\mathbf{x}^T = (x_1, x_2)$ no son linealmente separables. Para alcanzar dicho objetivo, no basta con una arquitectura que incluya únicamente una capa de entrada y una de salida, como ocurre en el Perceptrón simple; en cambio, se requiere una red que incorpore al menos una capa oculta.

En la actualidad, las redes neuronales emplean un modelo de neurona basado en funciones no lineales, en lugar del esquema tradicional de neuronas binarias. Este tipo de neuronas se distingue por aplicar una función de activación no lineal a la entrada ponderada. A diferencia de las neuronas binarias, cuya salida se limita a valores discretos, las neuronas con activación no lineal generan una salida continua que puede tomar cualquier valor real entre 0 y 1. Esta salida suele interpretarse como una probabilidad asociada a la pertenencia del dato de entrada a una determinada clase o como una probabilidad de éxito.

Por ejemplo, puede considerarse una red neuronal multicapa, es decir, una arquitectura que incorpora una o más capas ocultas con un número determinado de neuronas sigmoideas en cada una de ellas. Se asume que cada neurona perteneciente a las capas ocultas recibe información de todas las neuronas de la capa anterior y transmite su salida a cada una de las neuronas de la capa siguiente. Además, se establece que ninguna neurona toma como entrada la salida de neuronas ubicadas en capas posteriores ni de neuronas de su misma capa. Estos supuestos garantizan que la información fluya de manera unidireccional desde la capa de entrada hacia la capa de salida, característica propia de las redes *feedforward*, y que las capas estén densamente conectadas. A este tipo de redes se hará referencia en lo sucesivo dentro del presente trabajo. Una red con estas características se denomina Perceptrón Multicapa (MLP). Un ejemplo representativo de esta arquitectura se presenta en la Figura 3.4.

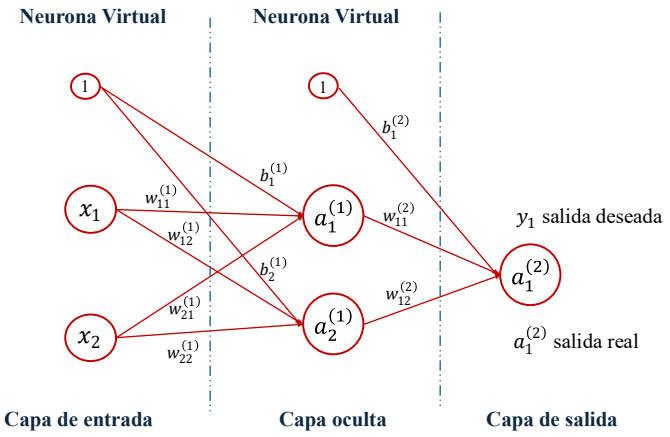


Figura 3.4. Representación esquemática de un perceptrón multicapa. Elaboración propia.

Al numerar las capas como $l = 0,1, \dots, L$, donde $l = 0$ corresponde a la capa de entrada y $l = L$ a la capa de salida, puede observarse que la red representada en la Figura 3.4 corresponde a un Perceptrón Multicapa (MLP). Esta red cuenta con una capa de entrada ($l = 0$) compuesta por dos neuronas que reciben y transmiten los datos de entrada x_1 y x_2 hacia las dos neuronas de la capa oculta ($l = 1$). A su vez, estas neuronas ocultas envían la información a la única neurona presente en la capa de salida o última capa ($l = L = 2$). La figura también incluye, tanto en la capa de entrada como en la capa oculta, una neurona con activación constante de valor 1, cuya salida se transmite a las neuronas de la capa siguiente ponderada por el sesgo o *bias* correspondiente a cada neurona de destino.

Por ejemplo, en la red representada en la Figura 3.4, la matriz de pesos sinápticos correspondiente a la primera capa oculta ($l = 1$) es la siguiente:

$$\mathbf{w}^{(1)} = \begin{pmatrix} w_{11}^{(1)} & w_{12}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} \end{pmatrix}, \tag{3.5}$$

el vector de preactivaciones correspondiente a la primera capa oculta es el siguiente:

$$\mathbf{z}^{(1)} = \begin{pmatrix} z_1^{(1)} \\ z_2^{(1)} \end{pmatrix}, \tag{3.6}$$

y el vector de activaciones correspondiente es el siguiente:

$$\mathbf{a}^{(1)} = \begin{pmatrix} a_1^{(1)} \\ a_2^{(1)} \end{pmatrix}. \tag{3.7}$$

Las redes neuronales del tipo MLP pueden ser entrenadas para llevar a cabo tareas de clasificación dicotómica, mediante el uso de una neurona en la capa de salida que aplica una función capaz de predecir la probabilidad de que los datos ingresados pertenezcan a una determinada clase. Asimismo, un MLP puede ser entrenado para clasificar datos en más de dos categorías, como ocurre en aplicaciones de reconocimiento de dígitos manuscritos en imágenes. Además, este tipo de red neuronal puede ser utilizado para predecir el valor de una variable continua, sin que dicha variable esté necesariamente restringida a un intervalo acotado, como en los casos en que se busca reproducir una función definida sobre valores reales. El desempeño del modelo depende de la disponibilidad de datos de entrenamiento adecuados, así como de la idoneidad de la arquitectura de la red y de la configuración de sus parámetros, incluyendo pesos y sesgos.

3.2.2. Redes neuronales convolucionales.

Las redes neuronales convolucionales o CNN, constituyen un tipo especializado de red neuronal diseñado para procesar datos con una estructura conocida, semejante a una cuadrícula. Entre los ejemplos más representativos se encuentran los datos de series temporales, que pueden interpretarse como una cuadrícula unidimensional con muestras

tomadas a intervalos regulares, así como los datos de imágenes, concebidos como una cuadrícula bidimensional de píxeles. Este tipo de redes ha demostrado una notable eficacia en diversas aplicaciones prácticas. La denominación red neuronal convolucional hace referencia al uso de una operación matemática denominada **convolución**, la cual representa una forma especializada de operación lineal. En esencia, las redes convolucionales son redes neuronales que incorporan la convolución en lugar de la multiplicación matricial general en al menos una de sus capas.

En su formulación más general, la convolución se define como una operación entre dos funciones reales de una variable, denotadas comúnmente como f y g. Esta operación se expresa mediante la siguiente integral:

$$x * w = \int_0^t x(u)w(t-u)du. \tag{3.8}$$

Esta definición representa la convolución continua en una dimensión, donde t es el parámetro de desplazamiento y u es la variable de integración. La operación de convolución permite combinar la información de ambas funciones, desplazando y ponderando sus valores a lo largo del dominio, lo que resulta fundamental en el procesamiento de señales e imágenes, así como en el funcionamiento de redes neuronales convolucionales (Goodfellow, 2017).

La función w debe asumir un valor nulo para todos los argumentos negativos; de lo contrario, estaría orientada hacia el futuro, lo cual excede las capacidades actuales de modelado. En términos generales, la operación de convolución puede definirse para cualquier par de funciones que permitan la evaluación de la integral correspondiente, y su aplicación no se limita únicamente al cálculo de medias ponderadas. En el contexto de las redes neuronales convolucionales, el primer argumento de la convolución, representado por la función x, se denomina comúnmente entrada, mientras que el segundo argumento, la función w, recibe el nombre de núcleo o kernel. El resultado de esta operación se conoce en ocasiones como mapa de características.

Cuando el índice temporal t está restringido a valores enteros, y se supone que tanto x como w están definidas únicamente para dichos valores, es posible establecer la formulación de la convolución discreta:

$$s(t) = (x * w)(t) = \sum_{u=-\infty}^{\infty} x(u)w(t-u). \tag{3.9}$$

En el ámbito de las aplicaciones de aprendizaje automático, la entrada suele representarse mediante una matriz multidimensional de datos, mientras que el kernel corresponde a una matriz multidimensional de parámetros que se ajustan a través del algoritmo de aprendizaje. Estas matrices multidimensionales se denominan comúnmente tensores. Dado que cada elemento tanto de la entrada como del kernel debe almacenarse de forma explícita e individual, se asume habitualmente que dichas funciones son nulas en todos los puntos, excepto en un conjunto finito donde se registran los valores. Esta suposición permite, en la

práctica, implementar la suma infinita como una suma sobre un número finito de elementos de la matriz.

Además, es frecuente aplicar convoluciones sobre múltiples ejes de manera simultánea. Por ejemplo, al utilizar una imagen bidimensional I como entrada, resulta conveniente emplear también un kernel bidimensional K:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n). \tag{3.10}$$

Una de las propiedades fundamentales de la operación de convolución es su conmutatividad, lo que implica que la expresión correspondiente a la ecuación 3.10 puede reescribirse de manera equivalente:

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n). \tag{3.11}$$

Por lo general, la formulación final resulta más sencilla de implementar en librerías de aprendizaje automático, debido a la menor variabilidad en el rango de valores válidos de los índices m y n.

La propiedad conmutativa de la convolución se origina en la inversión del kernel con respecto a la entrada; es decir, a medida que el índice m aumenta, el índice correspondiente en la entrada también lo hace, mientras que el índice en el kernel disminuye. Esta inversión se realiza exclusivamente para garantizar la conmutatividad de la operación. Aunque dicha propiedad puede ser útil en la elaboración de demostraciones teóricas, no suele representar un aspecto relevante en la implementación práctica de redes neuronales. En su lugar, muchas librerías especializadas en redes neuronales emplean una operación relacionada denominada **correlación cruzada**, la cual es equivalente a la convolución, pero sin invertir el kernel. En el presente trabajo, esta distinción resulta relevante dado que la función de preprocesamiento basada en polinomios de Chebyshev actúa sobre la entrada antes de la operación convolucional, por lo que su formulación es compatible con la implementación estándar de las librerías de aprendizaje profundo utilizadas.

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n). \tag{3.12}$$

En el ámbito del aprendizaje automático, el algoritmo de entrenamiento se encarga de ajustar los valores del kernel en las posiciones correspondientes. Por ello, un algoritmo que emplea convolución con inversión del kernel aprenderá un kernel invertido en relación con aquel que sería aprendido por un algoritmo que no realiza dicha inversión. Además, es poco común que la convolución se utilice de forma aislada en este contexto; por el contrario, suele aplicarse en conjunto con otras funciones.

La combinación de estas funciones no presenta conmutatividad, independientemente de si la operación de convolución incluye o no la inversión del kernel. La convolución discreta puede interpretarse como una operación equivalente a la multiplicación por una matriz. No obstante, dicha matriz presenta múltiples entradas que deben coincidir con otras dentro de su estructura. Cualquier algoritmo de red neuronal que opere mediante multiplicación matricial y que no dependa de propiedades particulares de la estructura de la matriz es compatible con la convolución, sin requerir modificaciones adicionales en la arquitectura de la red. Aunque las redes neuronales convolucionales suelen incorporar especializaciones adicionales para procesar de manera eficiente entradas de gran tamaño, estas optimizaciones no son estrictamente necesarias desde una perspectiva teórica.

Las aplicaciones contemporáneas de redes convolucionales suelen involucrar arquitecturas con un número elevado de unidades, lo que hace indispensable disponer de implementaciones eficientes que aprovechen al máximo los recursos computacionales disponibles. No obstante, en diversos casos, es posible acelerar el proceso de convolución mediante la selección de un algoritmo de convolución apropiado.

Desde una perspectiva computacional, la operación de convolución puede interpretarse como la transformación de la entrada y del kernel al dominio de la frecuencia mediante una transformada de Fourier, seguida de la multiplicación punto a punto de ambas señales, y posteriormente la reconversión al dominio temporal mediante una transformada de Fourier inversa. Para ciertos tamaños de problema, esta estrategia puede resultar más eficiente que la implementación directa de la convolución discreta.

El desarrollo de métodos más eficientes para ejecutar convoluciones o aproximaciones de convolución sin comprometer la precisión del modelo constituye una línea de investigación altamente activa. Incluso aquellas técnicas que se enfocan exclusivamente en mejorar la eficiencia de la propagación hacia adelante resultan valiosas, dado que en entornos comerciales es común asignar mayores recursos al despliegue de una red neuronal que a su fase de entrenamiento.

Las redes convolucionales han desempeñado un papel fundamental en el desarrollo histórico del aprendizaje profundo. Representan un ejemplo destacado de la aplicación exitosa de conocimientos derivados del estudio del cerebro a problemas propios del aprendizaje automático. Asimismo, constituyeron algunos de los primeros modelos de redes profundas que demostraron un rendimiento eficaz, mucho antes de que los modelos profundos arbitrarios se consideraran viables. Además, las redes convolucionales fueron pioneras en la resolución de aplicaciones comerciales relevantes y continúan ocupando una posición de liderazgo en el ámbito comercial del aprendizaje profundo en la actualidad. Una gráfica representativa de la arquitectura CNN se presenta en la Figura 3.5.

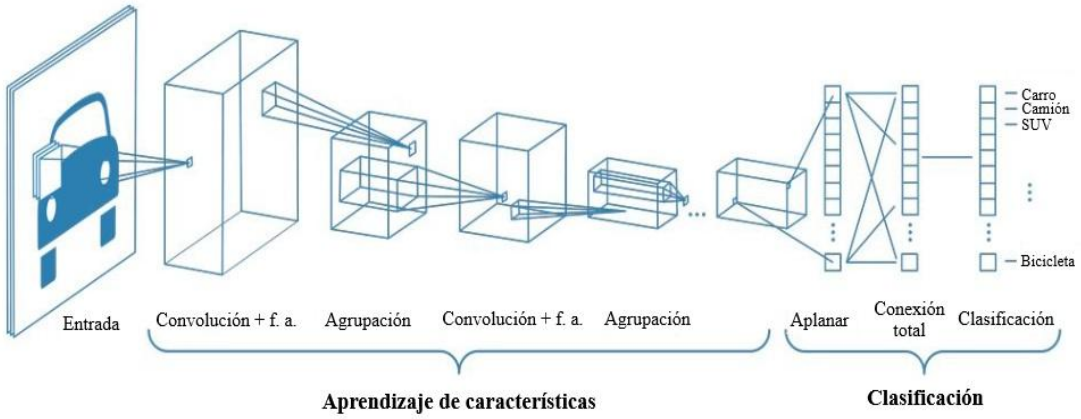


Figura 3.5. Representación esquemática de una red neuronal convolucional. Disponible en: https://la.mathworks.com/discovery/convolutional-neural-network.html Consultado en octubre 12, 2025.

Las redes convolucionales constituyeron algunos de los primeros modelos profundos funcionales que lograron ser entrenados mediante el algoritmo de retropropagación. No se ha determinado con certeza por qué estas redes obtuvieron resultados exitosos en una época en la que se consideraba que las redes de retropropagación generales no eran viables. Es posible que las redes convolucionales ofrecieran una mayor eficiencia computacional en comparación con las redes totalmente conectadas, lo que facilitó la realización de múltiples experimentos, así como el ajuste de sus implementaciones y de los hiperparámetros correspondientes. Además, se ha observado que las redes de mayor tamaño tienden a ser más fáciles de entrenar. Con el hardware moderno, las redes totalmente conectadas de gran escala muestran un rendimiento aceptable en diversas tareas, incluso cuando se emplean conjuntos de datos y funciones de activación que eran comunes en el periodo en que se cuestionaba la eficacia de dichas arquitecturas.

Las redes convolucionales ofrecen un mecanismo de especialización dentro del marco de las redes neuronales, orientado al procesamiento de datos que presentan una topología claramente estructurada en forma de cuadrícula. Esta arquitectura permite escalar los modelos a dimensiones considerablemente grandes. El enfoque ha demostrado ser particularmente exitoso en el tratamiento de imágenes con topología bidimensional. Para abordar el procesamiento de datos secuenciales unidimensionales, se recurre a otra especialización poderosa dentro del mismo marco: las redes neuronales recurrentes.

3.2.3. Redes neuronales recurrentes.

Durante la década de los noventa, ante las limitaciones de los modelos basados en n-gramas y los resultados prometedores obtenidos con redes neuronales tradicionales MLP (*propagación hacia adelante*), emergen las redes neuronales recurrentes, conocidas como **redes neuronales recurrentes** (RNN). El principio fundamental de estas redes radica en la incorporación de bucles dentro del flujo de información. A diferencia de las redes neuronales

convencionales, en las que el flujo se describe únicamente en términos de avance (*forward*) y retroceso (*backward*), las RNN introducen bucles representados gráficamente mediante líneas curveadas intermedias, como se observa en la Figura 3.6.

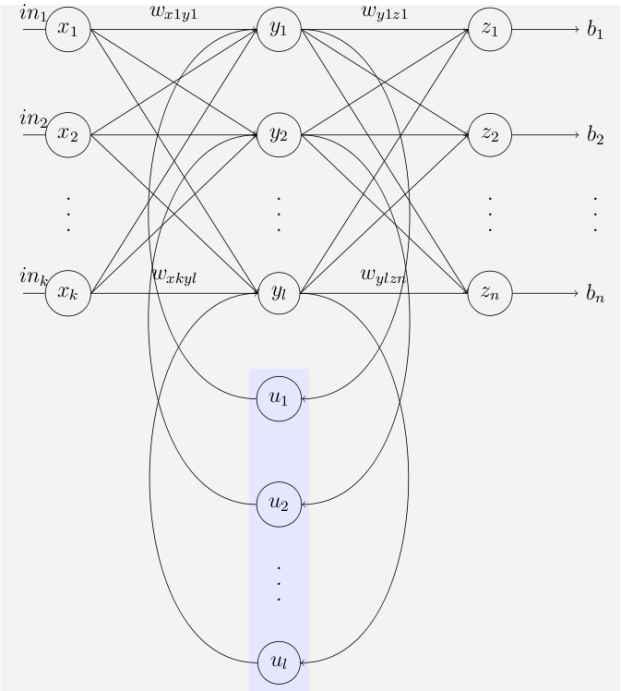


Figura 3.6. Estructura de la red neuronal recurrente simple de Elman. Disponible en: Elman srnn - Red neuronal recurrente - Wikipedia, la enciclopedia libre Consultado en octubre 09, 2025.

Estos bucles son interpretados como remanentes de información que se retroalimentan en la red junto con nuevos datos provenientes de la entrada, lo que permite la conservación y el procesamiento de información en el tiempo dentro de la propia arquitectura de la red. En síntesis, las redes neuronales recurrentes (RNN) emplean el razonamiento basado en experiencias previas para anticipar eventos futuros. Un ejemplo representativo de su aplicación es la traducción automática, en la cual una red neuronal procesa una oración de entrada en español y la convierte en una oración equivalente en inglés. Para ello, la red calcula la probabilidad de cada palabra en la oración de salida, considerando tanto la palabra en cuestión como la secuencia de palabras previamente generadas (Maqueda, D.E., 2021).

Una de las aplicaciones iniciales de las redes neuronales recurrentes (RNN) fue el modelado de datos de series temporales, tarea en la que estas redes conservan la información histórica mediante conexiones entre unidades ocultas y mecanismos de retraso temporal. Esta arquitectura les permite identificar correlaciones temporales entre eventos distantes dentro de la secuencia de datos. Aunque su propósito fundamental es capturar dependencias a largo plazo, tanto la evidencia teórica como los resultados experimentales han demostrado que procesar información de manera eficaz a lo largo de extensos intervalos temporales representa un desafío significativo (Bengio et al., 1994).

En términos generales, las redes neuronales recurrentes constituyen una familia de arquitecturas diseñadas específicamente para el procesamiento de datos secuenciales. De manera análoga a las redes convolucionales, que están especializadas en el tratamiento de cuadrículas de valores **X**, como ocurre en el caso de las imágenes, las redes recurrentes se orientan al procesamiento de secuencias de valores $x^{(1)}, x^{(2)}, \ldots, x^{(\tau)}$. Así como las redes convolucionales pueden adaptarse con facilidad a imágenes de gran anchura y altura —e incluso algunas variantes permiten manejar imágenes de tamaño variable—, las redes recurrentes presentan la capacidad de ajustarse a secuencias considerablemente extensas, lo cual resultaría impráctico para arquitecturas que no incorporan una especialización basada en la estructura secuencial. Además, la mayoría de las redes recurrentes están diseñadas para procesar secuencias de longitud variable, lo que amplía su aplicabilidad en contextos dinámicos y heterogéneos (Goodfellow, 2017).

Una idea relacionada con el procesamiento de secuencias es la aplicación de la convolución en series temporales unidimensionales. Esta operación permite que una red neuronal comparta parámetros a lo largo del tiempo, aunque de manera superficial. El resultado de dicha convolución es una secuencia en la que cada elemento de la salida depende únicamente de un número limitado de elementos vecinos de la entrada. La noción de compartición de parámetros se refleja en el uso repetido del mismo núcleo de convolución en cada paso temporal.

Por otro lado, las redes neuronales recurrentes implementan la compartición de parámetros de forma distinta. En este caso, cada elemento de la salida se calcula como una función de los elementos anteriores de la propia salida, utilizando una misma regla de actualización aplicada de manera recurrente. Esta formulación genera un intercambio de parámetros que se propaga a través de un gráfico computacional de gran profundidad, lo que permite capturar dependencias temporales más complejas en los datos secuenciales.

Las redes neuronales recurrentes (RNN) se describen como un operador sobre secuencias que contienen vectores $\boldsymbol{x}^{(t)}$, donde el índice temporal t varía desde 1 hasta τ. En la práctica, estas redes suelen procesar **minilotes** compuestos por múltiples secuencias, cada una de ellas con una longitud τ que puede diferir entre los distintos elementos del lote. Cabe destacar que el índice de paso temporal no necesariamente representa el tiempo real; en muchos casos, simplemente indica la posición relativa dentro de la secuencia.

Además, las RNN pueden extenderse a dos dimensiones para abordar datos espaciales, como imágenes. Incluso cuando se aplican a datos temporales, estas redes pueden incorporar conexiones que retroceden en el tiempo, siempre que la secuencia completa esté disponible antes de ser introducida en la red. Esta flexibilidad permite modelar dependencias tanto temporales como espaciales en una amplia variedad de contextos.

Algunos patrones de diseño relevantes en el contexto de las redes neuronales recurrentes incluyen los siguientes (Goodfellow, 2017):

- Redes recurrentes que generan una salida en cada paso temporal y presentan conexiones recurrentes entre las unidades ocultas, como se muestra en la Figura 3.7.

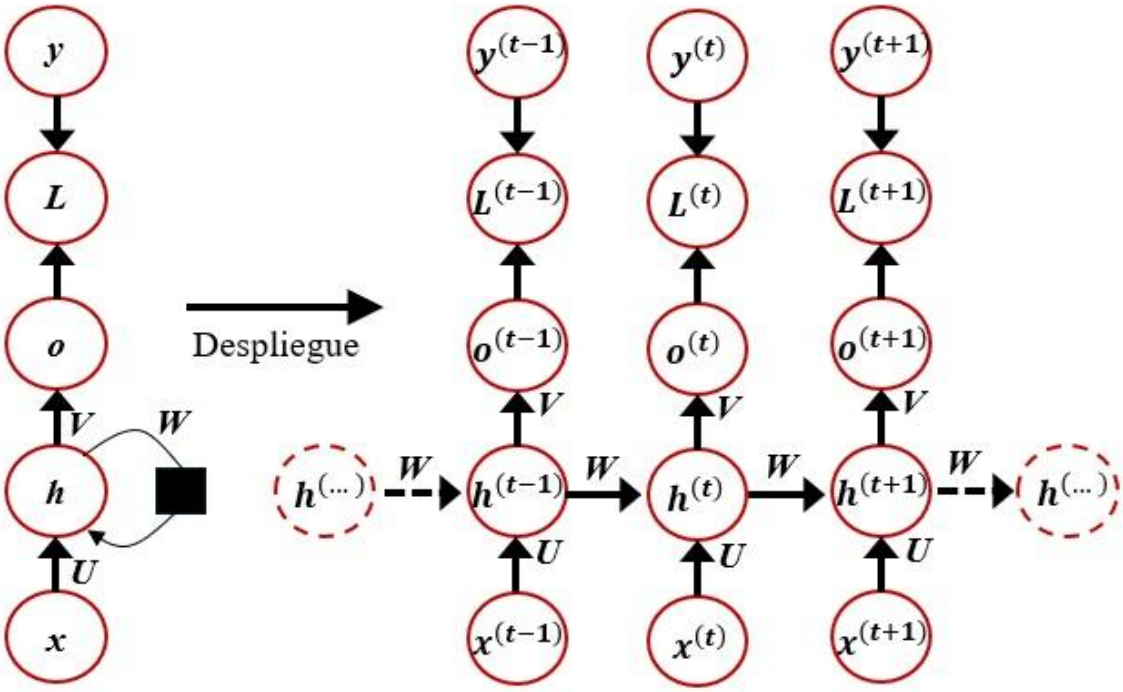


Figura 3.7. Gráfico computacional utilizado para calcular la pérdida de entrenamiento en una red neuronal recurrente, la cual realiza el mapeo de una secuencia de entrada compuesta por valores $\boldsymbol{x}$ hacia una secuencia correspondiente de valores de salida $\boldsymbol{o}$. Adaptado de *Deep Learning* (p. 378), por Goodfellow, Bengio y Courville, 2017, The MIT Press.

En la Figura 3.7 se observa que una pérdida L mide la distancia entre cada $\boldsymbol{o}$ y el objetivo de entrenamiento correspondiente $\boldsymbol{y}$. Cuando se utilizan salidas softmax, se supone que $\boldsymbol{o}$ son las probabilidades logarítmicas no normalizadas. La pérdida L calcula internamente $\hat{\boldsymbol{y}} = \text{softmax}(\boldsymbol{o})$ y lo compara con el objetivo $\boldsymbol{y}$. La RNN tiene entradas a conexiones ocultas parametrizadas por una matriz de pesos $\boldsymbol{U}$, conexiones recurrentes ocultas a ocultas parametrizadas por una matriz de pesos $\boldsymbol{W}$ y conexiones ocultas a salidas parametrizadas por una matriz de pesos $\boldsymbol{V}$. La ecuación 3.13 define la propagación hacia adelante en este modelo. La RNN y su pérdida dibujada con conexiones recurrentes (nodos izquierda). Lo mismo visto como un gráfico computacional desplegado en el tiempo, donde cada nodo está ahora asociado a una instancia temporal concreta (nodos derecha).

- Redes recurrentes que generan una salida en cada paso temporal y establecen conexiones recurrentes únicamente desde la salida de un paso temporal hacia las unidades ocultas del paso siguiente, como se observa en la Figura 3.8.

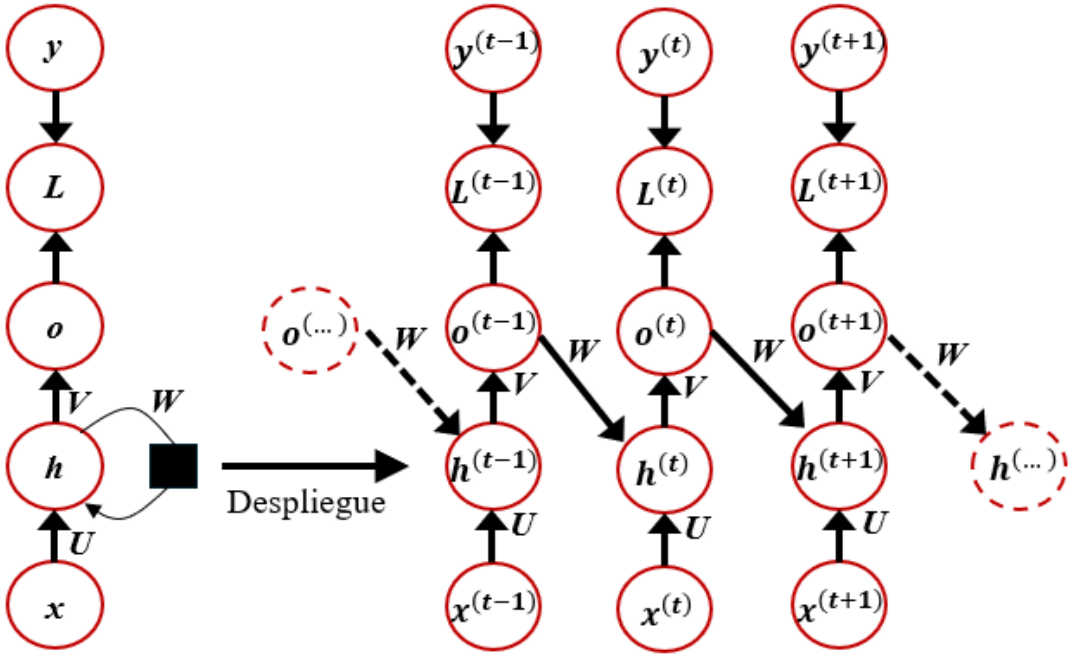


Figura 3.8. Representación de una RNN cuya única recurrencia consiste en una conexión de retroalimentación desde la salida hacia la capa oculta. En cada instante temporal t, la entrada está dada por $\boldsymbol{x}_t$, las activaciones de la capa oculta se denotan como $\boldsymbol{h}^{(t)}$, las salidas como $\boldsymbol{o}^{(t)}$, los objetivos como $\boldsymbol{y}^{(t)}$ y la pérdida correspondiente como $L^{(t)}$. Adaptado de *Deep Learning* (p. 380), por Goodfellow, Bengio y Courville, 2017, The MIT Press.

En la Figura 3.8 se presenta el diagrama del circuito (izquierda) junto con el gráfico computacional correspondiente (derecha). La RNN ilustrada en esta figura posee una arquitectura menos potente —en términos de capacidad para representar funciones— que aquella mostrada en la Figura 3.7. Mientras que la RNN de la Figura 3.7 puede incorporar cualquier información relevante del pasado en su representación oculta $\boldsymbol{h}$ y transmitir dicha representación $\boldsymbol{h}$ hacia el futuro, la RNN aquí representada está diseñada para generar un valor de salida específico en $\boldsymbol{o}$, siendo esta salida la única información que se permite propagar temporalmente.

No existen conexiones directas desde $\boldsymbol{h}$ hacia adelante; la representación oculta anterior $\boldsymbol{h}$ se vincula con el presente únicamente de forma indirecta, a través de las predicciones que contribuyeron a su generación. A menos que $\boldsymbol{o}$ sea altamente dimensional y contenga una riqueza informativa considerable, es probable que no conserve elementos significativos del pasado. Esta limitación reduce la potencia expresiva de la RNN, aunque puede facilitar su entrenamiento, dado que cada paso temporal puede ser optimizado de manera independiente, lo que permite una mayor paralelización durante el proceso de aprendizaje.

- Redes recurrentes con conexiones recurrentes entre unidades ocultas, diseñadas para leer una secuencia completa antes de producir una única salida, como se ilustra en la Figura 3.9.

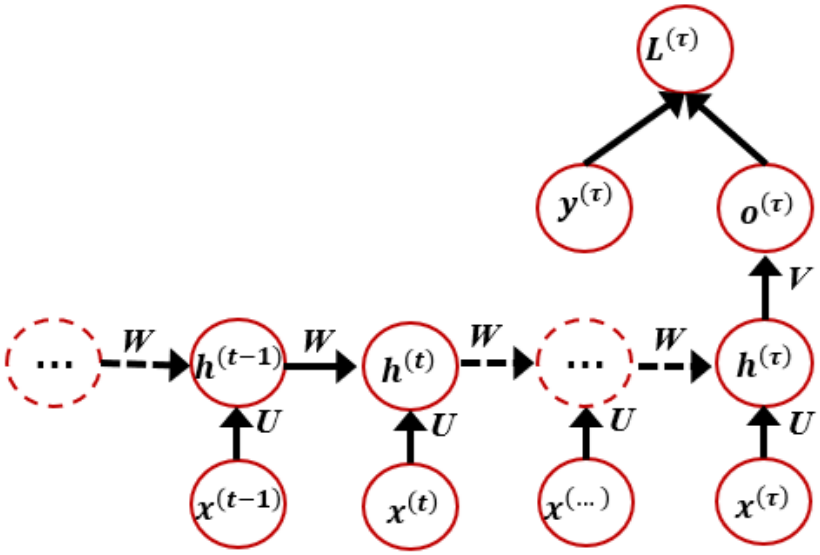


Figura 3.9. Representación de una red neuronal recurrente desplegada en el tiempo, caracterizada por generar una única salida al final de la secuencia procesada. Adaptado de *Deep Learning* (p. 382), por Goodfellow, Bengio y Courville, 2017, The MIT Press.

Una red neuronal como la representada en la Figura 3.9 puede emplearse para resumir una secuencia y generar una representación de tamaño fijo, la cual se utiliza como entrada en etapas posteriores de procesamiento. El objetivo puede situarse al final de la secuencia —como se muestra en dicha figura— o bien el gradiente en la salida $\boldsymbol{o}^{(t)}$ puede calcularse mediante retropropagación desde módulos ubicados más adelante en la arquitectura.

Al desarrollar las ecuaciones de propagación hacia adelante correspondientes a la red neuronal recurrente (RNN) representada en la Figura 3.7, se observa que dicha figura no especifica la elección de la función de activación para las unidades ocultas. En este contexto, se supone la utilización de la función de activación tangente hiperbólica. Asimismo, la figura no detalla con precisión la forma que adoptan la salida ni la función de pérdida. Por ello, se considera que la salida es discreta, como si la RNN estuviese destinada a la predicción de palabras o caracteres.

Una representación natural para variables discretas consiste en suponer que la salida $\boldsymbol{o}$ proporciona las probabilidades logarítmicas no normalizadas de cada valor posible de la variable discreta. Posteriormente, se aplica la operación *softmax* como etapa de posprocesamiento, con el fin de obtener un vector $\hat{\boldsymbol{y}}$ de probabilidades normalizadas sobre la salida. El proceso de propagación hacia adelante se inicia con la especificación del estado

inicial $\boldsymbol{h}^{(0)}$. A continuación, para cada paso temporal desde $t = 1$ hasta $t = \tau$, se aplican las siguientes ecuaciones de actualización:

$$\boldsymbol{a}^{(t)} = \boldsymbol{b} + \boldsymbol{W}\boldsymbol{h}^{(t-1)} + \boldsymbol{U}\boldsymbol{x}^{(t)}, \tag{3.13}$$

$$\boldsymbol{h}^{(t)} = \tanh\left(\boldsymbol{a}^{(t)}\right), \tag{3.14}$$

$$\boldsymbol{o}^{(t)} = \boldsymbol{c} + \boldsymbol{V}\boldsymbol{h}^{(t)}, \tag{3.15}$$

$$\hat{\boldsymbol{y}}^{(t)} = \text{softmax}\left(\boldsymbol{o}^{(t)}\right). \tag{3.16}$$

Los parámetros considerados incluyen los vectores de sesgo $\boldsymbol{b}$ y $\boldsymbol{c}$, así como las matrices de peso $\boldsymbol{U}$, $\boldsymbol{V}$ y $\boldsymbol{W}$, correspondientes respectivamente a las conexiones de entrada a capa oculta, de capa oculta a salida, y de capa oculta a sí misma.

Lo anterior constituye un ejemplo de una red neuronal recurrente que realiza el mapeo de una secuencia de entrada hacia una secuencia de salida de igual longitud. En este caso, la pérdida total asociada a una secuencia determinada de valores $\boldsymbol{x}$, emparejados con una secuencia correspondiente de valores $\boldsymbol{y}$, y se define como la suma de las pérdidas acumuladas en todos los pasos temporales. Por ejemplo, si $L^{(t)}$ representa el logaritmo de verosimilitud negativo del valor $y^{(t)}$ dado el historial de entradas $\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(t)}$, entonces:

$$L\left(\left\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(\tau)}\right\}, \left\{\boldsymbol{y}^{(1)}, \dots, \boldsymbol{y}^{(\tau)}\right\}\right) = \sum_t L^{(t)} \tag{3.17}$$

$$= -\sum_t \log p_{modelo}\left(y^{(t)} | \left\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(t)}\right\}\right). \tag{3.18}$$

La probabilidad del modelo $p_{modelo}\left(y^{(t)} | \left\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(\tau)}\right\}\right)$ se obtiene al leer la entrada correspondiente a $y^{(t)}$ desde el vector de salida generado por el modelo, denotado como $\hat{\boldsymbol{y}}^{(t)}$. El cálculo del gradiente de esta función de pérdida con respecto a los parámetros del modelo representa una operación computacionalmente costosa.

El cálculo del gradiente requiere realizar una propagación hacia adelante, desplazándose de izquierda a derecha a través de la representación del gráfico mostrada en la figura 10.3, seguida de una propagación hacia atrás, que se efectúa de derecha a izquierda sobre el mismo gráfico. El tiempo de ejecución asociado a este proceso es del orden de $O(\tau)$ y no puede reducirse mediante paralelización, debido a que la estructura del gráfico de propagación hacia adelante es inherentemente secuencial; cada paso temporal depende del cálculo del paso anterior. Los estados generados durante la pasada hacia adelante deben conservarse en memoria hasta que sean reutilizados en la fase de retropropagación, lo que implica un coste de almacenamiento también proporcional a $O(\tau)$. El algoritmo de retropropagación aplicado sobre el gráfico desenrollado, con coste computacional $O(\tau)$, se conoce como *retropropagación a través del tiempo* o *Backpropagation Through Time* (BPTT).

Otro algoritmo perteneciente a la familia de redes neuronales recurrentes es la Memoria a Largo y Corto Plazo (LSTM). En 1997, Hochreiter y Schmidhuber presentaron el primer modelo de este tipo, el cual incorpora unidades ocultas especializadas que tienden, de forma natural, a retener las entradas durante períodos prolongados (Hochreiter y Schmidhuber, 1997). Esta arquitectura es ampliamente reconocida por su capacidad para analizar y procesar de manera eficiente datos secuenciales con dependencias temporales extendidas. No obstante, a pesar de su popularidad, persiste el desafío de inicializar y optimizar adecuadamente los modelos RNN-LSTM, lo que puede afectar negativamente su rendimiento y precisión. En este contexto, uno de los objetivos fundamentales es examinar cómo las técnicas de inicialización y ajuste de pesos pueden contribuir a mejorar el desempeño de dichas arquitecturas (Al-Selwi et al., 2024).

Las redes neuronales recurrentes (RNN) se aplican ampliamente en diversos campos, tales como la asistencia sanitaria y la medicina, el sector energético, el procesamiento del lenguaje natural, la visión artificial, las comunicaciones y las redes, entre otros (Al-Selwi et al., 2024). Esta versatilidad ha permitido el desarrollo de investigaciones en múltiples disciplinas, en estrecha relación con las aplicaciones específicas del modelo LSTM, sus arquitecturas, los conjuntos de datos empleados y las métricas utilizadas para la evaluación del desempeño.

Asimismo, los lenguajes de programación, los marcos de desarrollo y las bibliotecas especializadas desempeñan un papel fundamental en la implementación y entrenamiento de modelos basados en RNN. El proceso de entrenamiento implica la inicialización y optimización iterativa de los pesos de la red, con el objetivo de alcanzar la convergencia hacia una matriz de pesos ideal y un valor mínimo de pérdida. Una inicialización adecuada de los parámetros resulta esencial para evitar fenómenos como la desaparición o explosión de gradientes, y para acelerar el rendimiento del modelo. El uso de redes neuronales recurrentes debe orientarse hacia una atención más rigurosa a la naturaleza de los experimentos y a sus requerimientos específicos, seleccionando algoritmos que se ajusten adecuadamente a dichas condiciones y que permitan alcanzar resultados sobresalientes.

3.3. Arquitecturas de redes neuronales convolucionales.

El desarrollo de arquitecturas avanzadas de redes neuronales convolucionales se ha visto impulsado por la necesidad de optimizar de manera simultánea la precisión, la eficiencia computacional y la capacidad de generalización en tareas de visión por computadora. Conforme los modelos han incrementado su profundidad y complejidad, también se han intensificado los desafíos vinculados con el entrenamiento, la escalabilidad y el despliegue en entornos con recursos limitados. Esta evolución ha puesto de manifiesto limitaciones estructurales en los enfoques tradicionales, lo que ha motivado la búsqueda de soluciones arquitectónicas más flexibles y eficientes.

Diversas propuestas recientes han replanteado los principios fundamentales de las CNN mediante estrategias que incluyen la optimización del flujo de gradientes, la redefinición de los mecanismos para modelar relaciones espaciales y de canal, y la incorporación de componentes capaces de equilibrar la capacidad representacional con la eficiencia operativa. Estas líneas de innovación han permitido superar cuellos de botella asociados al incremento de parámetros y al costo computacional, favoreciendo arquitecturas más robustas y adaptables a distintos escenarios de aplicación.

Como resultado, han surgido modelos que no solo alcanzan un rendimiento superior en tareas de clasificación y segmentación, sino que también se ajustan a las exigencias de aplicaciones en tiempo real y a dispositivos con restricciones de hardware. Este avance evidencia una tendencia hacia arquitecturas que combinan sofisticación teórica con viabilidad práctica, consolidando un campo en el que la eficiencia y la capacidad predictiva ya no se conciben como objetivos contrapuestos, sino como dimensiones complementarias del diseño de redes neuronales modernas.

3.3.1. Xception.

La arquitectura Xception, presentada por François Chollet en 2017 (Chollet, F., 2017), se ha consolidado como una de las contribuciones más influyentes en la evolución contemporánea de las redes neuronales convolucionales. Su importancia radica en la reformulación explícita de la hipótesis estructural que subyace a las CNN tradicionales: la suposición de que la extracción de características espaciales y la integración de información entre canales pueden descomponerse en operaciones independientes sin comprometer la capacidad representacional del modelo. Aunque esta idea ya se insinuaba en las arquitecturas Inception mediante la factorización de convoluciones, Xception lleva este principio a su expresión más extrema mediante el uso sistemático de convoluciones separables en profundidad.

El planteamiento central de Xception surge de un cuestionamiento directo al diseño convencional de las convoluciones estándar, las cuales combinan simultáneamente la correlación espacial y la correlación entre canales. Desde esta perspectiva, la arquitectura sostiene que dicha integración conjunta no es necesariamente óptima y que una separación explícita de ambas dimensiones puede mejorar tanto la eficiencia computacional como la capacidad de aprendizaje. Para ello, Xception divide la operación convolucional en dos etapas diferenciadas:

- **Convolución en profundidad (depthwise)**, procesa cada canal de entrada de manera independiente, capturando exclusivamente patrones espaciales.
- **Convolución puntual (pointwise)** (1×1), combina la información entre canales, permitiendo la interacción y recombinación de características.

Esta descomposición, como se muestra en la Figura 3.10, reduce de manera significativa el número de parámetros y el costo computacional, lo cual resulta especialmente relevante en aplicaciones donde la eficiencia constituye un requisito crítico, como en dispositivos móviles o sistemas embebidos.

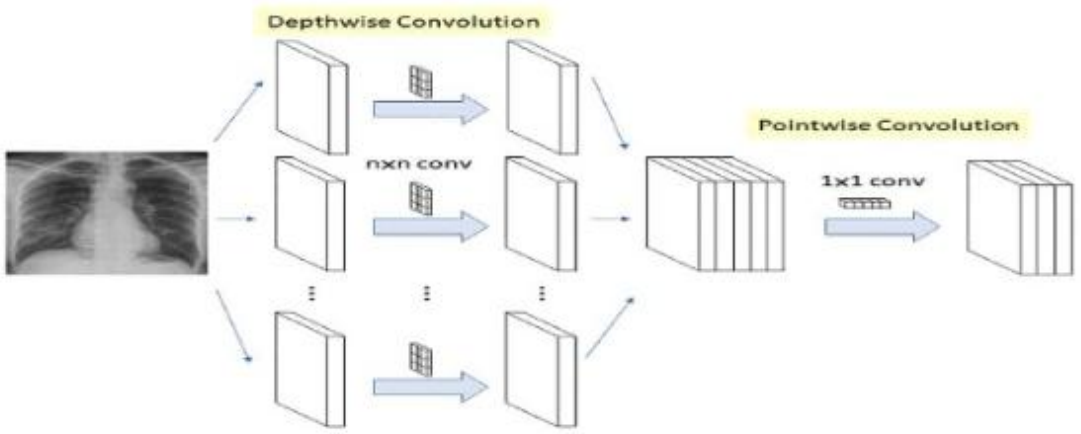


Figura 3.10. Representación de la arquitectura Xception. Adaptado y disponible en: viso.ai/deep-learning/xception-model/ Consultado en febrero 09, 2026.

La arquitectura se organiza en bloques modulares que integran convoluciones separables en profundidad junto con conexiones residuales, lo que facilita el flujo de gradientes y contribuye a la estabilidad del entrenamiento. Esta estructura modular favorece:

- **Eficiencia paramétrica**, al disminuir de forma drástica la cantidad de operaciones requeridas.
- **Mejor capacidad de generalización**, al separar explícitamente los patrones espaciales de la interacción entre canales.
- **Escalabilidad**, dado que la arquitectura puede adaptarse a distintas profundidades sin incrementar de manera desproporcionada la complejidad computacional.

En evaluaciones comparativas, Xception ha demostrado igualar o superar el rendimiento de arquitecturas contemporáneas más complejas, lo que evidencia que la factorización extrema no solo reduce costos computacionales, sino que también potencia la expresividad del modelo. Este resultado refuerza la idea de que la separación explícita entre procesamiento espacial y combinación de canales constituye un principio arquitectónico sólido para el diseño de redes neuronales modernas.

3.3.2. EfficientNet.

La arquitectura **EfficientNet**, desarrollada por Tan y Le en 2019 (Tan, M., & Le, Q. V. 2019), se reconoce como una de las propuestas más influyentes en la optimización de arquitecturas convolucionales debido a su planteamiento sistemático para el escalamiento de modelos. Su aporte central radica en el escalado compuesto (**compound scaling)**, un principio que sostiene que la profundidad, el ancho y la resolución deben ampliarse de manera conjunta y proporcional para mejorar el rendimiento sin incrementar de forma desmedida el costo computacional. Las estrategias tradicionales de escalamiento en redes convolucionales suelen modificar solo una dimensión del modelo, lo que conduce con frecuencia a arquitecturas sobredimensionadas o ineficientes. EfficientNet argumenta, en cambio, que el desempeño óptimo emerge cuando las tres dimensiones se ajustan de manera coordinada, sustentado en evidencia empírica que demuestra una mejor relación entre precisión y eficiencia.

El método de escalamiento compuesto permite generar una familia completa de modelos —de EfficientNet-B0 a EfficientNet-B7— que mantienen un equilibrio consistente entre capacidad predictiva y uso de recursos. Esta coherencia estructural posibilita alcanzar niveles competitivos de precisión con un número considerablemente menor de parámetros respecto a modelos contemporáneos. EfficientNet se fundamenta en el bloque **MBConv**, derivado de las convoluciones invertidas de MobileNetV2, el cual combina convoluciones separables en profundidad, expansiones de canales y atajos residuales para reducir el costo computacional sin sacrificar capacidad representacional. A ello se suma la integración de módulos **Squeeze-and-Excitation**, que recalibran dinámicamente la contribución de cada canal y fortalecen la identificación de características relevantes.

Las ventajas operativas asociadas a EfficientNet se manifiestan en diversos ámbitos del diseño y la implementación de modelos convolucionales. Entre los aspectos más relevantes destacan:

- **Eficiencia paramétrica**, que permite reducir de manera sustancial el número de parámetros sin comprometer la precisión.
- **Escalabilidad controlada**, posibilitada por el método de escalamiento compuesto que genera variantes ajustadas a distintas restricciones de hardware.
- **Desempeño superior en benchmarks**, donde supera a arquitecturas más voluminosas como ResNet o Inception con un costo computacional considerablemente menor.
- **Adecuación para dispositivos con recursos limitados**, lo que amplía su aplicabilidad en entornos móviles o embebidos.

El impacto de EfficientNet trasciende su propio diseño y ha orientado la evolución de arquitecturas posteriores. Su premisa de que la eficiencia debe constituir un objetivo explícito —y no un efecto secundario del proceso de optimización— ha sido retomada en propuestas como EfficientNetV2, que profundiza en la reducción del tiempo de entrenamiento y mejora el rendimiento en inferencia. Asimismo, como se muestra en la Figura 3.11 la integración de

bloques MBConv, convoluciones separables y mecanismos de atención ha servido como referencia para modelos enfocados en eficiencia energética y despliegue en tiempo real, consolidando un paradigma de diseño centrado en la relación.

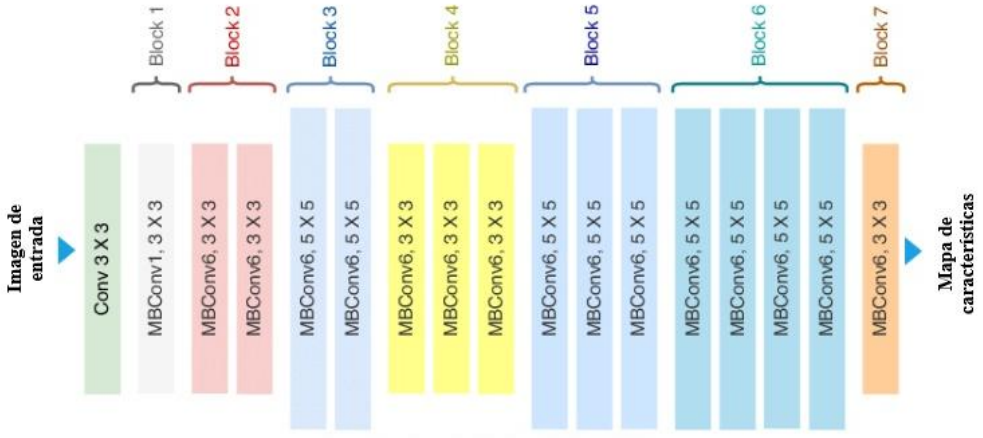


Figura 3.11. Representación de la arquitectura EfficientNet. Adaptado y disponible en: https://iq.opengenus.org/efficientnet/ Consultado en febrero 09, 2026.

En ámbitos donde resulta indispensable equilibrar precisión y costo computacional —como la segmentación médica, la clasificación de imágenes biomédicas o el procesamiento de grandes volúmenes de datos— EfficientNet se posiciona como una alternativa particularmente sólida. Su estructura modular y su capacidad de escalamiento permiten adaptarla a diversos escenarios experimentales sin comprometer la calidad de los resultados, lo que la convierte en una herramienta versátil para investigaciones que requieren modelos robustos y eficientes.

3.3.3. ResNet.

La arquitectura ResNet, propuesta por He et al. en 2015 (He, et al., 2016), marcó un punto de inflexión en el desarrollo de redes neuronales convolucionales al introducir el aprendizaje residual como estrategia para facilitar la optimización de modelos profundos. Esta propuesta surgió como respuesta al problema de *degradación del desempeño*, fenómeno observado en arquitecturas tradicionales donde el incremento en la profundidad no solo dejaba de aportar mejoras, sino que podía deteriorar la precisión del modelo. Para enfrentar esta limitación, ResNet incorpora conexiones de atajo que permiten el flujo directo de información entre capas, mitigando la desaparición del gradiente y posibilitando el entrenamiento estable de redes con cientos de capas.

El principio fundamental de ResNet consiste en redefinir la tarea que cada capa debe aprender. En lugar de aproximar directamente una transformación compleja $H(x)$, los bloques residuales se enfocan en modelar una función $F(x)$ tal que:

$$H(x) = F(x) + x. \tag{3.19}$$

Esta formulación permite que la red aprenda ajustes residuales sobre la identidad, lo que simplifica el proceso de optimización y mejora la estabilidad del entrenamiento. Las conexiones de atajo, al no introducir parámetros adicionales, contribuyen además a mantener la eficiencia del modelo, reforzando la idea de que la profundidad puede incrementarse sin comprometer la capacidad de aprendizaje. La adopción del aprendizaje residual transformó la manera en que se conciben las arquitecturas profundas, demostrando que es posible escalar la profundidad sin incurrir en los problemas que históricamente limitaban a las CNN. Este enfoque no solo permitió el desarrollo de modelos más robustos y precisos, sino que también influyó en arquitecturas posteriores que retomaron y extendieron el concepto residual para mejorar la eficiencia y la estabilidad del entrenamiento.

La arquitectura ResNet se estructura a partir de **bloques residuales** que integran convoluciones, normalización por lotes y funciones de activación ReLU, como se muestra en la Figura 3.12.

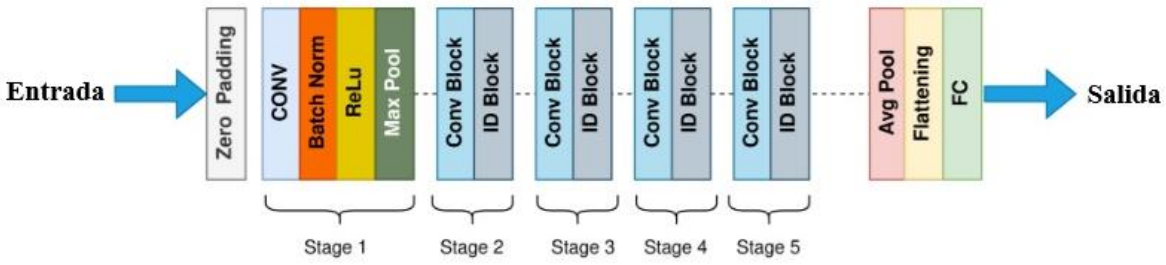


Figura 3.12. Representación de la arquitectura ResNet. Adaptado y disponible en: https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758/ Consultado en febrero 09, 2026.

En función de la profundidad del modelo, estos bloques pueden adoptar una configuración básica o una variante *bottleneck*, diseñada para disminuir el costo computacional en redes de gran escala. Esta organización modular permite ajustar la complejidad del modelo sin comprometer la estabilidad del entrenamiento ni la capacidad representacional. ResNet consolidó su relevancia en el campo del aprendizaje profundo gracias a un conjunto de ventajas que fortalecen su desempeño en tareas complejas:

- **Entrenamiento estable en redes profundas**, las conexiones de atajo facilitan el flujo del gradiente, evitando problemas de desaparición y permitiendo optimizar modelos con un número elevado de capas.

- **Mejor capacidad de generalización**, el aprendizaje de funciones residuales simplifica la optimización y favorece la obtención de representaciones más robustas.
- **Escalabilidad estructural**, variantes como ResNet-18, ResNet-50 o ResNet-152 permiten adaptar la arquitectura a diferentes restricciones computacionales sin alterar su principio fundamental.

Estas características posicionaron a ResNet como una arquitectura de referencia en **clasificación, detección y segmentación de imágenes**, influyendo de manera directa en el diseño de modelos contemporáneos. Su enfoque residual ha sido retomado y extendido en arquitecturas posteriores que buscan mejorar la eficiencia, la estabilidad del entrenamiento y la capacidad de generalización, consolidando así un paradigma que continúa vigente en el desarrollo de redes neuronales profundas.

3.3.4. MobileOne.

La arquitectura MobileOne, presentada por Apple y Meta en 2022 (Vasu et al., 2023), se concibió para superar las limitaciones que enfrentan los modelos de visión por computadora en contextos donde la inferencia debe ejecutarse en tiempo real y en dispositivos con recursos computacionales restringidos. A diferencia de arquitecturas previas orientadas a la eficiencia —como MobileNet o ShuffleNet—, MobileOne adopta un enfoque basado en reparametrización estructural, lo que permite entrenar modelos complejos y expresivos que, durante la fase de inferencia, se transforman en redes equivalentes significativamente más simples y rápidas. Este planteamiento responde a la necesidad de equilibrar capacidad representacional y eficiencia operativa en escenarios de despliegue móvil o embebido.

El principio conceptual que sustenta a MobileOne, como se muestra en la Figura 3.13 consiste en entrenar bloques convolucionales con múltiples ramas paralelas, entre las que se incluyen convoluciones de 3×3, convoluciones de 1×1 y capas de identidad. Estas ramas aportan diversidad de representaciones durante el entrenamiento, lo que facilita la optimización y mejora la capacidad del modelo para capturar patrones complejos. Una vez concluido el proceso de entrenamiento, dichas ramas se fusionan algebraicamente en una única convolución equivalente, generando un modelo compacto que reduce de manera sustancial el costo computacional asociado a la inferencia.

La estrategia de MobileOne demuestra que es posible entrenar arquitecturas ricas en capacidad expresiva sin sacrificar eficiencia en el despliegue, lo que la convierte en una alternativa particularmente relevante para aplicaciones que requieren baja latencia, consumo energético reducido y desempeño consistente en hardware limitado. Su enfoque de reparametrización ha influido en el desarrollo de modelos posteriores que buscan optimizar la brecha entre entrenamiento e inferencia, consolidando un paradigma de diseño centrado en la eficiencia estructural. La arquitectura MobileOne organiza su estructura interna mediante bloques reparametrizables que integran convoluciones, normalización y activaciones ReLU. Durante la fase de entrenamiento, estos bloques incorporan múltiples

ramas paralelas que enriquecen la capacidad representacional del modelo; no obstante, en la fase de despliegue dichas ramas se fusionan en una estructura lineal equivalente, lo que reduce de manera sustancial la complejidad operativa sin alterar la función aprendida. Esta separación explícita entre entrenamiento e inferencia constituye uno de los elementos distintivos de la propuesta.

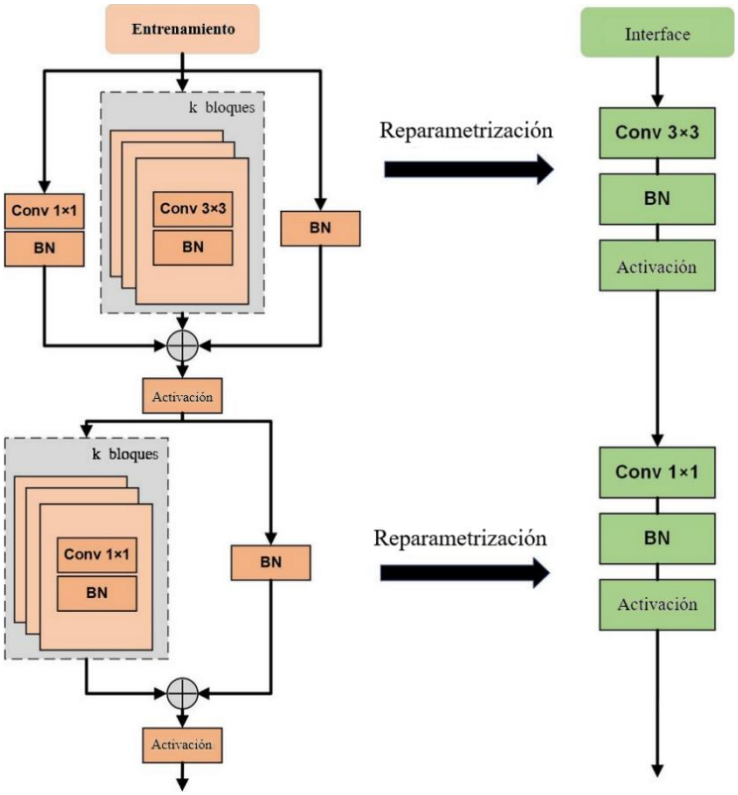


Figura 3.13. Representación del bloque MobileOne, tiene dos estructuras diferentes en el momento del entrenamiento y en el momento de la prueba. Izquierda: bloque MobileOne en el momento del entrenamiento con ramificaciones reparametrizables. Derecha: bloque MobileOne en la inferencia, donde las ramificaciones se reparametrizan. Se utiliza ReLU o SEReLU como activación. El factor de sobreparametrización trivial k es un hiperparámaetro que se ajusta para cada variante. Adaptado y disponible en https://liner.com/review/mobileone-improved-one-millisecond-mobile-backbone Consultado en febrero 09, 2026.

La estrategia de reparametrización adoptada por MobileOne ofrece beneficios concretos para aplicaciones que requieren eficiencia extrema:

- **Inferencia acelerada**, la eliminación de ramas paralelas disminuye el número de operaciones necesarias, reduciendo la latencia en tiempo de ejecución.
- **Compatibilidad con hardware móvil**, la arquitectura final se compone exclusivamente de convoluciones estándar, fácilmente optimizables por aceleradores presentes en dispositivos móviles y sistemas embebidos.

- **Escalabilidad estructural**, las variantes del modelo permiten ajustar profundidad y ancho según las restricciones computacionales del dispositivo objetivo.

Estas características posicionan a MobileOne como una alternativa especialmente adecuada para tareas en tiempo real, como reconocimiento visual en dispositivos móviles, análisis de video en el borde (*edge computing*) o procesamiento en sistemas embebidos donde la eficiencia energética y la baja latencia son requisitos críticos. MobileOne se inscribe en una tendencia creciente hacia arquitecturas que separan explícitamente las fases de entrenamiento e inferencia con el fin de maximizar la eficiencia sin sacrificar capacidad predictiva. Su enfoque de reparametrización ha influido en modelos posteriores y ha demostrado que es posible mantener un alto rendimiento sin renunciar a la simplicidad estructural ni a la velocidad de ejecución. En comparación con arquitecturas como MobileNetV2 o EfficientNet-Lite, MobileOne ofrece una alternativa competitiva cuando la prioridad es minimizar la latencia en dispositivos con recursos limitados, consolidándose como una opción robusta para aplicaciones de visión por computadora en entornos restringidos.

3.4. Modelos o arquitecturas para segmentación semántica.

Existe un amplio consenso en que el entrenamiento exitoso de redes neuronales artificiales profundas requiere muchos miles de muestras de entrenamiento. En esta sección se presentan arquitecturas que recurren intensivamente al aumento de datos para aprovechar de forma más eficiente las muestras disponibles. Estas arquitecturas suelen constar de una ruta de contracción destinada a capturar el contexto y de una ruta de expansión simétrica que posibilita una localización precisa. El uso típico de las redes convolucionales se centra en tareas de clasificación, donde el resultado de una imagen es una única etiqueta de clase. No obstante, en numerosas tareas visuales, especialmente en el procesamiento de imágenes biomédicas, el resultado deseado debe incluir la localización, de modo que se asigna una etiqueta de clase a cada píxel. Además, en las tareas biomédicas rara vez es posible disponer de miles de imágenes de entrenamiento (Ronneberger, 2015).

3.4.1 U-Net

En los últimos años, las redes neuronales artificiales convolucionales profundas han superado el estado del arte en numerosas tareas de reconocimiento visual. Aunque las redes convolucionales existen desde hace décadas, su desempeño estuvo limitado por el tamaño de los conjuntos de entrenamiento disponibles y por la modestia de las arquitecturas consideradas. El empleo más habitual de las redes convolucionales corresponde a tareas de clasificación, en las que la salida asociada a una imagen es una única etiqueta de clase. No obstante, en múltiples problemas visuales, y de manera destacada en el procesamiento de imágenes biomédicas, el objetivo incluye la localización, de forma que se requiere asignar

una etiqueta de clase a cada píxel. Además, en aplicaciones biomédicas rara vez es factible disponer de miles de imágenes de entrenamiento (Ronneberger, 2015).

La arquitectura de la red U-Net se muestra en la Figura 3.14. Dicha arquitectura está compuesta por una ruta de contracción en el lado izquierdo y una ruta de expansión en el lado derecho. La ruta de contracción sigue la arquitectura típica de una red convolucional.

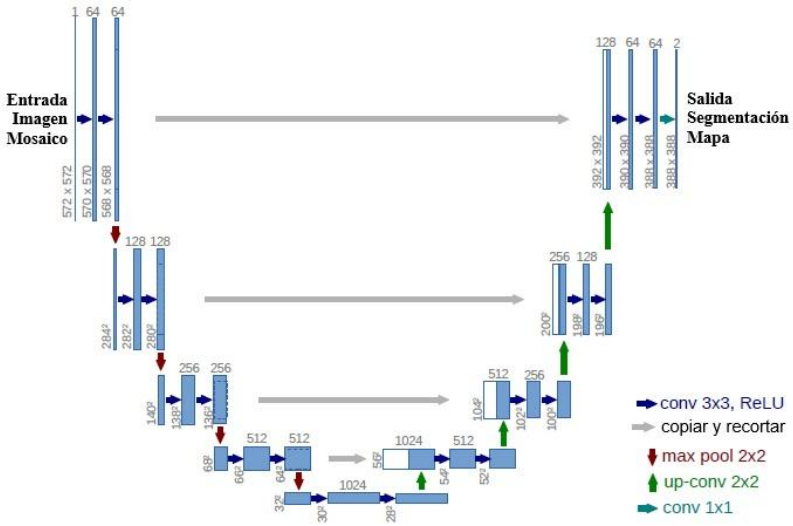


Figura. 3.14. Arquitectura U-net. Cada cuadro azul corresponde a un mapa de características multicanal. El número de canales se indica en la parte superior del cuadro. El tamaño $x - y$ se indica en el borde inferior izquierdo del cuadro. Los cuadros blancos representan mapas de características copiados. Las flechas indican las diferentes operaciones. Adaptado de https://doi.org/10.48550/arXiv.1505.04597.

De forma general, la arquitectura mostrada en la Figura 3.14 se basa en la aplicación repetida de dos convoluciones 3×3 sin relleno, cada una seguida por una función de activación y por una operación de agrupación máxima 2×2 con paso 2 para el submuestreo. En cada etapa de submuestreo se duplica el número de canales de características.

Cada etapa de la ruta expansiva incluye un muestreo ascendente del mapa de características seguido de una up-convolución 2×2 que reduce a la mitad el número de canales de características, la concatenación con el mapa de características recortado correspondiente de la ruta de contracción y dos convoluciones 3×3, cada una seguida de una función de activación. El recorte es necesario debido a la pérdida de píxeles en los bordes ocasionada por las convoluciones sin relleno. En la capa final se emplea una convolución 1×1 para mapear cada vector de características de 64 componentes al número deseado de clases. En total, la red contiene 23 capas convolucionales.

La U-Net combina un camino de contracción destinado a capturar contexto con un camino de expansión simétrico que permite una localización precisa. Emplea convoluciones 3×3 y operaciones de reducción espacial (max-pooling) para reducir la dimensionalidad y aumentar

progresivamente la profundidad de los mapas de características. En la fase de reconstrucción utiliza sobre muestreo seguido de la concatenación con los mapas de la ruta de contracción para recuperar detalle espacial y mejorar la resolución de salida. Para procesar imágenes de gran tamaño, la arquitectura implementa una estrategia de segmentación por superposición que extrapola regiones faltantes mediante reflejado de los bordes.

La U-Net, junto con su estrategia de entrenamiento, constituye una solución eficaz para la segmentación en escenarios con pocos datos; al integrar una arquitectura que preserva la información espacial y técnicas de entrenamiento que mejoran la generalización y la separación de instancias, resulta apropiada para una amplia gama de aplicaciones (Ronneberger, 2015). En el presente trabajo, esta arquitectura sirve como base para la propuesta metodológica: se incorporan funciones de preprocesamiento basadas en polinomios de Chebyshev en los bloques convolucionales del codificador, con el objetivo de mejorar la estabilidad en la generalización del modelo aplicado a imágenes dermatoscópicas de melanomas.

3.4.2 Redes piramidales de características FPN.

El reconocimiento de objetos en rangos de escala significativamente distintos constituye uno de los desafíos centrales en el campo de la visión artificial. En este contexto, las redes piramidales de características (Feature Pyramid Networks FPN) se han consolidado como la base de una solución ampliamente aceptada en la literatura especializada. Dichas arquitecturas se distinguen por su invarianza a la escala, entendida como la capacidad de compensar las variaciones en el tamaño de un objeto mediante el ajuste de su representación en un nivel diferente de la pirámide. Esta propiedad, de carácter estructural, permite que el modelo mantenga un desempeño robusto frente a cambios de escala, al posibilitar la detección de objetos a través de un proceso de escaneo que abarca tanto las posiciones espaciales como los distintos niveles jerárquicos de la pirámide (Lin, 2017).

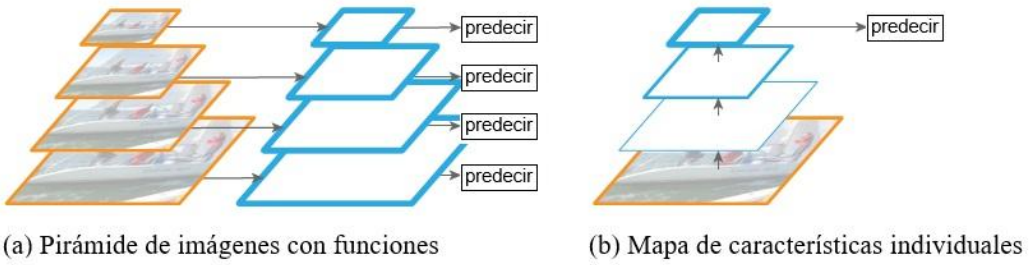


Figura 3.15. Arquitectura de redes piramidales de características. Adaptado de https://doi.org/10.1109/CVPR.2017.106

La Figura 3.15 presenta una representación esquemática de la arquitectura de una FPN. En el apartado (a) se ilustra el enfoque tradicional basado en el uso de una pirámide de imágenes para la construcción de una pirámide de características. En este esquema, las características se calculan de manera independiente en cada una de las escalas de imagen, lo que incrementa de forma considerable el costo computacional y ralentiza el proceso de detección. Por contraste, en el apartado (b) se muestra la estrategia adoptada por los sistemas de detección más recientes, los cuales han optado por emplear únicamente características derivadas de una escala única. Este cambio metodológico responde a la necesidad de optimizar la eficiencia del proceso, ya que permite acelerar la detección sin sacrificar de manera significativa la calidad de los resultados. De este modo, la comparación entre ambos enfoques evidencia la tensión existente entre la exhaustividad del muestreo multiescala y la viabilidad práctica de los sistemas de detección en escenarios de aplicación real.

Las FPN se han consolidado como un componente esencial en los detectores de objetos, dado que requieren un muestreo denso de la escala para alcanzar resultados satisfactorios. Su aplicación no se limita únicamente a la detección, sino que también se extiende a tareas de reconocimiento, en las cuales las características diseñadas manualmente han sido reemplazadas en gran medida por aquellas generadas automáticamente mediante redes convolucionales profundas. Este cambio responde a la capacidad de las FPN para representar semánticas de nivel superior, lo que las hace más robustas frente a la variación de escala y, en consecuencia, facilita el reconocimiento a partir de características calculadas en una única escala de entrada.

No obstante, incluso con dicha robustez, la utilización de pirámides continúa siendo necesaria para alcanzar los niveles más altos de precisión. La principal ventaja de caracterizar cada nivel de una pirámide de imágenes radica en la generación de una representación multiescala en la que todos los niveles mantienen una fortaleza semántica significativa, incluidos aquellos de mayor resolución. Esta propiedad resulta fundamental, ya que asegura que la información espacial detallada no se pierda en los niveles inferiores de la jerarquía, lo cual contribuye a mejorar la exactitud en la detección y el reconocimiento de objetos en contextos complejos.

No obstante, la caracterización de cada nivel en una pirámide de imágenes presenta limitaciones significativas que deben ser consideradas en el diseño de arquitecturas de visión por computadora. En primer lugar, el tiempo de inferencia se incrementa de manera considerable, lo que reduce la viabilidad de este enfoque en aplicaciones de carácter práctico o en escenarios que demandan procesamiento en tiempo real. Asimismo, el entrenamiento de redes profundas de extremo a extremo sobre una pirámide de imágenes resulta inviable desde el punto de vista de los requerimientos de memoria, lo que restringe su implementación en la fase de aprendizaje.

Como consecuencia, cuando se recurre a este tipo de estructuras, las pirámides de imágenes suelen emplearse únicamente durante la etapa de prueba, lo que introduce una discrepancia metodológica entre el proceso de entrenamiento y el de inferencia. Esta inconsistencia compromete la coherencia del modelo y plantea un desafío adicional en términos de generalización y estabilidad de los resultados, lo que refuerza la necesidad de explorar

estrategias alternativas que mitiguen dichas limitaciones sin sacrificar la precisión alcanzada (Lin, 2017).

3.4.3 Red de análisis de escenas piramidal PSPNet.

El análisis de escenas (scene parsing), fundamentado en la segmentación semántica, constituye un eje central en el campo de la visión por computador. Su propósito esencial radica en asignar a cada píxel de una imagen una etiqueta de categoría específica, lo que permite generar una interpretación exhaustiva del entorno visual. A través de este proceso, es posible predecir no solo la etiqueta correspondiente a cada elemento, sino también su ubicación y forma, lo que otorga al modelo una comprensión integral del escenario representado (Zhao, 2017).

Con este propósito, la arquitectura denominada *Pyramid Scene Parsing Network* (PSPNet) se constituye como una herramienta de gran relevancia. Su diseño permite abordar de manera más eficaz los desafíos inherentes al análisis de escenas, al integrar información contextual a múltiples escalas y, de este modo, enriquecer la representación semántica de la imagen. Esta capacidad resulta especialmente valiosa en escenarios complejos, donde la correcta interpretación de los elementos depende no solo de sus características locales, sino también de la relación que guardan con el contexto global. En consecuencia, la PSPNet se ha consolidado como una de las arquitecturas de referencia en segmentación semántica, al ofrecer un equilibrio entre precisión y robustez que la hace idónea para aplicaciones avanzadas en visión por computador.

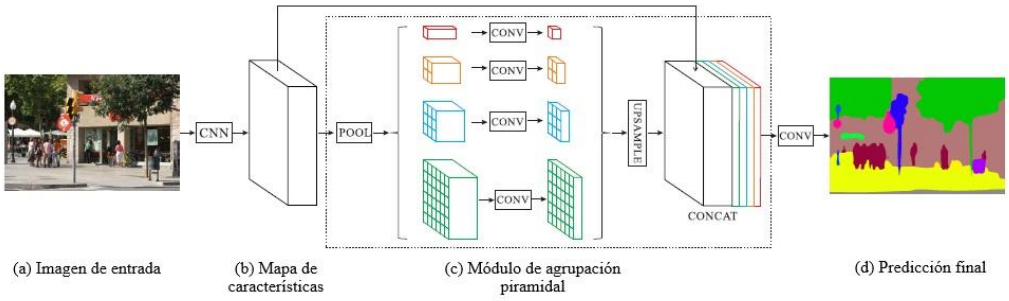


Figura 3.16. Esquema de arquitectura de red de escenas piramidal PSPNet. Adaptado de https://doi.org/10.48550/arXiv.1612.01105

La Figura 3.16 presenta una descripción general de la arquitectura PSPNet. En el apartado (a) se observa la imagen de entrada, la cual constituye el punto de partida del proceso. Posteriormente, en (b), se emplea una red neuronal convolucional (CNN) para extraer el mapa de características correspondiente a la última capa convolucional, lo que permite capturar representaciones de alto nivel. En la etapa (c), se incorpora un módulo de análisis piramidal destinado a recopilar representaciones de subregiones en distintas escalas, seguido

de operaciones de sobre muestreo y concatenación que generan una representación final enriquecida. Esta representación integra información contextual tanto local como global, lo que resulta fundamental para una interpretación más precisa de la escena. Finalmente, en (d), dicha representación se introduce en una capa convolucional adicional con el fin de producir la predicción final a nivel de píxel. Este flujo de procesamiento evidencia la relevancia del módulo piramidal como mecanismo para superar las limitaciones de las CNN convencionales, al dotar al modelo de la capacidad de integrar contexto global sin perder la precisión de los detalles locales. De este modo, la PSPNet se posiciona como una arquitectura robusta y eficiente para abordar los desafíos del análisis de escenas en visión por computador.

La importancia del análisis de escenas se evidencia en la amplitud de sus aplicaciones, que incluyen desde los sistemas de conducción autónoma hasta las plataformas de percepción en robótica, así como otros ámbitos emergentes de alto impacto tecnológico. No obstante, la complejidad que caracteriza a esta tarea se encuentra directamente asociada con la diversidad de contextos y la multiplicidad de etiquetas que deben ser consideradas, lo cual representa un desafío sustancial en términos de generalización y robustez de los modelos de segmentación semántica.

En este sentido, la necesidad de garantizar un desempeño consistente frente a escenarios heterogéneos ha impulsado la búsqueda de arquitecturas y metodologías más eficientes y precisas. De ahí que el desarrollo de enfoques innovadores en esta área se haya consolidado como una prioridad en la investigación contemporánea en visión artificial, dado que constituye un requisito indispensable para avanzar hacia sistemas más confiables y aplicables en entornos reales.

Los métodos basados en redes neuronales convolucionales profundas (CNN) han demostrado avances significativos en la comprensión de objetos dinámicos; sin embargo, aún enfrentan limitaciones derivadas de la diversidad de escenas y de la amplitud de un vocabulario no restringido. Un ejemplo ilustrativo de esta dificultad se observa en la posible confusión entre un bote y un automóvil debido a la similitud en su apariencia visual. No obstante, al considerar el contexto global —como la presencia de una casa flotante o un bote casa en un río— la predicción más adecuada corresponde a la categoría “bote”.

Este tipo de situaciones pone de manifiesto la necesidad de incorporar antecedentes contextuales globales para lograr una percepción de escena más precisa y coherente. El principal obstáculo de los modelos actuales radica en la ausencia de estrategias efectivas que permitan integrar de manera sistemática las pistas de categoría de escena a nivel global. En consecuencia, el desarrollo de mecanismos capaces de aprovechar dicha información contextual se presenta como un desafío prioritario para mejorar la robustez y la capacidad de generalización de los sistemas de segmentación semántica y análisis de escenas (Zhao, 2017).

3.4.4 SegFormer.

SegFormer se presenta como una arquitectura de segmentación semántica de última generación, fundamentada en el uso de transformadores, que integra un codificador jerárquico eficiente (*Mix Transformer*, MiT) junto con un decodificador perceptrón multicapa ligero MLP. Esta propuesta se distingue por dos características esenciales. En primer lugar, incorpora un codificador Transformer estructurado jerárquicamente, capaz de generar representaciones multiescala sin recurrir a codificación posicional. Esta decisión de diseño resulta estratégica, ya que evita la necesidad de interpolar códigos posicionales, un procedimiento que suele provocar una disminución del rendimiento cuando la resolución de prueba difiere de la utilizada en el entrenamiento.

En segundo lugar, SegFormer prescinde de decodificadores complejos y, en su lugar, implementa un decodificador MLP que agrega información proveniente de diferentes capas. De este modo, logra combinar de manera efectiva la atención local y la atención global, lo que se traduce en representaciones más robustas y expresivas. Esta simplicidad estructural, unida a su capacidad para integrar contexto multiescala, convierte a SegFormer en una arquitectura altamente competitiva, al ofrecer un equilibrio entre eficiencia computacional y precisión en la segmentación semántica (Xie, 2021).

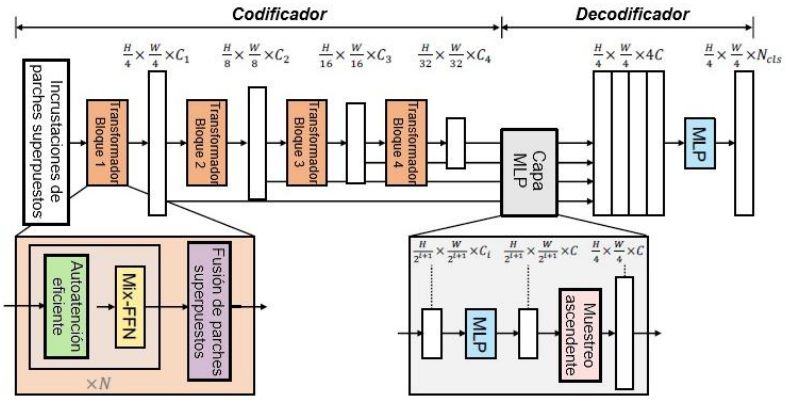


Figura 3.17. La arquitectura SegFormer consta de dos módulos principales: un codificador transformador jerárquico para extraer características generales y detalladas, y un decodificador MLP ligero para fusionar directamente estas características multinivel y predecir la máscara de segmentación semántica. (FFN) indica red de alimentación directa. Adaptado de https://doi.org/10.48550/arXiv.2105.15203

La Figura 3.17 ilustra el funcionamiento del modelo propuesto, en el cual el codificador presenta dos características fundamentales. En primer lugar, evita la interpolación de códigos posicionales al realizar inferencias sobre imágenes con resoluciones distintas a las empleadas durante el entrenamiento. Esta propiedad le permite adaptarse de manera flexible a resoluciones de prueba arbitrarias sin comprometer el rendimiento. Asimismo, la estructura jerárquica del codificador posibilita la generación simultánea de características finas de alta resolución y de características más gruesas de baja resolución, lo que contrasta con el *Vision Transformer* (ViT), limitado a la producción de mapas de características únicos de baja resolución y con resoluciones fijas.

En segundo lugar, el modelo incorpora un decodificador ligero basado en MLP, cuya idea central consiste en aprovechar las representaciones inducidas por el transformador. En este esquema, las atenciones de las capas inferiores tienden a concentrarse en regiones locales, mientras que las capas superiores capturan dependencias altamente no locales. Al integrar la información proveniente de diferentes niveles, el decodificador MLP logra combinar de manera efectiva la atención local y global. Como resultado, se obtiene un decodificador sencillo y directo, capaz de producir representaciones potentes que fortalecen el desempeño del sistema en tareas de segmentación semántica.

La segmentación semántica constituye una tarea esencial dentro del campo de la visión por computadora, ya que habilita un amplio espectro de aplicaciones posteriores. Esta disciplina mantiene una estrecha relación con la clasificación de imágenes, aunque se diferencia de ella en el nivel de granularidad de la predicción: mientras la clasificación asigna una categoría a la imagen completa, la segmentación semántica produce una etiqueta por cada píxel, lo que permite una interpretación más detallada del contenido visual.

Dada esta conexión, numerosos marcos de segmentación semántica de última generación se han desarrollado como variantes de arquitecturas originalmente diseñadas para la clasificación de imágenes en *ImageNet*. En consecuencia, el diseño de arquitecturas troncales continúa siendo un área de investigación activa y estratégica. La evolución histórica de estas estructuras, desde los primeros enfoques basados en VGG hasta las arquitecturas más recientes con bases significativamente más profundas y potentes, ha impulsado de manera decisiva los límites de rendimiento alcanzados en segmentación semántica.

Paralelamente, otra línea de investigación ha formulado la segmentación semántica como un problema de predicción estructurada, orientando sus esfuerzos hacia el diseño de módulos y operadores capaces de capturar de manera eficaz la información contextual. Este enfoque complementa el avance en arquitecturas base, al reconocer que la precisión en la segmentación no depende únicamente de la capacidad de representación de la red, sino también de la integración adecuada del contexto global y local en el proceso de inferencia (Xie, 2021).

3.5. Regularización.

La regularización comprende un conjunto de técnicas destinadas a moderar el sobreajuste en modelos de aprendizaje automático mediante la imposición de restricciones sobre la complejidad del modelo o el proceso de entrenamiento. En términos prácticos, la regularización busca mejorar la generalización del modelo a datos no vistos, aceptando con frecuencia una reducción marginal de la precisión durante el entrenamiento a cambio de una mayor robustez predictiva en escenarios reales. Este intercambio se justifica por la necesidad de priorizar la capacidad del modelo para capturar patrones reproducibles frente a la adaptación excesiva al ruido del conjunto de entrenamiento; por tanto, la regularización

constituye un componente metodológico esencial en diseños experimentales que emplean modelos de alta capacidad (Goodfellow, I., 2017).

La regularización comprende un conjunto de técnicas destinadas a corregir el sobreajuste en modelos de aprendizaje automático y, en consecuencia, a incrementar la capacidad de generalización de un modelo, es decir, su aptitud para generar predicciones precisas sobre conjuntos de datos no observados. Este objetivo se alcanza mediante la imposición de restricciones sobre la complejidad del modelo o sobre el proceso de optimización, lo que habitualmente implica aceptar un aumento del error en el conjunto de entrenamiento a cambio de una mejora en el desempeño sobre datos de prueba. En términos prácticos, los métodos de regularización tienden a producir predicciones menos ajustadas a las características del conjunto de entrenamiento, pero más robustas y fiables en escenarios reales; por tanto, su aplicación constituye una decisión metodológica crítica cuando se busca garantizar la validez externa de modelos de alta capacidad.

La regularización y la optimización constituyen dos conceptos diferenciados pero complementarios en el aprendizaje automático. La regularización se orienta a incrementar la generalización del modelo mediante la imposición de restricciones sobre su complejidad o el proceso de entrenamiento, con el propósito de reducir el sobreajuste a los datos de entrenamiento. La optimización, en cambio, se centra en mejorar la precisión del ajuste durante el entrenamiento, buscando minimizar la función de pérdida para que el modelo reproduzca con fidelidad las observaciones del conjunto de entrenamiento (Krogh, A., 1991).

Desde una perspectiva metodológica, ambos enfoques son indispensables: la optimización garantiza que el modelo aprenda patrones relevantes del conjunto de entrenamiento, mientras que la regularización asegura que ese aprendizaje sea transferible a datos no vistos. En consecuencia, la elección y el balance entre técnicas de optimización y estrategias de regularización constituyen decisiones críticas en el diseño experimental, pues determinan el compromiso entre ajuste y generalización y, por ende, la validez externa de los modelos desarrollados en ciencia de datos.

3.5.1 Compensación entre sesgo y varianza.

El otorgamiento de un mayor error en el conjunto de entrenamiento para reducir el error en datos no vistos se denomina **compensación entre sesgo y varianza**. Este equilibrio constituye un problema central y bien documentado en el campo del aprendizaje automático, pues determina la capacidad del modelo para generalizar más allá de los ejemplos observados.

El **sesgo** se refiere al error sistemático introducido por las suposiciones del modelo sobre la relación entre variables. Un modelo con alto sesgo es excesivamente simple y no captura la complejidad subyacente de los datos, lo que conduce a un rendimiento pobre tanto en entrenamiento como en prueba debido a un subajuste persistente.

La **varianza** cuantifica la sensibilidad del modelo a las fluctuaciones del conjunto de entrenamiento. Un modelo con alta varianza se ajusta demasiado a las particularidades del conjunto de entrenamiento (ruido y excepciones), mostrando un error de entrenamiento bajo pero un rendimiento inestable y degradado en datos nuevos por sobreajuste.

El compromiso entre sesgo y varianza es intrínseco: reducir el sesgo mediante modelos más complejos suele aumentar la varianza, mientras que imponer restricciones para disminuir la varianza incrementa el sesgo. La selección óptima del punto de equilibrio depende del problema concreto, de la cantidad y calidad de los datos disponibles (Hastie, 2009).

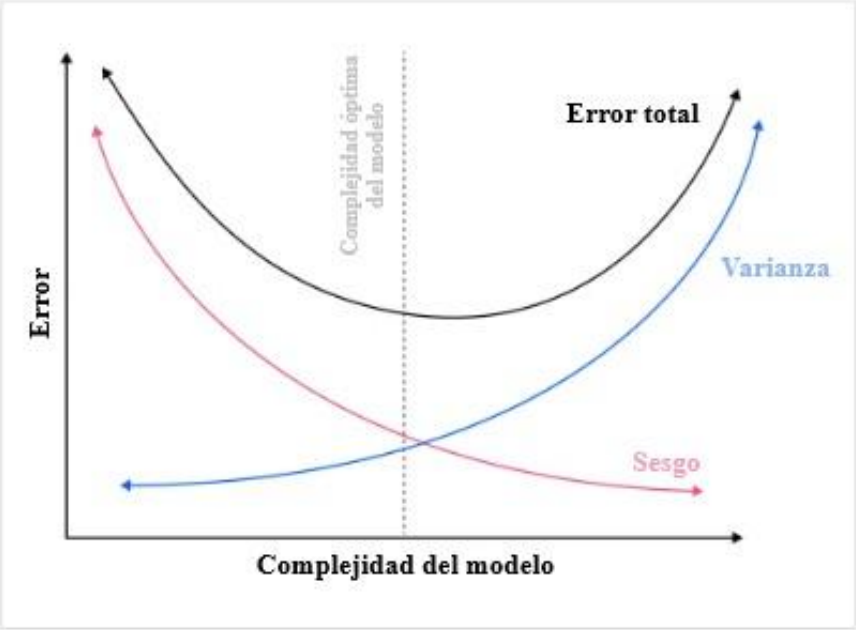


Figura 3.18. Grafica del punto de equilibrio de sesgo – varianza. Adaptado de www.ibm.com/mx-es/think/topics/bias-variance-tradeoff

La existencia de este compromiso se diagnostica comparando métricas de rendimiento entre conjuntos de entrenamiento, validación y prueba. Una brecha amplia entre entrenamiento y prueba sugiere alta varianza; un rendimiento bajo en ambos conjuntos indica alto sesgo. En contextos clínicos, un balance inadecuado puede traducirse en falsos negativos peligrosos o en sobrediagnóstico y procedimientos innecesarios.

Para mitigar el problema se emplean técnicas como regularización, aumento de datos, validación cruzada, selección de modelos y ajuste de la complejidad arquitectónica. Estas estrategias buscan desplazar el punto de operación hacia una región donde el error en datos no vistos sea mínimo, priorizando la robustez y la validez externa del modelo en aplicaciones reales (James, G., 2021).

3.5.2. Tipos de regularización con modelos lineales

La regresión lineal y la regresión logística constituyen modelos predictivos fundamentales en el campo del aprendizaje automático. La **regresión lineal**, frecuentemente estimada mediante mínimos cuadrados ordinarios, persigue cuantificar y predecir el efecto de uno o varios predictores sobre una variable de salida continua, mediante la determinación de la recta (o hiperplano) que minimiza la suma de los errores al cuadrado en el conjunto de entrenamiento. Por su parte, la **regresión logística** se orienta a la estimación de probabilidades de pertenencia a clases discretas; a partir de un conjunto de predictores, el modelo calcula la probabilidad de que una observación pertenezca a una categoría binaria y, en función de un umbral, asigna la etiqueta correspondiente. En términos operativos, la regresión lineal genera predicciones cuantitativas continuas, mientras que la regresión logística produce predicciones categóricas discretas, de modo que ambas técnicas ofrecen herramientas complementarias para tareas de predicción y clasificación en aplicaciones de ciencia de datos (Friedman, J., 2010).

En el contexto de la regresión lineal y del ajuste de modelos, la **regularización** se define como la incorporación de un término de penalización a la función de costo —habitualmente el **Error Cuadrático Medio**— con el propósito de restringir la magnitud de los coeficientes del modelo y, por ende, controlar su complejidad. Esta intervención metodológica resulta esencial cuando existe multicolinealidad entre las variables predictoras o cuando el número de predictores supera al número de observaciones, situaciones en las que las estimaciones ordinarias se vuelven inestables y propensas al sobreajuste. Desde una perspectiva argumentativa, la regularización ofrece un compromiso deseable: sacrificar una parte del ajuste en el conjunto de entrenamiento para obtener estimadores más estables y una mejor capacidad de generalización en datos no vistos, lo que incrementa la validez externa del modelo y su utilidad en aplicaciones empíricas.

Formas principales de regularización.

- **Regularización L2 (Regresión Ridge**): La regresión Ridge, concebida originalmente para mitigar problemas de multicolinealidad severa, incorpora a la función de costo una penalización proporcional al cuadrado de la magnitud de los coeficientes. Su mecanismo operativo consiste en contraer los coeficientes hacia valores próximos a cero sin forzarlos, en la práctica, a anularse de forma exacta. Esta característica produce modelos que conservan la totalidad de las variables (o de las funciones de base, como las de Chebyshev), pero atenúan la influencia de las características menos informativas, distribuyendo el peso entre predictores correlacionados y mejorando así el condicionamiento numérico del estimador.

  Desde una perspectiva metodológica, Ridge ofrece un compromiso favorable entre sesgo y varianza: al introducir un sesgo controlado, reduce la varianza del estimador y, por ende, la sensibilidad del modelo a fluctuaciones del conjunto de entrenamiento,

lo que resulta especialmente valioso en contextos de alta dimensionalidad o colinealidad entre predictores. La ecuación de penalización es:

$$\lambda \sum_{j=1}^{p} \beta_j^2 , \tag{3.20}$$

- **Regularización L1 (Regresión Lasso):** La regresión Lasso (Least Absolute Shrinkage and Selection Operator) incorpora en la función de pérdida una penalización basada en la suma de los valores absolutos de los coeficientes, lo que induce un efecto de contracción (shrinkage*)* con propiedades particulares desde el punto de vista de la selección de variables. A diferencia de la regresión Ridge, la penalización L1 característica de Lasso favorece soluciones **esparsas**, es decir, estimaciones en las que un número significativo de coeficientes β toma el valor exacto de cero; en la práctica, ello convierte a Lasso en un procedimiento de selección automática de características. En el marco de la presente tesis, esta propiedad implica que el modelo puede identificar y descartar por completo aquellos grados del polinomio de Chebyshev que resulten irrelevantes para la representación de la frontera lesionaria, simplificando la parametrización y mejorando la interpretabilidad del modelo. No obstante, conviene señalar que la selección por Lasso puede comportarse de forma inestable ante predictores altamente correlacionados, circunstancia que debe considerarse al diseñar la estrategia de regularización y al interpretar los coeficientes resultantes. La ecuación de penalización es:

$$\lambda \sum_{j=1}^{p} |\beta_j| , \tag{3.21}$$

- **Elastic Net:** Constituye una técnica híbrida de regularización que combina las penalizaciones L1 (Lasso) y L2 (Ridge) con el propósito de integrar la capacidad de selección de variables propia de la primera y la estabilidad numérica característica de la segunda. Su formulación introduce un término de penalización mixto que permite ajustar, mediante un parámetro de mezcla, la contribución relativa de ambas normas; de este modo, Elastic Net favorece soluciones parciales de esparsidad mientras evita la inestabilidad que suele aparecer cuando los predictores están altamente correlacionados. Elastic Net resulta especialmente apropiado en escenarios donde existen grupos de variables que deberían seleccionarse de forma conjunta, ya que tiende a retener conjuntos correlacionados en lugar de elegir arbitrariamente un solo miembro del grupo. La ecuación de penalización es:

$$\lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2, \tag{3.22}$$

La aplicación práctica de estas técnicas en regresión lineal o logística presenta matices importantes: la forma de la función de pérdida, la escala de las variables, el criterio para seleccionar el hiperparámetro de penalización y la interpretación de los coeficientes varían según el tipo de modelo y el objetivo predictivo. En consecuencia, la elección y el ajuste de la regularización deben considerarse decisiones experimentales informadas por validación cruzada, análisis de curvas de aprendizaje y criterios de interpretación clínica, dado que su implementación afecta tanto la precisión en datos de entrenamiento como la robustez en datos no vistos (Hastie, 2009)).

3.6. Medidas de calidad.

En el ámbito del procesamiento de datos visuales, las redes neuronales convolucionales se han consolidado como la arquitectura predominante en la mayoría de las aplicaciones relacionadas con imágenes. Su capacidad para extraer características espaciales jerárquicas las convierte en herramientas idóneas para tareas como la clasificación y segmentación de imágenes. Por otro lado, las redes neuronales recurrentes (RNN) han demostrado una eficacia notable en el tratamiento de datos secuenciales, siendo comúnmente empleadas en el análisis de flujos temporales, tales como secuencias de video o series de imágenes, donde la dependencia temporal entre los datos es crucial.

La selección de métricas de evaluación constituye un aspecto determinante en la interpretación de los resultados obtenidos mediante modelos de aprendizaje profundo. En el presente trabajo, se opta por utilizar las métricas *accuracy* y F_1-*score* para evaluar el desempeño en tareas de clasificación, dado que ambas permiten cuantificar la precisión global y el equilibrio entre precisión y exhaustividad, respectivamente (Liu et al., 2021). En el caso de la segmentación semántica, se recurre a los coeficientes *Intersection over Union* (IoU) y *Dice*, los cuales son ampliamente reconocidos por su capacidad para medir la superposición entre las regiones segmentadas y las referencias manuales, proporcionando así una evaluación más precisa de la calidad de segmentación (Li, 2023). La elección de estas métricas responde a su pertinencia metodológica y a su aceptación en la literatura especializada, lo que garantiza una comparación objetiva y reproducible de los resultados experimentales.

En la evaluación del rendimiento de algoritmos de aprendizaje profundo aplicados al análisis de imágenes, resulta fundamental la selección adecuada de criterios métricos, dado que estos inciden directamente en la interpretación y validez de los resultados obtenidos. Entre los enfoques predominantes, destaca el uso de la notación clásica derivada de la matriz de

confusión, la cual proporciona una base sólida para cuantificar el desempeño en tareas de clasificación y segmentación.

Esta estrategia permite capturar tanto el equilibrio entre clases como la sensibilidad a distribuciones desbalanceadas, lo cual es especialmente relevante en dominios como la medicina, donde ciertas clases pueden representar condiciones críticas, pero poco frecuentes. Así, la elección de estas métricas no solo responde a una tradición metodológica consolidada, sino que también refleja una necesidad analítica de adaptar la evaluación a las características específicas del problema abordado.

En el presente trabajo se adoptan las siguientes notaciones, las cuales permiten representar de manera rigurosa las variables, operadores y métricas involucradas en los distintos procesos de clasificación y segmentación. Esta formalización responde a la necesidad de mantener consistencia terminológica a lo largo del documento, así como de alinear el tratamiento matemático con los estándares utilizados en la literatura especializada.

- TP— verdaderos positivos (True Positives).
- TN— verdaderos negativos (True Negatives).
- FP— falsos positivos (False Positives).
- FN— falsos negativos (False Negatives).
- N— total de muestras/píxeles: $N = TP + TN + FP + FN$.

3.6.1 Exactitud (Accuracy).

La métrica de exactitud (*accuracy*) constituye uno de los indicadores más utilizados en la evaluación del rendimiento de modelos de clasificación en aprendizaje profundo. Esta medida cuantifica la proporción de predicciones correctas respecto al total de instancias evaluadas, ofreciendo así una visión general del desempeño del algoritmo en términos de aciertos globales.

$$Accuracy = \frac{TP + TN}{N}, \tag{3.23}$$

La métrica de exactitud es útil en contextos con clases equilibradas, ya que refleja el porcentaje de clasificaciones correctas. Sin embargo, en escenarios con clases desbalanceadas, como en aplicaciones médicas, puede sobreestimar el rendimiento del modelo al favorecer las clases mayoritarias, lo que limita su capacidad para representar adecuadamente el comportamiento real del sistema.

3.6.2 F_1-Score

El *F1-score* es una métrica que combina precisión (precision) y sensibilidad (recall) mediante su media armónica, y se utiliza para evaluar modelos de clasificación en contextos donde es crucial equilibrar falsos positivos y falsos negativos. A diferencia de otras métricas sensibles a la distribución de clases, el F_1-*score* proporciona una evaluación más robusta del rendimiento cuando se busca un equilibrio entre aciertos y errores en la identificación de elementos relevantes.

$$Recall = \frac{TP}{TP + FN}, \tag{3.24}$$

$$Precision = \frac{TP}{TP + FP}, \tag{3.25}$$

$$F_1 = 2 \cdot \frac{(Precision)(Recall)}{Precision + Recall} = \frac{2\,TP}{2\,TP + FP + FN}, \tag{3.26}$$

El F_1-*score* se presenta como una métrica esencial para evaluar modelos de clasificación en contextos donde la exactitud resulta insuficiente o engañosa. Al combinar precisión y exhaustividad, permite valorar el equilibrio entre la correcta identificación de casos relevantes y la minimización de errores críticos. Su utilidad se acentúa en dominios sensibles como la medicina, la seguridad y la detección de fraudes, donde los falsos positivos y negativos pueden tener consecuencias significativas. Por ello, el F_1-score se ha consolidado como un estándar técnico confiable para comparar modelos y fundamentar decisiones operativas en entornos exigentes.

3.6.3 Intersección sobre Unión IoU

La función Dice Loss (Zhao, 2020) es crucial en la arquitectura U-Net, ya que optimiza la segmentación semántica al manejar mejor el desequilibrio de clases en las imágenes. En general, mide la superposición entre la segmentación predicha y la real, lo que la hace ideal para conjuntos de datos con regiones pequeñas o desbalanceadas. A diferencia de la entropía cruzada, que se utiliza comúnmente para tareas de clasificación, Dice Loss mejora la precisión en los bordes y reduce los falsos negativos.

El coeficiente de Dice (DSC - Dice Similarity Coefficient) se define de la siguiente manera:

$$DSC = \frac{2\left|y_{pred} \cap y_{true}\right|}{\left|y_{pred}\right| + \left|y_{true}\right|}, \tag{3.27}$$

donde,

- y_{pred} son los valores predichos por el modelo.
- y_{true} son los valores reales.

- La multiplicación $y_{pred} \cdot y_{true}$ representa la intersección entre la predicción y la verdad.

El coeficiente de Dice toma valores entre 0 y 1, para los cuales:

- **0** significa que no hay superposición entre la predicción y la verdad.
- **1** significa que hay una coincidencia perfecta.

La función de pérdida se define como:

$$Dice\ Loss = 1 - DSC, \tag{3.28}$$

o de manera equivalente:

$$Dice\ Loss = 1 - \frac{2|y_{pred} \cap y_{true}| + \epsilon}{|y_{pred}| + |y_{true}| + \epsilon}, \tag{3.29}$$

donde ϵ es un pequeño valor para evitar divisiones por cero.

La métrica IoU (Intersection over Union) (Cheng. 2021) es utilizada para evaluar la precisión de modelos de segmentación y detección de objetos en visión por computadora. Se calcula como el área de superposición entre la región predicha por el modelo y la región real en los datos, dividida por el área total combinada de ambas regiones. Matemáticamente, se expresa como:

$$IoU = \frac{|y_{pred} \cap y_{true}|}{|y_{pred}| \cup |y_{true}|} = \frac{TP}{TP + FP + FN}, \tag{3.30}$$

donde la intersección es la zona en común entre la predicción y la región real, y la unión es la suma de ambas áreas, restando la intersección para evitar contarla dos veces.

Su valor varía entre 0 y 1, donde 1 indica una coincidencia perfecta. Un IoU alto significa mejor rendimiento del modelo, generalmente, un umbral de $IoU \geq 0.5$ se considera aceptable. IoU es una métrica fundamental para evaluar la detección de objetos y segmentación de imágenes con CNN. Un valor alto de IoU indica una mejor precisión del modelo, lo que es clave para aplicaciones en el análisis médico de imágenes.

3.7. Polinomios de Chebyshev.

En los últimos años, el interés de la comunidad científica por funciones alternativas a las clásicas ReLU, tanh y sigmoide ha experimentado un incremento significativo. Este fenómeno se encuentra motivado por las limitaciones teóricas y prácticas que dichas funciones presentan en relación con problemas de estabilidad, expresividad, eficiencia y robustez en modelos de aprendizaje profundo. En este marco, las funciones construidas a partir de polinomios de Chebyshev se han consolidado como una alternativa prometedora,

dado que poseen propiedades de aproximación ortogonales ampliamente reconocidas en matemáticas aplicadas, análisis numérico y teoría de la aproximación. Asimismo, estas funciones ofrecen la capacidad de representar comportamientos complejos con un control numérico más preciso sobre las oscilaciones y el condicionamiento, lo que las convierte en un recurso valioso para el diseño de arquitecturas neuronales avanzadas.

3.7.1 Definición.

Los polinomios de Chebyshev pueden definirse de manera rigurosa mediante diversas aproximaciones equivalentes. Entre las más relevantes se encuentran la definición trigonométrica, que establece una relación directa con funciones cosenoidales, y la formulación basada en relaciones de recurrencia, la cual permite construir de manera sistemática cada polinomio a partir de los anteriores. Estas caracterizaciones no solo garantizan la consistencia matemática de la familia de polinomios, sino que también facilitan su análisis en contextos de ortogonalidad, aproximación de funciones y aplicaciones en métodos numéricos.

- **Definición trigonométrica.**

Los polinomios de Chebyshev de **primer tipo**, denotados por $T_n(x)$, se definen de manera formal mediante una relación trigonométrica que establece su conexión con la función coseno. En particular, para $x \in [-1,1]$ y $n \in \mathbb{N}$, se cumple que:

$$T_n(x) = \cos(n \cdot \arccos x). \tag{3.31}$$

Esta formulación permite caracterizar a la familia de polinomios $T_n(x)$ como funciones que preservan propiedades de ortogonalidad en el intervalo $[-1,1]$, además de facilitar su análisis en contextos de aproximación de funciones y métodos numéricos.

De manera análoga, los polinomios de Chebyshev de **segundo tipo**, denotados por $U_n(x)$, se definen mediante una relación trigonométrica que vincula su estructura con la función seno. En particular, para $x \in [-1,1]$ y $n \in \mathbb{N}$, se cumple que:

$$U_n(x) = \frac{\sin\big((n+1) \cdot \arccos x\big)}{\sqrt{1-x^2}}. \tag{3.32}$$

Esta formulación permite caracterizar a los polinomios $U_n(x)$ como una familia ortogonal en el intervalo $[-1,1]$, con propiedades que resultan fundamentales en teoría de la aproximación y en métodos de integración numérica basados en cuadraturas de Gauss-Chebyshev (Szegő, 1975).

- **Relación recursiva.**

La construcción recursiva matemática de los polinomios de Chebyshev de primer tipo se establece mediante una relación que permite generar cada término a partir de los anteriores, garantizando así una definición sistemática de la sucesión polinómica. Dicha relación se expresa de la siguiente manera:

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_{n+1}(x) &= 2T_n(x) - T_{n-1}(x), \ \ \forall\, n \geq 1 \end{aligned} \tag{3.33}$$

Dicha recurrencia permite representar cualquier función continua en el intervalo como una combinación lineal de polinomios de Chebyshev:

$$f(x) \approx \sum_{k=0}^{N} a_k T_{k(x)}, \tag{3.34}$$

donde los coeficientes a_k pueden obtenerse de manera eficiente mediante métodos como la transformada discreta del coseno o la transformada rápida de Fourier (FFT) (Belianskaya, 2024).

La formulación recursiva de los polinomios de Chebyshev de segundo tipo se define a través de una relación matemática que posibilita la obtención de cada término a partir de los precedentes, lo que garantiza una construcción ordenada y coherente de la sucesión polinómica. Esta relación se presenta de la siguiente forma:

$$\begin{aligned} U(x) &= 1 \\ T_1(x) &= 2x \\ U_{n+1}(x) &= 2xU_n(x) - U_{n-1}(x), \ \ \forall\, n \geq 1 \end{aligned} \tag{3.35}$$

Esta formulación recursiva constituye la base para la construcción de los polinomios $U_n(x)$, los cuales presentan propiedades de ortogonalidad en el intervalo $[-1,1]$ y desempeñan un papel relevante en la teoría de aproximación y en métodos numéricos asociados a cuadraturas de Gauss-Chebyshev.

La Figura 3.19 presenta la representación gráfica correspondiente a los primeros cinco polinomios de Chebyshev de primer tipo, ilustrando su comportamiento en el intervalo $[-1,1]$ y evidenciando las características distintivas de cada uno de ellos.

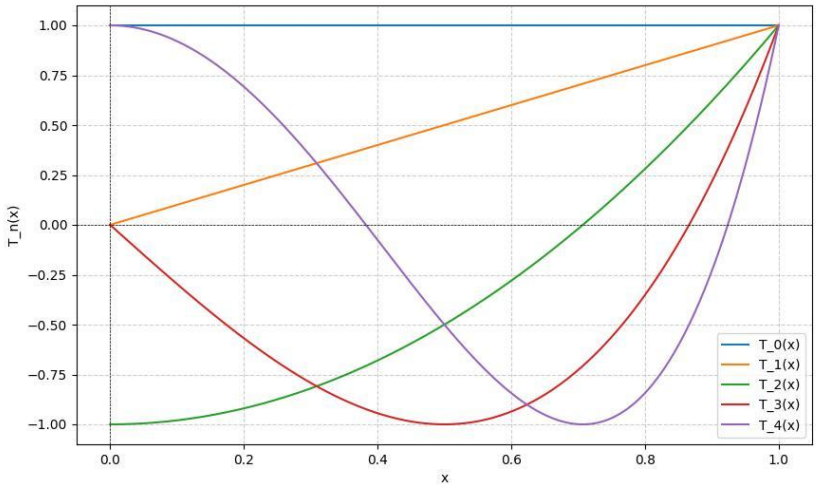


Figura 3.19. Representación gráfica de la función de Chebyshev.

3.7.2 Propiedades.

La ortogonalidad constituye una propiedad esencial de los polinomios de Chebyshev, característica que los posiciona como una base eficiente para la aproximación de funciones en el intervalo $[-1,1]$. Esta propiedad se manifiesta en la existencia de relaciones de ortogonalidad respecto a funciones de peso específicas, las cuales garantizan la independencia lineal y la utilidad de dichos polinomios en métodos numéricos y en la teoría de aproximación.

En el caso de los polinomios de Chebyshev de primer tipo $T_n(x)$, la ortogonalidad se establece con respecto a la función de ponderación $p(x) = \frac{1}{\sqrt{1-x^2}}$, de acuerdo con la siguiente relación:

$$\int_{-1}^{1} \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}dx = \begin{cases} 0, & si\ \ m \neq n \\ \frac{\pi}{2}, & si\ \ m = n \neq 0 \\ \pi, & si\ m = n = 0 \end{cases} \tag{3.36}$$

Las raíces de los polinomios de Chebyshev de primer tipo, $T_n(x)$, se denominan ceros o nodos de Chebyshev y constituyen un elemento esencial en la interpolación polinómica. Estos nodos poseen la propiedad de minimizar el fenómeno de Runge, lo que permite obtener una interpolación más estable y cercana a la mejor aproximación polinómica bajo la norma máxima, conforme al criterio minimax. De manera explícita, las raíces de Chebyshev se definen como:

$$x_k = \cos\left(\frac{2k-1}{2n}\pi\right), \quad k = 1,2,\dots,n. \tag{3.37}$$

La distribución de estas raíces en el intervalo $[-1,1]$ concentra mayor densidad en los extremos, lo que contribuye a reducir las oscilaciones indeseadas en la interpolación y asegura una representación más precisa de la función aproximada.

Por su parte, los polinomios de Chebyshev de segundo tipo $U_n(x)$ son ortogonales respecto a la ponderación $p(x) = \sqrt{1-x^2}$, cumpliendo la relación:

$$\int_{-1}^{1} U_m(x)U_n(x)\sqrt{1-x^2}\,dx = \begin{cases} 0, & si\ \ m \neq n \\ \dfrac{\pi}{2}, & si\ \ m = n \end{cases} \tag{3.38}$$

Estas propiedades de ortogonalidad confieren a los polinomios de Chebyshev un papel central en la construcción de sistemas de aproximación óptimos, en la formulación de cuadraturas de Gauss-Chebyshev y en la resolución de problemas asociados a la expansión de funciones en series polinómicas.

Las raíces de los polinomios de Chebyshev de segundo tipo, $U_n(x)$, se reconocen como nodos de Chebyshev y desempeñan un papel relevante en la teoría de interpolación y en la formulación de cuadraturas de Gauss-Chebyshev. Al igual que en el caso de los polinomios de primer tipo, dichos nodos favorecen la estabilidad de la interpolación polinómica y contribuyen a la reducción de las oscilaciones indeseadas en el intervalo $[-1,1]$.

De manera explícita, los nodos de Chebyshev de segundo tipo se definen como:

$$x_k = \cos\left(\frac{k}{n+1}\pi\right), \quad k = 1,2,\dots,n. \tag{3.39}$$

La distribución de estos nodos presenta una mayor concentración hacia los extremos del intervalo, lo que favorece la precisión en la aproximación de funciones y asegura una representación más robusta en aplicaciones numéricas.

En redes neuronales, el uso de polinomios de Chebyshev como base funcional asigna a cada neurona una activación del tipo:

$$f(x) = \sum_{k=0}^{n} w_k T_{k(x)}, \tag{3.40}$$

donde $T_{k(x)}$ denota el polinomio de Chebyshev de primer tipo y los coeficientes w_k codifican los pesos de la combinación lineal.

En enfoques basados en interpolación polinómica, el valor de salida se obtiene mediante interpolación de Lagrange sobre nodos de Chebyshev. Existen variantes en las que la activación adopta la forma de un Modelo Lineal en Parámetros (LIP), con pesos aprendidos sobre la expansión en polinomios de Chebyshev, así como aproximaciones espectrales empleadas para calcular derivadas de orden superior. La ortogonalidad de los polinomios de

Chebyshev garantiza que las integrales cruzadas sean nulas para polinomios de distinto grado, lo que tiene profundas implicaciones en la estabilidad y el acondicionamiento numérico. (Wang, 2021).

Las funciones basadas en Chebyshev se posicionan como una alternativa avanzada y robusta frente a las funciones clásicas en redes neuronales artificiales. El desarrollo y la aplicación de estas funciones constituyen un campo emergente y fértil, asentado en fundamentos sólidos de la teoría de aproximación, la estabilidad numérica y la eficiencia computacional, con repercusiones en tareas computacionalmente sofisticadas y de creciente demanda en la inteligencia artificial contemporánea. Se observa un aumento en la adopción y la investigación de variantes de Chebyshev, impulsado por la necesidad del aprendizaje automático de abordar aplicaciones cada vez más exigentes en precisión, estabilidad y flexibilidad. En el contexto de la presente investigación, las propiedades de ortogonalidad, recurrencia y minimización del error máximo de los polinomios de Chebyshev de primer tipo fundamentan su integración como función de preprocesamiento en los bloques convolucionales de la arquitectura U-Net, con el propósito de mejorar la estabilidad en la generalización del modelo de segmentación semántica aplicado al diagnóstico de melanomas.

Capítulo 4. Implementación del modelo.

En este capítulo se presenta una propuesta destinada a coadyuvar en la segmentación semántica de imágenes dermatoscópicas de lesiones cutáneas, orientada a favorecer el diagnóstico temprano del melanoma, cuya atención se ve limitada por la escasez de especialistas, el tiempo disponible y el tamaño de las muestras a analizar. Se expone la metodología implementada, compuesta por cuatro pasos, cada uno descrito en una sección independiente, así como la formulación matemática de la función de preprocesamiento basada en polinomios de Chebyshev integrada en los bloques convolucionales de la arquitectura U-Net, y se detallan las tecnologías empleadas a lo largo de la investigación.

4.1. Metodología.

La metodología diseñada para la implementación de la propuesta de solución, estructurada en cuatro pasos se ilustran en la Figura 4.1. Cada etapa ha sido concebida para desarrollar una estrategia por fases que integre los retos técnicos y experimentales vinculados al problema planteado: la primera fase establece los criterios de preprocesamiento y adquisición de datos; la segunda define la arquitectura y los parámetros de entrenamiento; la tercera contempla la validación y ajuste de hiperparámetros; y la cuarta se centra en la evaluación final y la extracción de resultados clínicos relevantes para el apoyo al diagnóstico. La organización en pasos permite una trazabilidad metodológica clara y facilita la replicabilidad del estudio, al tiempo que justifica las decisiones técnicas adoptadas en función de objetivos medibles y criterios de calidad previamente definidos.

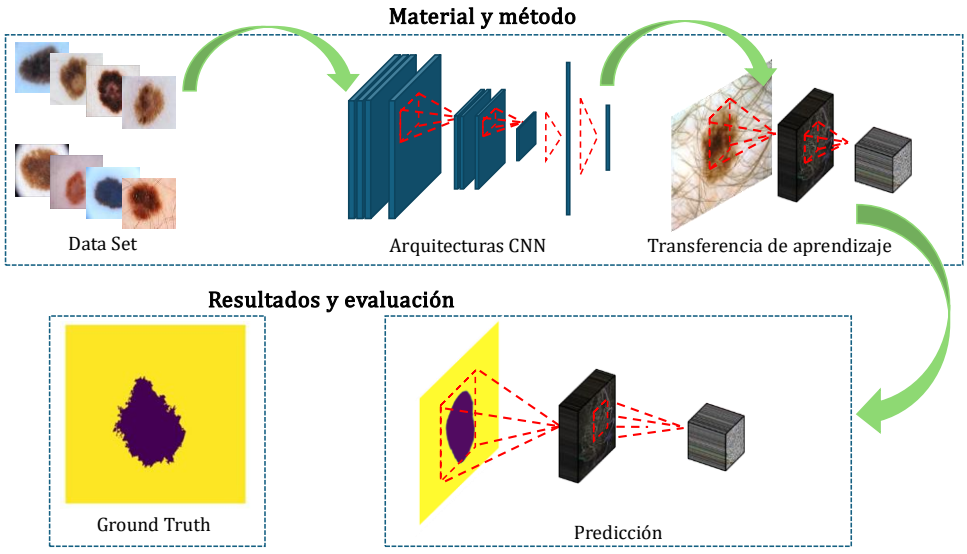


Figura 4.1. Metodología de desarrollo del proyecto.

4.2. Adquisición del conjunto de datos.

El conjunto de datos seleccionado para la presente investigación fue el "International Skin Imaging Collaboration. SIIM-ISIC 2017 Challenge Dataset" (Codella, 2017). Este conjunto incluye **2,600 imágenes** dermatoscópicas de lesiones cutáneas benignas y malignas, divididas en **2,000 imágenes de entrenamiento** y **600 de prueba**. Las resoluciones originales varían desde **540 × 576** hasta **4499 × 6748**; para el procesamiento se aplicó un reescalamiento a **512 × 512**. Se mantuvo la normalización indicada por cada modelo empleado. Las imágenes de entrada se registraron en formato **.jpg** y las máscaras de salida en formato **.png**.

Cada imagen se asocia a un individuo mediante un identificador único de paciente. Los diagnósticos malignos fueron confirmados por histopatología, mientras que los diagnósticos benignos se corroboraron mediante el consenso de expertos, el seguimiento longitudinal o, en algunos casos, histopatología. Existe una publicación exhaustiva que describe las características del conjunto de datos en forma de preimpresión, la cual aún no ha sido sometida a revisión por pares. El conjunto fue generado por la International Skin Imaging Collaboration ISIC y las imágenes provienen de las siguientes instituciones: Hospital Clínic de Barcelona; Universidad Médica de Viena; Memorial Sloan Kettering Cancer Center; Instituto de Melanoma de Australia; Universidad de Queensland; y la Facultad de Medicina de la Universidad de Atenas.

La tabla 02 presenta la cantidad de imágenes destinadas a entrenamiento y prueba agrupadas por año, lo cual permite evaluar la evolución temporal de la base de datos y considerar posibles sesgos cronológicos en el muestreo. Esta desagregación anual facilita el análisis de la representatividad de los datos frente a cambios en protocolos de adquisición o en la prevalencia de lesiones, y constituye un insumo relevante para justificar la validez externa de los resultados de los modelos de segmentación.

Tabla 02. Imágenes de prueba y entrenamiento SIIM-ISIC.

Año	Prueba	Entrenamiento	Datos de salida del entrenamiento
2016	379	900	Sí
2017	600	2,000	Sí
2018	2,512	10,015	No
2019	8,238	25,331	No
2020	10,982	33,126	No

4.3. Análisis y selección del modelo de segmentación semántica.

La segmentación semántica se presenta como un pilar esencial en los sistemas de diagnóstico asistido por computadora (CAD) como lo muestra la Figura 4.2, especialmente en el contexto del análisis de imágenes dermatoscópicas, donde la delimitación exacta de la lesión resulta determinante para la generación de representaciones cuantificables que apoyen la clasificación clínica. En este sentido, la calidad de la segmentación condiciona de manera

directa la fiabilidad de las métricas morfológicas y texturales empleadas en la clasificación clínica, de modo que deficiencias en la delimitación comprometen la validez diagnóstica y la capacidad de generalización del modelo ante nuevas imágenes.



Figura. 4.2. Segmentación semántica.

No obstante, las imágenes dermatoscópicas plantean desafíos técnicos complejos que dificultan la obtención de máscaras segmentadas robustas: bordes difusos, bajo contraste entre la lesión y la piel circundante y la presencia de artefactos como vello o burbujas de gel. Estas características incrementan la variabilidad y reducen la eficacia de los métodos clásicos de procesamiento de imágenes, lo que justifica la búsqueda de enfoques más adaptativos y generalizables.

4.3.1. Arquitectura empleada.

En este escenario, las redes neuronales convolucionales han mostrado un rendimiento superior frente a las técnicas tradicionales, al ofrecer capacidad de aprendizaje jerárquico de características y robustez frente a variaciones de apariencia. Dentro del conjunto de arquitecturas basadas en CNN, la U-Net, propuesta por Ronneberger et al. (2015), se ha consolidado como referencia para la segmentación de imágenes biomédicas, dado su diseño que facilita la recuperación de detalles espaciales y la integración de contextos locales y globales, lo que la posiciona como la opción preferente en estudios orientados a la segmentación dermatoscópica.

la U-Net se caracteriza por una topología en forma de U que integra una **ruta de contracción** —el codificador— destinada a capturar el contexto global, y una **ruta de expansión** simétrica —el decodificador— orientada a la localización precisa de estructuras en la imagen. En el contexto de la dermatoscopía, la preservación de la integridad de los bordes de las lesiones pigmentadas constituye un requisito indispensable para evaluar asimetría y estructura. La U-Net, al conservar y reintroducir información de alta resolución, contribuye a mantener la definición de dichos bordes frente a artefactos y bajo contraste.

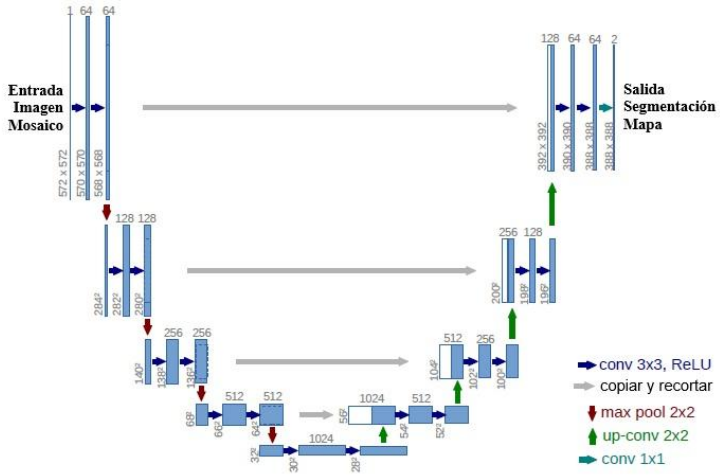


Figura. 4.3. Arquitectura U-net. Adaptado de https://doi.org/10.48550/arXiv.1505.04597.

4.3.1.1. Entrada U-Net.

La entrada de la red U-Net consiste en una imagen representada como una matriz de píxeles, la cual puede variar en dimensiones espaciales y en el número de canales de color según la aplicación concreta. Estas variaciones no son meros detalles formales, sino que condicionan tanto el diseño arquitectónico como las etapas de preprocesamiento y normalización, dado que la estructura de los filtros y la compatibilidad entre capas dependen directamente de las dimensiones y canales de la entrada.

Para la aplicación considerada, la entrada corresponde a una imagen dermatoscópica con resolución de 512×512 píxeles en formato RGB, como se muestra en la Figura 4.4, por lo que la matriz de entrada presenta tres canales.

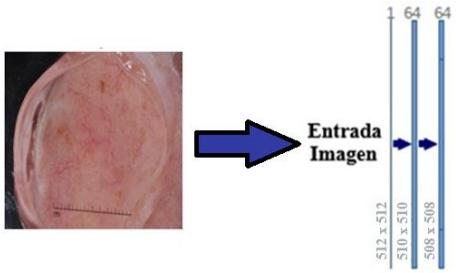


Figura. 4.4. Imagen representativa de la entrada a la arquitectura U-Net.

4.3.1.2. Parte contractiva (codificador).

La ruta de contracción (codificador) de la arquitectura U-Net tiene como función principal la captura de características y del contexto espacial de la imagen de entrada. Esta etapa se estructura en una secuencia de **bloques convolucionales** que aplican filtros o kernel para extraer representaciones a distintos niveles de abstracción, incrementando progresivamente la complejidad semántica de las características extraídas.

Cada bloque suele combinar **operaciones de convolución**, **función de activación** y una **regularización**, lo que permite estabilizar el aprendizaje y mejorar la discriminación de patrones relevantes, mediante operaciones de **submuestreo**.

La **innovación propuesta** consiste en introducir, antes de la aplicación del bloque convolucional en la arquitectura U-Net, una función basada en **polinomios de Chebyshev** que actúa sobre cada matriz de entrada. En este esquema, cada matriz de píxeles se evalúa sucesivamente en un polinomio de Chebyshev, de modo que la transformación genera representaciones polinomiales que capturan información espectral y estructural de la señal de entrada. Esta incorporación busca mejorar la estabilidad en la generalización del modelo al introducir una base funcional ortogonal con propiedades de minimización del error máximo, lo que contribuye a reducir la propensión al sobreajuste durante el entrenamiento.

La operación subsecuente se diseña explícitamente para garantizar que la función resultante permanezca dentro del dominio [-1,1]. Esta restricción se implementa mediante una normalización adecuada de las evaluaciones polinomiales que preserva la información relativa entre coeficientes mientras asegura la compatibilidad con las condiciones de convergencia y las hipótesis de estabilidad de los polinomios de Chebyshev.

Definimos la siguiente función:

$$\hat{A}_{in} = \tanh(A_{in}), \tag{4.1}$$

donde A_{in} son las características del bloque anterior, esto es para garantizar que la función este en el dominio $[1,1]$. La función empleada está dada por:

$$\mathcal{C} = \frac{\sum_{k=1}^{K} T_k(\hat{A}_{in})}{K}, \tag{4.2}$$

donde, T_k son los polinomios de Chebyshev para el grado $\mathrm{k} = 1,2,3$. El cociente entre K ajusta el dominio de la función. Con la ecuación anterior aplicamos la convolución:

$$A_{out} = \mathcal{C} * W, \tag{4.3}$$

donde, $W = [w_1,\ w_2, \dots, w_n]$ es la matriz de pesos. La forma continua de la convolución está dada por:

$$A_{out}(t) = \int_{-\infty}^{\infty} \mathcal{C}(t) * W(t-\tau)d\tau. \tag{4.4}$$

Mantener la salida en dicho intervalo facilita la integración con funciones de activación y esquemas de inicialización empleados en las capas convolucionales, reduciendo la probabilidad de saturación y mejorando la estabilidad del entrenamiento. Metodológicamente, la incorporación de esta función previa al bloque convolucional persigue los siguientes objetivos:

1. Aprovecha las propiedades matemáticas de los polinomios para una mejor aproximación de funciones, contribuyendo a que el modelo logre representaciones espectrales y texturales más estables y generalizables ante la variabilidad propia de las imágenes dermatoscópicas..
2. Son una secuencia de polinomios ortogonales conocidos por su propiedad «minimax», lo que significa que minimizan el error máximo al aproximar una función dada.
3. Permitir a la red modelar relaciones no lineales complejas en los datos de una manera más estructurada y controlada.
4. Las propiedades de ortogonalidad y recurrencia de los polinomios de Chebyshev contribuyen a una mejor estabilidad numérica durante el entrenamiento, lo que ayuda a abordar problemas como la desaparición de gradientes o la explosión del flujo de datos que pueden afectar a otros métodos basados en polinomios.

En consecuencia, esta etapa preliminar se plantea como un mecanismo de preprocesamiento aprendido o parametrizado que contribuye a una segmentación más robusta y reproducible en contextos biomédicos con variabilidad de contraste y presencia de elementos diferentes a las lesiones cutáneas (vello, burbujas de gel, reflejos o marcas de calibración).

El siguiente paso es la operación de **convolución** en la ruta de contracción de la U-Net, constituye el mecanismo fundamental mediante el cual la red extrae características locales y construye representaciones jerárquicas de la imagen de entrada.

Definición: Sea $\boldsymbol{I} \in M_{m,n}$ una matriz de valores que representa una imagen de entrada, y sea $\boldsymbol{K} \in M_{p,q}$ un filtro o kernel convolucional, donde $p < m$ y q$< n$. La operación de convolución discreta entre $\boldsymbol{I}$ y $\boldsymbol{K}$, denotada por $\boldsymbol{I} * \boldsymbol{K}$, produce una matriz de salida $\boldsymbol{O} \in M_{(m-p+1),(n-q+1)}$, definida por:

$$\boldsymbol{O}_{i,j} = \sum_{u=0}^{p-1} \sum_{v=0}^{q-1} \boldsymbol{I}_{(i+u),(j+v)} \cdot \boldsymbol{K}_{u,v}\text{, para } 0 \le i \le m-p,\ 0 \le j \le n-q. \qquad (4.5)$$

La expresión de la definición corresponde a una convolución válida sin **relleno** (**padding**), y con **paso** (**stride)** unitario. Son dos parámetros fundamentales que controlan cómo el filtro (kernel) "camina" sobre la imagen y cómo cambia el tamaño de la imagen resultante.

El relleno es el proceso de agregar píxeles extra (generalmente con valor 0) alrededor de los bordes de la imagen de entrada antes de pasar el filtro. El paso es la cantidad de píxeles que se desplaza el filtro en cada movimiento sobre la imagen. En caso de aplicar relleno o modificar el paso, la dimensión de salida se ajusta según la siguiente ecuación:

$$\dim(\boldsymbol{O}) = \left(\frac{m - p + 2s}{t}\right) \times \left(\frac{n - q + 2s}{t}\right), \tag{4.6}$$

con relleno $s \geq 0$ y paso $t \geq 1$.

En la figura 4.5 se presenta un ejemplo representativo de la operación de convolución empleando un **filtro 3×3**, **sin relleno** y con **paso unitario**, ilustrando el desplazamiento del núcleo sobre una matriz de entrada y la formación del mapa de activación resultante.

La operación central del modelo es la convolución, la cual actúa como un extractor de características automatizado. Matemáticamente, consiste en el deslizamiento de un filtro (kernel) K de dimensiones 3×3 sobre la imagen de entrada, calculando el producto punto en cada posición espacial.

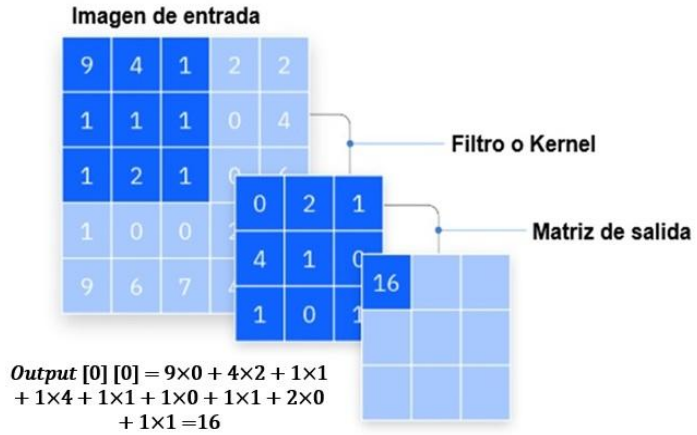


Figura. 4.5. Convolución. Creación propia.

La convolución es una operación lineal que combina multiplicaciones y sumas para producir mapas de activación a partir de la entrada. Si se apilaran exclusivamente capas convolucionales lineales, la composición resultante seguiría siendo una transformación lineal global; en términos algebraicos, la concatenación de operadores lineales equivale a un único operador lineal. En consecuencia, una red compuesta únicamente por convoluciones sería análoga a una regresión lineal de gran dimensión y su capacidad discriminativa se limitaría a separar clases mediante fronteras afines o, en el plano, mediante líneas rectas.

Tras la aplicación de la operación de convolución en la ruta contractiva de la U-Net, se emplea la **función de activación ReLU** figura 4.6. La limitación de la convolución adquiere relevancia práctica en el dominio de la dermatoscopia, donde las lesiones cutáneas (nevos, melanomas) exhiben contornos curvos, formas irregulares y variabilidad morfológica compleja que no pueden ser descritas adecuadamente por separadores lineales simples. Por

tanto, la arquitectura debe incorporar mecanismos que permitan modelar relaciones no lineales entre las características extraídas y las etiquetas de interés para representar contornos y estructuras orgánicas con fidelidad.

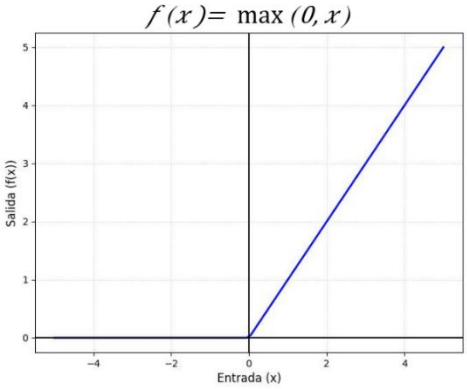


Figura. 4.6. Función ReLU. Creación propia.

La función de activación ReLU introduce la no linealidad necesaria. Al anular las activaciones negativas y conservar las positivas, en términos geométricos, la presencia de ReLU permite que la red "pliegue o curve" el espacio de características, combinando regiones lineales locales para aproximar contornos curvos y estructuras irregulares; de este modo, la red puede delinear perímetros complejos alrededor de una lesión cuya forma no se ajusta a figuras geométricas simples. Por estas razones, la incorporación de la función de activación no lineales ReLU es un requisito metodológico para que las arquitecturas convolucionales alcancen la expresividad necesaria en tareas de segmentación biomédica donde la representación de contornos y formas orgánicas es determinante para la validez diagnóstica.

Tras la aplicación de la función de activación, se procede a una **reducción de dimensionalidad** mediante la implementación de la operación de **muestreo máximo** (**Max Pooling**). Esta capa opera deslizando una ventana 2×2 sobre el mapa de características con un paso de 2 Figura 4.7, seleccionando en cada posición el valor de activación máximo y descartando las demás muestras.

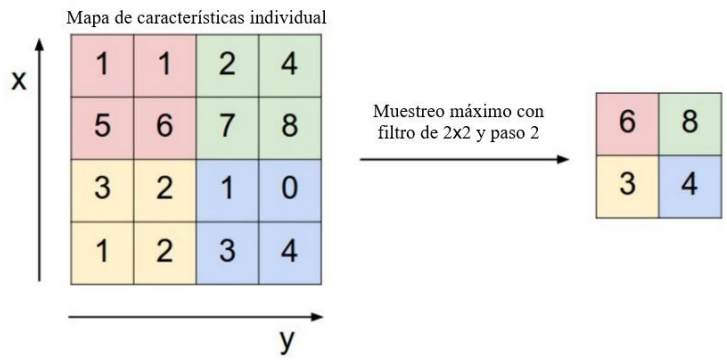


Figura. 4.7. Imagen representativa del muestreo máximo.

Desde una perspectiva metodológica, el muestreo máximo cumple dos funciones complementarias: por un lado, reduce la resolución espacial de los mapas de activación, lo que disminuye la carga computacional y la sensibilidad a pequeñas traslaciones; por otro, incrementa el campo receptivo efectivo de las neuronas, favoreciendo la integración de contexto local. Matemáticamente, si x es la entrada de una región local, la salida y se define como $y = \max(x_{i,j})$.

La elección de Max Pooling sobre otras técnicas de submuestreo se justifica por su capacidad para preservar las características texturales y de bordes más prominentes de la lesión dermatológica, proporcionando a su vez invarianza a pequeñas traslaciones y distorsiones en la imagen de entrada. Además, esta operación expande progresivamente el campo receptivo de la red, permitiendo que las capas profundas del codificador capturen el contexto global semántico necesario para distinguir la lesión del tejido cutáneo circundante.

La arquitectura U-Net aborda de manera efectiva la compensación entre abstracción semántica y precisión de localización mediante la incorporación de **conexiones de salto** (**skip connections**) Figura 4.8, las cuales permiten transferir información espacial de bajo nivel desde la ruta de contracción hacia la ruta de expansión, facilitando así la recuperación de detalles finos sin sacrificar la capacidad de representación global. En la segmentación computarizada de lesiones cutáneas, uno de los desafíos más relevantes consiste en la delimitación precisa de bordes difusos y en la identificación de patrones texturales sutiles cuando existe un bajo contraste respecto a la piel circundante. Esta dificultad compromete tanto la exactitud morfológica como la robustez del modelo frente a variaciones intrínsecas de la imagen.

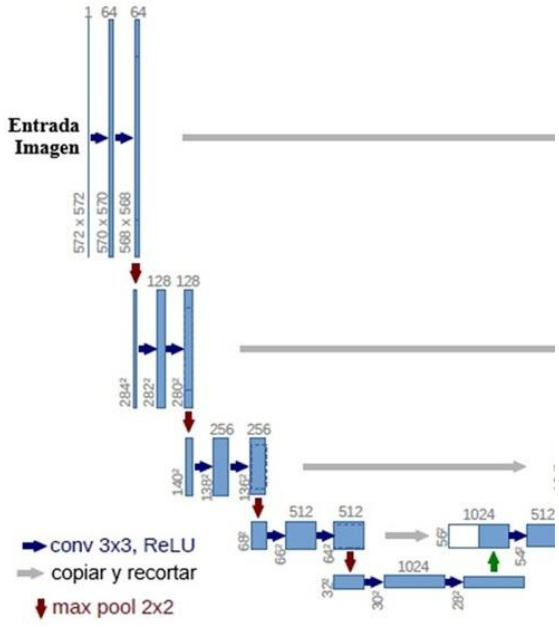


Figura. 4.8. Imagen representativa de la conexión de salto (líneas grises).

Las conexiones de salto actúan como puentes que transfieren mapas de características de alta resolución desde las capas de codificación hacia las etapas expansivas (**decodificador**). En el ámbito del diagnóstico de cáncer de piel, este mecanismo adquiere una importancia crítica, puesto que posibilita la recuperación de información granular que se pierde durante el submuestreo, como las irregularidades en los bordes de la lesión y las variaciones pigmentarias más sutiles. Al concatenar dicha información espacial de bajo nivel con el contexto semántico global proporcionado por las capas profundas, el modelo mejora sustancialmente la precisión de las máscaras de segmentación, obteniendo márgenes más exactos; esta mejora no es meramente técnica, sino que constituye un factor determinante para la evaluación clínica de asimetrías y de áreas sospechosas de malignidad en melanomas y carcinomas.

4.3.1.3. Cuello de botella (Bottleneck).

El cuello de botella Figura 4.9, situado entre la fase de contracción y la de expansión en la arquitectura U-Net, constituye el componente en el que se alcanza la máxima compresión de la información y el mayor grado de abstracción semántica. Se puede sostener que, en esta etapa, la reducción extrema de la resolución espacial —concomitante a un incremento en la profundidad de los mapas de características— obliga al modelo a condensar la entrada en una representación de la información comprimida y esencial que queda después de filtrar todo lo innecesario. Esta compresión no es un mero efecto colateral del diseño arquitectónico, sino un requisito funcional, al limitar la capacidad espacial, el cuello de botella promueve la extracción de rasgos de alto nivel y la supresión de detalles redundantes, facilitando así tareas posteriores de reconstrucción y segmentación.

En el contexto del procesamiento de imágenes dermatoscópicas, el **cuello de botella** adquiere un papel decisivo para la comprensión contextual de la patología, pues constituye la etapa en la que la red sintetiza la información visual en su forma más abstracta. Al operar sobre una representación espacialmente reducida y semánticamente enriquecida, esta fase facilita la identificación de patrones globales y características de alto nivel —por ejemplo, la asimetría estructural o la distribución general del pigmento— que resultan determinantes para la caracterización clínica de las lesiones.

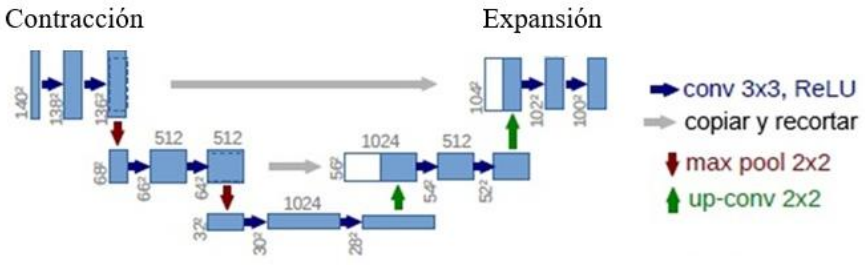


Figura. 4.9. Imagen representativa del cuello de botella.

Asimismo, al priorizar rasgos de baja frecuencia semántica frente a detalles de alta frecuencia, el cuello de botella contribuye a la supresión de artefactos y ruido irrelevante —como vellos, burbujas de gel o texturas cutáneas no diagnósticas—, lo que incrementa la robustez de la representación latente frente a variaciones superficiales de la imagen. Esta capacidad de abstracción no solo mejora la discriminación entre patrones patológicos y no patológicos, sino que también reduce la propensión del modelo a sobre ajustarse a detalles espurios presentes en el conjunto de entrenamiento. Desde una perspectiva metodológica, ello implica que el diseño y la parametrización de esta etapa deben orientarse explícitamente a preservar la información semántica relevante y a mitigar la influencia de ruido, dado que dichas decisiones arquitectónicas condicionan de manera directa la validez clínica y la generalización del sistema.

4.3.1.4. Parte expansiva (decodificador).

La **fase expansiva** se describe como el componente encargado de la reconstrucción espacial de la imagen y de la generación de la máscara de segmentación final en tareas de segmentación de lesiones cutáneas en imágenes dermatoscópicas. Su objetivo fundamental consiste en restaurar de manera progresiva la resolución espacial original de la entrada a partir de representaciones profundas y de baja resolución obtenidas en la fase de codificación. Este proceso no es meramente inverso al de contracción; implica decisiones arquitectónicas y operativas que condicionan la fidelidad espacial y la coherencia semántica de la salida.

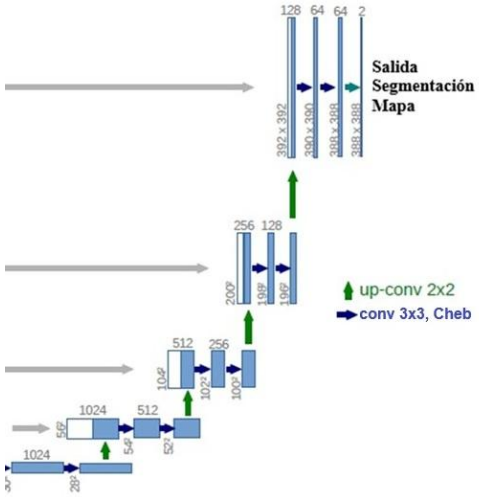


Figura. 4.10. Imagen representativa de la parte expansiva de la U-Net.

En el ámbito del análisis dermatoscópico, la arquitectura U-Net se emplea preferentemente para la segmentación automática debido a su capacidad para combinar información contextual y detalles locales, y la fase expansiva desempeña un papel central en esta integración.

Al fusionar mapas de características de distinta resolución mediante operaciones de **sobremuestreo** (**upsampling**) y concatenación, la fase expansiva transforma representaciones abstractas en una máscara precisa a nivel de píxel, lo que resulta determinante para la extracción posterior de rasgos clínicos relevantes como asimetría, bordes, color y diámetro (Celebi et al., 2015).

Desde una perspectiva metodológica, la eficacia de la fase expansiva depende de la configuración de sus bloques (método de sobremuestreo, uso de conexiones de salto, normalización y funciones de activación) y del equilibrio entre preservación de detalles y supresión de artefactos. Por tanto, la parametrización de esta etapa debe justificarse en función de los objetivos diagnósticos y de las características del conjunto de datos, dado que decisiones subóptimas pueden inducir errores de segmentación que comprometan la validez clínica y la generalización del modelo.

Cada bloque de la fase expansiva comienza con una operación de sobremuestreo, habitualmente implementada mediante **convoluciones transpuestas** (**transposed convolutions**). Dicha operación tiene por finalidad incrementar la resolución espacial de los mapas de características para compensar la pérdida de información derivada del submuestreo efectuado durante la fase de codificación (Long et al., 2015). No se trata únicamente de recuperar dimensiones espaciales, sino de reconstituir una estructura morfológica coherente que permita la fusión efectiva de información contextual y detalles locales en etapas posteriores. En consecuencia, la selección y parametrización del método de sobremuestreo constituyen decisiones arquitectónicas críticas, pues condicionan la fidelidad espacial de la reconstrucción y la propensión a introducir artefactos que pueden afectar la precisión de la máscara de segmentación.

En los enfoques de segmentación semántica que operan a nivel de píxel, resulta deseable mantener la congruencia entre las dimensiones espaciales de la entrada y la salida. En consecuencia, la arquitectura U-Net y el diseño del flujo de datos priorizar mecanismos que conservan la alineación espacial entre entrada y salida para garantizar una supervisión más directa y una mayor fidelidad en la segmentación. En particular, la convolución transpuesta es un mecanismo para invertir las operaciones de submuestreo implementadas por la convolución. La convolución transpuesta realiza un mapeo que aprende y que incrementa la resolución mediante parámetros entrenables, lo que la diferencia de técnicas de sobremuestreo no paramétricas y le confiere mayor flexibilidad para modelar estructuras espaciales complejas.

La exposición considera la **operación elemental de convolución transpuesta** con paso igual a uno y sin relleno. Se parte de un tensor de entrada de dimensiones $n_h \times n_w$, y de un núcleo convolucional de tamaño $k_h \times k_w$. Al deslizar la ventana del núcleo con un paso de 1 a lo largo de cada fila y columna, se generan $n_h n_w$ posiciones de aplicación del núcleo, lo que da lugar a un conjunto de resultados intermedios cuya naturaleza y tamaño se describen a continuación.

Cada uno de los $n_h n_w$ resultados intermedios se representa por un tensor de dimensiones $(n_h + k_h - 1) \times (n_w + k_w - 1)$ inicializado en cero. El cálculo de un tensor intermedio concreto se efectúa multiplicando cada elemento del tensor de entrada por el núcleo; el producto entre un escalar de entrada y el núcleo produce un bloque de tamaño $k_h \times k_w$ que sustituye una subregión del tensor intermedio correspondiente. Finalmente, la salida de la operación se obtiene mediante la suma elemento a elemento de todos los tensores intermedios generados. Este procedimiento de acumulación integra las contribuciones locales producidas por cada elemento de entrada y por el núcleo, reconstruyendo así una salida de mayor resolución espacial que invierte, en términos operativos, el efecto de submuestreo introducido por convoluciones con paso mayor que uno.

La Figura 4.11 ilustra el cálculo de la convolución transpuesta empleando un núcleo de 2×2 sobre un tensor de entrada 2×2. Al acumular las cuatro contribuciones mediante una suma elemento a elemento, se obtiene una salida de dimensiones 3×3, coherente con la relación $(n_h + k_h - 1) \times (n_w + k_w - 1)$.

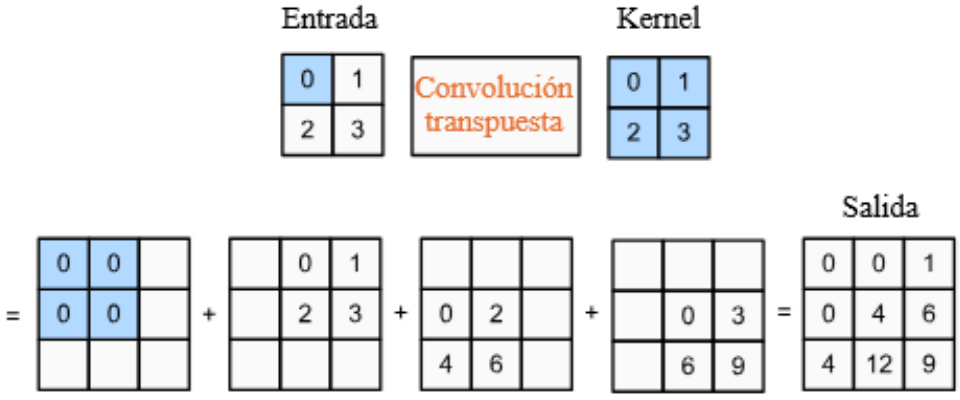


Figura. 4.11. Imagen representativa de la convolución transpuesta con un kernel de 2×2. Las partes sombreadas son una parte de un tensor intermedio, así como los elementos del tensor de entrada y del núcleo utilizados para el cálculo

En imágenes dermatoscópicas, el **sobremuestreo** desempeña un papel esencial al proyectar patrones semánticos de carácter global sobre una representación espacial más fina, lo que facilita la reconstrucción precisa de estructuras complejas como bordes irregulares y regiones de pigmentación heterogénea (Codella et al., 2018). Al incrementar la resolución de los mapas de características, el sobremuestreo permite que la información contextual extraída en niveles profundos se traduzca en contornos y texturas coherentes a nivel de píxel, favoreciendo así una segmentación que refleje la morfología clínica de la lesión en lugar de limitarse a la continuidad local de intensidad.

La arquitectura U-Net se caracteriza por la incorporación de **conexiones de salto** (**skip connections**) que enlazan de forma directa los mapas de características generados por el codificador con los niveles homólogos del decodificador que mantienen la misma resolución espacial (Ronneberger et al., 2015). Estas conexiones permiten preservar información espacial fina perdida durante las operaciones de reducción de resolución, facilitando la fusión de rasgos de bajo nivel con representaciones semánticas de alto nivel. Desde una perspectiva funcional, las conexiones de salto mejoran la capacidad de localización y reconstrucción del modelo y contribuyen a una optimización más estable al favorecer el flujo de gradiente entre las etapas profundas y superficiales de la red. Por tanto, su presencia constituye un elemento estructural clave que explica la eficacia de U-Net en tareas de segmentación y reconstrucción de imágenes.

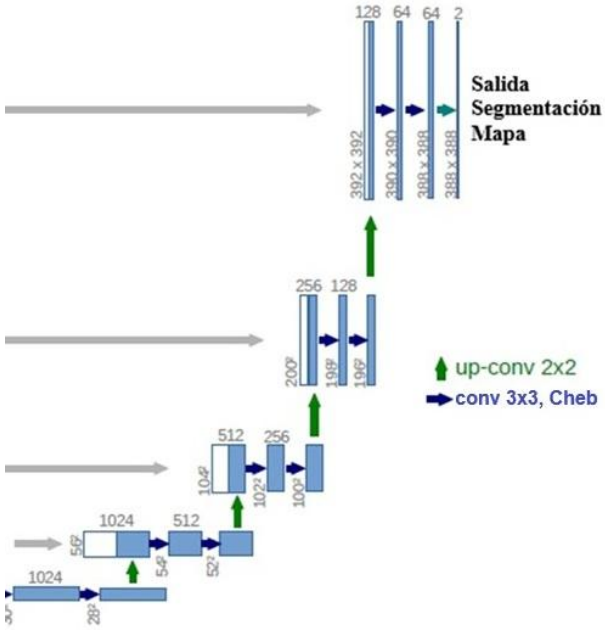


Figura. 4.12. Imagen representativa de la conexión de salto (líneas grises).

Las **conexiones de salto** permiten conservar detalles espaciales de bajo nivel, como bordes y texturas locales, los cuales son determinantes para la precisión en la segmentación de lesiones cutáneas. Esta preservación de información fina resulta especialmente relevante en zonas donde el contraste entre la lesión y la piel sana es reducido, puesto que facilita la diferenciación de límites y la detección de patrones texturales sutiles. En consecuencia, la integración de mapas de características de baja y alta resolución mediante estas conexiones contribuye de manera decisiva a la robustez y fiabilidad del proceso de segmentación en contextos clínicos con baja distinción visual (Kassani & Kassani, 2019).

4.3.1.5. Salida U-Net.

La fase expansiva culmina con una **convolución 1×1** destinada a proyectar los mapas de características finales al espacio de clases definido. Esta operación, al reducir la dimensionalidad de los canales sin alterar la resolución espacial, facilita la transformación de representaciones profundas en valores que indican la pertenencia a cada clase, lo que permite una interpretación directa de la salida de la red. En el presente trabajo, dicha capa posibilita la obtención de una **máscara de segmentación binaria** que discrimina entre lesión cutánea y piel sana Figura 4.13; dicha máscara se emplea posteriormente como insumo para la extracción de características clínicas y morfológicas. Por tanto, la inclusión de esta etapa no solo cumple una función de mapeo final, sino que constituye un componente esencial para garantizar la coherencia espacial de la segmentación y la utilidad diagnóstica de las representaciones derivadas en análisis posteriores.

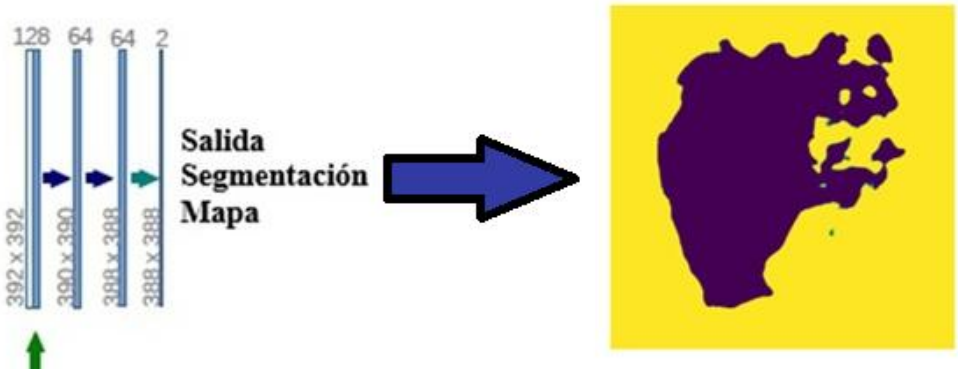


Figura. 4.13. Imagen representativa de la obtención de una máscara de segmentación semántica.

El **decodificador** en tareas de **segmentación semántica** cumple la función de traducir la información semántica de alto nivel, obtenida en las etapas profundas de la red, a una representación densa y espacialmente coherente que posibilite la asignación de una etiqueta a cada píxel. Al reconstruir la estructura espacial perdida durante la codificación, el decodificador integra características contextuales y detalles locales mediante operaciones de sobremuestreo y la fusión con mapas de resolución intermedia, lo que permite preservar la coherencia morfológica de la lesión cutánea.

Desde una perspectiva clínica, esta transformación es determinante: una representación latente correctamente densificada facilita una interpretación diagnóstica adecuada, ya que la máscara resultante refleja tanto la distribución global de rasgos patológicos como la delimitación precisa de los bordes. En términos metodológicos, la eficacia del decodificador depende de decisiones de diseño —por ejemplo, el tipo de interpolación, el uso de conexiones de salto y las estrategias de normalización— que deben justificarse en función de la necesidad de equilibrar fidelidad espacial y robustez frente a ruido, dado que errores en esta etapa comprometen la validez clínica y la utilidad diagnóstica del sistema (Ronneberger et al., 2015).

En síntesis, la arquitectura U-Net constituye un modelo convolucional diseñado para segmentación de imágenes, estructurado en una parte contractiva o codificador que extrae características y contexto, y una parte expansiva o decodificador que reconstruye la representación segmentada; las conexiones de salto permiten fusionar información de distintos niveles de abstracción, preservando detalles finos y mejorando la precisión a nivel de píxel. Su entrenamiento supervisado se realiza con conjuntos anotados y funciones de pérdida apropiadas, y su desempeño se cuantifica mediante métricas y protocolos de validación que son determinantes para la selección de hiperparámetros y la prevención del sobreajuste. Aunque la U-Net ha demostrado eficacia en múltiples dominios de segmentación y se ha consolidado como referencia en visión por computador, su adopción exige una evaluación crítica de la calidad de las anotaciones, la diversidad del conjunto de entrenamiento y las estrategias de preprocesamiento y posprocesamiento para garantizar su robustez y aplicabilidad en contextos reales.

4.4. Visualización de resultados.

4.4.1. Entrenamiento.

La red **U-Net** se entrena a partir de un conjunto de datos compuesto por imágenes etiquetadas, en el que cada imagen está acompañada por una máscara de segmentación que especifica la asignación de etiquetas a nivel de píxel. Esta correspondencia pixel-a-pixel entre la imagen y su máscara constituye la referencia supervisada necesaria para que el modelo aprenda a discriminar regiones de interés y a reproducir la estructura espacial presente en las anotaciones humanas.

Durante el proceso de entrenamiento se emplea una función de pérdida adecuada —por ejemplo, la pérdida de **entropía cruzada**— con el fin de cuantificar la discrepancia entre las etiquetas predichas por la red y las etiquetas reales del conjunto de entrenamiento. Dicha función de pérdida actúa como criterio objetivo que guía la optimización de los parámetros mediante algoritmos de descenso por gradiente y retropropagación, permitiendo evaluar la convergencia del modelo y comparar alternativas de diseño o regularización en términos de su capacidad para reducir el error de segmentación.

En la fase de optimización de parámetros se emplea un algoritmo —por ejemplo, **descenso de gradiente estocástico (SGD)** o **Adam**— para actualizar los pesos y sesgos de la red en función de la función de pérdida. Estos optimizadores determinan la dirección y magnitud de las modificaciones de los parámetros mediante la estimación del gradiente, de modo que la iteración sucesiva de actualizaciones converge hacia mínimos locales o globales del paisaje de pérdida. La selección del optimizador y de sus configuraciones influye directamente en la velocidad de convergencia y en la estabilidad del proceso de entrenamiento, por lo que su elección debe justificarse en términos de la naturaleza del problema y de las características del conjunto de datos.

El objetivo primordial del procedimiento de optimización es minimizar la función de pérdida para mejorar la precisión de la segmentación a nivel de píxel. Para evaluar de manera objetiva el rendimiento y evitar decisiones sesgadas por el conjunto de entrenamiento, se reserva un **conjunto de validación** independiente que permite controlar la generalización del modelo durante el entrenamiento. La evaluación periódica sobre este conjunto facilita la detección temprana de fenómenos indeseables, como el sobreajuste, y sirve como criterio para técnicas de control como la **parada anticipada** (**early stopping)**.

Asimismo, se ajustan sistemáticamente los **hiperparámetros** —entre ellos la tasa de aprendizaje, el tamaño del lote y el número de épocas— mediante procedimientos de búsqueda y validación que pueden incluir búsqueda en malla, búsqueda aleatoria o estrategias bayesianas. La tabla 4.2 resume los hiperparámetros del modelo. El ajuste riguroso de estos hiperparámetros contribuye a optimizar el compromiso entre sesgo y varianza, mejorando el rendimiento final del modelo y reduciendo la probabilidad de sobreajuste; por tanto, constituye una etapa esencial en la metodología experimental.

Tabla 03. Hiperparámetros de entrenamiento.

Hiperparámetros	Valores
Algoritmo de optimización	Adam
Épocas	500
Tamaño de lote	8
Función de pérdida	Entropía cruzada
Tasa de aprendizaje	0.001
β_1, β_2	0.9, 0.999
ϵ	$1e^{-7}$
Semilla	42
Grado K del polinomio de Chebyshev	3

4.4.2. Evaluación y predicción.

Una vez entrenada, la arquitectura **U-Net** puede emplearse para efectuar segmentación semántica sobre imágenes inéditas, aprovechando la correspondencia aprendida entre patrones visuales y etiquetas a nivel de píxel. Al introducir una imagen de entrada en la red, el modelo produce como salida una máscara segmentada que asigna categorías semánticas a cada píxel, lo que permite transformar información visual en representaciones estructuradas y etiquetadas.

La utilización de la U-Net en escenarios prácticos trasciende a la generación de máscaras: la imagen segmentada constituye un insumo para tareas posteriores, tales como la detección y localización de objetos, el análisis cuantitativo y cualitativo en imágenes médicas, y la interpretación de escenas en visión por computador. En contextos clínicos, por ejemplo, la segmentación facilita la delimitación de estructuras anatómicas y la cuantificación de lesiones; en aplicaciones de percepción ambiental, mejora la comprensión espacial y semántica necesaria para la toma de decisiones automatizada.

Desde una perspectiva metodológica, la aplicación de la red a nuevos datos exige considerar la robustez y la generalización del modelo, así como la adecuación de los preprocesamientos y el tratamiento posterior para garantizar la utilidad práctica de las máscaras generadas. En consecuencia, la evaluación sistemática sobre conjuntos de prueba representativos y la integración de la salida segmentada en flujos de trabajo analíticos es un paso indispensable para validar su eficacia en las distintas aplicaciones mencionadas.

4.5. Métricas del modelo.

La tabla 04 expone las cinco arquitecturas de redes neuronales convolucionales que fueron sometidas a evaluación mediante métricas de rendimiento en dos contextos: el conjunto de entrenamiento y el conjunto de prueba. El análisis comparativo de los resultados obtenidos permite identificar patrones relevantes en cuanto a la capacidad de generalización, la eficiencia computacional y el desempeño en el aprendizaje. Estos resultados no solo evidencian las fortalezas y limitaciones particulares de cada arquitectura, sino que también ofrecen elementos críticos para la selección informada de modelos en tareas específicas de segmentación semántica de imágenes biomédicas.

Tabla 04. Métrica del modelo.

Modelo	IoU conjunto de prueba	IoU conjunto de entrenamiento	Número de parámetros (M)	Tamaño de archivo (MB)	Diferencia entre entrenamiento y prueba
Xception	**0.8972**	**0.9979**	**22.9**	**88**	**0.1007**
Xception-Cheb	**0.7836**	**0.8777**	**22.9**	**88**	**0.0941**
EfficienNet-b3	**0.9137**	**0.9961**	**12.2**	**48**	**0.0824**
ResNet-18	**0.8854**	**0.9980**	**11.7**	**45**	**0.1126**
MobileOne-s0	**0.8907**	**0.9955**	**5.3**	**8**	**0.1048**

La Figura 4.14 muestra la comparación visual de los valores de **IoU** obtenidos en los conjuntos de entrenamiento y prueba, junto con la brecha de generalización correspondiente a cada una de las cinco arquitecturas analizadas. La representación gráfica evidencia que **Xception-Cheb** presenta la menor distancia entre ambos conjuntos, lo que confirma la consistencia del comportamiento previamente documentado en la Tabla 04. Esta visualización facilita la interpretación comparativa entre modelos al exponer simultáneamente tanto el desempeño absoluto como el nivel de estabilidad alcanzado por cada arquitectura.

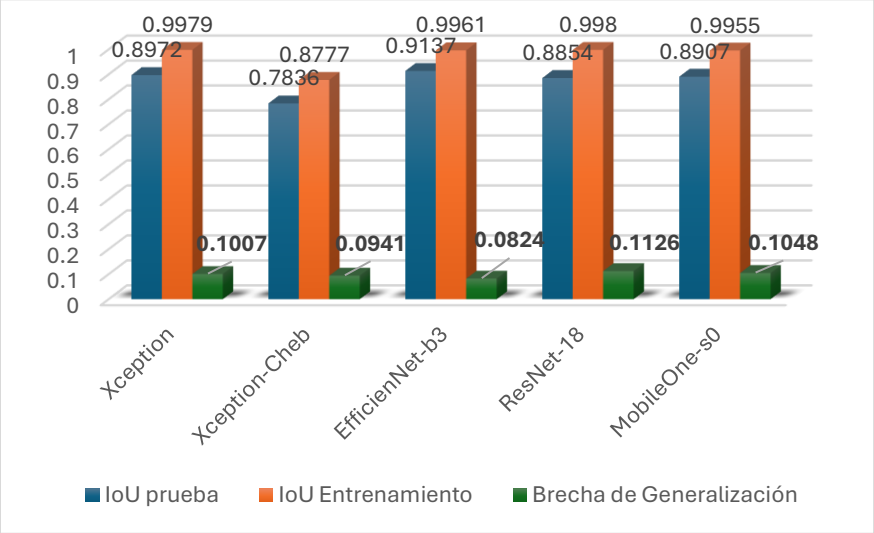


Figura 4.14. Comparativa visual de IoU en entrenamiento, prueba y brecha de generalización por arquitectura.

La Tabla 05 amplía el análisis previamente presentado en la Tabla 04 al incorporar tres indicadores derivados: el **IoU por millón de parámetros**, el **IoU por megabyte de tamaño del modelo** y la **estabilidad relativa**, definida como el complemento de la brecha de generalización. Estos descriptores permiten extender la evaluación más allá del rendimiento absoluto en segmentación, integrando en el análisis la eficiencia computacional y los recursos requeridos por cada arquitectura. La interpretación conjunta de estos indicadores evidencia que **Xception-Cheb** mantiene una posición competitiva en términos de estabilidad relativa, atributo que adquiere especial relevancia en entornos clínicos donde la **robustez**, la **consistencia predictiva** y la **reproducibilidad** constituyen criterios prioritarios frente a la mera optimización de métricas puntuales.

Tabla 05. Métricas de eficiencia computacional y rendimiento por parámetro.

Modelo	IoU conjunto de prueba	IoU conjunto de entrena-miento	Número de paráme-tros (M)	Tamaño de archivo (MB)	Diferencia entre entrenamie-nto y prueba	IoU prueba por millón de paráme-tros	IoU prueba por MB de tamaño de archivo	estabilidad relativa (1 - brecha)
Xception	**0.8972**	**0.9979**	**22.9**	**88**	**0.1007**	**0.0391**	**0.0101**	**0.8993**
Xception-Cheb	**0.7836**	**0.8777**	**22.9**	**88**	**0.0941**	**0.0342**	**0.0089**	**0.9059**
EfficienNet-b3	**0.9137**	**0.9961**	**12.2**	**48**	**0.0824**	**0.0748**	**0.0190**	**0.9176**
ResNet-18	**0.8854**	**0.9980**	**11.7**	**45**	**0.1126**	**0.0756**	**0.0196**	**0.8874**
MobileOne-s0	**0.8907**	**0.9955**	**5.3**	**8**	**0.1048**	**0.1680**	**0.1113**	**0.8952**

La figura 4.15 presentada ilustra la evolución del valor de intersección sobre unión (IoU) a lo largo de las distintas épocas de entrenamiento para cinco arquitecturas de redes neuronales convolucionales: Xception, Xception-Chev, EfficientNet, ResNet y MobileOne. El IoU, reconocido como una métrica fundamental en tareas de segmentación semántica y detección de objetos, se define como la razón entre el área de intersección y el área de unión de las regiones predichas por el modelo respecto a las anotaciones de verdad de referencia. Esta métrica constituye un indicador robusto de la precisión espacial de las predicciones, ya que penaliza tanto las omisiones como las inclusiones erróneas en la segmentación. La

comparación de la evolución del IoU entre las distintas arquitecturas permite evaluar de manera objetiva la capacidad de cada modelo para aprender representaciones discriminativas y generalizar sobre el conjunto de datos, proporcionando así una base sólida para la selección de la arquitectura óptima en aplicaciones de segmentación semántica.

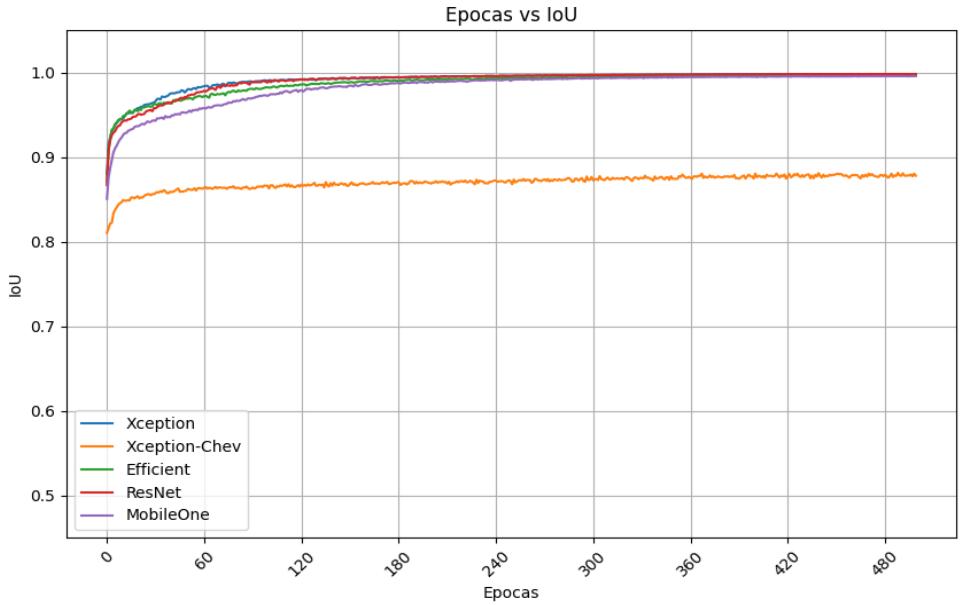


Figura. 4.15. Imagen representativa de la obtención de una máscara de segmentación semántica.

4.6. Análisis de los resultados de los modelos.

La arquitectura Xception-Cheb presenta la **menor diferencia entre el rendimiento de entrenamiento y prueba** (0.0941), lo que evidencia una notable estabilidad en el proceso de generalización. Este valor, aunque ligeramente superior al reportado para EfficientNet-b3 (0.0824), resulta particularmente relevante cuando se considera el punto de partida y las condiciones experimentales asociadas a cada modelo. La consistencia observada en Xception-Cheb sugiere una menor propensión al sobreajuste en comparación con arquitecturas como ResNet-18 (0.1126) y MobileOne-s0 (0.1048). Esta característica adquiere especial importancia en aplicaciones donde la robustez predictiva frente a datos no vistos constituye un requisito fundamental, como en sistemas de diagnóstico médico asistido por inteligencia artificial, ya que garantiza un desempeño confiable y estable en escenarios reales. En consecuencia, los resultados obtenidos posicionan a Xception-Cheb como una alternativa preferente en contextos que demandan alta estabilidad en la generalización y resistencia al sobreajuste, superando en este aspecto a modelos ampliamente utilizados en la literatura.

La interpretación del valor de IoU obtenido en el conjunto de prueba (0.7836), aunque pudiera parecer moderado frente al desempeño alcanzado por otras arquitecturas, requiere situarlo dentro del marco de su capacidad de mejora marginal. La diferencia registrada de 0.0941 entre entrenamiento y prueba evidencia que Xception-Cheb mantiene un comportamiento más conservador y estable, evitando la memorización excesiva de patrones particulares del conjunto de entrenamiento. Esta propiedad resulta especialmente relevante en contextos donde los datos de producción presentan una variabilidad sustancial respecto a los datos utilizados durante el entrenamiento, ya que favorece una mayor robustez predictiva y reduce la susceptibilidad al sobreajuste.

La Figura 4.17 representa la relación entre el **IoU obtenido en el conjunto de prueba** y la **estabilidad de generalización** para cada una de las arquitecturas evaluadas, entendida esta última como el complemento de la brecha entre los desempeños de entrenamiento y prueba. Esta visualización permite identificar el **frente de Pareto** conformado por las arquitecturas no dominadas, dentro del cual **Xception-Cheb** se posiciona como una solución óptima en la dimensión de estabilidad, mientras que **EfficientNet-b3** se sitúa en el extremo opuesto al maximizar la precisión absoluta. La gráfica respalda la argumentación previamente desarrollada al mostrar que la selección de un modelo no puede reducirse a la optimización de una única métrica, sino que debe abordarse como un **problema de decisión multiobjetivo**, donde la precisión y la estabilidad constituyen dimensiones complementarias y, en ocasiones, contrapuestas.

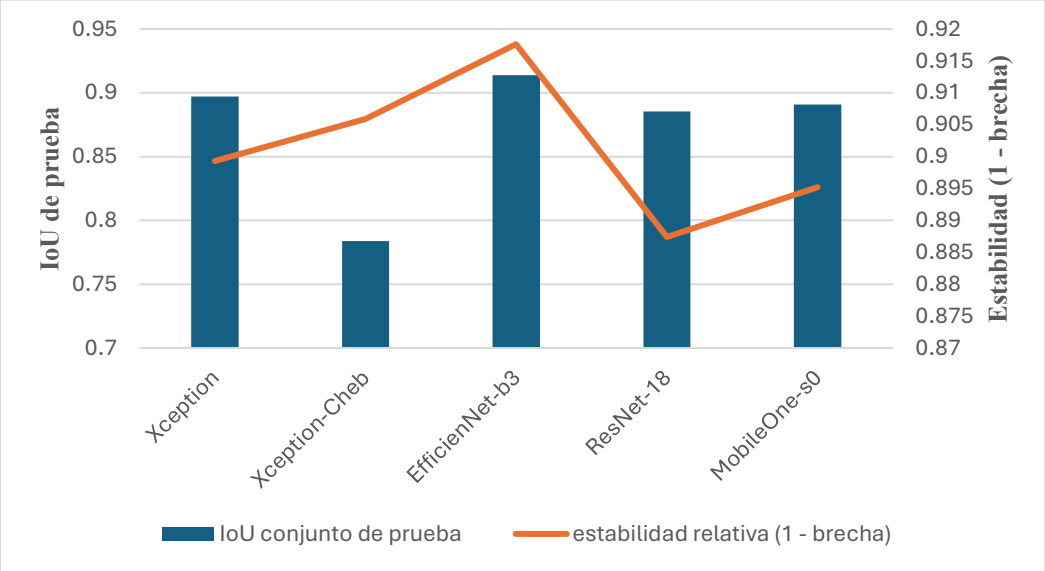


Figura. 4.16. Relación entre el IoU obtenido en el conjunto de prueba y la estabilidad de generalización.

Considerando como criterio central de evaluación la estabilidad en la generalización del modelo, alineada con el objetivo general de la presente investigación. La arquitectura Xception-Cheb, al conservar la misma configuración paramétrica que la versión estándar de Xception (22.9 millones de parámetros y un tamaño aproximado de 88 MB), incorpora la **posibilidad de emplear convoluciones basadas en polinomios de Chebyshev**. Este tipo de operación introduce ventajas teóricas en la representación de características espectrales y en la modelación de dependencias espaciales mediante

aproximaciones polinomiales, lo que constituye un fundamento matemático distintivo frente a las convoluciones tradicionales. Dicho enfoque puede traducirse en una **mayor interpretabilidad** de las representaciones aprendidas y en una mejor adaptabilidad a dominios donde las transformaciones espectrales desempeñan un papel relevante, fortaleciendo así el potencial de la arquitectura para abordar problemas especializados con mayor rigor y coherencia estructural.

Finalmente, la brecha de rendimiento observada frente a arquitecturas de mayor complejidad, como EfficientNet-b3, puede ser compensada mediante la aplicación de **estrategias de regularización, técnicas de aumento de datos o ajustes finos de hiperparámetros**, aprovechando la menor propensión al sobreajuste que caracteriza a Xception-Cheb. Estas líneas de mejora se identifican como prioritarias en los trabajos futuros derivados de la presente investigación, donde la estabilidad en la generalización demostrada constituye una base metodológica sólida sobre la cual incrementar el rendimiento absoluto del modelo.

En síntesis, Xception-Cheb se configura como una arquitectura que prioriza la estabilidad, la consistencia predictiva y una sólida fundamentación matemática, atributos que resultan esenciales tanto en entornos académicos como en aplicaciones donde la confiabilidad del modelo adquiere mayor relevancia que la maximización aislada de métricas de rendimiento. Este enfoque equilibrado refuerza su pertinencia en escenarios donde la robustez metodológica y la capacidad de generalización constituyen criterios centrales para la evaluación y selección de modelos.

Capítulo 5. Resultados.

El presente capítulo reúne los principales resultados obtenidos en la investigación y expone de manera clara los aportes teóricos, metodológicos y prácticos del modelo desarrollado. A partir del análisis de los experimentos y de la base conceptual que sustenta la integración de funciones de preprocesamiento basadas en polinomios de Chebyshev en los bloques convolucionales de la arquitectura U-Net, se presenta una reflexión general sobre la pertinencia, el alcance y las limitaciones del enfoque propuesto en relación con el objetivo central de mejorar la estabilidad en la generalización del modelo de segmentación semántica aplicado al diagnóstico de melanomas. Asimismo, se destacan las contribuciones más relevantes en relación con el estado del arte y se señalan las áreas de mejora y las oportunidades de investigación futura. Con ello, el capítulo consolida el valor científico del trabajo y establece una guía para desarrollos posteriores y para su posible aplicación en entornos clínicos reales.

5.1 Conclusiones.

El objetivo central de la presente investigación consistió en el desarrollo de un modelo de segmentación semántica aplicado a imágenes dermatoscópicas de melanomas, mediante el diseño de una arquitectura de red neuronal tipo U-Net que integra funciones de preprocesamiento basadas en polinomios de Chebyshev en los bloques convolucionales, con el fin de incrementar la estabilidad en la generalización para mejorar el apoyo diagnóstico en el contexto de la salud. Los resultados obtenidos permiten afirmar que la incorporación de estas funciones ortogonales en las etapas iniciales del codificador representa una estrategia metodológicamente válida para mejorar la estabilidad en la generalización y reducir la propensión al sobreajuste. No obstante, dicha aproximación presenta un rendimiento absoluto moderado en términos de IoU , según lo evidenciado por las métricas de segmentación evaluadas.

En términos cuantitativos, la arquitectura Xception-Cheb alcanzó un coeficiente de Intersección sobre Unión (IoU) de 0.7836 en el conjunto de prueba y de 0.8777 en el conjunto de entrenamiento, lo que se traduce en una diferencia de apenas 0.0941 entre ambos escenarios. Esta brecha, considerablemente menor que la observada en otras arquitecturas evaluadas —como ResNet-18 (0.1126) y MobileOne-s0 (0.1048)—, sugiere una mayor capacidad de generalización y una menor dependencia de patrones específicos del conjunto de entrenamiento. Aunque el IoU obtenido en prueba es inferior al reportado por EfficientNet-b3 (0.9137), la estabilidad en la generalización demostrada por Xception-Cheb constituye una ventaja estratégica en contextos clínicos, donde la robustez predictiva y la reproducibilidad adquieren mayor relevancia que la maximización aislada de métricas puntuales.

Desde una perspectiva teórica, la integración de polinomios de Chebyshev en la función de preprocesamiento previa a cada bloque convolucional introdujo una base funcional sustentada en propiedades matemáticas ampliamente reconocidas, entre ellas la ortogonalidad, la propiedad minimax y la estabilidad numérica. Estas características facilitaron la representación de variaciones de intensidad y la modelación de texturas complejas presentes en las imágenes dermatoscópicas, lo que a su vez contribuyó a una mayor estabilidad en la generalización del modelo y a una mayor resistencia frente a artefactos comunes, tales como vello, burbujas de gel y fluctuaciones en las condiciones de iluminación. Asimismo, la normalización de las evaluaciones polinomiales al intervalo $[-1,1]$ aseguró la compatibilidad con las funciones de activación subsecuentes y disminuyó la probabilidad de saturación durante el proceso de entrenamiento, reforzando así la viabilidad técnica y conceptual de la propuesta metodológica.

En el contexto de la salud pública mexicana, donde el diagnóstico tardío del melanoma se asocia con tasas de mortalidad cercanas al 30% y la insuficiencia de especialistas en dermatología constituye una barrera estructural para garantizar una atención oportuna, la implementación de herramientas de segmentación automática que sean robustas y generalizables adquiere una relevancia crítica. Bajo este escenario, la arquitectura propuesta, al priorizar la estabilidad y la consistencia predictiva por encima de la mera optimización de métricas puntuales, se posiciona como una alternativa técnicamente viable para su integración en sistemas de diagnóstico asistido por computadora (CAD) orientados a la detección temprana de lesiones cutáneas malignas. Esta pertinencia resulta especialmente significativa en unidades de primer nivel con acceso limitado a tecnologías especializadas, donde la disponibilidad de modelos confiables puede contribuir a reducir brechas diagnósticas y fortalecer la capacidad de respuesta del sistema sanitario.

No obstante, es necesario reconocer las limitaciones inherentes al presente estudio. En primer lugar, el desempeño moderado en términos de IoU absoluto sugiere que la arquitectura Xception-Cheb, en su configuración actual, podría beneficiarse de la incorporación de estrategias adicionales de regularización, técnicas más amplias de aumento de datos y un ajuste fino más exhaustivo de los hiperparámetros, con el fin de competir en precisión con arquitecturas de referencia como EfficientNet-b3. En segundo lugar, la validación del modelo se llevó a cabo exclusivamente sobre el conjunto de datos ISIC 2017, lo que restringe la evaluación de su capacidad de generalización frente a bases de datos independientes, poblaciones con características fenotípicas diversas o condiciones de adquisición de imágenes distintas. Finalmente, aunque la integración de polinomios de Chebyshev aporta ventajas teóricas relevantes, la cuantificación de su impacto diferencial sobre la estabilidad en la generalización en escenarios con mayor variabilidad de adquisición, diversidad fenotípica y condiciones clínicas heterogéneas requiere una evaluación más profunda en escenarios de aplicación real, donde las implicaciones diagnósticas adquieren un peso determinante.

Para finalizar, la presente investigación demostró que la incorporación de funciones de preprocesamiento basadas en polinomios de Chebyshev en la arquitectura U-Net constituye una estrategia metodológica sólida para mejorar la estabilidad y la capacidad de generalización de modelos de segmentación semántica aplicados a imágenes dermatoscópicas. Los resultados obtenidos respaldan la hipótesis de que el uso de funciones ortogonales como preprocesamiento en las capas convolucionales favorece una regularización efectiva del modelo, reduciendo la brecha entre el desempeño en entrenamiento y en prueba. La arquitectura propuesta ofrece un equilibrio favorable entre precisión, estabilidad en la generalización y fundamentación matemática, lo que justifica su consideración en aplicaciones clínicas donde la robustez y la reproducibilidad constituyen criterios centrales para la toma de decisiones diagnósticas, especialmente en escenarios donde la variabilidad de las condiciones de adquisición y la heterogeneidad de los pacientes demandan modelos menos susceptibles al sobreajuste.

5.2 Trabajos futuros.

Los resultados obtenidos en la presente investigación abren diversas líneas de trabajo orientadas tanto a la optimización técnica del modelo propuesto como a su validación clínica y eventual implementación en entornos reales de atención médica. En este sentido, se identifican varias direcciones que, por su relevancia metodológica y su potencial impacto en la práctica clínica, se consideran prioritarias para la continuidad y profundización del estudio. A continuación, se presentan dichas líneas de investigación, acompañadas de una argumentación que sustenta su pertinencia y necesidad dentro del marco de desarrollo futuro del modelo.

5.2.1. Optimización del modelo y exploración de bases funcionales ortogonales alternativas.

En primer lugar, resulta pertinente explorar configuraciones alternativas de la función de preprocesamiento basada en polinomios de Chebyshev, con el fin de maximizar su impacto en la capacidad representacional del modelo sin comprometer la estabilidad del entrenamiento. En particular, se propone analizar el efecto de modificar el grado K de los polinomios empleados, así como evaluar esquemas adaptativos en los que dicho grado se ajuste dinámicamente en función de la profundidad de la capa o de la complejidad de las características extraídas. Del mismo modo, la incorporación de polinomios de Chebyshev de segundo tipo u otras bases ortogonales —como los polinomios de Legendre o Hermite— podría aportar ventajas adicionales en términos de aproximación funcional y condicionamiento numérico, lo que justifica su consideración en evaluaciones comparativas dentro de estudios futuros.

5.2.2. Estrategias avanzadas de regularización y aumento de datos.

Dado que la arquitectura Xception-Cheb mostró una notable estabilidad en su capacidad de generalización, pero un rendimiento moderado en términos de IoU absoluto, se plantea la necesidad de incorporar técnicas avanzadas de regularización y aumento de datos orientadas a mejorar la precisión sin comprometer la robustez del modelo. Entre las estrategias potenciales destacan el uso de una regularización de abandono adaptativo (adaptive dropout) para prevenir el sobreajuste (overfitting), la normalización por lotes condicional y la implementación de funciones de pérdida compuestas que integren términos de regularización espacial y espectral. Asimismo, la aplicación de técnicas de aumento de datos específicas para imágenes dermatoscópicas —como la simulación de vello, la variación controlada de iluminación o la generación sintética de lesiones mediante redes generativas adversariales (GANs)— podría incrementar la diversidad del conjunto de entrenamiento y fortalecer la capacidad del modelo para generalizar en escenarios clínicos heterogéneos.

5.2.3. Integración con sistemas de apoyo al diagnóstico clínico (CAD).

La arquitectura propuesta debe entenderse como un componente dentro de un sistema integral de apoyo al diagnóstico clínico, en el cual la segmentación semántica constituye únicamente la primera etapa de un flujo de trabajo más amplio que incluye la extracción de características morfológicas y texturales, la clasificación de lesiones en categorías diagnósticas —benignas, sospechosas o malignas— y la generación de reportes interpretables para el personal médico. En esta línea, se plantea el desarrollo de módulos complementarios que integren la salida del modelo de segmentación con algoritmos orientados al análisis de asimetría, irregularidad de bordes, variabilidad cromática y otras características derivadas de la regla ABCD, así como con modelos de clasificación basados en aprendizaje profundo que aprovechen las representaciones latentes generadas por la arquitectura U-Net. Del mismo modo, resulta fundamental diseñar interfaces de usuario intuitivas y clínicamente pertinentes que faciliten la interpretación de los resultados por parte de dermatólogos y médicos de atención primaria, favoreciendo así la adopción efectiva del sistema en entornos clínicos reales.

5.2.4. Exploración de arquitecturas emergentes y técnicas de aprendizaje por transferencia.

Adicionalmente, se considera pertinente explorar arquitecturas emergentes en el ámbito de la segmentación semántica, tales como SegFormer, Swin-UNet y otros modelos basados en transformadores, los cuales han demostrado un desempeño competitivo en diversas tareas de visión por computadora y podrían beneficiarse de la integración de funciones ortogonales en sus mecanismos de atención. Del mismo modo, la aplicación de técnicas de aprendizaje por transferencia a partir de modelos preentrenados en grandes bases de datos de imágenes naturales o biomédicas podría acelerar la convergencia del entrenamiento y mejorar la

capacidad de generalización del sistema, especialmente en escenarios con disponibilidad limitada de datos anotados. La combinación de estas estrategias con la propuesta de regularización mediante polinomios de Chebyshev constituye una línea de investigación prometedora que merece ser abordada en trabajos futuros.

En conclusión, los trabajos futuros derivados de la presente investigación deben orientarse tanto a la optimización técnica del modelo propuesto como a su validación clínica rigurosa y a su eventual implementación en sistemas de salud reales. La integración de estrategias avanzadas de regularización, técnicas de aumento de datos, esquemas de validación cruzada en bases de datos independientes, así como la evaluación de métricas clínicas específicas y la realización de estudios de impacto prospectivo, constituyen pasos indispensables para consolidar la propuesta como una herramienta confiable, robusta y clínicamente útil en el diagnóstico temprano del melanoma. Estos esfuerzos deben partir de la base metodológica establecida en la presente investigación: una arquitectura U-Net con funciones de preprocesamiento basadas en polinomios de Chebyshev que ha demostrado estabilidad en la generalización, cualidad esencial para el desarrollo de sistemas de apoyo diagnóstico confiables en el contexto de la salud pública mexicana. El avance en estas líneas de investigación podría contribuir de manera significativa a la reducción de la mortalidad asociada a esta enfermedad en el contexto de la salud pública mexicana, fortaleciendo la capacidad del sistema sanitario para ofrecer diagnósticos oportunos y precisos.

Referencias

Adegun, A. A., & Viriri, S. (2020). Deep learning-based system for automatic melanoma detection. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 7160–7172. doi:10.1109/access.2019.2962812

Agostinelli, F., Hoffman, M., Sadowski, P., & Baldi, P. (2015). Learning activation functions to improve deep neural networks. arXiv preprint arXiv:1412.6830. https://doi.org/10.48550/arXiv.1412.6830

Al-Selwi, S. M., Hassan, M. F., Abdulkadir, S. J., Muneer, A., Sumiea, E. H., Alqushaibi, A., & Ragab, M. G. (2024). RNN-LSTM: From applications to modeling techniques and beyond—Systematic review. *Journal of King Saud University - Computer and Information Sciences*, *36*(5), 102068. doi:10.1016/j.jksuci.2024.102068

American Cancer Society. (s.f.). Radiación UV y cáncer de piel. https://www.cancer.org/es/cancer/prevencion-del-riesgo/sol-y-uv/radiacion-uv.html

Banasode, P., Patil, M., & Ammanagi, N. (2021). A melanoma skin cancer detection using machine learning technique: Support vector machine. *IOP Conference Series. Materials Science and Engineering*, *1065*(1), 012039. doi:10.1088/1757-899x/1065/1/012039

Belianskaya, V., Rodríguez Gonzales, A., Savoye, P., & Villegas Poquechoque, M. (2024). Polinomios de Chebyshev del primer tipo: Fundamentos teóricos y aplicaciones. Revista de Ciencia, Tecnología e Innovación, 22(32). https://doi.org/10.56469/rcti.v22i32.945

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166. doi:10.1109/72.279181

Carlos-Ortega, B., Ángeles-Garay, U., Fernández-Martínez, M., Gómez-Campos, G., & Roldán-Marín, R. (2024). Characteristics of melanoma in Mexicans seen at 'La Raza' National Medical Center. *Revista medica del Instituto Mexicano del Seguro Social*, *62*(6), 1–7. doi:10.5281/zenodo.13306762

Celebi, M. E., Wen, Q., Iyatomi, H., Shimizu, K., Zhou, H., & Schaefer, G. (2015). **A state-of-the-art survey on lesion border detection in dermoscopy images**. *Dermatology*, 231(2), 83–98. https://doi.org/10.1159/000381589

Cheng, B. et al. (2021) 'Boundary IoU: Improving Object-Centric Image Segmentation Evaluation', in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society, pp. 15329–15337. Available at: https://doi.org/10.1109/CVPR46437.2021.01508.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1251–1258. https://doi.org/10.1109/CVPR.2017.195

Codella, N. C. F., et al. (2018). **Skin lesion analysis toward melanoma detection: A challenge at the 2017 ISIC workshop**. *IEEE Journal of Biomedical and Health Informatics*, 23(2), 501–512. https://doi.org/10.1109/JBHI.2018.2868377

Cohen, S., Gunasekar, S., & Srebro, N. (2022). *Understanding gradient descent on the edge of stability in deep learning*. arXiv. https://doi.org/10.48550/arXiv.2205.09745

Cox Figueroa, E. J., Espinoza Sánchez, M. G., Falcone Figueroa, F. F., & Coox Zambrano, R. E. (2023). *Radiación electromagnética y su incidencia en los seres humanos*. Polo del Conocimiento, 8(12), 756–768. https://doi.org/10.23857/pc.v8i12.6311

Datapeaker. (s.f.). *Función de activación sigmoide*. https://datapeaker.com/wiki/funcion-de-activacion-sigmoide/

F. Nachbar et al., 'The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions', Journal of the American Academy of Dermatology, vol. 30, no. 4, pp. 551–559, 1994.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Gaceta de la Facultad de Medicina (UNAM). (2023, 28 de agosto). *Panorama del cáncer de piel en México*. Universidad Nacional Autónoma de México. https://gaceta.facmed.unam.mx/index.php/2023/08/28/panorama-del-cancer-de-piel-en-mexico/

Gallegos, J. F. (20 de julio de 2024). *En verano, reforzar cuidados para reducir riesgos de cáncer de piel*. Rastreator. https://www.gob.mx/imss/articulos/en-verano-reforzar-cuidados-para-reducir-riesgos-de-cancer-de-piel?idiom=es

García, R. A., & Lepera, A. F. (2024). Multilayer Perceptron – Mathematical Model and Implementation in Python. Zenodo. https://doi.org/10.5281/zenodo.10676991

Goodfellow, I., Bengio, Y. & Courville, A. (2017). *Deep learning*. Editorial The MIT Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://arxiv.org/abs/1512.03385

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural* Computation, 9(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735

Instituto Nacional de Cancerología (INCan). (2025). *Programa Institucional 2025-2030 del Instituto Nacional de Cancerología*. Diario Oficial de la Federación. https://sidof.segob.gob.mx/notas/5772404

Instituto Nacional de Estadística y Geografía (INEGI). (2025). *Estadísticas de Defunciones Registradas (EDR) 2024*. Comunicado de prensa 39/25. https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2025/EAP_DMvsCancer25.pdf

International Skin Imaging Collaboration (ISIC). (2024). *Digital Imaging in Dermatology: Standards and Datasets for AI Training*. ISIC Archive.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R* (2nd ed.). Springer. https://www.statlearning.com

Kandhro, I. A., Manickam, S., Fatima, K., Uddin, M., Malik, U., Naz, A., & Dandoush, A. (2024). Performance evaluation of E-VGG19 model: Enhancing real-time skin cancer detection and classification. *Heliyon*, *10*(10), e31488. doi:10.1016/j.heliyon.2024.e31488

Kaur, R., GholamHosseini, H., Sinha, R., & Lindén, M. (2022). Automatic lesion segmentation using atrous convolutional deep neural networks in dermoscopic skin cancer images. *BMC Medical Imaging*, *22*(1), 103. doi:10.1186/s12880-022-00829-y

Kroese, D. P., Botev, Z. I., Taimre, T., & Vaisman, R. (2019). *Data science and machine learning: Mathematical and statistical methods*. Chapman & Hall/CRC.

Krogh, A., & Hertz, J. A. (1991). A simple weight decay can improve generalization. Advances in Neural Information Processing Systems (NIPS), 4, 950–957.

Li, M., Jiang, Y., Zhang, Y., & Zhu, H. (2023). Medical image analysis using deep learning algorithms. Frontiers in Public Health, 11, 1273253. https://doi.org/10.3389/fpubh.2023.1273253

Lifeder. (13 de marzo de 2020). Espectro electromagnético: características, bandas, aplicaciones. Recuperado de: https://www.lifeder.com/espectro-electromagnetico/.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2117–2125. https://doi.org/10.1109/CVPR.2017.106

Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., Mahendiran, T., Moraes, G., Shamdas, M., Kern, C., Ledsam, J. R., Schmid, T., Balaskas, K., Topol, E. J., & Denniston, A. K. (2021). Advances in deep learning-based medical image analysis. *Health Data Science*, *2021*, 8786793. https://doi.org/10.34133/2021/8786793

Long, J., Shelhamer, E., & Darrell, T. (2015). **Fully convolutional networks for semantic segmentation**. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. Proceedings of the 30th International Conference on Machine Learning (ICML). https://web.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf

Maqueda Bojórquez, DE (2021). De redes neuronales recurrentes a modelos de lenguaje: la evolución del PLN en la generación de textos. *TIES, Revista de Tecnología e Innovación En Educación Superior* , (4), 1–10. doi:10.22201/dgtic.26832968e.2021.4.1

Qiu, H., Zhang, Y., & Wang, Y. (2022). *Leaky ReLU revisited: Theoretical insights and practical improvements*. Proceedings of the 36th AAAI Conference on Artificial Intelligence. https://ojs.aaai.org/index.php/AAAI/article/view/20202

Ramachandran, P., Zoph, B., & Le, Q. V. (2022). *Searching for activation functions*. arXiv preprint arXiv:1710.05941v4. https://arxiv.org/abs/1710.05941

Roncali, A., & Martínez, J. (2025). Inteligencia Artificial y diagnóstico oncológico en América Latina: Retos y adopción regional en México. *IECCMEXICO Journal, 12*(2), 45-60.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. arXiv preprint arXiv:1505.04597. https://doi.org/10.48550/arXiv.1505.04597.

Salinas, M. P., et al. (2024). A systematic review and meta-analysis of artificial intelligence versus clinicians for skin cancer diagnosis. *npj Digital Medicine, 7*(103). https://doi.org/10.1038/s41746-024-01103-x

Sara, U., Das, U. K., Sikder, J., & Chowdhury, S. R. (2025). Automated skin cancer classification and detection using convolutional neural networks and dermoscopy images. *Procedia Computer Science*, *252*, 108–117. doi:10.1016/j.procs.2024.12.012

Secretaría de Salud (México). (2024, 24 de mayo). *196. Falta de prevención y detección tardía ocasiona 30% de fallecimientos en pacientes con melanoma*. Gobierno de México. https://www.gob.mx/salud/prensa/196-falta-de-prevencion-y-deteccion-tardia-ocasiona-30-de-fallecimientos-en-pacientes-con-melanoma

Shihab, A., Salah, H., & Mocanu, M. (2021, May). Detection and diagnosis of skin cancer based on K-means cluster and convolutional neural network. *2021 23rd International Conference on Control Systems and Computer Science (CSCS)*. Presented at the 2021 23rd International Conference on Control Systems and Computer Science (CSCS), Bucharest, Romania. doi:10.1109/cscs52396.2021.00031

Szegő, G. (1975). *Orthogonal polynomials* (4th ed.). American Mathematical Society.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning (ICML). https://arxiv.org/abs/1905.11946

Torres, J. (2020). *Python deep learning: Introducción práctica con Keras y TensorFlow 2*. Alpha Editorial.

Ultralytics. (s.f). *Tanh (tangente hiperbólica)*. https://www.ultralytics.com/es/glossary/tanh-hyperbolic-tangent

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). *Searching for activation functions* (arXiv:1710.05941v4). arXiv. https://arxiv.org/abs/1710.05941

Vasu, P. K. A., Gabriel, J., Zhu, J., Tuzel, O., & Ranjan, A. (2023). MobileOne: An improved one millisecond mobile backbone. Apple Machine Learning Research. https://machinelearning.apple.com/research/mobileone

Wang, Y., & Li, X. (2021). *Chebyshev activation functions for deep neural networks: Theory and applications*. Neural Networks, 139, 1–12. https://doi.org/10.1016/j.neunet.2021.02.003

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with transformers. In Advances in Neural Information Processing Systems (NeurIPS 2021). arXiv. https://doi.org/10.48550/arXiv.2105.15203

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning requires rethinking generalization. *Communications of the ACM, 64*(3), 62–71. https://doi.org/10.1145/3446776.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). *Pyramid scene parsing network*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 6230–6239). IEEE. https://doi.org/10.48550/arXiv.1612.01105

Zhao, R. et al. (2020) 'Rethinking Dice Loss for Medical Image Segmentation', in 2020 IEEE International Conference on Data Mining (ICDM), pp. 851–860. Available at: https://doi.org/10.1109/ICDM50108.2020.00094.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Anexos.

A. Articulo congreso CITCA.

Deep Learning-Based Semantic Segmentation of Skin Lesions Using Pre-Trained Models.

García Ramírez, Fernando*, González Huitrón, Víctor Alejandro*, Guzmán Rivera, Miguel Ángel*, Montante Garza, Mario Alberto*, Gorrostieta Hurtado, Efrén**

**Tecnológico Nacional de México campus Querétaro, Av. Tecnológico S/N, Centro Histórico, Centro, 76000 Santiago de Querétaro, Qro.*
e-mail: fernando.gr@queretaro.tecnm.mx, victor.gh@queretaro.tecnm.mx, miguel.gr@queretaro.tecnm.mx, mario.mg@queretaro.tecnm.mx

*** Universidad Autónoma de Querétaro, Cerro de las Campanas s/n, C.P. 76010, Querétaro, Qro.*

e-mail: efrengorrostieta@gmail.com

Abstract: The largest organ of the human body is the skin, which performs multiple physiological functions. Among these, one of the primary roles is to act as a protective barrier against external agents that may pose threats to the organism. Consequently, the early detection of skin lesions is of significant importance to maintain health and prevent the progression of severe conditions, such as skin cancer. In this context, automated image analysis systems offer promising support tools for the early identification and assessment of dermatological abnormalities. This research presents a methodology based on convolutional neural networks for the semantic segmentation of skin lesion images, utilizing data from The ISIC 2017 Challenge Dataset. The study proposes a comparative evaluation involving the U-Net architecture and a set of pre-trained classification models originally trained on the ImageNet dataset. The selected models—based on the EfficientNet, MobileOne, ResNet, and Xception architectures—were chosen due to their diversity in terms of parameter complexity and their potential for practical application and deployment in resource-constrained environments. The evaluation of these models was carried out using the Intersection over Union (IoU) metric, which is commonly employed in image segmentation tasks to quantify the overlap between predicted and ground truth regions. The analysis emphasizes the relative strengths and limitations of each architecture in terms of segmentation accuracy and computational efficiency. The results of this study aim to guide the selection of optimal models by balancing predictive performance, resource requirements, and feasibility of implementation in real-world clinical or diagnostic systems.

Keywords: Deep Learning, Computer Vision, Skin Cancer, Semantic Segmentation.

1. INTRODUCCIÓN

El cáncer de piel, y especialmente el melanoma, está experimentando un aumento en su incidencia en México, debido a factores como la exposición excesiva a la radiación ultravioleta y cambios en el comportamiento social. Detectar estos cánceres de manera temprana es crucial para mejorar las tasas de supervivencia, pero el diagnóstico se ve limitado por la variabilidad en la interpretación de las imágenes dermatoscópicas y la falta de especialistas en ciertas áreas del país.

Los algoritmos de detección médica son una de las primeras herramientas utilizadas para determinar si una lesión cutánea es maligna o benigna (Nachbar, 1994; Zalaudek, 2006; Henning, 2007). Nachbar (1994) desarrollaron un método subjetivo basado en la percepción visual de la lesión. La regla ABCD se basa en la asimetría, bordes, color, forma y diámetro que aparecen en las lesiones cutáneas. Debido a la simplicidad del algoritmo, es uno de los más practicados para evaluar una lesión mediante un examen a simple vista o utilizando un dermatoscopio.

El algoritmo médico ABCD se compone de las siguientes partes:

- **Asimetría A**: La lesión se divide en dos ejes perpendiculares a 90° entre sí, de forma que se obtenga la menor puntuación de asimetría posible. En otras palabras, se determina si la lesión es simétrica o no. Por cada eje en el que se detecta asimetría, se añade un punto.
- **Bordes B**: La lesión se divide en cortes por ocho ejes que determinan si una lesión presenta bordes abruptos. Si un segmento presenta un borde abrupto, se añade un punto.
- **Colores C**: La lesión puede presentar uno o varios de los siguientes colores: blanco, parduzco, marrón oscuro, negro, azul y rojo. Son generados por los vasos y las concentraciones de melanina, por lo que por cada color encontrado se suma un punto por color.
- **Estructuras Dermatoscópica D**: La lesión presenta el aspecto de las siguientes estructuras: puntos, manchas, redes pigmentadas y zonas no estructuradas. Se añade un punto por cada estructura encontrada en la lesión.

Las características descritas se ponderan del siguiente modo:

$$TDS = (1.3 \cdot A) + (0.1 \cdot B) + (0.5 \cdot C) + (0.5 \cdot D) \quad (1)$$

donde TDS es la Puntuación Dermatoscópica Total si es inferior a 4.75 se concluye que la lesión es benigna si la puntuación está entre 4.75 y 5.45 la lesión se considera sospechosa si es superior a 5.45 se considera maligna. Los valores de ponderación de la regla ABCD han sido determinados empíricamente a través de estudios clínicos y análisis estadísticos (Nachbar, 1994).

Un ejemplo de imágenes RGB aplicando los criterios ABCD se muestra en la Fig. 1.

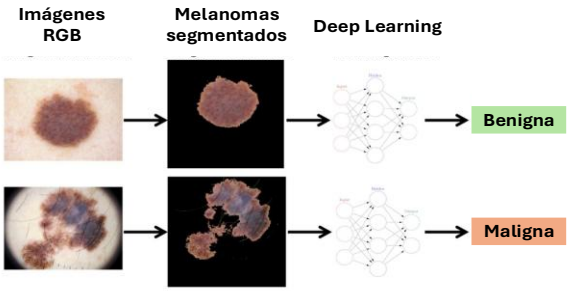
Fig. 1. Imágenes cutáneas con características para la regla ABCD



Los algoritmos del aprendizaje profundo más comunes son: redes neuronales profundas (DNN) (Sauter, 2023), redes neuronales convolucionales (CNN), redes neuronales recurrentes (RNN) (Popescu, 2022), redes neuronales de transformadores (ViT) (Xiao, 2023), Redes Neuronales de Autoencoders (Bhatt, 2023), Redes Neuronales Generativas Antagónicas (GANs), Redes Neuronales de Grafos (GNN) (Dildar, 2021).

Es importante señalar que el aprendizaje profundo es un subcampo del aprendizaje automático usado en la inteligencia artificial el cual es fundamental para el desarrollo de un sistema de inteligencia artificial. Un ejemplo común de manera generalizada se presenta en la Fig. 2. como diagrama de flujo.

Fig. 2. Diagrama de flujo para un clasificador de melanomas.

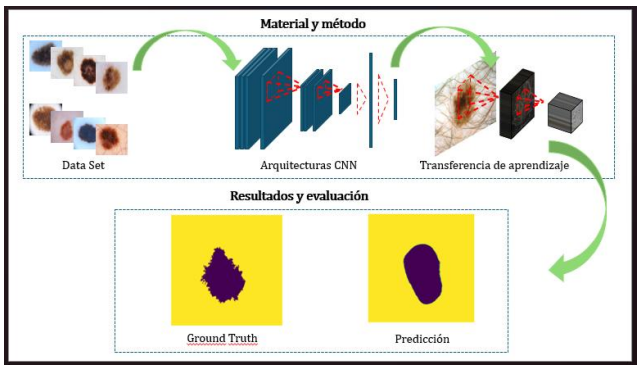


En la actualidad la inteligencia artificial enfrenta desafíos en las imágenes médicas, ya que depende del intercambio de datos, el desarrollo de algoritmos y la regulación ética. El éxito requiere grandes volúmenes de datos con un etiquetado preciso, realizado por expertos, ya que un etiquetado deficiente afecta el aprendizaje del sistema. Debido a la expansión del uso de la IA en el diagnóstico médico por imagen y a los numerosos retos a los que se enfrenta, este trabajo pretende evaluar y resumir algunas funciones de activación para la segmentación de imágenes dermatoscópicas.

2. METODOLOGÍA

Se plantea un estudio comparativo con la arquitectura U-Net y las CNN; EfficientNet (Tan, 2019), MobileOne (Vasu, 2023), ResNet (He, 2016) y Xception (Chollet, 2017) para evaluar su rendimiento en segmentación de imágenes. El análisis considerará la métrica IoU, destacando sus ventajas y desventajas. Como se muestra en la figura 3, los resultados permitirán determinar la mejor arquitectura según el equilibrio entre exactitud y costo computacional.

Fig. 3. Diagrama de bloques.



2.1 Segmentación semántica y aprendizaje profundo.

La segmentación semántica es una tarea de visión por computadora que asigna una etiqueta a cada píxel de una imagen, permitiendo la identificación de objetos y regiones (Hao, 2020). El aprendizaje profundo ha revolucionado este campo mediante CNN, mejorando la precisión y automatización del proceso. Modelos como U-Net (Huang, 2020), DeepLab (Chen, 2018) y SegNet (Badrinarayanan, 2017) utilizan arquitecturas profundas para extraer características y segmentar con alta precisión. Gracias a estas técnicas, la segmentación semántica se aplica en medicina y otros campos de la ciencia e ingeniería.

La arquitectura U-Net es fundamental, ya que segmenta imágenes con alta precisión. Su estructura en forma de "U", con un codificador – decodificador, que incluye conexiones de salto en los enlaces directos entre capas, estas conexiones permiten combinar características de baja y alta resolución, haciendo que el modelo sea más preciso en tareas de segmentación semántica. Gracias a su eficiencia y capacidad de aprendizaje en conjuntos de datos pequeños, U-Net se ha convertido en un estándar en aplicaciones médicas, satelitales y de visión por computadora.

La función Dice Loss (Zhao, 2020) es crucial en la arquitectura U-Net, ya que optimiza la segmentación semántica al manejar mejor el desequilibrio de clases en las imágenes. En general, mide la superposición entre la segmentación predicha y la real, lo que la hace ideal para conjuntos de datos con regiones pequeñas o desbalanceadas. A diferencia de la entropía cruzada, que se utiliza comúnmente para tareas de clasificación, Dice Loss mejora la precisión en los bordes y reduce los falsos negativos.

El coeficiente de Dice (DSC - Dice Similarity Coefficient) se define de la siguiente manera:

$$DSC = \frac{2|y_{pred} \cap y_{true}|}{|y_{pred}| + |y_{true}|}, \quad (2)$$

donde,

- y_{pred} son los valores predichos por el modelo.
- y_{true} son los valores reales.

- La multiplicación $y_{pred} \cdot y_{true}$ representa la intersección entre la predicción y la verdad.

El coeficiente de Dice toma valores entre 0 y 1, para los cuales:

- **0** significa que no hay superposición entre la predicción y la verdad.
- **1** significa que hay una coincidencia perfecta.

La función de pérdida se define como:

$$Dice\ Loss = 1 - DSC, \tag{3}$$

o de manera equivalente:

$$Dice\ Loss = 1 - \frac{2|y_{pred} \cap y_{true}| + \epsilon}{|y_{pred}| + |y_{true}| + \epsilon}, \tag{4}$$

donde ϵ es un pequeño valor para evitar divisiones por cero.

2.2 Arquitecturas CNN.

El núcleo de una CNN es la capa de convolución, donde se aplican filtros o "kernels" a la imagen de entrada. Estos filtros recorren la imagen, detectando características locales como bordes, texturas o formas simples. A medida que la información avanza por las capas de la red, las características simples se combinan para formar representaciones más complejas y abstractas, permitiendo que la red reconozca objetos completos o escenas específicas.

Además de las capas de convolución, las CNN incorporan capas de pooling (submuestreo) que reducen la dimensionalidad de las representaciones intermedias, conservando las características más relevantes y disminuyendo el costo computacional. Las funciones de activación no lineales, como la ReLU (Rectified Linear Unit), se utilizan para introducir no linealidades en el modelo, permitiendo que la red capture relaciones complejas en los datos (Zhao, 2024).

Xception es un tipo de red neuronal convolucional que optimiza la arquitectura de las redes Inception al reemplazar sus módulos convencionales por convoluciones separables en profundidad. Esta técnica divide la operación de convolución tradicional en dos pasos: primero, una convolución en profundidad que aplica filtros de manera independiente a cada canal de entrada, y luego, una convolución 1x1 que combina la información obtenida. Este enfoque reduce de manera significativa la cantidad de parámetros y el costo computacional, sin comprometer el rendimiento en tareas de clasificación de imágenes e incluso mejorándolo en algunos casos. Además, Xception incorpora conexiones residuales, lo que facilita el entrenamiento de redes más profundas y permite una propagación más eficiente del gradiente.

EfficientNet es una familia de arquitecturas de redes neuronales convolucionales diseñada para lograr un equilibrio óptimo entre precisión y eficiencia computacional. Su innovación principal radica en un método llamado "compounding scaling", que ajusta de manera equilibrada la profundidad, el ancho de las capas y la resolución de entrada de la red para mejorar el rendimiento sin aumentar excesivamente el número de parámetros. En lugar de escalar arbitrariamente uno de estos factores, EfficientNet optimiza su crecimiento de manera conjunta, lo que permite entrenar modelos más precisos y eficientes en términos de memoria y cómputo.

ResNet es una arquitectura de redes neuronales convolucionales que aborda el problema del desvanecimiento del gradiente al entrenar modelos profundos. Su innovación clave son las conexiones residuales o skip connections, que permiten que la información fluya directamente entre capas sin necesidad de pasar por todas las transformaciones intermedias. Esto facilita el entrenamiento de redes más profundas sin que el gradiente se desvanezca o se degrade con el tiempo.

MobileOne es una arquitectura optimizada para su uso en dispositivos móviles y entornos con recursos computacionales limitados. Su diseño se basa en una estrategia de reparametrización estructural, lo que significa que durante el entrenamiento utiliza una estructura más compleja con múltiples caminos de convolución, pero en la fase de inferencia se transforma en una arquitectura completamente secuencial y eficiente. Esto permite alcanzar un equilibrio entre precisión y velocidad, manteniendo un bajo consumo de memoria y mejorando el rendimiento en comparación con modelos anteriores.

3. RESULTADOS Y/O DISCUSIÓN

El conjunto de datos seleccionado para la presente investigación es "International Skin Imaging Collaboration. SIIM-ISIC 2017 Challenge Dataset" (Codella, 2017) El conjunto de datos contiene 2,600 imágenes dermatoscópicas de lesiones cutáneas benignas y malignas, que se dividen en 2,000 de entrenamiento y 600 de prueba. La resolución varia desde 540x576 hasta 4499x6748 y aplicando un reescalamiento a 512x512. Se mantuvo la normalización especificada por cada modelo. Las imágenes de entrada están en formato jpg y su correspondiente salida en png. Cada imagen se asocia a uno de estos individuos mediante un identificador único de paciente. Todos los diagnósticos malignos se han confirmado mediante histopatología y los diagnósticos benignos se han confirmado mediante el acuerdo de expertos, el seguimiento longitudinal o la histopatología. Una publicación exhaustiva que describe todas las características de este conjunto de datos está disponible en forma de preimpresión que aún no ha sido sometida a revisión por pares.

El conjunto de datos fue generado por la Colaboración Internacional de Imágenes de la Piel (ISIC) y las imágenes provienen de las siguientes fuentes: Hospital Clínic de Barcelona, Universidad Médica de Viena, Memorial Sloan Kettering Cancer Center, Instituto de Melanoma de Australia, Universidad de Queensland y la Facultad de Medicina de la Universidad de Atenas.

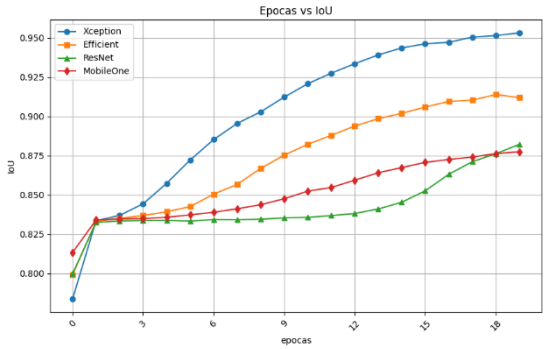
La métrica IoU (Intersection over Union) (Cheng. 2021) es utilizada para evaluar la precisión de modelos de segmentación y detección de objetos en visión por computadora. Se calcula como el área de superposición entre la región predicha por el modelo y la región real en los datos, dividida por el área total combinada de ambas regiones. Matemáticamente, se expresa como:

$$IoU = \frac{|y_{pred} \cap y_{true}|}{|y_{pred}| \cup |y_{true}|}, \qquad (5)$$

donde la intersección es la zona en común entre la predicción y la región real, y la unión es la suma de ambas áreas, restando la intersección para evitar contarla dos veces.

Su valor varía entre 0 y 1, donde 1 indica una coincidencia perfecta. Un IoU alto significa mejor rendimiento del modelo, generalmente, un umbral de $IoU \geq 0.5$ se considera aceptable. IoU es una métrica fundamental para evaluar la detección de objetos y segmentación de imágenes con CNN. Un valor alto de IoU indica una mejor precisión del modelo, lo que es clave para aplicaciones en el análisis médico de imágenes.

Fig. 4. Valores de IoU por época obtenidos en el entrenamiento.



En la figura 4 se evalúa cómo varía el desempeño de los modelos en términos de IoU a medida que avanza el entrenamiento en función de las épocas. La arquitectura Xception tiene el mejor desempeño en general para la segmentación a pesar de que parte con un IoU bajo, pero su crecimiento es rápido y constante lo que nos indica que aprende características útiles muy temprano en el entrenamiento, presentando una tendencia ascendente y estabilizándose en las últimas épocas lo que nos indica que podría mejorar y evita el sobreajuste.

EfficientNet es eficiente en términos de precisión contra el tamaño de la muestra, llega a un buen rendimiento sin la complejidad de la arquitectura de Xception, aunque su crecimiento después de la mitad de las épocas es lenta, lo que indica una posible cercanía al punto de saturación. En contraparte ResNet y MobileOne son las que tiene un progreso más lento y tienen un rendimiento casi similar, aunque MobileOne tiene una mejora más continua que indica que con más entrenamiento podría mejorar y en el caso de ResNet puede deberse a la tasa de aprendizaje.

Tabla 1. Resultado IoU y características de los modelos.

Modelos	IoU conjunto de prueba	Número de parámetros (M)	Tamaño de archivo (MB)
Xception	0.7789	20	88
EfficientNet-b3	0.7755	10	48
ResNet-152	0.7805	58	232
MobileOne-s0	0.7806	4	4.8

En la tabla 1 muestra como la arquitectura ResNet es la de mayor complejidad y su tamaño es mucho mayor que MobileOne, lo que se traduce en tiempo de entrenamientos más largos o un mayor consumo de GPU y no sería ideal para dispositivos con recursos limitados sin alguna optimización previa. En el modelo EfficientNet tiene un límite de aprendizaje más cercano y aunque es más liviano, aun puede ser más costoso que modelos diseñados para una eficiencia externa. La arquitectura Xception aunque es un modelo clásico, el desempeño es limitado para la segmentación. En el caso de MobileOne su rendimiento es modesto y no es el modelo más robusto, pero es el que presenta mejor posibilidad de rendimiento en implementaciones de hardware limitado.

4. CONCLUSIONES

La revisión del estado del arte demuestra los avances que ha tenido el aprendizaje profundo en diferentes ramas de la ciencia y tecnología, aunado a esto el diagnóstico médico por imágenes dermatoscópicas en lesiones cutáneas puede beneficiarse enormemente del aprendizaje profundo. Nuestra investigación se centra en la segmentación semántica de imágenes dermatoscópicas y coadyuvar en un modelo de inteligencia artificial para la detección de melanomas. Al tratarse de una aplicación médica, la precisión y la fiabilidad del modelo se vuelven especialmente críticas, ya que los errores pueden tener implicaciones serias en diagnósticos, tratamientos o decisiones clínicas. En la segmentación semántica de lesiones cutáneas es vital tener modelo que segmente con altísima fidelidad y un IoU alto es esencial, porque una pequeña diferencia en la imagen puede representar un diagnóstico incorrecto.

Si queremos una máxima precisión sin importar el costo computacional, Xception cumple el requerimiento, en algunos ambientes clínicos se dispone de dichos recursos como servidores o entornos en la nube. La arquitectura EfficientNet tiene un buen balance entre precisión y eficiencia; lo que puede ayudar en caso de que el hardware sea más limitado, por ejemplo, en quirófanos o estaciones móviles. Si se requiere ejecutar el modelo en un dispositivo móvil, como un sistema portátil tele médico o de asistencia remota, MobileOne es el más viable, por su tamaño, sin embargo, su precisión es riesgosa para tareas críticas. ResNet se puede usar como un modelo base o para investigación básica, su desempeño es bajo y para medicina esto no es suficiente sin ajustes.

La disponibilidad y el intercambio de datos siguen siendo una de las mayores preocupaciones que podrían obstaculizar el desarrollo y la aplicación con éxito del aprendizaje profundo en el campo del diagnóstico médico por imagen. Es fundamental que la arquitectura empleada generalice lo mejor posible y no solo aprenda del conjunto de datos. Es deseable usar modelos cuya arquitectura sea bien entendida y documentada en entornos clínicos, para facilitar la aprobación regulatoria y la validación por pares médicos.

REFERENCIAS

F. Nachbar et al., ‘The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions’, Journal of the American Academy of Dermatology, vol. 30, no. 4, pp. 551–559, 1994.

I. Zalaudek et al. (2006) ‘Three-point checklist of dermoscopy: an open internet study’, British Journal of Dermatology, vol. 154, no. 3, pp. 431–437.

J. S. Henning et al., ‘The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy’, Journal of the American Academy of Dermatology, vol. 56, no. 1, pp. 45–52, 2007.

D. Sauter, G. Lodde, F. Nensa, D. Schadendorf, E. Livingstone, and M. Kukuk, (2023) ‘Deep learning in computational dermatopathology of melanoma: A technical systematic literature

review', Computers in Biology and Medicine, vol. 163, p. 107083.

D. Popescu, M. El-Khatib, H. El-Khatib, and L. Ichim, (2022) 'New Trends in Melanoma Detection Using Neural Networks: A Systematic Review', Sensors, vol. 22, no. 2.

Xiao, H., Li, L., Liu, Q., Zhu, X., & Zhang, Q. (2023). Transformers in medical image segmentation: A review. Biomedical Signal Processing and Control, 84, 104791. doi:10.1016/j.bspc.2023.104791

H. Bhatt, V. Shah, K. Shah, R. Shah, and M. Shah, (2023) 'State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review', Intelligent Medicine, vol. 03, no. 03, pp. 180–190.

M. Dildar et al. (2021) 'Skin Cancer Detection: A Review Using Deep Learning Techniques', International Journal of Environmental Research and Public Health, vol. 18, no. 10.

Zhao, R. et al. (2020) 'Rethinking Dice Loss for Medical Image Segmentation', in 2020 IEEE International Conference on Data Mining (ICDM), pp. 851–860. Available at: https://doi.org/10.1109/ICDM50108.2020.00094.

Zhao, X. et al. (2024) 'A review of convolutional neural networks in computer vision', Artificial Intelligence Review, 57(4), p. 99. Available at: https://doi.org/10.1007/s10462-024-10721-6

Chollet, F. (2017) 'Xception: Deep Learning with Depthwise Separable Convolutions', in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society, pp. 1800–1807. Available at: https://doi.org/10.1109/CVPR.2017.195.

Tan, M. and Le, Q.V. (2019) 'EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks', CoRR, abs/1905.11946. Available at: http://arxiv.org/abs/1905.11946.

He, K. et al. (2016) 'Deep Residual Learning for Image Recognition', in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Vasu, P.K.A. et al. (2023) 'MobileOne: An Improved One millisecond Mobile Backbone', in 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7907–7917. Available at: https://doi.org/10.1109/CVPR52729.2023.00764.

Hao, S., Zhou, Y., & Guo, Y. (2020). A Brief Survey on Semantic Segmentation with Deep Learning. Neurocomputing, 406, 302–321. doi:10.1016/j.neucom.2019.11.118

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., … Wu, J. (2020). UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1055–1059. doi:10.1109/ICASSP40776.2020.9053405

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 834–848. doi:10.1109/TPAMI.2017.2699184

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495. doi:10.1109/TPAMI.2016.2644615

Cheng, B. et al. (2021) 'Boundary IoU: Improving Object-Centric Image Segmentation Evaluation', in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society, pp. 15329–15337. Available at: https://doi.org/10.1109/CVPR46437.2021.01508.

Codella, N.C.F. et al. (2017) 'Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC)', CoRR, abs/1710.05006. Available at: http://arxiv.org/abs/1710.05006.

Educación
Secretaría de Educación Pública

TECNOLÓGICO NACIONAL DE MÉXICO

EL TECNOLÓGICO NACIONAL DE MÉXICO
A TRAVÉS DEL CENTRO NACIONAL DE INVESTIGACIÓN
Y DESARROLLO TECNOLÓGICO

OTORGA EL PRESENTE

RECONOCIMIENTO

A

FERNANDO GARCÍA RAMÍREZ | VÍCTOR ALEJANDRO GONZÁLEZ HUITRÓN | MIGUEL ÁNGEL GUZMÁN RIVERA | EFRÉN GORROSTIETA HURTADO | MARIO ALBERTO MONTANTE GARZA

Por su destacada participación con la Contribución: ARTÍCULO PARA CONGRESO, en el área: CIENCIAS DE LA COMPUTACIÓN, con el Título:
SEGMENTACIÓN SEMÁNTICA DE LESIONES CUTÁNEAS BASADA EN APRENDIZAJE PROFUNDO MEDIANTE MODELOS PREENTRENADOS.
presentada en el marco del 4.º Congreso Internacional de Tecnología y Ciencia Aplicada (CITCA), celebrado en el Centro Nacional de Investigación y Desarrollo Tecnológico del 21 al 23 de mayo de 2025.

Cuernavaca, Morelos, 23 de mayo de 2025

LORENZO O. HERNÁNDEZ
Director del CENIDET

MANUEL ADAM MEDINA
Coordinador General
4.º CITCA

cenidet
Centro Nacional de Investigación y Desarrollo Tecnológico

CITCA
4to. Congreso Internacional de Tecnología y Ciencia Aplicada

Sello Digital:
phAwDRcbcMW8KDyGyAhTop+LBK/6RcoVL1By+DegnDwAK32hsKcn6xxbS77u2w1lpD6AwO4
QHUTTNpmX/2DPqm7KO2u2v24/ww640GP3RpvqRqQ4FnDRUbhbWG24Mcc++v9AcRkDe/MMqK
RvHUDuVf2RPLPuR1VokORNqdD9eEn7UDJBWQFmDwQFmDaEr9qMWMcn+cLaGG3ab+aRCv/P7
w76uIN/4PYr1MZDwKe1OQQDwA74OqMK2qUj1wOG2+xLfxKccQpwkPLZOvKcPoG3Rmv+8IKAQ
J1tAAwTCKVS04ZGTRdcqYfogc6/jSkKYBrn+G/fNP1+4GPG2ZPA74Q==

CITCA256A25 http://constancias.cenidet.tecnm.mx

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

B. Articulo Eureka ENC.

DETECCIÓN DE MELANOMAS EN IMÁGENES MEDIANTE APRENDIZAJE PROFUNDO.

F. García-Ramírez
División de Estudios de Posgrado e Investigación
Tecnológico Nacional de México campus Querétaro
Querétaro, México
fernando.gr@queretaro.tecnm.mx

M. A. Guzmán-Rivera
División de Estudios de Posgrado e Investigación
Tecnológico Nacional de México campus Querétaro
Querétaro, México
miguel.gr@queretaro.tecnm.mx

V. A. González-Huitrón
División de Estudios de Posgrado e Investigación
Tecnológico Nacional de México campus Querétaro
Querétaro, México
victor.gh@queretaro.tecnm.mx

E. Gorrostieta-Hurtado
Universida Autonoma de Querétaro
Facultad de ingeníeria
Querétaro, México
efrengorrostieta@gmail.com

M. A. Montante-Garza
Departamento de Ciencias Báicas
Tecnológico Nacional de México campus Querétaro
Querétaro, México
mario.mg@queretaro.tecnm.mx

***Abstract*. El aprendizaje automático basado en redes neuronales aplicado a la detección de melanomas es un tema complejo, ya que las redes neuronales pueden utilizarse para aproximar funciones en alta dimensionalidad con una eficacia y precisión sobresaliente. El propósito de este trabajo es repasar los principales logros en los casos de éxito y explicar las sutilezas de tales éxitos, así como discutir los principales interrogantes pendientes. Una vez estudiado lo anterior, obtendremos un algoritmo de aprendizaje profundo el cual aplicaremos en una base de datos de imágenes de lesiones de piel. Se analizarán los resultados y se validarán con resultados conocidos que apliquen otros algoritmos. Por último, al discutir los resultados y para trabajos futuros se podrían proponer nuevos algoritmos matemáticos que tengan el mismo éxito.**

Keywords. Aprendizaje automático, aprendizaje profundo, redes neuronales.

I. INTRODUCCIÓN

El aprendizaje automático basado en redes neuronales es extremadamente poderoso y delicado. Por un lado, puede utilizarse para aproximar funciones en espacios de alta dimensión con una eficacia y exactitud sin precedentes. Lo cual ha abierto posibilidades totalmente nuevas en un amplio espectro de diferentes disciplinas. El éxito depende de la matemática aplicada, y el ajuste de los parámetros.

Lo principal de un estudio matemático en el aprendizaje automático es:

1. Explicar las razones de éxito y sutilezas,
2. Proponer nuevos modelos matemáticos que tengan el mismo éxito.

El propósito de este trabajo es repasar los principales logros en la consecución del primer objetivo y discutir los principales enigmas pendientes. Siguiendo la tradición de las matemáticas aplicadas, no sólo se prestará atención a los resultados matemáticos rigurosos, sino que también se discutirán los conocimientos que se obtengan a partir de experimentos numéricos, así como del análisis de modelos simulados.

Por el momento se ha prestado menos atención al segundo objetivo, por la rigurosidad de la matemática. La idea en general es formular modelos continuos de problemas de aprendizaje automático y luego discretizar para obtener algoritmos concretos.

Los algoritmos del aprendizaje profundo más comunes son: redes neuronales profundas (DNN) [1], redes neuronales convolucionales (CNN) [2], redes neuronales recurrentes (RNN) [2], redes neuronales de transformadores (Transformers) [3], Redes Neuronales de Autoencoders [3], Redes Neuronales Generativas Antagónicas (GANs) [4], Redes Neuronales de Grafos (GNN) [4]. Es importante señalar que el aprendizaje profundo es un subcampo del aprendizaje automático usado en la inteligencia artificial el cual es fundamental para el desarrollo de un sistema de inteligencia artificial.

Los algoritmos de detección médica son una de las primeras herramientas utilizadas para determinar si una lesión cutánea es maligna o benigna [5-7]. Nachbar et al. [5] desarrollaron un método subjetivo basado en la percepción visual de la lesión. La regla ABCD se basa en la asimetría, bordes, color, forma y diámetro que aparecen en las lesiones cutáneas. Debido a la simplicidad del algoritmo, es uno de los más practicados para evaluar una lesión mediante un examen a simple vista o utilizando un dermatoscopio.

El algoritmo médico ABCD se compone de las siguientes partes:

- **Asimetría A**: La lesión se divide en dos ejes perpendiculares a 90° entre sí, de forma que se obtenga la menor puntuación de asimetría posible. En otras palabras, se determina si la lesión es simétrica o no. Por cada eje en el que se detecta asimetría, se añade un punto.
- **Bordes B**: La lesión se divide en cortes por ocho ejes que determinan si una lesión presenta bordes abruptos. Si un segmento presenta un borde abrupto, se añade un punto.
- **Colores C**: La lesión puede presentar uno o varios de los siguientes colores: blanco, parduzco, marrón oscuro, negro, azul y rojo. Son generados por los vasos y las concentraciones de melanina, por lo que por cada color encontrado se suma un punto por color.
- **Estructuras Dermatoscópica D**: La lesión presenta el aspecto de las siguientes estructuras: puntos, manchas, redes pigmentadas y zonas no estructuradas. Se añade un punto por cada estructura encontrada en la lesión.

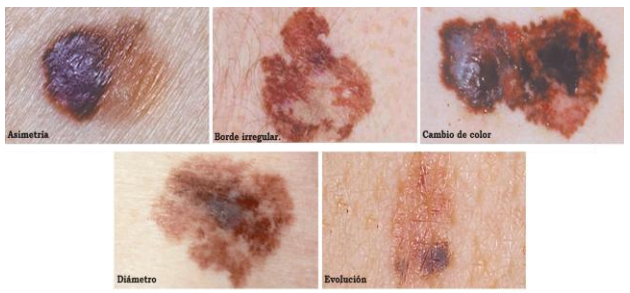


Figura 1. Imágenes cutáneas con características para la regla ABCD [8].

Las características descritas se ponderan del siguiente modo:

$$TDS = (1.3 \cdot A) + (0.1 \cdot B) + (0.5 \cdot C) + (0.5 \cdot D) \qquad (1)$$

donde TDS es la Puntuación Dermatoscópica Total si es inferior a 4.75 se concluye que la lesión es benigna si la puntuación está entre 4.75 y 5.45 la lesión se considera sospechosa si es superior a 5.45 se considera maligna.

II. ANTECEDENTES

El problema de un modelo del aprendizaje automático en el que se centra este trabajo puede formularse del siguiente modo. Dado un conjunto de datos,

$$S = \{(x_i, y_i = f^*(x_i)),\ i \in [n]\}, \qquad (2)$$

es aproximar a f^* con la mayor precisión posible. Si f^* toma valores continuos, se trata de un problema de regresión, y si f^* toma valores discretos, se trata de un problema de clasificación, n es el tamaño del conjunto de datos de entrenamiento.

Este trabajo se concentra en el problema de clasificación, el cual se trata de un problema de aproximación de funciones. Como tal, puede considerarse un problema de análisis numérico o un problema estadístico. Se adoptará el primer punto de vista, ya que se ajusta más a las cuestiones algorítmicas y de análisis que se trabajarán.

El procedimiento estándar para el aprendizaje automático es el siguiente:

- Se elije un espacio de hipótesis, el cual es el conjunto de funciones de prueba, que se denotará por H_m. En el análisis numérico clásico, se suelen utilizar polinomios o polinomios a trozos. En el aprendizaje automático moderno, se utilizan modelos de redes neuronales. El subíndice m caracteriza el tamaño del modelo, puede ser el número de parámetros o de neuronas, y se especifica para cada modelo.
- Se elige una función de pérdida. El objetivo principal es ajustar los datos, por lo tanto, la elección es la función riesgo empírico:

$$\hat{R}_n(f) = \frac{1}{n}\sum_i \big(f(\boldsymbol{x}_i) - f^*(\boldsymbol{x}_i)\big)^2. \qquad (3)$$

  Algunas veces se añaden algunos términos de regularización.
- Elegir un algoritmo de optimización y los hiperparámetros. Las opciones más populares son el descenso de gradiente (GD), el descenso de gradiente estocástico (SGD) y optimizadores avanzados como *Adam* o *RMSprop*.

III. OBJETIVO GENERAL

Implementar un modelo matemático que minimice a la función de riesgo de población, también conocido como error de generalización:

$$R_n(f) = E_{x\sim P}\big(f(\boldsymbol{x}_i) - f^*(\boldsymbol{x}_i)\big)^2. \qquad (4)$$

Donde $E_{x\sim P}$ denota el error, $f(\boldsymbol{x}_i)$ la función de los datos de entrada y $f^*(\boldsymbol{x}_i)$ los datos de salida.

IV. METODOLOGÍA

La metodología propuesta para llevar a cabo la investigación se muestra en el siguiente diagrama.

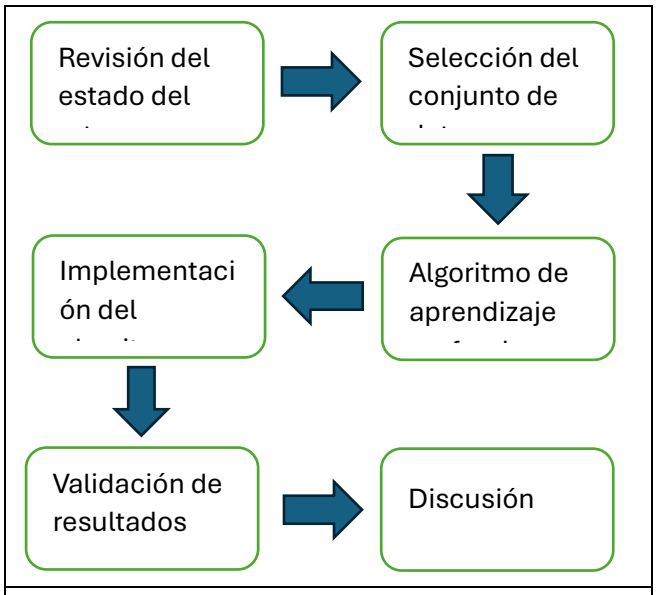


Figura 2. Diagrama de flujo de la metodológica del proceso de investigación.

Revisión del estado del arte. En dicha etapa se explorará la literatura existente para conocer de manera general los avances de los algoritmos matemáticos dentro del aprendizaje automático para coadyuvar en el análisis de imágenes para la detección, predicción y diagnóstico de melanomas. A continuación, se presenta un cuadro comparativo de algunos artículos con un impacto relevante en la investigación de interés.

TABLE I. COMPARACIÓN DE ACCURACY DE ALGUNOS TRABAJOS DE INTERÉS.

Referencia	**Autor**	**Base de datos**	**Accuracy**
[9]	J. A. Almaraz *et al.*	ISIC 2018 [10]	92.40
[11]	F. Adjed *et al.*	Mendonca, T. *et al.* [12]	86.07
[13]	Q. Abbas y M. E. Celebi	N. H. Quang *et al.* [14]	95
[15]	X. Li *et al.*	ISIC 2018 [10]	85.55

Selección del conjunto de datos. El conjunto de datos seleccionado para la presente investigación es <<International Skin Imaging Collaboration. SIIM-ISIC 2020 Challenge Dataset [16]>> El conjunto de datos contiene 33,126 imágenes de entrenamiento dermatoscópicas de lesiones cutáneas con 26,199 benignas y 6,927 malignas de más de 2,056 pacientes. Cada imagen se asocia a uno de estos individuos mediante un identificador único de paciente. Todos los diagnósticos malignos se han confirmado mediante histopatología y los diagnósticos benignos se han confirmado mediante el acuerdo de expertos, el seguimiento longitudinal o la histopatología. Una publicación exhaustiva que describe todas las características de este conjunto de datos está disponible en forma de preimpresión que aún no ha sido sometida a revisión por pares. El conjunto de datos fue generado por la Colaboración Internacional de Imágenes de la Piel (ISIC) y las imágenes provienen de las siguientes fuentes: Hospital Clínic de Barcelona, Universidad Médica de Viena, Memorial Sloan Kettering Cancer Center, Instituto de Melanoma de Australia, Universidad de Queensland y la Facultad de Medicina de la Universidad de Atenas.

Arquitecturas o algoritmos de aprendizaje profundo. Analizados los algoritmos de aprendizaje profundo antes mencionados, se elegirá el que se adapte mejor al desarrollo de la investigación.

Implementación del algoritmo. Las características de aprendizaje profundo se obtendrán utilizando una arquitectura de una red neuronal seleccionada, que se entrenara en una tarea de clasificación de imágenes; esta red neuronal se empleara como extractor de características. Todas las características extraídas más importantes se concatenan en un vector, que posteriormente se fusiona según el criterio la métrica de la regla ABCD. El clasificador seleccionado se entrena en el conjunto de datos, compuesto por imágenes de lesiones cutáneas malignas y benignas. Por último, el modelo clasificador entrenado se utiliza para predecir cada imagen de lesión cutánea pigmentada no vista como lesión benigna o maligna.

Validación de resultados. Una vez concluida la fase de análisis dada por el clasificador, se procederá a validar los resultados obtenidos para determinar si son los esperados o no. Dichos resultados se compararán con algún trabajo previo para comprobar si dicho algoritmo matemático mejoro o no la optimización de los parámetros del sistema.

En general se aplican métricas de rendimiento, como:

- El valor de exactitud mide las clasificaciones adecuadas sobre el total de elementos evaluados,

$$Accuracy = \frac{TP+tn}{tp+tn+fp+fn}. \quad (5)$$

- El valor de sensibilidad, también conocido como *recall*, mide el número de elementos positivos clasificados correctamente,

$$Recall = \frac{tp}{tp+fn}. \quad (6)$$

- El valor de precisión mide el número de elementos correctamente clasificados entre todos los elementos positivos a evaluar,

$$Precision = \frac{tp}{tp+fp}. \quad (7)$$

- El valor de la puntuación F mide la media armónica entre precisión y sensibilidad (*recall*),

$$F_{Score} = \frac{2tp}{2tp+fp+fn}. \quad (8)$$

Estos criterios se describen en términos de tp, tn, fp y fn, que denotan verdadero positivo, verdadero negativo, falso positivo y falso negativo, respectivamente.

Discusión. Para concluir, los resultados obtenidos se discutirán para contrastar con los algoritmos matemáticos usados actualmente y obtener nuevo conocimiento dentro del aprendizaje profundo.

V. RESULTADOS ESPERADOS

En la implementación se espera la mejora del algoritmo matemático propuesto que tiene como objetivo el minimizar el error de generalización y aplicarlo a un algoritmo de aprendizaje profundo, una vez realizadas estas etapas con los estudios y ajustes necesarios, se aplicara a una base de datos de imágenes de lesiones cutáneas para detectar de manera más eficientes los melanomas en la piel.

VI. CONCLUSIONES

La revisión del estado del arte demuestra los avances que ha tenido el aprendizaje profundo en diferentes ramas de la ciencia y tecnología, aunado a ello hay una gran oportunidad en un amplio espectro de la matemática ya que algunos problemas se encuentran abiertos en la actualidad. Nuestra investigación se enfoca en un problema matemático en particular y coadyubar en el modelo de inteligencia artificial para la detección de melanomas.

REFERENCIAS

[1] D. Sauter, G. Lodde, F. Nensa, D. Schadendorf, E. Livingstone, and M. Kukuk, 'Deep learning in computational dermatopathology of melanoma: A technical systematic literature review', *Computers in Biology and Medicine*, vol. 163, p. 107083, 2023.

[2] D. Popescu, M. El-Khatib, H. El-Khatib, and L. Ichim, 'New Trends in Melanoma Detection Using Neural Networks: A Systematic Review', *Sensors*, vol. 22, no. 2, 2022.

[3] H. Bhatt, V. Shah, K. Shah, R. Shah, and M. Shah, 'State-of-the-art machine learning techniques for melanoma skin cancer detection and classification: a comprehensive review', *Intelligent Medicine*, vol. 03, no. 03, pp. 180–190, 2023.

[4] M. Dildar *et al.*, 'Skin Cancer Detection: A Review Using Deep Learning Techniques', *International Journal of Environmental Research and Public Health*, vol. 18, no. 10, 2021.

[5] F. Nachbar *et al.*, 'The ABCD rule of dermatoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions', *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551–559, 1994.

[6] I. Zalaudek *et al.*, 'Three-point checklist of dermoscopy: an open internet study', *British Journal of Dermatology*, vol. 154, no. 3, pp. 431–437, 03 2006.

[7] J. S. Henning *et al.*, 'The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy', *Journal of the American Academy of Dermatology*, vol. 56, no. 1, pp. 45–52, 2007.

[8] Escrito por el personal de Mayo Clinic, "Fotos de melanomas para identificar Cáncer de Piel," Mayo Clinic, 16-Mar-2024. [Online]. Available: https://www.mayoclinic.org/es/diseases-conditions/melanoma/in-depth/melanoma/art-20546856. [Accessed: 02-Aug-202

[9] J.-A. Almaraz-Damian, V. Ponomaryov, S. Sadovnychiy, and H. Castillejos-Fernandez, 'Melanoma and Nevus Skin Lesion Classification Using Handcraft and Deep Learning Feature Fusion via Mutual Information Measures', *Entropy*, vol. 22, no. 4, 2020.

[10] N. Codella *et al.*, 'Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)', *arXiv preprint arXiv:1902. 03368*, 2019.

[11] F. Adjed, S. J. Safdar Gardezi, F. Ababsa, I. Faye, and S. Chandra Dass, 'Fusion of structural and textural features for melanoma recognition', *IET Computer Vision*, vol. 12, no. 2, pp. 185–195, 2018.

[12] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, 'PH 2-A dermoscopic image database for research and benchmarking', in *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2013, pp. 5437–5440.

[13] Q. Abbas and M. E. Celebi, 'DermoDeep-A classification of melanoma-nevus skin features and deep neural network', *Multimedia Tools and Applications*, vol. 78, no. 16, pp. 23559–23580, 2019.

[14] N. H. Quang and Others, 'Automatic skin lesion analysis towards melanoma detection', in *2017 21st Asia Pacific symposium on intelligent and evolutionary systems (IES)*, 2017, pp. 106–111.

[15] X. Li, J. Wu, H. Jiang, E. Z. Chen, X. Dong, and R. Rong, 'Skin lesion classification via combining deep learning features and clinical criteria representations', BioRxiv, p. 382010, 2018.

[16] "The ISIC 2020 Challenge Dataset," ISIC 2020 Challenge Dataset. [Online]. Available: https://challenge2020.isic-archive.com/. [Accessed: 02-Aug-202

EL INSTITUTO TECNOLÓGICO DE CD. MADERO Y
LA SOCIEDAD MEXICANA DE CIENCIA DE LA COMPUTACIÓN
OTORGAN EL PRESENTE

RECONOCIMIENTO

A

Fernando García Ramírez, Miguel Ángel Guzmán Rivera, Víctor Alejandro González Huitrón, Efrén Gorrostieta Hurtado, Mario Alberto Montante Garza

Por haber presentado el artículo:

DETECCIÓN DE MELANOMAS EN IMÁGENES MEDIANTE APRENDIZAJE PROFUNDO

en el Coloquio de Estudiantes del Encuentro Nacional de Computación, Eureka ENC 2024, llevado a cabo del 23 al 25 de septiembre en Cd. Madero, Tamaulipas.

Dra. Yasmín Hernández Pérez
Presidenta de la SMCC

Dra. Laura Cruz Reyes
Organizadora Eureka ENC 2024

Dra. Claudia Gómez Santillán
Organizadora Eureka ENC 2024



C. Trabajo en el Priemer Foro Nacional de la Agenda Estratégica de la Salud del Tecnológico Nacional de México.

Diseño de un Sistema de Apoyo al Diagnóstico para la Detección de Melanomas en Imágenes Dermatoscópicas mediante Inteligencia Artificial

Resumen: Este proyecto tiene como objetivo desarrollar un sistema de apoyo al diagnóstico basado en inteligencia artificial para la detección temprana de melanomas a partir de imágenes dermatoscópicas. El melanoma es uno de los tipos de cáncer de piel más agresivos y con alta mortalidad si no se detecta a tiempo. En México, donde la incidencia de cáncer de piel está en aumento debido a la exposición prolongada al sol y otros factores ambientales, el diagnóstico temprano y preciso es fundamental. Sin embargo, el diagnóstico de melanoma sigue siendo un desafío debido a la alta dependencia de la experiencia del dermatólogo y la variabilidad en la interpretación de las imágenes. El sistema propuesto busca agilizar el proceso de diagnóstico, proporcionando una herramienta precisa y rápida que asista a los especialistas en la identificación de lesiones sospechosas, optimizando los recursos médicos y reduciendo los costos asociados al diagnóstico.

Justificación: El cáncer de piel, y especialmente el melanoma, está experimentando un aumento en su incidencia en México, debido a factores como la exposición excesiva a la radiación ultravioleta y cambios en el comportamiento social. Detectar estos cánceres de manera temprana es crucial para mejorar las tasas de supervivencia, pero el diagnóstico se ve limitado por la variabilidad en la interpretación de las imágenes dermatoscópicas y la falta de especialistas en ciertas áreas del país. Este proyecto propone la implementación de un sistema basado en inteligencia artificial que pueda analizar imágenes dermatoscópicas y asistir a los dermatólogos en la identificación de melanomas. Al automatizar una parte del proceso de diagnóstico, se pretende reducir el margen de error, acelerar la toma de decisiones y disminuir el uso de materiales costosos como biopsias innecesarias, al mismo tiempo que se mejora la cobertura del diagnóstico, especialmente en regiones con menos acceso a especialistas.

Productos obtenidos a la fecha:

- Artículo con diseño de la propuesta presentado en el Encuentro Nacional de Computación 2024 celebrado en IT Cd Madero
- Modelos entrenados mediante aprendizaje profundo a partir de datos reales del estado del arte

Relación Agenda salud TecNM: Este proyecto se alinea con la Agenda Estratégica de la Salud del Tecnológico Nacional de México (TecNM) al fomentar la sinergia entre las áreas de ingeniería y el sector salud, contribuyendo al fortalecimiento del sistema de salud del país. Al integrar tecnologías avanzadas como la inteligencia artificial en el diagnóstico de melanoma, se potencia la capacidad de los especialistas médicos para tomar decisiones más rápidas y precisas, lo que mejora la eficiencia y calidad de la atención. La colaboración entre ingenieros y médicos en proyectos como este refuerza la idea de una agenda estratégica que promueva la innovación tecnológica aplicada a la salud, generando soluciones que impacten directamente en el diagnóstico, tratamiento y prevención de enfermedades. Este enfoque multidisciplinario es clave para enfrentar los desafíos de salud pública en México, y representa un aliciente importante para proponer y consolidar la Agenda Estratégica de la Salud del TecNM, que busca integrar soluciones tecnológicas de vanguardia en el sector salud, optimizando recursos y mejorando la cobertura y accesibilidad de la atención médica.

Asesores principales:

MTE Miguel Angel Guzman Rivera

Correo: miguel.gr@queretaro.tecnm.mx

Dr. Víctor alejandro González Huitrón

Correo: victor.gh@queretaro.tecnm.mx

Departamento de Sistemas y Computación

Estudiante:

Lic. Fernando García Ramírez

Programa académico Maestría en Ciencia de Datos

Correo: fernando.gr@queretaro.tecnm.mx

Educación
Secretaría de Educación Pública

TECNOLÓGICO NACIONAL DE MÉXICO

EL TECNOLÓGICO NACIONAL DE MÉXICO

A TRAVÉS DE LA

SECRETARÍA DE PLANEACIÓN, EVALUACIÓN Y DESARROLLO INSTITUCIONAL

OTORGA EL PRESENTE

RECONOCIMIENTO

A

Fernando García Ramírez

Por su participación como ponente en el **Primer Foro Nacional de la Agenda Estratégica de la Salud del TecNM**, celebrado los días 10 y 11 de abril de 2025, presentando la ponencia titulada: **"Diseño de un Sistema de Apoyo al Diagnostico para la Detección de Melanomas en Imágenes Dermatoscópicas mediante Inteligencia Artificial"**

JORGE SANTOS VALENCIA
SECRETARIO DE PLANEACIÓN, EVALUACIÓN
Y DESARROLLO INSTITUCIONAL

CIUDAD DE MÉXICO A, 11 DE ABRIL DE 2025