



UNIVERSIDAD ESAN
FACULTAD DE INGENIERÍA
INGENIERÍA DE TECNOLOGÍAS DE INFORMACIÓN Y SISTEMAS

Segmentación Avanzada de Características Morfológicas en Imágenes de Piel Facial usando
Redes Neuronales Convolucionales: Enfoque en Arrugas y Poros

Tesis para optar por el Bachiller de Ingeniería de Tecnologías de Información y Sistemas

Victor Manuel Ormeño Salazar
Jeferson Joseph Sandoval Díaz
Asesor: Junior Fabián Arteaga

Lima, 5 de abril de 2025

Esta tesis denominada:

DISEÑO DE UN MODELO DE REDES NEURONALES CONVOLUCIONALES PARA LA
SEGMENTACIÓN DE CARACTERÍSTICAS MOFOLÓGICAS EN IMÁGENES
FACIALES

.....

Jurado 1

.....

Jurado 2

Universidad ESAN
2025

Segmentación Avanzada de Características Morfológicas en Imágenes de Piel Facial usando
Redes Neuronales Convolucionales: Enfoque en Arrugas, Poros y Manchas

Dedicatoria

Thx.

Índice general

Resumen	1
Introducción	3
Capítulo I: Planteamiento del Problema	4
1.1 Descripción de la Realidad Problemática	4
1.2 Formulación del Problema	5
1.2.1 Problema General	5
1.2.2 Problemas Específicos	5
1.3 Objetivos de la Investigación	5
1.3.1 Objetivo General	6
1.3.2 Objetivos Específicos	6
1.4 Justificación de la Investigación	6
1.4.1 Teórica	6
1.4.2 Práctica	6
1.4.3 Metodológica	7
1.5 Delimitación del Estudio	7
1.5.1 Espacial	7
1.5.2 Temporal	7
1.5.3 Conceptual	8
Capítulo II: Marco Teórico	9
2.1 Antecedentes de la investigación	9
2.2 Bases Teóricas	22
2.2.1 Inteligencia Artificial	22
2.2.2 Aprendizaje Automático	22
2.2.3 Aprendizaje Profundo	25
2.2.4 Aprendizaje Profundo Multimodal	26
2.2.5 Inteligencia Artificial Generativa	28
2.2.6 Segmentación de Imágenes	30
2.2.7 Redes Neuronales Convolucionales (CNN)	33
2.2.8 Redes Generativas Adversarias (GANs)	35
2.2.9 Modelos Avanzados de Segmentación en Imágenes Médicas	39
2.3 Marco Conceptual	42
2.3.1 Características Morfológicas de la Piel	42
2.3.2 Redes de Atención	45

2.4	Hipótesis	48
2.4.1	Hipótesis General	48
2.4.2	Hipótesis Específicas	48
Capítulo III:	Metodología de la Investigación	49
3.1	Diseño de la investigación	49
3.1.1	Alcance de la investigación	49
3.1.2	Enfoque de la investigación	49
3.1.3	Población	50
3.1.4	Muestra	50
3.2	Metodología de Implementación de la Solución	50
3.2.1	Adquisición y Preparación de los Datos	50
3.3	Diagrama de la Metodología	51
3.3.1	Desarrollo de los Modelos de Segmentación	52
3.3.2	Evaluación y Validación del Sistema	53
3.3.3	Implementación y Pruebas Finales	53
3.4	Metodología para la Medición de Resultados de la Implementación	53
3.5	Cronograma de actividades y presupuesto	55
Capítulo IV:	Desarrollo de la Solución	57
4.1	Determinación y evaluación de alternativas de solución	57
4.2	Propuesta solución	57
4.2.1	Planeamiento y descripción de Actividades	57
4.2.2	Desarrollo de actividades	57
4.3	Medición de la solución	57
4.3.1	Análisis de Indicadores cuantitativo	57
Capítulo V:	Conclusiones y Recomendaciones	58
Referencias	61
Anexos	65
A	Árbol de Problemas	66
B	Árbol de Objetivos	67
C	Matriz de Consistencia	68

Índice de Figuras

Figura 1.	Comparación del modelo propuesto con otros 6 modelos de redes neuronales	13
Figura 2.	Estructura de red de DeepLabV3+	14
Figura 3.	El proceso de convolución separable en profundidad	14
Figura 4.	Diagrama de flujo del sistema propuesto	15
Figura 5.	Modelo totalmente convolucional propuesto para la segmentación de lesiones cutáneas	16
Figura 6.	Propuesta basada en RethNet RE-Xception y el decodificador de DeepLabv3+	17
Figura 7.	Detección de edad y género	19
Figura 8.	Análisis del tono de piel	19
Figura 9.	Análisis de arrugas	20
Figura 10.	Entrenamiento en dos etapas para la segmentación de arrugas faciales . . .	21
Figura 11.	El algoritmo de K medias	24
Figura 12.	Relación entre IA, ML y DL	25
Figura 13.	Modelo de aprendizaje profundo	26
Figura 14.	Modelo que combina imágenes y texto	27
Figura 15.	Un modelo multimodal para las señales de la vista	28
Figura 16.	Modelos de difusión	29
Figura 17.	Autocodificadores variacionales	30
Figura 18.	Arquitectura de un modelo CNN	34
Figura 19.	Red Generativa Antagónica de Imágenes	36
Figura 20.	Red Generativa Antagónica de Imágenes	37
Figura 21.	Imágenes de Ruido Gaussiano	37
Figura 22.	Diagrama de la Metodología	51
Figura 23.	Cronograma de actividades	55

Índice de Tablas

Tabla 1.	Comparación de Modelos: Precisión de Píxeles, IoU y Puntuación F1	10
Tabla 2.	Comparación de diferentes métodos de segmentación de lesiones cutáneas en el conjunto de datos ISIC 2018.	10
Tabla 3.	Performance Metrics of Different Models	11
Tabla 4.	Presupuesto	56

Índice de Ecuaciones

Índice de Algoritmos

Resumen

La industria cosmética y de cuidado de la piel ha experimentado un crecimiento acelerado en las últimas décadas. Sin embargo, a pesar de la diversidad de productos y tratamientos disponibles, persisten desafíos importantes en la evaluación precisa y personalizada de problemas estéticos, como arrugas, poros dilatados y manchas faciales. Uno de los mayores obstáculos es la falta de herramientas tecnológicas avanzadas que permitan un análisis profundo y objetivo de las características morfológicas de la piel.

Este trabajo de investigación busca desarrollar un sistema de segmentación basado en redes neuronales convolucionales (CNN) para detectar y evaluar características morfológicas clave de la piel facial, como arrugas, poros y manchas. El sistema propuesto mejorará la precisión en la evaluación estética y la personalización de tratamientos cosméticos. Se utilizarán arquitecturas de CNN adaptadas para segmentar con precisión las imágenes de la piel facial, basándose en un conjunto de datos representativo que permita entrenar y validar el modelo. Las métricas de desempeño, como la precisión, el recall y el F1-score, serán empleadas para evaluar la eficiencia del sistema.

Este avance contribuirá tanto al ámbito de la dermatología computacional como al de la cosmética, optimizando los procesos de diagnóstico y recomendación de productos personalizados. La implementación de este sistema permitirá mejorar la precisión de los diagnósticos y personalizar tratamientos de manera más efectiva, beneficiando a consumidores que buscan soluciones específicas para sus necesidades estéticas.

Palabras clave: Segmentación, redes neuronales convolucionales, arrugas, poros, manchas, imagen de la piel, cosmética, tratamiento personalizado.

Abstract

The cosmetics and skincare industry has seen rapid growth in recent decades. However, despite the variety of products and treatments available, there remain significant challenges in the precise and personalized evaluation of aesthetic issues, such as wrinkles, enlarged pores, and skin spots. One of the major obstacles is the lack of advanced technological tools that allow for a deep and objective analysis of the skin's morphological features.

This research aims to develop a segmentation system based on Convolutional Neural Networks (CNN) to detect and evaluate key morphological features of facial skin, such as wrinkles, pores, and spots. The proposed system will enhance the accuracy of aesthetic evaluations and the personalization of cosmetic treatments. CNN architectures will be used to segment facial skin images accurately, using a representative dataset to train and validate the model. Performance metrics such as accuracy, recall, and F1-score will be used to assess the system's efficiency.

This advancement will contribute to both computational dermatology and cosmetics by optimizing diagnostic processes and personalized product recommendations. The implementation of this system will improve diagnostic precision and allow for more effective treatment personalization, benefiting consumers seeking solutions tailored to their specific aesthetic needs.

Keywords: Segmentation, convolutional neural networks, wrinkles, pores, spots, skin image, cosmetics, personalized treatment.

Introducción

En resumen, el principal objetivo de la presente investigación es el siguiente: desarrollar un sistema avanzado de segmentación de características morfológicas de la piel facial, es decir, las arrugas, los poros y las manchas, utilizando redes neuronales convolucionales (CNN). En otras palabras, este sistema tiene como objetivo mejorar la precisión de la valoración estética y, por lo tanto, fortalecer la personalización del tratamiento cosmético y favorecer el desarrollo de la industria de la piel. El sector cosmético se ha desarrollado de forma sostenida en las últimas décadas, a medida de una creciente preocupación en torno a problemas estéticos y el interés de mantener la piel sana y joven. No obstante, la mayor parte del diagnóstico de enfermedades cutáneas se basa en la observación manual o utilizando herramientas tradicionales, que son a menudo inexactas y no objetivas. Por lo tanto, se necesita urgentemente el desarrollo de nuevas tecnologías para evaluar con mayor detalle y precisión las características de la piel. En este contexto, las CNN han demostrado ser altamente efectivas para tareas de segmentación y análisis de imágenes, permitiendo identificar patrones complejos en datos visuales. Este estudio se enfoca en aplicar estas técnicas para desarrollar un modelo capaz de detectar y segmentar características clave de la piel, como arrugas, poros dilatados y manchas faciales, con el fin de brindar un análisis objetivo y detallado. La investigación sigue un enfoque metodológico que incluye la recopilación de datos, el desarrollo y prueba de modelos basados en CNN, y la evaluación de su desempeño mediante métricas como precisión, recall, F1-score y AUC-ROC. Este enfoque asegura la validación y confiabilidad del sistema desarrollado, destacando su aplicabilidad en el sector cosmético.

Capítulo I: Planteamiento del Problema

1.1 Descripción de la Realidad Problemática

La industria cosmética y de cuidado de la piel ha visto un crecimiento exponencial en las últimas décadas. Según datos de Statista, el mercado global de productos para el cuidado de la piel superó los 130 mil millones de dólares en 2023, con una tasa de crecimiento anual compuesta de aproximadamente el 0,045 (Statista, 2023). Este auge refleja la creciente demanda por soluciones que permitan a los consumidores mejorar su apariencia, retardar los signos del envejecimiento y solucionar problemas estéticos como las arrugas, los poros dilatados y las manchas faciales.

A pesar de la variedad de productos y tratamientos disponibles en el mercado, persisten varios desafíos relacionados con la evaluación precisa y personalizada de los problemas de la piel. Uno de los mayores obstáculos en la actualidad es la falta de herramientas tecnológicas que permitan realizar un análisis profundo y cuantitativo de las características morfológicas de la piel. Problemas como las arrugas, los poros dilatados y las manchas son difíciles de evaluar de manera objetiva, ya que la mayoría de los diagnósticos aún dependen de la observación manual o de sistemas de análisis que no logran captar las sutilezas y complejidades de la piel facial (Phillips & Zhang, 2020).

La precisión en la detección de estas características es crucial para el desarrollo de tratamientos más efectivos. Por ejemplo, la identificación temprana de las arrugas incipientes permitiría la aplicación de productos antiarrugas de manera preventiva, antes de que las líneas se profundicen. Sin embargo, las evaluaciones actuales a menudo son subjetivas y pueden variar según el profesional o las herramientas utilizadas. En este sentido, estudios recientes han demostrado que las redes neuronales convolucionales (CNN) tienen el potencial de revolucionar la manera en que se realiza la segmentación y análisis de imágenes de piel (Esteva et al., 2017). Estas redes permiten procesar grandes volúmenes de datos visuales y detectar patrones morfológicos con un nivel de precisión que supera las técnicas tradicionales.

Otro desafío significativo es la evaluación de poros y manchas. Los poros dilatados son una preocupación estética común, especialmente entre personas con piel grasa. La falta de herramientas capaces de cuantificar y analizar adecuadamente el tamaño y la densidad de los poros limita las recomendaciones de tratamiento personalizadas (Jia et al., 2019). De manera similar, las manchas faciales, que pueden aparecer debido a la edad, la exposición solar o factores hormonales, requieren una evaluación temprana para evitar su progresión. Las CNN, al especializarse en la segmentación de imágenes, pueden proporcionar una solución eficaz para mapear la distribución y evolución de estas imperfecciones cutáneas (Esteva et al., 2017).

Estadísticamente, se ha encontrado que más del 0,7 de las personas mayores de 25 años muestran algún signo de envejecimiento facial, como arrugas y manchas, lo que impulsa la demanda de productos antiarrugas y despigmentantes (American Academy of Dermatology, 2022). Sin embargo, el éxito de estos productos depende en gran medida de una evaluación precisa del estado de la piel. Actualmente, los diagnósticos imprecisos o subjetivos pueden llevar a la aplicación de productos inapropiados, lo que no solo afecta la satisfacción del consumidor, sino que también reduce la efectividad de los tratamientos (Khatri & Maibach, 2018).

Este contexto evidencia la necesidad urgente de desarrollar tecnologías avanzadas que mejoren la precisión en la evaluación estética de la piel. Las redes neuronales convolucionales ofrecen una herramienta prometedora para abordar este problema, al permitir una segmentación detallada de las características morfológicas clave de la piel, como arrugas, poros y manchas. El uso de estas tecnologías no solo permitiría mejorar los diagnósticos estéticos, sino que también contribuiría a la creación de tratamientos personalizados más efectivos, incrementando la satisfacción del usuario final.

1.2 Formulación del Problema

1.2.1 Problema General

PG: ¿Cómo afecta la falta de un sistema de segmentación de características morfológicas de la piel facial en la detección de arrugas, poros y manchas?

1.2.2 Problemas Específicos

- PE1: ¿Cómo se medirá la eficiencia del sistema de segmentación morfológica en la detección de arrugas, poros y manchas?
- PE2: ¿Cómo se desarrollará el sistema de segmentación basado en redes neuronales convolucionales (CNN)?
- PE3: ¿De dónde se obtendrá la data para entrenar y validar el sistema de segmentación?
- PE4: ¿Cómo se determinará cuál es el mejor modelo de segmentación para la detección de características morfológicas?

1.3 Objetivos de la Investigación

A continuación, se presentan el objetivo general y los objetivos específicos.

1.3.1 Objetivo General

OG: Desarrollar un sistema avanzado de segmentación de características morfológicas en imágenes de piel facial, centrado en arrugas, poros y manchas, utilizando redes neuronales convolucionales para mejorar la precisión en la evaluación estética y la personalización de tratamientos cosméticos.

1.3.2 Objetivos Específicos

- OE1: Desarrollar métricas de evaluación como precisión, recall, F1-score y AUC-ROC para medir la eficiencia del sistema de segmentación en la identificación de características morfológicas de la piel facial.
- OE2: Desarrollar e implementar un sistema de segmentación utilizando redes neuronales convolucionales, adaptando sus arquitecturas para la detección y diferenciación de características morfológicas de la piel como arrugas, poros y manchas.
- OE3: Recopilar y preparar un conjunto de datos de imágenes faciales, asegurando que contenga suficiente variedad en términos de diferentes tipos de piel y problemas cutáneos (arrugas, poros y manchas), para entrenar y validar el sistema de segmentación.
- OE4: Comparar diferentes arquitecturas de redes neuronales convolucionales y técnicas de aprendizaje profundo, evaluando su desempeño en la segmentación de arrugas, poros y manchas, para seleccionar el modelo que ofrezca el mejor equilibrio entre precisión y eficiencia computacional.

1.4 Justificación de la Investigación

1.4.1 Teórica

Este estudio formula el desarrollo de un modelo de redes neuronales convolucionales (CNN), que han demostrado ser altamente eficaces para tareas de segmentación y clasificación de imágenes. El desarrollo de un sistema de segmentación morfológica permitirá avanzar en el campo de la dermatología computacional y los sistemas inteligentes aplicados a la cosmética y el cuidado de la piel. Además, se contribuirá a la literatura sobre la intersección entre las ciencias de la salud y la inteligencia artificial, explorando el uso de técnicas como el deep learning para detectar y segmentar características faciales como arrugas, poros y manchas.

1.4.2 Práctica

El sistema de segmentación propuesto puede tener aplicaciones directas en la industria cosmética, dermatológica y médica. La detección temprana y precisa de imperfecciones en la

piel es esencial para tratamientos preventivos y correctivos. Este sistema permitirá automatizar y optimizar la evaluación cutánea, mejorando la precisión y reduciendo el tiempo necesario para diagnósticos o recomendaciones personalizadas de tratamiento. Esto también podría ser útil para mejorar la personalización de productos de belleza, apoyando la creación de soluciones adaptadas a las necesidades específicas de la piel de cada persona.

1.4.3 Metodológica

Utilizar arquitecturas probadas de CNN para segmentar características específicas como arrugas, poros y manchas proporciona una metodología efectiva y replicable. Al emplear técnicas de deep learning, se puede mejorar la precisión en la segmentación de imágenes de la piel facial, lo que facilitará comparaciones entre distintas aproximaciones y modelos. La experimentación con distintos modelos también ayudará a determinar cuál es el enfoque más eficiente y preciso para este tipo de segmentación.

1.5 Delimitación del Estudio

Este estudio se centrará exclusivamente en la segmentación de arrugas, poros y manchas en imágenes de piel facial. No se abordarán otras imperfecciones cutáneas ni se desarrollarán sistemas de diagnóstico clínico. El enfoque está dirigido hacia el sector cosmético y de belleza, en lugar de la dermatología médica. Además, el estudio se delimitará a imágenes en dos dimensiones (2D) y no incluirá análisis en 3D.

1.5.1 Espacial

El estudio utilizará imágenes faciales provenientes de bases de datos públicas con imágenes de alta calidad. Se concentrará en muestras representativas de diferentes tipos de piel, obtenidas principalmente de poblaciones de distintas regiones geográficas para asegurar la diversidad en la segmentación de las características faciales.

1.5.2 Temporal

El desarrollo del estudio y la validación del sistema se llevarán a cabo en un periodo de aproximadamente 6 a 12 meses, incluyendo fases de recolección de datos, desarrollo del modelo, pruebas y evaluación del sistema. Las imágenes utilizadas para el entrenamiento y pruebas corresponderán a muestras recolectadas en los últimos cinco años para asegurar la actualidad de las características cutáneas.

1.5.3 Conceptual

Este estudio aborda conceptos fundamentales de la segmentación de imágenes, redes neuronales convolucionales, características morfológicas de la piel y técnicas de procesamiento de imágenes. Conceptos clave como segmentación, CNN, arrugas, poros y manchas se definirán claramente para establecer el marco teórico y práctico del trabajo. Además, se discutirán términos como detección automática y análisis morfológico en el contexto del cuidado de la piel.

Capítulo II: Marco Teórico

2.1 Antecedentes de la investigación

En esta parte de la investigación se presentan algunos antecedentes relacionados a la detección y pre-diagnóstico de nódulos en distintos órganos y a través de diversas metodologías. Estos ayudarán a entender el enfoque y obtener bases para un correcto desarrollo del proyecto en cuestión.

La investigación de Firdaus et al. Firdaus et al. (2023) presenta un sistema de segmentación de lesiones cutáneas basado en redes neuronales convolucionales (CNN), utilizando específicamente la arquitectura U-Net, ampliamente reconocida en la segmentación de imágenes biomédicas. Este enfoque busca abordar los desafíos inherentes al análisis de imágenes de dermatoscopia, donde las características de las lesiones cutáneas, como bordes borrosos, variaciones en el contraste, y residuos (cabello y marcadores de regla), dificultan la segmentación precisa y la posterior detección de patologías.

La segmentación precisa de lesiones cutáneas es crucial para el diagnóstico temprano de enfermedades como el melanoma, un tipo de cáncer de piel con alta mortalidad. El estudio emplea el conjunto de datos HAM10000, que contiene más de 10,000 imágenes de dermatoscopia, cubriendo siete categorías diagnósticas de lesiones cutáneas. Estas imágenes, recopiladas en múltiples ubicaciones a lo largo de 20 años, representan una amplia variedad de casos, lo que fortalece la robustez y aplicabilidad del modelo.

El modelo U-Net propuesto se optimizó mediante técnicas de preprocesamiento de imágenes, tales como redimensionamiento, escalado de características, y aumentación de datos, con el objetivo de mejorar la capacidad de generalización y reducir el riesgo de sobreajuste. Durante el entrenamiento, se experimentó con diferentes combinaciones de hiperparámetros, como funciones de pérdida (entropía cruzada binaria y coeficiente de Dice), tasas de aprendizaje y tamaños de lotes. El modelo final alcanzó resultados sobresalientes, con una precisión de píxeles del 95.89 %, un índice de intersección sobre unión (IoU) de 90.37 %, y una puntuación F1 de 92.54 %, lo que evidencia su efectividad y precisión en la segmentación de lesiones.

En comparación con otros métodos de segmentación previos, como campos aleatorios de Markov, bosques aleatorios y máquinas de soporte vectorial (SVM), el modelo U-Net superó a estos enfoques al no requerir extracción de características manual y al ofrecer una segmentación más precisa. La arquitectura U-Net, diseñada con capas de convolución y pooling, permite capturar características complejas de las lesiones, logrando una segmentación de alta calidad que puede ser fundamental para la detección temprana y precisa de enfermedades cutáneas.

En conclusión, el modelo U-Net desarrollado por Firdaus et al. demuestra ser una herramienta eficaz para la segmentación de lesiones cutáneas en imágenes de dermatoscopia. Aunque el estudio reconoce limitaciones en términos de recursos computacionales y la necesidad de ajuste fino de hiperparámetros, plantea futuras mejoras que podrían optimizar aún más la segmentación automática de imágenes médicas Firdaus et al. (2023).

Model	Pixel Accuracy	IoU	F1 Score
Model 1	95.86	90.29	92.53
Model 2	95.20	88.82	90.94
Model 3	95.61	89.70	91.73
Model 4	95.67	89.79	92.10
Model 5	95.62	89.72	92.06
Model 6	95.89	90.37	92.54

Tabla 1

Comparación de Modelos: Precisión de Píxeles, IoU y Puntuación F1

Además se realizó una comparación con otras investigaciones relevantes y se obtuvieron los siguientes resultados.

Research	Method	Dataset	Acc	IoU	F1 Score
Salih & Viriri Salih y Viriri (2020)	SRM+MRF	ISIC 2018	0.92	0.79	0.88
Jin et al. Jin et al. (2021)	CKDNet	ISIC 2018	0.93	0.79	0.87
Arora et al. Arora et al. (2021)	Attn_U-Net+GN	ISIC 2018	0.95	0.83	0.91
Proposed*	U-Net	ISIC 2018	0.95	0.90	0.92

Tabla 2

Comparación de diferentes métodos de segmentación de lesiones cutáneas en el conjunto de datos ISIC 2018.

El estudio de Yoon et al. Yoon et al. (2023) propone un modelo de segmentación de características morfológicas de la piel facial, específicamente arrugas y poros, mediante técnicas de aprendizaje profundo. Este enfoque es relevante para la dermatología estética y el cuidado de la piel, ya que permite realizar análisis detallados de la piel y personalizar recomendaciones de productos cosméticos.

Técnicas Utilizadas: El modelo está basado en la arquitectura U-Net, que ha demostrado ser efectiva en la segmentación de imágenes biomédicas. En este trabajo, U-Net se complementa con mecanismos de atención que mejoran el enfoque en zonas clave, como las áreas faciales donde arrugas y poros son más frecuentes. Además, se implementa una técnica de codificación posicional que aprovecha la disposición típica de estas características en el rostro, mejorando así la precisión del modelo al reducir los falsos positivos y centrarse en las regiones de interés.

Metodología: Para optimizar la precisión de la segmentación, se desarrolla un método de generación de “ground truth” (GT) adaptado a la naturaleza específica de las arrugas y

poros. Este GT se obtiene utilizando mapas de textura específicos: un filtro de alta frecuencia que realza los detalles de las arrugas y un método de pirámide laplaciana para destacar los poros. El conjunto de datos incluyó 314 imágenes faciales obtenidas mediante dispositivos de diagnóstico dermatológico, de las cuales 264 fueron empleadas para entrenamiento y 50 para validación. Las imágenes fueron preprocesadas y anotadas manualmente por especialistas.

Resultados: Los resultados obtenidos demostraron que el modelo propuesto superó a otros métodos tradicionales de procesamiento de imágenes y arquitecturas de aprendizaje profundo, como U-Net++. Específicamente, el modelo alcanzó un valor de Intersección sobre Unión (IoU) de 0.2341 para arrugas y de 0.4032 para poros, superando los valores de 0.2160 y 0.3669 obtenidos con U-Net++ en estas mismas categorías. En términos de precisión de píxeles, el modelo alcanzó un 95.89

Conclusiones: Yoon et al. concluyen que la integración de mecanismos de atención y codificación posicional en la arquitectura U-Net proporciona una segmentación más precisa de arrugas y poros, con potencial de aplicación en tareas avanzadas como la estimación de la edad de la piel y el análisis de su elasticidad y rugosidad. Este enfoque innovador podría facilitar diagnósticos estéticos y médicos de la piel, permitiendo mejorar la personalización en el cuidado cutáneo Yoon et al. (2023).

Tabla 3
Performance Metrics of Different Models

Models	#Params	Loss	IoU of Wrinkle	IoU of Pore
U-Net	17.3 M	1.243	0.2078	0.3601
Reduced U-Net	4.3 M	1.250	0.2147	0.3646
Reduced U-Net, Attentions	5.2 M	1.242	0.2250	0.3714
Reduced U-Net, Attentions, Zero-padding (Proposed)	5.2 M	1.145	0.2341	0.4032

El artículo de Thanh et al. (2021) Thanh et al. (2021) presenta un método avanzado de segmentación de lesiones cutáneas en imágenes dermoscópicas, diseñado para facilitar la detección temprana de melanoma. Esta técnica utiliza la arquitectura U-Net en combinación con el codificador VGG-16, mejorando la segmentación en áreas de baja intensidad, un aspecto crítico en las imágenes dermoscópicas. El método propuesto es notablemente eficaz en sistemas de cómputo con recursos limitados, como los que carecen de GPU potentes, y ofrece una precisión superior al 95 % tras el entrenamiento. El estudio emplea el conjunto de datos ISIC para evaluar su rendimiento, aplicando métricas de similitud Sorensen-Dice y Jaccard. Los resultados experimentales demuestran que esta técnica supera a otros enfoques basados en redes profundas, especialmente en la segmentación precisa de regiones de baja intensidad en

las imágenes.

El enfoque presentado evita la necesidad de preprocesamiento de la imagen, como la eliminación de cabello, la extracción de regiones de interés (ROI) o la mejora del contraste. Este método permite el procesamiento directo de imágenes en color sin la conversión a escala de grises ni la segmentación en canales separados. La implementación de este método se realizó en MATLAB, logrando buenos resultados en términos de sensibilidad y especificidad, con valores promedio de 0.92 y 0.86 para las métricas Dice y Jaccard, respectivamente, en las imágenes de prueba.

En el artículo de Kim y Song (2023) M. Kim y Song (2023), se propone un sistema mejorado para el diagnóstico de problemas de piel facial mediante una versión refinada de Mask R-CNN combinada con una red generativa adversarial de superresolución (SR-GAN). La piel facial es un factor crucial en la percepción de la edad, salud y belleza de una persona. Para abordar los desafíos técnicos inherentes al diagnóstico de problemas de piel, como acné, manchas y poros, los autores identifican cinco obstáculos técnicos principales: (1) la detección de problemas de pequeño tamaño, (2) la variabilidad en la apariencia de un mismo problema entre diferentes individuos, (3) la similitud visual entre distintos tipos de problemas, (4) la dificultad para detectar múltiples tipos de problemas en la misma imagen y (5) las segmentaciones erróneas en áreas no faciales.

Para superar estos desafíos, se implementan cinco tácticas que mejoran significativamente el rendimiento. En primer lugar, el modelo Mask R-CNN se optimiza mediante capas de fusión y deconvolución, lo que permite detectar características de pequeño tamaño, como poros y arrugas. En segundo lugar, se emplea un SR-GAN para aumentar la resolución de las imágenes de baja calidad, mejorando la precisión en la detección de problemas pequeños. Tercero, se entrenan modelos de segmentación específicos para cada tipo de problema, lo que optimiza la detección al reducir las interferencias de clases no relacionadas. La cuarta táctica consiste en utilizar modelos de segmentación específicos para cada dirección facial (frontal, lateral izquierda y derecha), ya que la posición y visibilidad de ciertos problemas varía según la orientación del rostro. Finalmente, la quinta táctica emplea un modelo de detección de landmarks faciales para descartar segmentaciones en áreas no faciales, como ojos, cejas y cabello, evitando falsos positivos.

Los resultados experimentales muestran que estas tácticas incrementan el rendimiento diagnóstico en un 32.58 % respecto a los modelos CNN convencionales, alcanzando una precisión de 83.38 %. Este enfoque no solo es preciso, sino que es adecuado para implementarse en dispositivos de bajo costo y en aplicaciones móviles, proporcionando una alternativa económica a las visitas clínicas. Los autores sugieren que este sistema podría ser de utilidad en clínicas

de cuidado de la piel o como una herramienta accesible para el diagnóstico domiciliario.

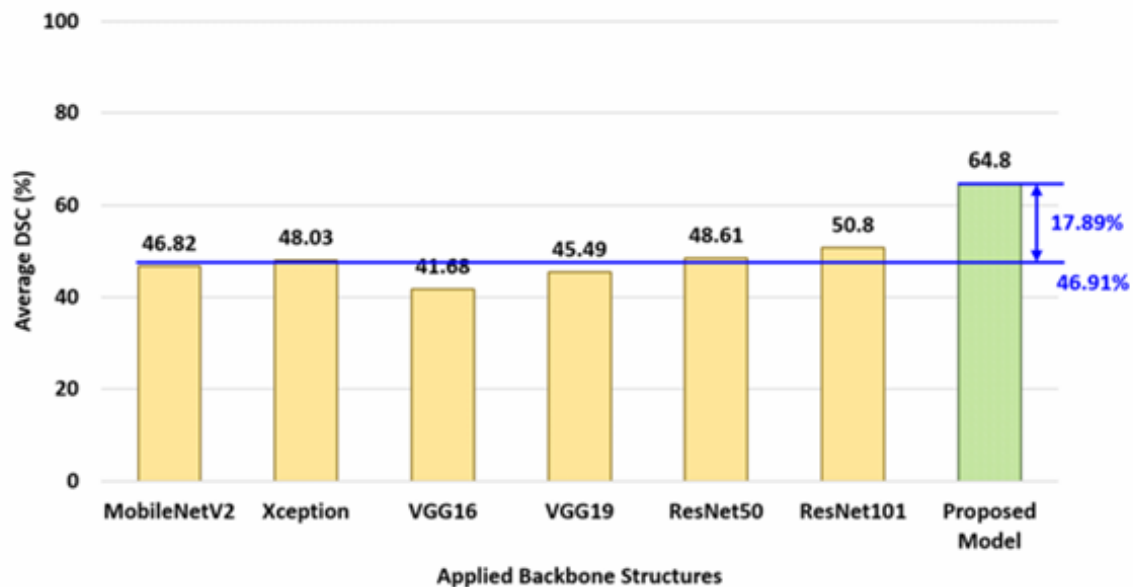


Figura 1. Comparación del modelo propuesto con otros 6 modelos de redes neuronales.

Fuente: M. Kim y Song (2023). «High Performing Facial Skin Problem Diagnosis with Enhanced Mask R-CNN and Super Resolution GAN». (p. 8)

El artículo titulado «Facial Wrinkle Detection Based on DeepLabV3+ and Semi-Automatic Labelling Strategy» de Zhong et al. (2024) presenta un enfoque innovador para la detección de arrugas faciales, abordando las limitaciones de métodos tradicionales que se ven afectados por interferencias con otras características faciales. Para mejorar la precisión, los autores proponen el uso del modelo DeepLabV3+ junto con una estrategia de etiquetado semi-automática, lo que permite generar datos de entrenamiento más representativos y mejorar la segmentación de arrugas. La investigación, como podemos ver en las Figuras ?? y 2, emplea técnicas avanzadas como DeepLabV3+, una red optimizada para segmentación de imágenes, y MobileNetV2 para reducir la carga computacional. Además, se desarrolla una estrategia de etiquetado semi-automática combinando anotaciones dermatológicas con mapas de textura generados mediante filtros Wiener y umbralización adaptativa. La precisión del modelo se evalúa mediante el Índice de Similitud Jaccard (JSI).

El conjunto de datos se construyó con 300 imágenes de Flickr-Face-HQ, divididas en entrenamiento (225), validación (25) y pruebas (50). Las pruebas se realizaron en un sistema con hardware de alto rendimiento (Intel Core i9-12900K, GPU RTX 3070 y 32GB RAM) y utilizando PyTorch como framework principal.

Los resultados demostraron que el método propuesto supera a enfoques tradicionales

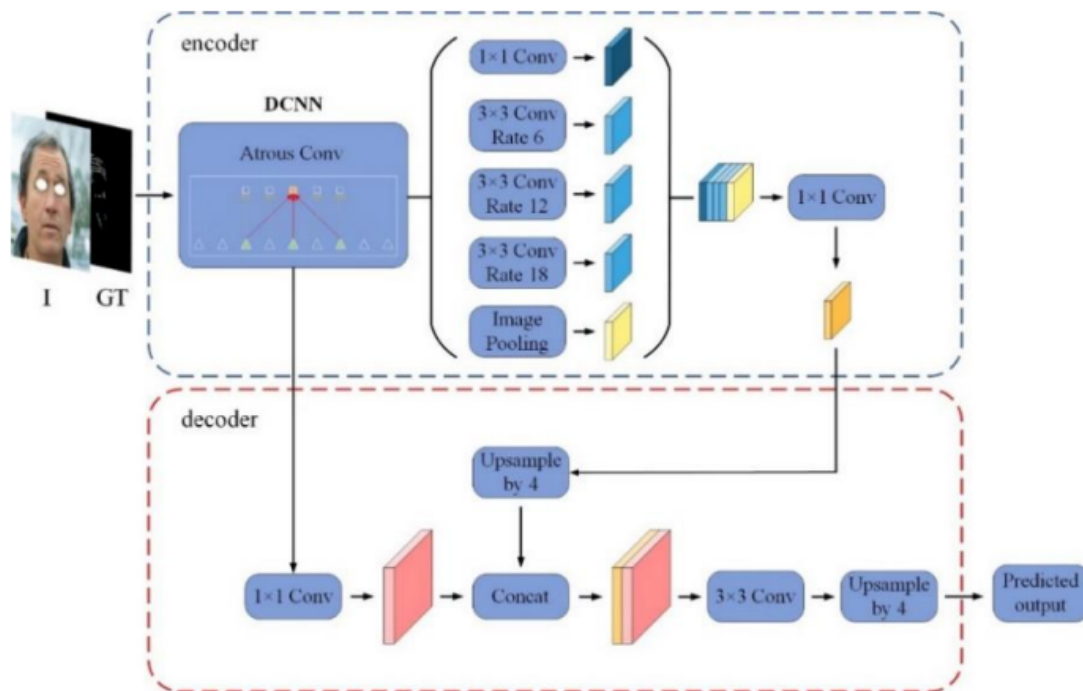


Figura 2. Estructura de red de DeepLabV3+.

Fuente: Zhong et al. (2024). «Facial Wrinkle Detection Based on DeepLabV3+ and Semi-Automatic Labelling Strategy». (p. 4)

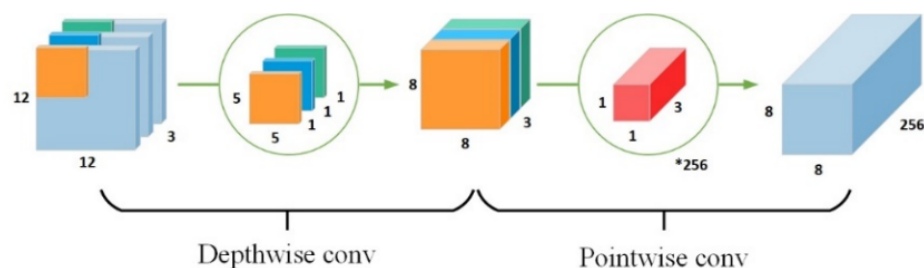


Figura 3. El proceso de convolución separable en profundidad.

Fuente: Zhong et al. (2024). «Facial Wrinkle Detection Based on DeepLabV3+ and Semi-Automatic Labelling Strategy». (p. 4)

como el filtro Hessiano y el modelo U-Net, obteniendo mayores valores de JSI en la frente (0.62) y área ocular (0.64). Además, redujo la cantidad de falsos positivos y mejoró la segmentación de bordes, logrando un mejor desempeño en la detección de arrugas finas.

En conclusión, el modelo DeepLabV3+ con etiquetado semi-automático se mostró más efectivo en la detección de arrugas faciales. Sin embargo, aún enfrenta desafíos en la diferen-

ciación de arrugas muy finas y cabellos, lo que sugiere la necesidad de mejoras futuras para incrementar la precisión del sistema.

En el artículo de Karshiev et al. (2020) llamado «Improved U-Net: Fully Convolutional Network Model for Skin-Lesion Segmentation» analiza el problema de la segmentación de lesiones cutáneas, una tarea fundamental para el diagnóstico temprano del melanoma. Aunque el modelo U-Net ha sido ampliamente utilizado en segmentación médica, presenta limitaciones como ralentización en el entrenamiento y problemas con la función de activación ReLU. Para mejorar su desempeño, los autores proponen una versión optimizada que incorpora interpolación bilineal para el upsampling y la función de activación PReLU, lo que mejora la precisión y evita problemas como el sobreajuste.

La investigación emplea un diagrama el cual se puede ver en la Figura 4 y a su vez varias técnicas clave: la interpolación bilineal sustituye la deconvolución tradicional para mejorar la segmentación de bordes, la función PReLU reemplaza ReLU para prevenir "neuronas muertas" optimizar la convergencia, y el dropout se utiliza después de cada bloque convolucional para reducir el sobreajuste. Estas modificaciones permiten mejorar la estabilidad y eficiencia del entrenamiento.

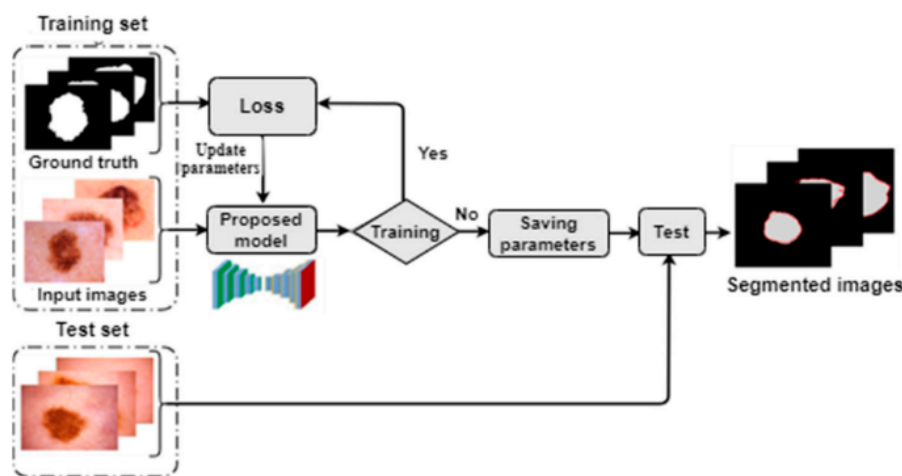


Figura 4. Diagrama de flujo del sistema propuesto.

Fuente: Karshiev et al. (2020). «Improved U-Net: Fully Convolutional Network Model for Skin-Lesion Segmentation». (p. 4)

El modelo propuesto como se ve en la Figura 5, se basa en una arquitectura U-Net modificada con capas convolucionales optimizadas, PReLU, dropout y upsampling mediante interpolación bilineal. Fue entrenado en un sistema con un procesador Intel Core i7-9700K, 32 GB de RAM y una GPU NVIDIA GeForce RTX 2060 SUPER, garantizando un entorno

adecuado para el procesamiento intensivo de imágenes médicas.

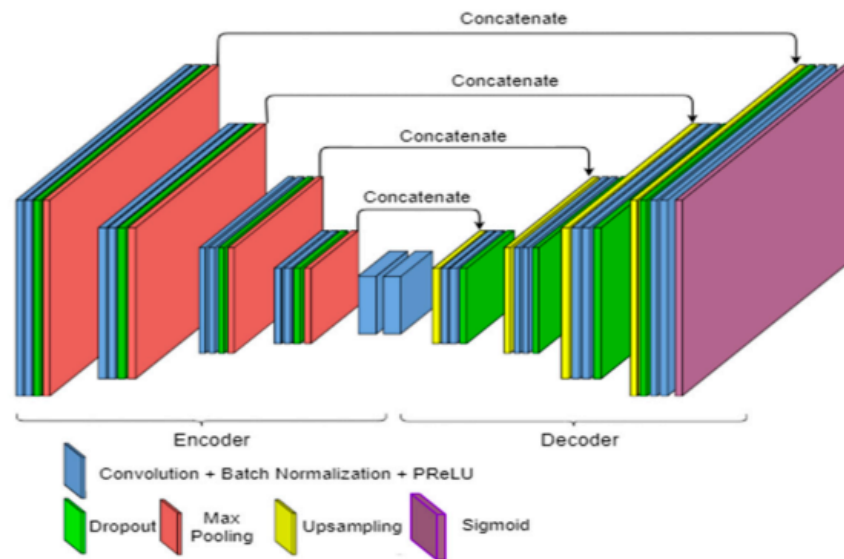


Figura 5. Modelo totalmente convolucional propuesto para la segmentación de lesiones cutáneas.

Fuente: Karshiev et al. (2020). «Improved U-Net: Fully Convolutional Network Model for Skin-Lesion Segmentation». (p. 4)

Para el entrenamiento y prueba del modelo se utilizó un conjunto de datos de imágenes dermoscópicas, con 2594 imágenes etiquetadas para entrenamiento y 1000 para prueba. Todas las imágenes fueron preprocesadas, redimensionadas a 256×256 píxeles y convertidas a escala de grises, lo que facilitó la normalización y mejoró la eficiencia del modelo.

Los resultados muestran un alto rendimiento del modelo mejorado, alcanzando una precisión por píxel del 94% y un coeficiente Dice del 88%. Estas mejoras permiten reducir artefactos en la segmentación y aumentar la eficiencia computacional en comparación con la U-Net estándar, consolidándose como una alternativa más precisa y robusta para la segmentación de lesiones cutáneas.

En conclusión, la versión mejorada de U-Net supera las limitaciones del modelo original al abordar problemas como gradientes débiles y artefactos en la segmentación. La integración de interpolación bilineal, PReLU y dropout permite lograr una mayor precisión y eficiencia, posicionando este enfoque como una herramienta prometedora para la segmentación de imágenes médicas en el ámbito dermatológico.

El artículo titulado «Object-by-Object Learning for Facial Skin Lesion Detection Using

Semantic Segmentation Methods» de los autores Bekmirzaev et al. (2021), presenta una novedosa técnica de aprendizaje "objeto por objeto" para detectar 11 tipos de lesiones cutáneas faciales mediante segmentación semántica. Los autores destacan que muchas clases de lesiones, como arrugas y manchas, presentan relaciones visuales concurrentes, y explotan esta característica para abordar ambigüedades en la apariencia de las imágenes, mejorando así la precisión en el diagnóstico automatizado.

Entre las técnicas utilizadas, introducen los módulos REthinker, formados por capas convLSTM/Conv3D locales y un módulo SE como mecanismo de atención. Además, integran estos bloques en redes CNN estándar, junto con técnicas avanzadas como convolución atrous, agrupación espacial piramidal y convolución separable en profundidad, lo que permite una representación más rica y contextual de las lesiones cutáneas.

La metodología como se observa en la Figura 6, se basa en modelar explícitamente las relaciones contextuales entre clases de objetos mediante el aprendizaje "objeto por objeto". Los módulos REthinker se diseñan para mejorar la sensibilidad de la red tanto en contextos locales como globales. Las CNN se modifican para incorporar estos módulos, y los modelos resultantes se entrenan y validan sobre datos dermatológicos cuidadosamente anotados.

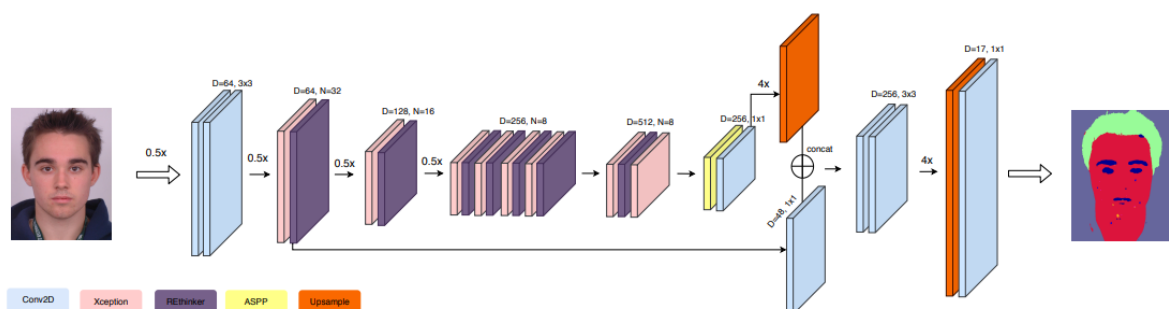


Figura 6. Propuesta basada en RethNet RE-Xception y el decodificador de DeepLabv3+.

Fuente: Bekmirzaev et al. (2021). «Object-by-Object Learning for Facial Skin Lesion Detection Using Semantic Segmentation Methods». (p. 5)

Para validar su enfoque, los autores construyen la base de datos propia **MSLD**, que contiene 27,790 imágenes faciales, de las cuales 412 están anotadas con 11 tipos de lesiones cutáneas y 6 clases adicionales. También evalúan sus modelos con el conjunto de datos ISIC 2018, ampliamente utilizado en tareas de segmentación dermatológica.

Los resultados muestran que su método supera en un 15.34% en MIOU a redes de segmentación de última generación en el conjunto de datos MSLD, y obtiene buenos resultados

en ISIC 2018, lo que demuestra su eficacia en contextos diversos.

En conclusión, el enfoque propuesto basado en aprendizaje ^objetivo por objetos los módulos Rethinker permiten capturar eficazmente relaciones contextuales entre diferentes tipos de lesiones. Esta estrategia mejora considerablemente la precisión de los sistemas de segmentación facial, destacándose como una alternativa poderosa frente a los métodos convencionales.

Tamilkodi et al. (2024) elaboraron el artículo titulado «A New Approach to Facial Skin Analyzer», que presenta un software avanzado para el análisis de la piel facial que permite identificar características clave como arrugas, poros, manchas y textura, además de predecir edad y género. El sistema compara los resultados con datos de personas del mismo grupo etario, clasifica el tono de piel, genera máscaras visuales para resaltar zonas específicas del rostro y ofrece recomendaciones personalizadas. El objetivo central es desarrollar un modelo robusto capaz de realizar un análisis facial completo en diversos contextos.

Para lograrlo, el software utiliza redes neuronales convolucionales profundas (D-CNN), apoyadas por modelos preentrenados para la detección de edad y género. También emplea técnicas como KMeans para clasificar el tono de piel y gradientes de Sobel para identificar arrugas. Estas herramientas se integran para ofrecer un análisis visual y cuantitativo de cada característica facial.

La metodología del sistema incluye la captura de imágenes faciales desde múltiples ángulos, el preprocesamiento mediante técnicas como desenfoque gaussiano y conversión de formatos, seguido del análisis específico de textura, tono y arrugas. Finalmente, los resultados se visualizan en gráficos y máscaras y se utilizan para generar recomendaciones personalizadas para el usuario.

Como base de datos, se empleó la ****Indian Face Age Database (IFAD)****, compuesta por 3,296 imágenes de 55 personas con diversidad de edad, poses, expresiones e iluminación. Esta variedad permite entrenar modelos con mayor capacidad de generalización y robustez en escenarios reales de análisis facial.

Los resultados como se pueden ver en las Figuras 7, 8 y 9, muestran una alta precisión en la detección de características faciales y predicciones fiables de edad y género. Además, el sistema logra una visualización efectiva mediante gráficos y máscaras que resaltan áreas clave del rostro, lo que mejora la comprensión de los resultados.

En conclusión, el enfoque propuesto demuestra ser eficaz para detectar estructuras faciales importantes incluso en condiciones variables. Aunque aún no alcanza una precisión perfecta, se reconoce el valor de expandir el conjunto de datos para mejorar el rendimiento. Este modelo tiene un gran potencial para aplicaciones automatizadas en análisis dermatológico y

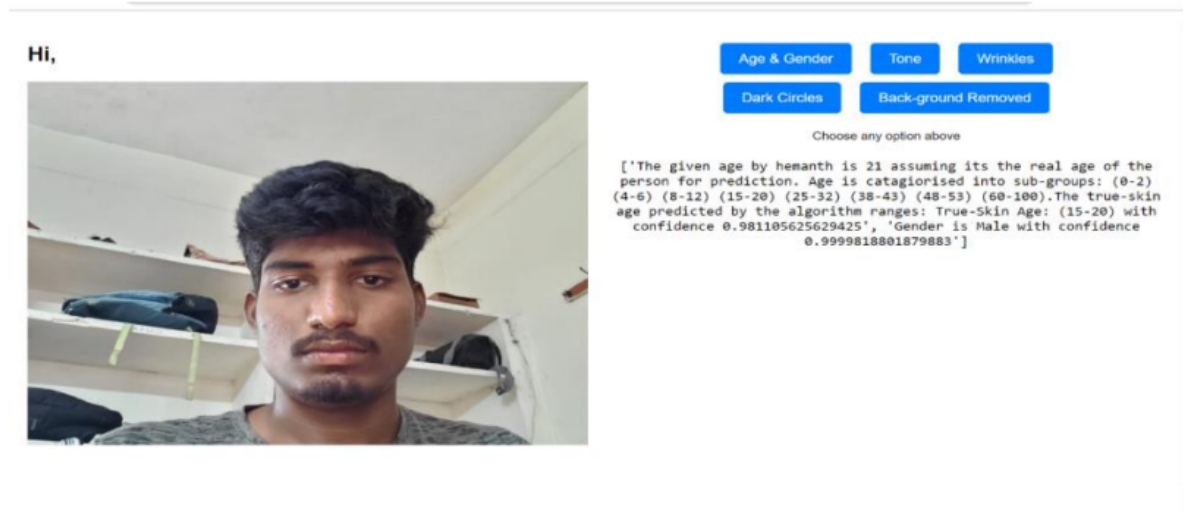


Figura 7. Detección de edad y género.

Fuente: Tamilkodi et al. (2024). «A New Approach to Facial Skin Analyzer». (p. 6)

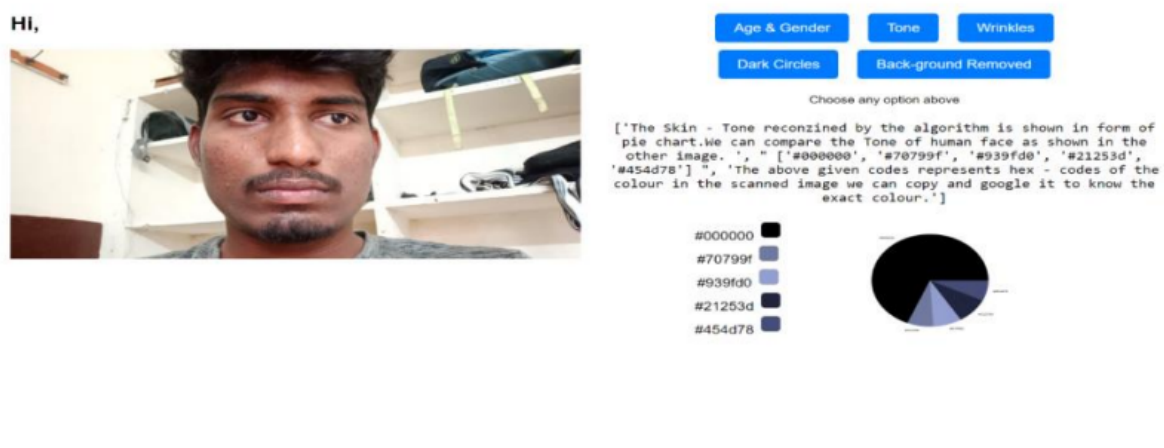


Figura 8. Análisis del tono de piel.

Fuente: Tamilkodi et al. (2024). «A New Approach to Facial Skin Analyzer». (p. 6)

cuidado personalizado de la piel.

El artículo de Moon et al. (2024) llamado «Dermatology: Pretraining with Texture Map-Based Weak Supervision» se enfoca en la detección automática de arrugas faciales, un tema relevante en dermatología estética, debido a que los métodos manuales de segmentación son subjetivos, costosos y laboriosos. Para abordar este problema, los autores desarrollan un enfoque basado en aprendizaje profundo que incluye la creación del primer conjunto de datos público especializado en segmentación de arrugas (FFHQ-Wrinkle), y una estrategia de en-



Figura 9. Análisis de arrugas.

Fuente: Tamilkodi et al. (2024). «A New Approach to Facial Skin Analyzer». (p. 6)

trenamiento en dos etapas que combina un preentrenamiento débilmente supervisado con un ajuste fino supervisado, optimizando así el rendimiento del modelo con un uso eficiente de los recursos de etiquetado.

Se aplican dos enfoques complementarios: primero, un preentrenamiento débilmente supervisado que genera automáticamente mapas de textura mediante filtros Gaussianos sobre 50,000 imágenes no etiquetadas; y segundo, un ajuste fino supervisado que utiliza 1,000 imágenes con máscaras de arrugas anotadas manualmente por expertos. Para la segmentación se emplean arquitecturas de redes neuronales como U-Net y Swin UNETR, que permiten capturar tanto características locales como globales de las arrugas faciales.

La metodología, como podemos ver en la Figura 10, consiste en generar etiquetas débiles mediante mapas de textura derivados de imágenes faciales, eliminando regiones no relevantes con un modelo de segmentación facial (BiSeNet). Luego, tres expertos generan manualmente máscaras de arrugas en regiones específicas como la frente y las líneas nasolabiales. El modelo se entrena en dos etapas: primero aprende a predecir texturas con pérdida MSE y luego realiza un ajuste fino con datos manualmente etiquetados y los mapas de textura como entrada adicional, fortaleciendo así su capacidad de generalización.

El estudio introduce el conjunto FFHQ-Wrinkle, derivado del conjunto de datos FFHQ, compuesto por 50,000 imágenes con etiquetas débiles generadas automáticamente y 1,000 imágenes con anotaciones manuales precisas. Este conjunto incluye una amplia variedad de

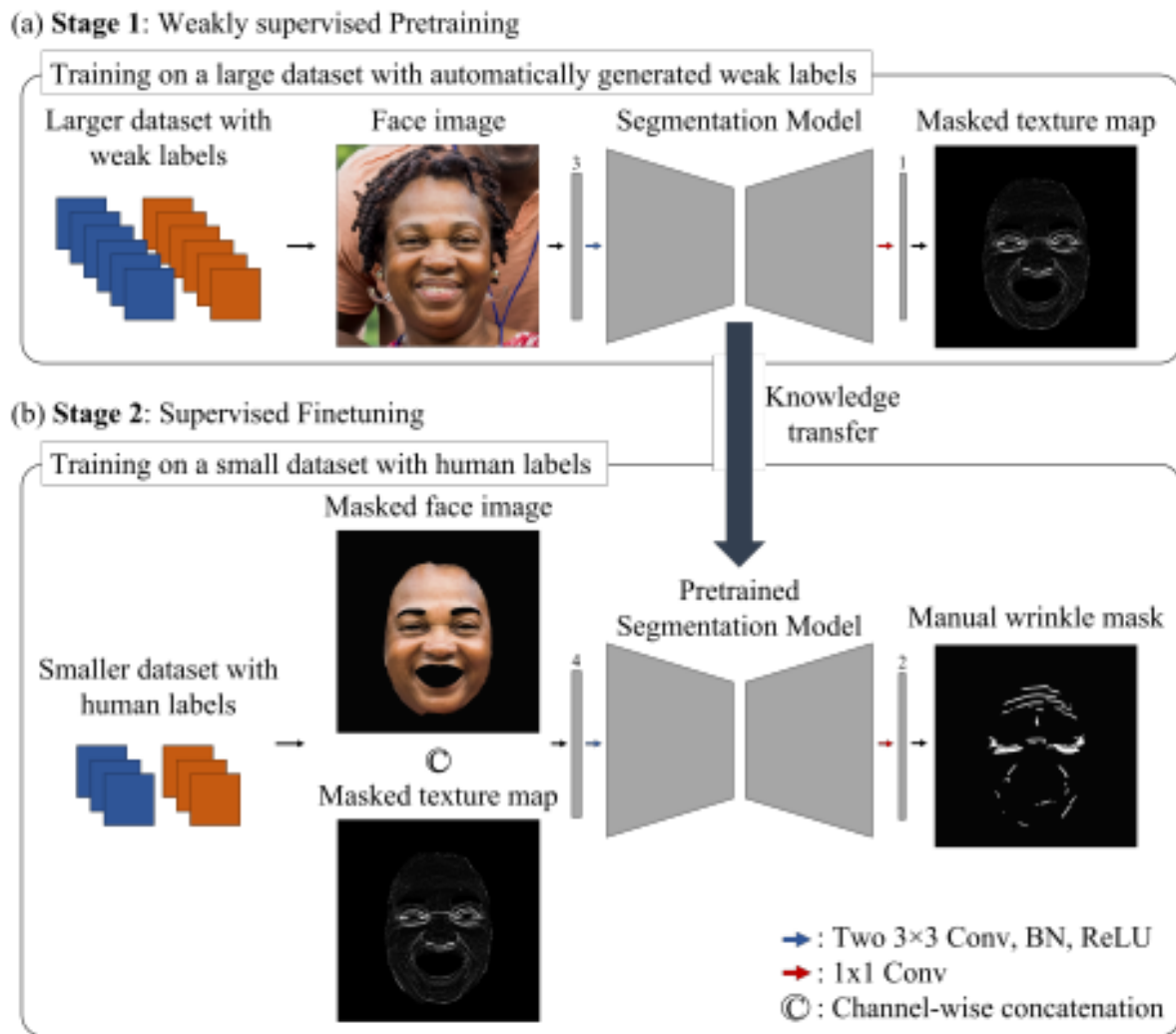


Figura 10. Entrenamiento en dos etapas para la segmentación de arrugas faciales.

Fuente: Moon et al. (2024). «Dermatology: Pretraining with Texture Map-Based Weak Supervision». (p. 3)

edades, géneros y etnias, lo que contribuye a la robustez del modelo entrenado y favorece su aplicación generalizada en diversos contextos.

Los modelos entrenados mediante la estrategia propuesta lograron un rendimiento superior en comparación con métodos anteriores, tanto en métricas cuantitativas como cualitativas. La combinación de datos etiquetados automáticamente y manualmente permitió reducir significativamente los recursos necesarios para el etiquetado, manteniendo una alta precisión en la segmentación de arrugas faciales, especialmente en regiones clave del rostro.

La investigación demuestra que la combinación de aprendizaje débilmente supervisado con ajuste fino supervisado mejora sustancialmente la segmentación automática de arrugas

faciales. Además, el nuevo conjunto de datos FFHQ-Wrinkle representa un aporte valioso al campo, ya que ofrece una base estandarizada y diversa para futuras investigaciones sobre análisis facial automatizado y aplicaciones en dermatología estética.

2.2 Bases Teóricas

2.2.1 Inteligencia Artificial

El método racional fusiona la ingeniería y las matemáticas basándose en las "leyes del pensamiento", las cuales tienen su origen en la antigua Grecia y han sido influenciadas por filósofos como Aristóteles. Durante el siglo XIX, se diseñaron programas capaces de resolver problemas de lógica. Por consiguiente, el propósito de la Inteligencia Artificial en la vida real es crear sistemas inteligentes que posean estas habilidades. Aun en situaciones de incertidumbre, un "Agente Racional" toma acciones con el fin de obtener el mejor resultado posible. La inteligencia artificial se apoya en diversas disciplinas, tales como la ingeniería computacional, la teoría de control, la cibernética, la lingüística, la filosofía, la economía, la psicología, la neurociencia y las matemáticas, de acuerdo con Russell y Norvig (2004).

Dos investigadores en neurociencia crearon el primer modelo de IA basado en neuronas artificiales en 1943, dando inicio al análisis de la Inteligencia Artificial. McCulloch y Pitts idearon el prototipo que permitía que las neuronas fueran «activadas» o «desactivadas», lo que demostró que una red de neuronas era capaz de realizar cualquier tarea computacional. Posteriormente, Donald Hebb propuso la «Regla de Aprendizaje Hebbiano». John McCarthy, Allen Newell y Herbert Simon desarrollaron un programa que podía tener el pensamiento no numérico en el taller de Dartmouth, aunque no se publicó. El término «Inteligencia Artificial» fue acuñado por McCarthy, (Russell & Norvig, 2004).

La IA comenzó a entrar en la industria en los años 80, especialmente en grandes empresas de países desarrollados, a través de la investigación en sistemas expertos y el desarrollo de computadoras más poderosas.

2.2.2 Aprendizaje Automático

El Machine Learning es un área de la Inteligencia Artificial enfocada en técnicas que permiten a las computadoras aprender a través de algoritmos, convirtiendo muestras de datos en programas sin requerir programación explícita. Según **bk`russell2009intart<empty citation>**, el aprendizaje automático es una división de la inteligencia artificial. Estos algoritmos emplean tecnologías como el procesamiento del lenguaje natural, el aprendizaje profundo y las redes neuronales. Tanto el aprendizaje supervisado como el no supervisado se fundamentan en lecciones extraídas de los datos. La creación de algoritmos capaces de recibir datos de entrada y

utilizar análisis estadístico para prever una salida, la cual se ajusta conforme se obtienen nuevos datos, constituye el fundamento del aprendizaje automático Alpaydin (2014).

Se puede clasificar en cuatro tipos principales de la siguiente manera según el objetivo que se desea alcanzar mediante el uso de ML:

- **Aprendizaje Supervisado:** El Aprendizaje Supervisado se ganó su nombre porque los científicos de datos actúan como una guía para enseñarle al algoritmo las conclusiones a las que debe llegar. Es similar a la forma en que un estudiante aprende aritmética básica de un maestro. Este tipo de aprendizaje requiere datos etiquetados con las respuestas correctas que se esperan del resultado del algoritmo. Para problemas de clasificación y regresión, el aprendizaje supervisado demostró ser preciso y rápido según (Zambrano, 2018).

Los dos tipos de Aprendizaje Supervisado son:

- **La Clasificación:** es la predicción del valor categórico de salida que permite dividir los datos en clases específicas. La clasificación se puede usar para varios propósitos, como determinar el clima, determinar si un correo electrónico es spam o no o identificar tipos de animales después de recibir una educación adecuada, un conjunto de datos con etiquetas de imágenes que incluyen la especie y algunas identificaciones características, según (Zambrano, 2018).
- **La Regresión:** es un tipo de problema en el que la predicción de un valor de respuesta continua es necesaria, como los precios de las acciones y la vivienda, según (Zambrano, 2018).

Por lo tanto, funciona modelando las relaciones y dependencias entre las características de entrada y la salida de predicción objetivo, lo que permite predecir los valores de salida para nuevos datos utilizando las relaciones que aprendió de conjuntos de datos anteriores, según (Alpaydin, 2014).

- **Aprendizaje No Supervisado:** Por otro lado, el Aprendizaje No Supervisado se asemeja más a lo que algunos expertos llaman Inteligencia Artificial real: la idea de que una máquina puede aprender a identificar patrones y procesos complejos sin la supervisión de humanos. Cuando los expertos no saben qué buscar en los datos y los datos en sí no incluyen objetivos, este método es particularmente útil. La agrupación de k-means, el análisis de componentes principales e independientes y las reglas de asociación según (Zambrano, 2018) son algunos de los muchos casos de uso del Aprendizaje Automático No Supervisado.

- **Agrupación K-means:** es un tipo de problema en el que cosas similares están agrupadas, como se muestra en la Figura 11. Comparte el mismo concepto con la clasificación, pero no se proporcionan etiquetas, por lo que el sistema entenderá los datos y los agrupará. Un uso de esto sería agrupar los artículos y las noticias según su género y contenido, según (Sancho Caparrini, 2018)

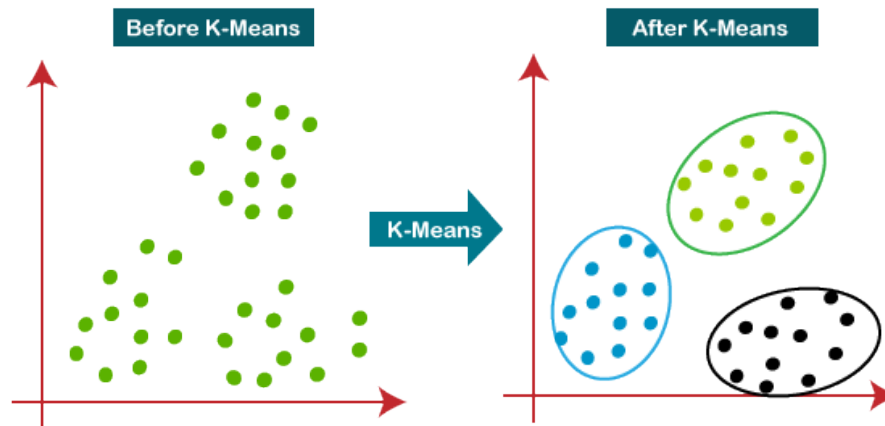


Figura 11. El algoritmo de K medias.

Fuente: Sancho Caparrini (2018). *Clasificación Supervisada y No Supervisada*.

Debido a su complejidad y dificultad de implementación, este tipo de Aprendizaje Automático no se utiliza tan frecuentemente como el Aprendizaje Supervisado, a pesar de que abre las puertas a la resolución de problemas que los humanos normalmente no abordarían, según (Sancho Caparrini, 2018)

- **Aprendizaje Semisupervisado:** Hasta el momento, todos los datos enviados han sido etiquetados con el resultado deseado o no han sido etiquetados en absoluto. El Aprendizaje Automático Semisupervisado utiliza ambos. El costo de etiquetar es bastante alto en muchas situaciones prácticas y, en el caso de grandes conjuntos de datos, se vuelve aburrido y requiere mucho tiempo. Además, proporcionar demasiados datos etiquetados puede hacer que el modelo tenga sesgos humanos. A pesar de que los datos sin etiquetar son desconocidos para la red, ofrecen información útil sobre los parámetros del grupo objetivo, que conduce a la conclusión de que se puede mejorar la precisión del modelo al incluir datos sin etiquetar y, al mismo tiempo, ahorrar tiempo y dinero en su construcción. Por ejemplo, la clasificación de páginas web, el reconocimiento de voz o la secuenciación genética pueden usar Aprendizaje Automático Semisupervisado. En esos casos, los científicos de datos pueden acceder a grandes cantidades de datos sin etiquetarlos, y la tarea de etiquetarlos todos llevaría mucho tiempo, según (Zambrano, 2018).

Se puede comparar estos tres tipos de Aprendizaje Automático para el mismo uso, como clasificación, utilizando los datos recopilados hasta ahora:

- **Clasificación supervisada:** el algoritmo clasificará los tipos de páginas web según las etiquetas proporcionadas desde el principio, según (Zambrano, 2018).
- **Agrupación no supervisada:** el algoritmo buscará patrones y características que ayuden a agrupar páginas web en grupos, según (Zambrano, 2018).
- **Clasificación semi no supervisada:** identificará varios grupos de páginas web utilizando los datos etiquetados, luego utilizará los datos no etiquetados para establecer los límites de esos grupos de páginas web y buscar otros tipos que posiblemente no aparezcan en los datos etiquetados, según (Zambrano, 2018).

2.2.3 Aprendizaje Profundo

Desde que llegó la Inteligencia Artificial hace un tiempo, tiene una amplia gama de aplicaciones y se divide en muchas ramas, como se menciona en (SAS Institute, s.f.). El Aprendizaje Profundo es un subconjunto del Aprendizaje Automático, que es en sí mismo un subcampo de la IA. La Figura 12 es una representación visual de la relación entre AI, ML y DL.

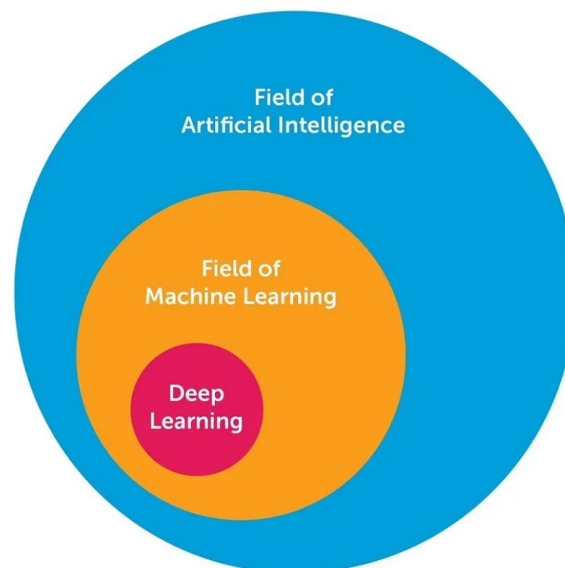


Figura 12. Relación entre IA, ML y DL.

Fuente: House of Bots (2018). *Most Popular 20 Free Online Courses to Learn Deep Learning*.

El Aprendizaje Profundo no solo permite representar datos de la manera correcta, sino

que también permite que la computadora aprenda programas informáticos de varios pasos al incluir el concepto de profundidad en sus modelos. Como se muestra en la Figura 13, cada capa de representación puede interpretarse como el estado de la memoria de la computadora. Las computadoras interpretan las imágenes como una colección de valores de píxeles que representan escenas de nuestra realidad. Según (House of Bots, 2018), identificar un objeto o mapear su identidad a partir de esos valores es una tarea difícil para las máquinas y puede resultar casi imposible cuando se intenta aprender este mapeo directamente.

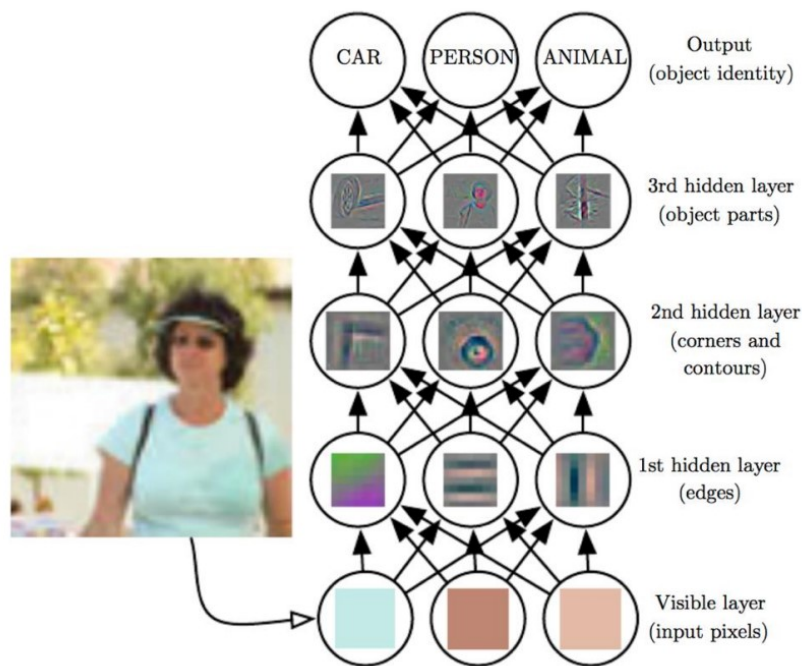


Figura 13. Modelo de aprendizaje profundo.

Fuente: House of Bots (2018). *Most Popular 20 Free Online Courses to Learn Deep Learning*.

2.2.4 Aprendizaje Profundo Multimodal

Podemos ver en la Figura 14, el fundamento del Aprendizaje Profundo Multimodal es la integración de diversas modalidades a través del uso de redes neuronales profundas. (Deng & Liu, 2018).

Las señales provenientes de distintas fuentes brindan datos adicionales sobre diferentes aspectos de un tema, lo cual capacita a los métodos que emplean varias técnicas para efectuar inferencias más sólidas. En el ámbito de las técnicas multimodales existen diversas estrategias como las estrategias de integración y combinación, los sistemas computacionales, el proceso de unir diferentes modelos y las estructuras de redes neuronales con múltiples capas. Al combinar

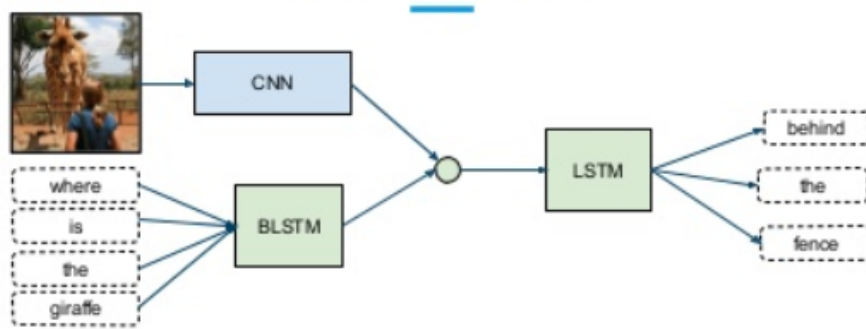


Figura 14. Modelo que combina imágenes y texto.

Fuente: Nishida y Nakayama (2015). *Multimodal gesture recognition using multi-stream recurrent neural network*.

estas características para la toma de decisiones, se emplean enfoques aditivos que recopilan información relevante y mejoran conjuntamente la precisión de las predicciones Deng y Liu (2018).

Baheti (2020) afirman que el uso de modelos multimodales mejora el rendimiento de las redes neuronales y permite una extracción más efectiva de características, lo que favorece predicciones mayores. Uno de las ventajas es que los datos de diferentes fuentes ofrecen información complementaria que revela patrones ocultos que no se pueden ver cuando las modalidades se analizan individualmente, mejorando así la precisión de las predicciones.

Según Brownlee (2018), se puede mejorar la calidad del promedio del modelo al considerar de manera ponderada las aportaciones individuales de cada submodelo a la predicción global. Este enfoque se conoce como «generalización apilada» y se divide en dos niveles.

- Nivel 0: Este nivel utiliza los datos de entrada del conjunto de entrenamiento para enseñar a los modelos a hacer predicciones.
- Nivel 1: Aquí, los datos provienen de las salidas de los modelos del nivel anterior, que se utilizan como entradas para instruir a los modelos en esta etapa, también conocidos como meta-aprendices o generalizadores.

En el contexto de los modelos apilados, se utilizan dos métodos diferentes:

- Enfoque de apilamiento por separado: Implica crear los datos de entrada conjuntos utilizando los pronósticos amalgamados de modelos adicionales. Estos datos se emplean para instruir a un modelo extra que realiza la predicción final.

- **Modelo apilado integrado:** Esta estrategia facilita la utilización de arquitecturas neuronales y un modelo de múltiples cabezas conformado por otros modelos preentrenados. Estos modelos no requieren conservar la misma configuración, dado que sus niveles se etiquetan como "no susceptibles al entrenamiento" para prevenir la modificación de sus valores antes de efectuar predicciones. Las salidas de los submodelos se fusionan directamente en el modelo principal.

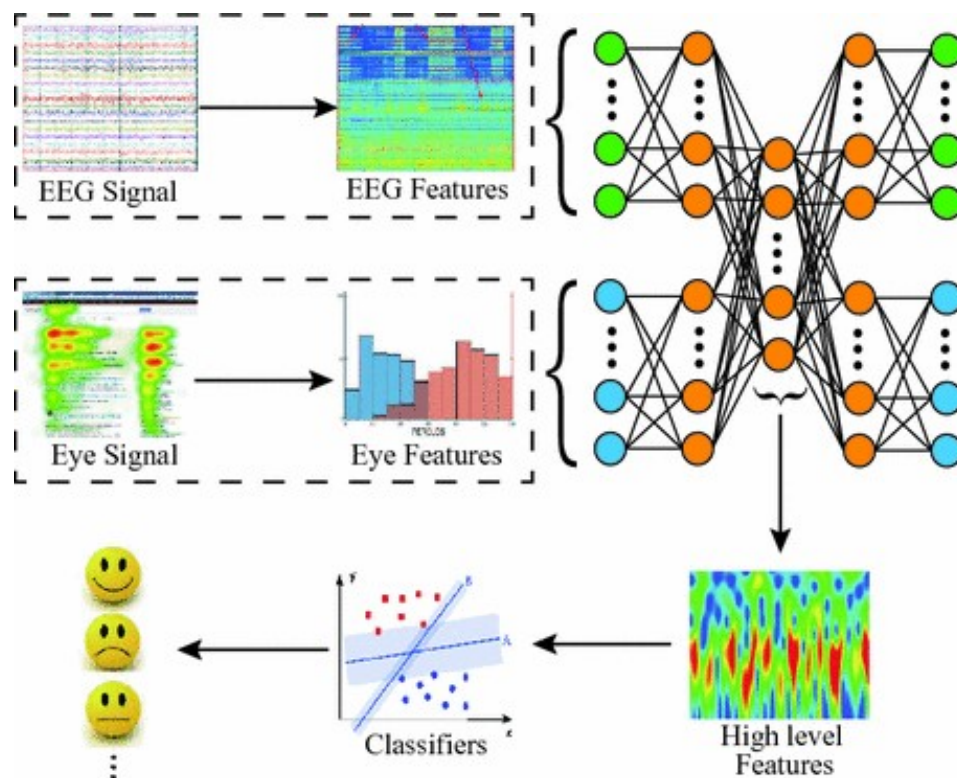


Figura 15. Un modelo multimodal para las señales de la vista.

Fuente: Baheti (2020). «Introduction to Multimodal Deep Learning».

2.2.5 Inteligencia Artificial Generativa

La Inteligencia Artificial Generativa es el campo de la ciencia que estudia cómo crear inteligencia totalmente automatizada. Esto contrasta con el campo de la Inteligencia Artificial Moderna, que investiga cómo los humanos comprenden y construyen la inteligencia. La construcción manual es lo que hacen los investigadores, pero la parte «por humanos» de la definición de IA suele ser una variable oculta en la construcción de los sistemas de IA contemporáneos. Aunque la diferencia puede parecer sutil, incluso innecesaria, la construcción automatizada requiere una perspectiva completamente nueva, que no está disponible en la literatura sobre IA actual. No importa si los humanos comprenden cómo funcionan los mecanismos internos de una máquina en la Inteligencia Artificial Generativa, aunque podría ayudar a los investigadores

en su búsqueda de la creación de máquinas inteligentes. Lo que importa es que la máquina pueda controlar y dirigir el proceso de creación de estructuras internas en la dirección deseada. Es como criar hijos: los seres humanos no son conscientes de los procesos mentales, pero interactúan con sus hijos en un nivel diferente al hablar de sus sustratos neuronales. En lugar de eso, utilizan una variedad de máquinas clasificadoras para diferenciar los comportamientos positivos de los negativos, que podrían ser beneficiosos para el niño cuando sea mayor. Los castigos corporales, los institutos educativos, los medios de comunicación y otros lugares son algunos de los usos de estas máquinas clasificadoras, según (th'zant2010genai)

A continuación, ofrecemos una variedad de categorías de modelos de Inteligencia Artificial Generativa.

- **Modelos de difusión:** crean nuevos datos iterando cambios aleatorios controlados en una muestra de datos previa. Empezan con los datos originales y luego agregan cambios sutiles, también conocidos como ruido, que hacen que gradualmente pierdan la similitud con el original. Este ruido se controla minuciosamente para garantizar que los datos generados sigan siendo consistentes y realistas. El modelo de difusión invierte el proceso después de agregar ruido en varias iteraciones. Como se muestra en la Figura 16, la eliminación de ruido inversa crea una muestra de datos nueva que se asemeja a la original, según (Amazon, 2023).

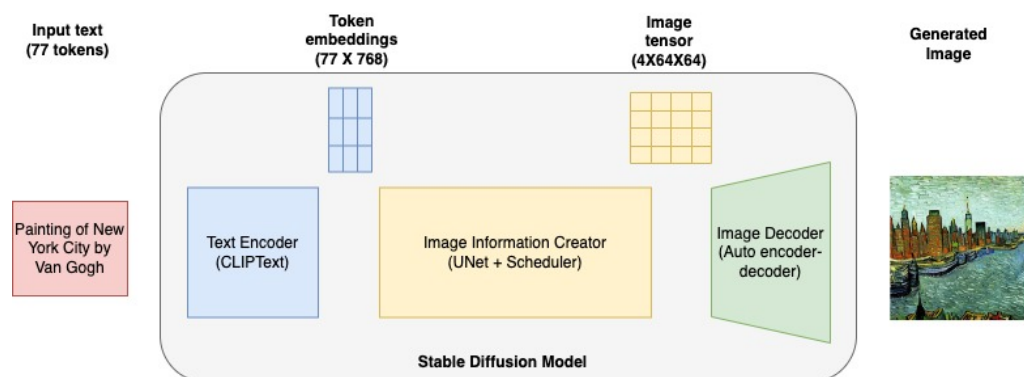


Figura 16. Modelos de difusión.

Fuente: Amazon (2023). «¿Qué es la IA generativa?»

- **Redes generativas adversativas:** compiten con dos redes neuronales. La primera red, también conocida como generador, agrega ruido aleatorio para crear muestras de datos falsas. La segunda red, conocida como discriminador, ayuda a distinguir entre los datos reales y falsos generados por el generador. El generador mejora continuamente su capacidad de generar datos realistas, mientras que el discriminador mejora su capacidad de

distinguir entre lo real y lo falso. Hasta que el generador produzca datos tan persuasivos que el discriminador no pueda diferenciarlos de los datos reales, este proceso adversativo termina, según (Amazon, 2023).

- **Autocodificadores variacionales:** aprenden sobre el espacio latente, una pequeña representación de datos. La representación matemática de los datos se conoce como espacio latente. Puede verse como un código único que representa los datos en función de cada característica. Por ejemplo, cuando se estudian los rostros, el espacio latente contiene números que representan la forma de las orejas, los pómulos, la nariz y los ojos. Las dos redes neuronales utilizadas por los VAE son el codificador y el decodificador. Para cada dimensión del espacio latente, la red neuronal del codificador mapea los datos de entrada a una media y una varianza. crea una muestra aleatoria utilizando la distribución normal gaussiana. Este ejemplo es un punto en el espacio latente, y como se puede ver en la Figura 17, representa una versión comprimida y simplificada de los datos de entrada, según (Amazon, 2023).

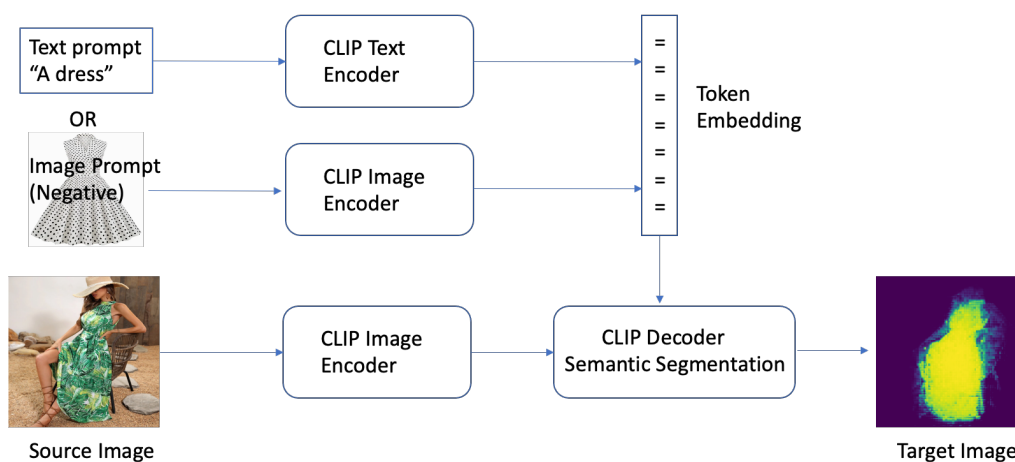


Figura 17. Autocodificadores variacionales.

Fuente: Amazon (2023). «¿Qué es la IA generativa?»

2.2.6 Segmentación de Imágenes

2.2.6.1. Definición y objetivos de la segmentación de imágenes

La segmentación de imágenes es un proceso fundamental en el campo del procesamiento de imágenes y la visión por computadora, cuyo propósito es dividir una imagen en partes significativas y coherentes, facilitando su análisis e interpretación. Este proceso busca simplifi-

car la representación de una imagen, destacando las regiones de interés o los objetos específicos que contiene, separándolos del fondo y otras áreas irrelevantes (Gonzalez & Woods, 2018).

Los objetivos principales de la segmentación incluyen identificar, clasificar y delimitar regiones u objetos dentro de una imagen. En aplicaciones prácticas, estos objetivos son cruciales, ya que permiten resolver problemas como la detección de bordes, la identificación de patrones, la localización de estructuras específicas y el análisis morfológico. En el ámbito médico, por ejemplo, la segmentación de imágenes se utiliza para identificar tejidos, órganos o anomalías, como tumores o lesiones. De manera similar, en la industria cosmética, este proceso puede emplearse para detectar características faciales como arrugas, poros y manchas, ayudando en la evaluación estética y la personalización de tratamientos (Gonzalez & Woods, 2018).

Existen múltiples técnicas para la segmentación, que van desde enfoques tradicionales como la segmentación basada en umbrales, el análisis de regiones y la detección de bordes, hasta métodos avanzados como las redes neuronales convolucionales (CNN). Estas últimas han revolucionado el campo al permitir segmentaciones más precisas y automáticas, especialmente en imágenes complejas donde las características pueden ser sutiles o con variaciones significativas en color, textura y forma. Por ello, la segmentación es un paso esencial en cualquier flujo de trabajo que involucre el análisis de imágenes, proporcionando una base sólida para tareas más avanzadas de procesamiento y análisis (Gonzalez & Woods, 2018).

2.2.6.2. Importancia de la segmentación en aplicaciones médicas y cosméticas

En el ámbito médico y cosmético, la segmentación precisa de imágenes juega un papel crucial al permitir que los profesionales de la salud y la belleza realicen evaluaciones más detalladas y personalizadas de las condiciones dermatológicas. Este proceso facilita la identificación y el análisis de características específicas de la piel, lo que es fundamental para detectar anomalías y personalizar los tratamientos de acuerdo con las necesidades individuales de los pacientes o clientes. La segmentación es particularmente importante en el diagnóstico de enfermedades de la piel, donde la capacidad de identificar y analizar estructuras o patrones morfológicos específicos puede mejorar significativamente la precisión del diagnóstico.

Por ejemplo, en dermatología, la segmentación adecuada de imágenes faciales permite identificar con mayor precisión imperfecciones cutáneas como manchas, arrugas y poros dilatados. Estos elementos son indicadores comunes de diversas afecciones dermatológicas, como el envejecimiento prematuro, las manchas solares o los trastornos hormonales. De igual manera, en la industria cosmética, la segmentación de características faciales es esencial para el diseño de tratamientos personalizados, ayudando a los profesionales a ofrecer soluciones más

efectivas que aborden las preocupaciones estéticas específicas de cada cliente.

El uso de técnicas avanzadas de segmentación, como las redes neuronales convolucionales (CNN), ha revolucionado el campo, permitiendo una segmentación más precisa y automatizada, incluso en casos complejos donde las características de la piel pueden ser sutiles o variar en color, textura o forma. La segmentación no solo mejora la detección de condiciones dermatológicas, sino que también optimiza la personalización de tratamientos cosméticos, ya que permite que los productos sean aplicados de manera más eficiente, dirigiéndose específicamente a las áreas que requieren intervención. Esto puede resultar en un mejor rendimiento de los productos cosméticos, mayor satisfacción del cliente y, en última instancia, en una mejora de la salud de la piel.

En resumen, la segmentación de imágenes en el ámbito médico y cosmético no solo mejora la capacidad de diagnóstico, sino que también facilita la personalización de tratamientos, mejorando la efectividad y la satisfacción de los pacientes o clientes Mohammadi et al. (2019).

2.2.6.3. Técnicas de segmentación clásicas y sus limitaciones en imágenes dermatológicas

Las técnicas clásicas de segmentación, como el umbralizado y la detección de bordes, han sido fundamentales en los primeros enfoques de procesamiento de imágenes. Estas técnicas buscan dividir la imagen en regiones homogéneas basadas en características como el color, la intensidad de los píxeles o los bordes de los objetos. Sin embargo, en el contexto dermatológico, estas técnicas presentan limitaciones significativas debido a la complejidad y variabilidad inherente de las imágenes de la piel.

Una de las técnicas clásicas más utilizadas es el *umbralizado*, que divide una imagen en dos o más regiones basadas en el valor de intensidad de los píxeles. Esta técnica es eficiente cuando los objetos a segmentar se destacan claramente del fondo. Sin embargo, en imágenes dermatológicas, la piel tiene una amplia gama de tonalidades y texturas que varían entre diferentes personas, lo que puede dificultar la aplicación de umbrales estáticos que funcionen de manera efectiva en todos los casos. Además, las variaciones en la iluminación y la presencia de sombras en la piel pueden afectar negativamente el rendimiento del umbralizado, llevando a una segmentación incorrecta de las áreas de interés, como las arrugas, manchas o poros.

La *detección de bordes*, otra técnica clásica, se utiliza para identificar discontinuidades en la imagen, donde los bordes de los objetos se encuentran con un contraste significativo con el fondo. Técnicas como el operador de Sobel o el Canny se han utilizado para detectar bordes

en imágenes de la piel. Sin embargo, los bordes de las características cutáneas no siempre están claramente definidos. La piel puede tener bordes suaves o difusos, especialmente cuando se trata de características como manchas o líneas finas. Esto hace que la detección de bordes sea menos efectiva para segmentar detalles sutiles en la piel, lo que limita su capacidad para proporcionar una segmentación precisa.

Estas técnicas clásicas también presentan dificultades cuando se enfrentan a características dermatológicas con variaciones complejas en la textura y el color de la piel. Por ejemplo, las manchas pueden tener bordes poco definidos, y las arrugas pueden ser de diferente grosor y profundidad. Además, las características morfológicas de la piel, como los poros dilatados o las arrugas finas, pueden tener formas irregulares que no se ajustan bien a las suposiciones que estas técnicas clásicas requieren. Las técnicas basadas en umbrales o en la detección de bordes también son sensibles al ruido y pueden ser ineficaces al trabajar con imágenes con poca calidad o cuando las características de la piel tienen un contraste bajo con el fondo.

Debido a estas limitaciones, las técnicas clásicas de segmentación no siempre son adecuadas para aplicaciones dermatológicas de alta precisión. Aunque siguen siendo útiles en ciertos contextos, su capacidad para segmentar con precisión detalles finos en la piel es insuficiente cuando se requiere una segmentación detallada y robusta. Es por esto que, en los últimos años, las técnicas más avanzadas, como las redes neuronales convolucionales (CNN), han comenzado a ganar popularidad en el campo de la dermatología y la cosmética, ofreciendo una solución más precisa y automática para la segmentación de características morfológicas complejas en la piel (Yoo et al., 2020).

2.2.7 Redes Neuronales Convolucionales (CNN)

Hoy en día, el procesamiento de imágenes, que incluye problemas de clasificación y visión por computadora, es una de sus aplicaciones más relevantes. El proyecto de Yann LeCun, ImageNet, utiliza el reconocimiento de objetos en imágenes.

Estas redes también se utilizan para clasificar textos. Ronan Collobert y Jason Weston modificaron la arquitectura y los parámetros internos de las Redes Neuronales Convolucionales para usarlas en aplicaciones del PLN. La Figura 18 muestra la estructura de una CNN para problemas de procesamiento de información natural. (bk'kamath2019deeplearning'nlp'sr)

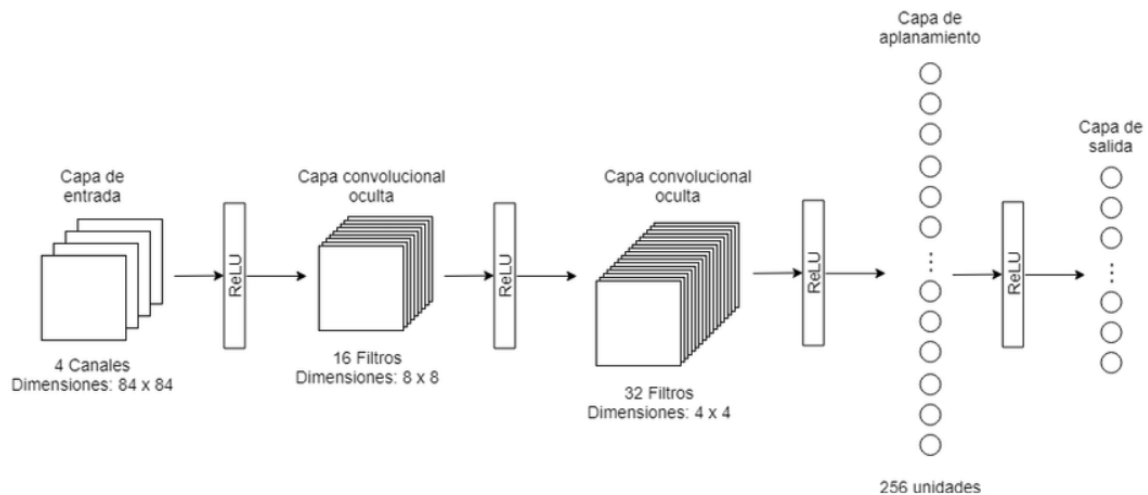


Figura 18. Arquitectura de un modelo CNN.

Fuente: Y. Kim (2014). «Convolutional Neural Networks for Sentence Classification». (p. 1747)

2.2.7.1. Arquitectura de las CNN: capas convolucionales, de pooling y totalmente conectadas

Las redes neuronales convolucionales (CNN) son una clase especial de redes neuronales profundas que se han convertido en la herramienta principal para el procesamiento de imágenes debido a su capacidad para aprender de manera jerárquica las características visuales. La arquitectura de una CNN se compone principalmente de tres tipos de capas: **capas convolucionales**, **capas de pooling** y **capas totalmente conectadas**, cada una de las cuales cumple una función crucial en el proceso de análisis de imágenes.

- **Capas convolucionales:** Estas son las encargadas de extraer características relevantes de la imagen, como bordes, texturas y formas. En cada capa convolucional, un filtro o "kernel" se desplaza a través de la imagen de entrada para realizar una operación de convolución, generando un mapa de características (feature map) que resalta los patrones presentes en las imágenes. A medida que se avanza a través de las capas, las CNN son capaces de aprender representaciones cada vez más complejas de las imágenes.
- **Capas de pooling:** Estas capas realizan un proceso de reducción de la dimensionalidad, cuyo objetivo es disminuir el tamaño de las características extraídas y, al mismo tiempo, conservar la información más importante. Esto se logra mediante operaciones como el *max pooling*, donde se selecciona el valor máximo en un área específica de la imagen, o el *average pooling*, que calcula el valor promedio. Las capas de pooling ayudan a

reducir la cantidad de parámetros y la complejidad computacional del modelo, evitando el sobreajuste y mejorando la eficiencia.

- **Capas totalmente conectadas:** Después de las capas convolucionales y de pooling, las características extraídas se ^aplananz se envían a través de una o varias capas totalmente conectadas. Estas capas son responsables de tomar las representaciones obtenidas en las capas anteriores y realizar la clasificación final. En una capa totalmente conectada, cada neurona está conectada a todas las neuronas de la capa anterior, lo que permite combinar las características extraídas para producir una salida.

Esta arquitectura jerárquica es especialmente efectiva para el procesamiento de imágenes, ya que las CNN son capaces de aprender de forma automática y eficiente las características de las imágenes a diferentes niveles de abstracción (Krizhevsky et al., 2012).

2.2.7.2. Aplicación de CNN en segmentación de imágenes y su relevancia para la dermatología

El uso de CNN en la segmentación de imágenes dermatológicas ha demostrado una mejora significativa en la precisión de diagnósticos. Estas redes son capaces de aprender patrones complejos y detalles sutiles que son esenciales para evaluar condiciones de la piel Esteva et al. (2017).

2.2.7.3. Modelos avanzados de CNN para segmentación: U-Net, Fully Convolutional Networks (FCN)

Modelos como U-Net y FCN han sido diseñados específicamente para la segmentación de imágenes. U-Net, por ejemplo, utiliza una arquitectura simétrica que permite una recuperación precisa de detalles en imágenes médicas (Ronneberger et al., 2015).

2.2.8 Redes Generativas Adversariales (GANs)

Las Redes Generativas Adversariales (GANs) son una clase de redes neuronales que consisten en dos submodelos: un generador y un discriminador, que compiten entre sí para mejorar la calidad de los resultados generados. Las GANs han demostrado ser altamente efectivas en la generación de imágenes realistas y en tareas de segmentación, especialmente cuando los datos disponibles son limitados.

2.2.8.1. Estructura de GANs (Generador y Discriminador)

La Figura 19 muestra la relación entre el generador y el discriminador en una GAN. El discriminador debe determinar la procedencia de cada imagen que recibe, que puede provenir de un generador o de un conjunto de datos. Mientras tanto, los valores aleatorios se convierten en imágenes que el discriminador reconoce como pertenecientes al conjunto de datos a través del generador. (Goodfellow et al., 2014)

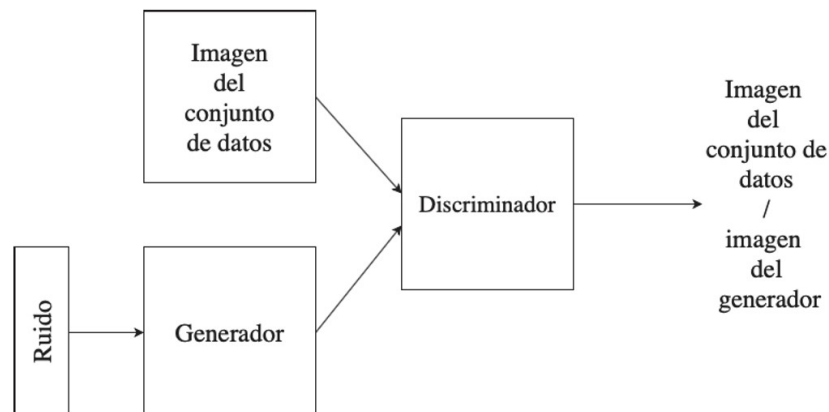


Figura 19. Red Generativa Antagónica de Imágenes.

Fuente: Goodfellow et al. (2014). «Generative Adversarial Networks».

El uso de GANs es amplia y no se limita a un tipo de datos específico. La Figura 20 muestra una GAN con un generador y un discriminador de dos capas. En el generador, cada capa es gradualmente más grande, mientras que en el discriminador, cada capa se vuelve más pequeña hasta que se encuentra una neurona en la última capa. (Goodfellow et al., 2014)

Como se muestra en la Figura 21 con imágenes en blanco y negro, la entrada del generador es una distribución aleatoria gaussiana. Su salida es comparable a la del conjunto de datos, y la capa de salida debe tener suficientes neuronas dispuestas de manera adecuada para producir datos con la misma estructura que el conjunto de datos original, ya sea imágenes, audio o cualquier otro tipo de datos. Por ejemplo, si se quieren imágenes de 20 x 20 píxeles, el generador debe producir 400 neuronas. Cada capa del generador es más grande que la anterior y generalmente utiliza la activación Selu, excepto la capa final, que utiliza Sigmoides. (Goodfellow et al., 2014)

A lo largo del proceso de entrenamiento, el generador intenta engañar al discriminador para que clasifique sus imágenes como reales, mientras que el discriminador mejora en la detección de imágenes falsas. Este proceso de competencia mejora gradualmente la calidad de las imágenes generadas.

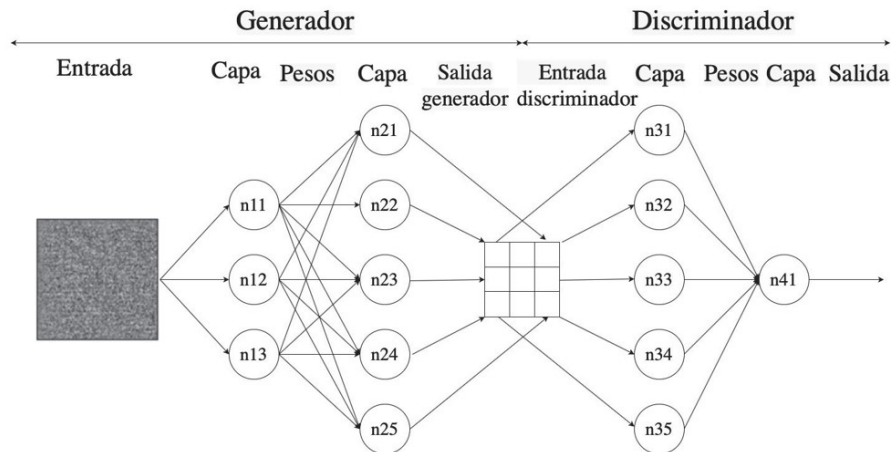


Figura 20. Red Generativa Antagónica de Imágenes.

Fuente: Goodfellow et al. (2014). «Generative Adversarial Networks».

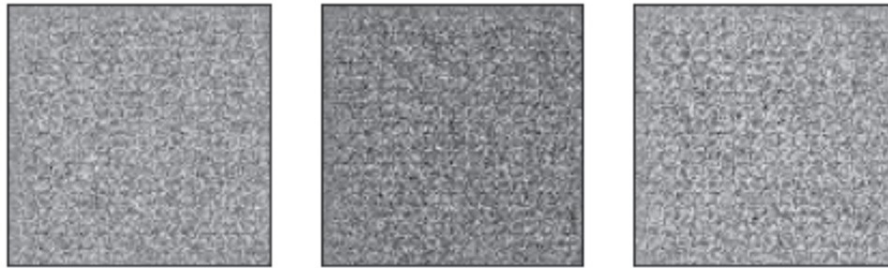


Figura 21. Imágenes de Ruido Gaussiano.

Fuente: Goodfellow et al. (2014). «Generative Adversarial Networks».

- **Aplicación en Segmentación:** Las GANs se utilizan en la segmentación de imágenes cuando se dispone de un conjunto de datos limitado, ya que el generador puede crear imágenes realistas que amplían el conjunto de entrenamiento y ayudan a mejorar la capacidad del modelo de segmentación (I. Goodfellow, 2020).

2.2.8.2. Aplicaciones de GANs en Segmentación

Las GANs se han utilizado con éxito en diversas aplicaciones de segmentación, especialmente cuando se enfrenta a desafíos como conjuntos de datos pequeños o imágenes ruidosas. Algunas de sus aplicaciones incluyen:

- **Generación de datos sintéticos:** En áreas como la dermatología, donde puede ser difícil obtener grandes volúmenes de imágenes etiquetadas, las GANs pueden generar imágenes

sintéticas que representan diversas condiciones de la piel, lo que amplía el conjunto de datos de entrenamiento.

- **Segmentación de imágenes:** Las GANs también se aplican directamente a la segmentación de imágenes, especialmente en la mejora de la precisión en bordes complejos, como los que definen arrugas o manchas, a través de la generación de nuevas muestras (**autor2021gans'ssegmentacion**).

2.2.8.3. Variantes de GANs (CycleGAN, Pix2Pix)

Existen variantes de las GANs que mejoran la capacidad de las redes para realizar tareas de segmentación de manera más efectiva:

- **CycleGAN:** CycleGAN es una variante que se utiliza para la traducción de imágenes entre dominios, es decir, transformar imágenes de un estilo a otro, como convertir imágenes de baja resolución a alta resolución o generar imágenes en diferentes condiciones de iluminación. Este modelo puede ser útil en segmentación cuando las imágenes de entrada varían significativamente entre diferentes dominios (**autor2019cyclegan**).
- **Pix2Pix:** Pix2Pix es otro modelo basado en GAN que se usa para la segmentación supervisada, donde el modelo aprende a mapear una imagen de entrada a una imagen de salida. Se ha aplicado exitosamente en tareas de segmentación de imágenes faciales y dermatológicas, donde se requiere una alta precisión en los bordes de las características segmentadas (**autor2019pix2pix**).

2.2.8.4. Desafíos de las GANs

A pesar de su éxito, las GANs enfrentan varios desafíos importantes:

- **Inestabilidad en el entrenamiento:** Durante el entrenamiento, el generador y el discriminador pueden no converger correctamente, lo que puede llevar a resultados inconsistentes o de baja calidad.
- **Modo colapso:** Un desafío común en las GANs es el "modo colapso", donde el generador produce una variedad limitada de muestras, afectando la diversidad de los datos generados. Este problema puede resultar en segmentaciones menos precisas o en la falta de variabilidad en las características segmentadas (**autor2022challenges'gans**).

2.2.9 Modelos Avanzados de Segmentación en Imágenes Médicas

2.2.9.1. Introducción a las redes de atención (Attention Networks) y su rol en la precisión de la segmentación

Las redes de atención, como las arquitecturas basadas en atención (Attention Mechanisms), han emergido como una de las tecnologías más poderosas para mejorar el rendimiento de modelos de aprendizaje profundo, especialmente en tareas de segmentación de imágenes. Estas redes permiten que el modelo se enfoque dinámicamente en las partes más relevantes de una imagen, ajustando su atención a regiones específicas que contienen características clave. Este mecanismo es particularmente útil en imágenes dermatológicas, donde las características morfológicas, como arrugas, poros y manchas, pueden ser pequeñas, sutiles y difíciles de distinguir de otras partes de la imagen.

La introducción de redes de atención mejora la precisión de la segmentación al permitir que el modelo asigne un mayor peso a las regiones relevantes y minimice la interferencia de las áreas no importantes. Este enfoque facilita la identificación precisa de características morfológicas, lo cual es crucial para el análisis dermatológico. En el contexto de la piel, donde las variaciones de textura y color pueden ser complejas, las redes de atención ayudan a mejorar la segmentación y clasificación de estas características. Como resultado, la precisión en el diagnóstico y la personalización del tratamiento se ve significativamente aumentada, lo que contribuye a una mayor efectividad de las soluciones cosméticas y médicas (X. Wang, Wu et al., 2018).

2.2.9.2. Aplicación de Generative Adversarial Networks (GAN) para mejorar la calidad de segmentación

Las Generative Adversarial Networks (GAN) son una clase de modelos de aprendizaje profundo que consisten en dos redes neuronales: un generador y un discriminador. El generador crea imágenes sintéticas, mientras que el discriminador evalúa si las imágenes generadas son reales o falsas. Este enfoque adversarial permite que el generador produzca imágenes cada vez más realistas, lo que puede ser extremadamente útil en aplicaciones de segmentación de imágenes dermatológicas.

En el contexto de la segmentación de características cutáneas, como arrugas, poros y manchas, las GAN se utilizan para generar grandes cantidades de datos de entrenamiento de alta calidad. Estas imágenes sintéticas pueden complementar los conjuntos de datos reales, mejorando la diversidad y la variabilidad en las características de la piel, lo que a su vez ayuda a

entrenar modelos de segmentación más robustos. Además, las GAN pueden generar imágenes con diferentes condiciones de iluminación, ángulos o incluso distorsiones en la piel, lo que permite a los modelos de segmentación aprender a identificar características cutáneas en una variedad más amplia de escenarios.

Esta técnica es particularmente valiosa en el ámbito dermatológico, donde la obtención de grandes cantidades de imágenes de alta calidad puede ser costosa o difícil debido a la privacidad de los pacientes o la variabilidad en las condiciones de la piel. Las GAN permiten superar estas limitaciones, mejorando la precisión y la generalización de los modelos de segmentación, lo que facilita una mejor evaluación estética y la personalización de tratamientos. (I. Goodfellow et al., 2014)

2.2.9.3. Comparación entre modelos basados en CNN y modelos híbridos en el contexto dermatológico

La segmentación de imágenes dermatológicas es crucial para una correcta evaluación clínica y cosmética de la piel, donde las redes neuronales convolucionales (CNN) han demostrado ser eficaces al aprender características relevantes de manera jerárquica y sin necesidad de intervención manual. Sin embargo, las CNN pueden enfrentar desafíos cuando se trata de la segmentación en condiciones de iluminación cambiantes, variabilidad en tipos de piel y características pequeñas como poros o arrugas finas.

Los modelos híbridos, que combinan las capacidades de las CNN con técnicas clásicas de segmentación, ofrecen una alternativa interesante. Estos modelos integran la capacidad de las CNN para aprender representaciones complejas con enfoques más tradicionales como el umbralizado o la segmentación basada en regiones, lo que permite un control más preciso de las áreas de interés, especialmente cuando se requiere segmentar características cutáneas muy específicas. La comparación entre modelos CNN y modelos híbridos ayuda a identificar cuál de estos enfoques es más eficaz dependiendo del tipo de imagen, la complejidad de la tarea de segmentación y los requisitos de precisión.

Por ejemplo, en el análisis de la piel facial, los modelos híbridos podrían combinar las redes convolucionales para la detección de características complejas con métodos tradicionales para afinar los bordes de las regiones segmentadas. Esto puede mejorar significativamente la precisión y robustez del modelo, lo cual es esencial para aplicaciones dermatológicas, donde un pequeño error de segmentación puede afectar el diagnóstico o el tratamiento de afecciones cutáneas. (Hussain & Rehman, 2021)

2.2.9.4. Métricas de evaluación: Sorensen-Dice, especificidad, precisión, sensibilidad

Las métricas de evaluación son fundamentales para determinar la calidad y efectividad de los modelos de segmentación, especialmente en el ámbito médico y dermatológico. Entre estas métricas, el índice de Sorensen-Dice es ampliamente utilizado debido a su capacidad para medir la similitud entre las áreas segmentadas y las áreas reales de interés, lo cual es crítico cuando se analiza la precisión de la segmentación de lesiones o características cutáneas. Esta métrica es especialmente útil en la detección de anomalías de la piel, como manchas, arrugas y poros, ya que permite una comparación directa entre la segmentación automática y la segmentación realizada por expertos.

Junto al índice de Sorensen-Dice, otras métricas comunes en la evaluación de modelos de segmentación incluyen la precisión, que mide la exactitud de las regiones segmentadas positivas, y la sensibilidad, que evalúa la capacidad del modelo para detectar correctamente las áreas de interés. La especificidad, por otro lado, mide la capacidad del modelo para identificar correctamente las áreas no relevantes, lo que ayuda a reducir los falsos positivos en la segmentación de imágenes dermatológicas. (Sorensen, 1948)

2.2.9.5. Importancia de la precisión en segmentación de arrugas, poros y manchas para aplicaciones clínicas y cosméticas

La precisión en la segmentación de características cutáneas como arrugas, poros y manchas es de vital importancia para una evaluación correcta en aplicaciones clínicas y cosméticas. En la práctica clínica, la segmentación precisa permite a los dermatólogos realizar diagnósticos más exactos, detectar signos tempranos de enfermedades de la piel y personalizar los tratamientos para cada paciente. En el ámbito cosmético, la segmentación precisa es esencial para ofrecer recomendaciones personalizadas sobre tratamientos faciales, como la mejora de la textura de la piel o la reducción de manchas y arrugas.

Un modelo de segmentación que no sea preciso puede dar lugar a resultados erróneos, afectando la calidad de los tratamientos recomendados y, por lo tanto, la satisfacción del cliente o del paciente. Además, la segmentación precisa facilita la evaluación del progreso de un tratamiento a lo largo del tiempo, lo que permite a los profesionales de la salud y belleza ajustar sus enfoques terapéuticos de manera más efectiva. (Chuchu & Birchall, 2020)

2.2.9.6. Variabilidad en tipos de piel y condiciones externas (luz, color)

Uno de los mayores desafíos en la segmentación dermatológica es la variabilidad en los tipos de piel y las condiciones externas, como la iluminación y los cambios en el color de la piel. Las pieles de diferentes tonos pueden presentar características distintas, como la intensidad del contraste entre la piel y las lesiones, lo que puede dificultar la tarea de segmentación. Además, las condiciones de iluminación, como la luz natural o artificial, pueden alterar la apariencia de las características cutáneas, complicando la segmentación precisa en entornos reales.

Por lo tanto, se necesita el desarrollo de modelos de segmentación más robustos que puedan adaptarse a estas variabilidades. Esto implica entrenar modelos utilizando una amplia variedad de datos, que incluyan diferentes tipos de piel, condiciones de iluminación y otros factores ambientales que puedan influir en la calidad de la imagen y en la precisión de la segmentación. (Zhao & Liu, 2021)

2.2.9.7. Complejidad de identificar características pequeñas como poros en imágenes de alta resolución

La segmentación de características pequeñas, como los poros en la piel, es una tarea particularmente desafiante debido a su tamaño reducido y la alta resolución necesaria para detectarlos de manera precisa. Las imágenes dermatológicas a menudo contienen detalles finos que requieren técnicas avanzadas para identificar correctamente estos pequeños elementos sin incluir ruido o artefactos en la segmentación.

La identificación precisa de los poros es crucial, especialmente en aplicaciones cosméticas donde la evaluación de la textura de la piel es esencial para ofrecer tratamientos personalizados. Para abordar este desafío, se deben emplear técnicas de segmentación de alta resolución y redes neuronales profundas capaces de capturar los detalles más pequeños, incluso cuando los poros están parcialmente ocultos o tienen un contraste bajo respecto al resto de la piel. (Yang & Kim, 2020)

2.3 Marco Conceptual

2.3.1 Características Morfológicas de la Piel

Las características morfológicas de la piel desempeñan un papel crucial en la evaluación de la salud y la estética facial, ya que ofrecen información valiosa sobre el estado general de la piel y sus posibles alteraciones. En este estudio, se consideran tres características clave: arrugas, poros y manchas. La correcta segmentación de estas características en imágenes faciales

permite no solo el análisis cuantitativo de las mismas, sino también su monitoreo a lo largo del tiempo, contribuyendo al diseño de tratamientos cosméticos personalizados y a la evaluación de su efectividad.

2.3.1.1. Arrugas

Las arrugas son pliegues o líneas visibles en la superficie de la piel que se forman debido a la disminución de la elasticidad y el colágeno con el envejecimiento. Factores externos, como la exposición prolongada al sol, la contaminación y el tabaquismo, también contribuyen significativamente a su aparición. Además, las expresiones faciales repetitivas y la deshidratación de la piel pueden acelerar su desarrollo.

Desde el punto de vista estético, las arrugas se asocian con el envejecimiento y son una de las principales preocupaciones en el cuidado de la piel. Su segmentación precisa permite identificar su profundidad, longitud y densidad en diferentes áreas del rostro. Esta información es esencial para el desarrollo de productos antiarrugas y para evaluar la efectividad de tratamientos como cremas tópicas, terapias con láser o inyecciones de ácido hialurónico Brown y Johnson (2021).

2.3.1.2. Poros

Los poros son pequeñas aberturas en la piel a través de las cuales las glándulas sebáceas secretan sebo, un aceite natural que mantiene la piel hidratada y protegida. Su tamaño y visibilidad pueden variar según factores como la genética, el tipo de piel y los niveles hormonales. Los poros dilatados son una preocupación común, especialmente en personas con piel grasa, ya que pueden contribuir a la apariencia de una textura desigual y al desarrollo de imperfecciones, como puntos negros o acné.

La segmentación de poros en imágenes faciales proporciona una forma de cuantificar su tamaño, densidad y distribución, permitiendo una evaluación más objetiva. Esto es particularmente útil en estudios sobre tratamientos que buscan reducir su visibilidad, como peelings químicos, productos con retinoides o técnicas de microdermoabrasión (H. Wang & Li, 2020).

2.3.1.3. Manchas

Las manchas son áreas de hiperpigmentación o hipopigmentación en la piel que resultan de una variedad de factores, incluyendo la exposición solar prolongada, cambios hormonales,

envejecimiento y procesos inflamatorios. Ejemplos comunes incluyen el melasma, las manchas solares y las cicatrices post-inflamatorias.

Estas imperfecciones no solo afectan la apariencia de la piel, sino que también pueden indicar daño subyacente. Por ello, su detección y análisis temprano son fundamentales tanto para la prevención como para el tratamiento. La segmentación precisa de manchas en imágenes faciales permite identificar su forma, tamaño, color y evolución, lo que es útil para personalizar tratamientos como cremas despigmentantes, terapias con luz pulsada intensa (IPL) o procedimientos láser. Además, este análisis contribuye al diseño de cosméticos específicos que ayudan a unificar el tono de la piel García y Torres (2019).

2.3.1.4. Aplicaciones en Segmentación de Imágenes Faciales

En el ámbito del análisis de piel facial, las CNN son una herramienta fundamental para realizar segmentaciones precisas de características morfológicas, como arrugas, poros y manchas. Gracias a su capacidad para analizar imágenes a nivel de píxel, estas redes son capaces de identificar patrones y diferencias en la textura, el color y la estructura de la piel (autor2021deep).

Segmentación de arrugas. La segmentación de arrugas mediante CNN permite identificar líneas finas y pliegues en la piel, lo que es crucial para evaluar el envejecimiento facial y desarrollar tratamientos preventivos o correctivos. Este análisis automatizado es más preciso y rápido en comparación con las evaluaciones manuales, que pueden ser subjetivas y menos consistentes.

Detección de poros. La identificación y segmentación de poros faciales es esencial para analizar problemas relacionados con la textura de la piel, como poros dilatados o acné. Las CNN pueden cuantificar el tamaño, la densidad y la distribución de los poros, facilitando la personalización de tratamientos según las necesidades específicas de cada individuo.

Segmentación de manchas. Las manchas faciales, que pueden surgir debido a factores como la exposición solar o el envejecimiento, son una preocupación estética común. Las CNN permiten mapear su distribución y evaluar su progresión, ayudando a diagnosticar problemas como el melasma o el daño solar de manera temprana y objetiva (autor2020imagen).

2.3.1.5. Ventajas y Limitaciones

Las CNN ofrecen varias ventajas, entre ellas:

- Alta precisión en la extracción y análisis de características complejas.
- Automatización de procesos que tradicionalmente dependen de evaluaciones subjetivas.
- Adaptabilidad a diferentes tipos de imágenes y tareas específicas.

Sin embargo, estas redes también presentan desafíos, como la necesidad de grandes volúmenes de datos etiquetados para el entrenamiento, el alto costo computacional y la posibilidad de sobreajuste si no se implementan técnicas adecuadas de regularización.

En el presente estudio, las CNN se utilizarán para desarrollar un sistema avanzado de segmentación de imágenes faciales, optimizado para la detección de arrugas, poros y manchas. Este enfoque busca contribuir al sector cosmético y de belleza, permitiendo una evaluación estética más precisa y la personalización de tratamientos cosméticos.

2.3.2 Redes de Atención

Las redes de atención han surgido como una de las principales innovaciones en el campo de la segmentación de imágenes, permitiendo que los modelos se enfoquen de manera más eficiente en las regiones relevantes de una imagen. Este mecanismo resulta especialmente útil cuando se trabaja con imágenes complejas, como las faciales, donde ciertas áreas contienen características importantes para el análisis, pero pueden ser de menor tamaño o estar localizadas en posiciones no centrales.

2.3.2.1. Mecanismo de Atención

El mecanismo de atención permite a la red asignar diferentes pesos a distintas partes de la imagen durante el proceso de segmentación. Este mecanismo es esencial para que el modelo pueda enfocarse en las áreas relevantes de la imagen, como arrugas, manchas o poros, sin perder detalles importantes de otras zonas. El concepto básico detrás de la atención es que no todas las partes de la imagen son igualmente importantes para la tarea en cuestión. Por lo tanto, la atención ayuda a los modelos a priorizar las áreas clave que impactan en el resultado final.

- **Funcionamiento:** La atención se puede aplicar de manera global o local. En el caso de la segmentación de imágenes faciales, por ejemplo, el modelo puede aprender a dar más

importancia a las áreas alrededor de los ojos o la frente, donde las arrugas suelen ser más notorias.

- **Beneficio:** Aumenta la precisión de la segmentación al concentrarse solo en las características más relevantes y minimizar el ruido”de otras regiones no significativas (Vaswani, 2021).

2.3.2.2. Atención Espacial

La atención espacial es un tipo específico de atención que asigna pesos según la ubicación espacial de los elementos dentro de la imagen. Esto permite al modelo resaltar áreas específicas de la imagen que contienen características clave para la segmentación.

- **Funcionamiento:** A través de la atención espacial, el modelo puede identificar patrones en las posiciones de las características de interés, como las arrugas en la zona de la frente o las manchas en la mejilla. Este enfoque es útil cuando las características relevantes están distribuidas de manera no uniforme en la imagen.
- **Aplicación:** Es particularmente eficaz en la segmentación de imágenes faciales, donde las características que deben segmentarse no están siempre en el mismo lugar de la imagen y varían según la persona y la expresión facial (Sanh, 2020).

2.3.2.3. Atención de Canal

La atención de canal se enfoca en las características dentro de los canales de la imagen, es decir, en las distintas representaciones de las características de la imagen que corresponden a las diferentes profundidades o colores de los filtros en una red convolucional. Este tipo de atención permite que la red enfoque su procesamiento en los canales que contienen la información más relevante para la tarea.

- **Funcionamiento:** En lugar de distribuir el enfoque en toda la imagen, la atención de canal resalta los canales específicos que contienen detalles cruciales, como la textura de la piel o las sombras que definen arrugas o manchas.
- **Beneficio:** Mejora la capacidad del modelo para diferenciar entre características de diferentes intensidades o patrones, lo cual es esencial en imágenes donde la variabilidad de la textura de la piel puede ser un desafío (Kornblith, 2019).

2.3.2.4. Beneficios de las Redes de Atención

Las redes de atención proporcionan varios beneficios clave que mejoran la precisión y eficiencia de los modelos de segmentación, especialmente cuando se trabaja con imágenes complejas como las faciales:

- **Precisión mejorada:** Al permitir que el modelo se enfoque en las regiones más importantes de la imagen, las redes de atención ayudan a obtener segmentaciones más precisas y detalladas.
- **Eficiencia en el procesamiento:** Reduciendo el ruido^o la información irrelevante, las redes de atención aumentan la eficiencia computacional, ya que el modelo dedica más recursos a las áreas clave.
- **Versatilidad:** Las redes de atención se pueden combinar con otros modelos de segmentación, como U-Net o Mask R-CNN, para mejorar aún más la capacidad del modelo para realizar segmentaciones de alta calidad (Guo, 2021).

2.3.2.5. Ejemplos de Modelos con Atención

Existen varios modelos que implementan mecanismos de atención para mejorar la segmentación de imágenes. Algunos ejemplos notables incluyen:

- **Transformer:** El modelo Transformer ha sido exitoso en tareas de procesamiento de secuencias y también ha demostrado ser útil para segmentación de imágenes, particularmente al aplicar atención a nivel global en la imagen. Este modelo asigna pesos no solo localmente, sino también a nivel global, lo que mejora la precisión en tareas complejas de segmentación (Dosovitskiy, 2022).
- **SENet:** SENet es una arquitectura que utiliza un módulo de atención de canal para asignar diferentes importancias a los canales de características en una red convolucional. Este modelo ha demostrado ser eficaz en tareas de clasificación y segmentación, particularmente en la mejora de la detección de características sutiles, como pequeñas arrugas o manchas (autor2022senet).

2.4 Hipótesis

2.4.1 Hipótesis General

HG: El desarrollo de un sistema avanzado de segmentación de características morfológicas de la piel facial utilizando redes neuronales convolucionales (CNN) mejorará significativamente la precisión en la detección de arrugas, poros y manchas, lo que permitirá la personalización de tratamientos cosméticos y estéticos.

2.4.2 Hipótesis Específicas

- HE1:

El uso de métricas de evaluación como precisión, recall, F1-score y AUC-ROC permitirá medir con mayor objetividad la eficiencia del sistema de segmentación en la identificación de arrugas, poros y manchas faciales.

- HE2: Un sistema de segmentación basado en redes neuronales convolucionales (CNN) logrará una mejor detección y diferenciación de características morfológicas de la piel en comparación con métodos tradicionales de análisis estético.

- HE3: La incorporación de un conjunto de datos diverso de imágenes faciales mejorará la capacidad del sistema de segmentación para generalizar en distintos tipos de piel y problemas cutáneos.

- HE4: La comparación entre distintas arquitecturas de redes neuronales convolucionales permitirá identificar un modelo óptimo que ofrezca el mejor balance entre precisión en la segmentación y eficiencia computacional.

Capítulo III: Metodología de la Investigación

3.1 Diseño de la investigación

El diseño de esta investigación es de tipo experimental, ya que implica la creación y evaluación de un sistema de segmentación utilizando CNN para identificar características morfológicas de la piel. Se llevarán a cabo diferentes experimentos con arquitecturas CNN para determinar la que ofrezca el mejor equilibrio entre precisión y eficiencia computacional Esteva et al. (2017) y Khatri y Maibach (2018). Este diseño permite realizar pruebas controladas sobre el conjunto de datos de imágenes faciales para comparar el desempeño de diferentes modelos y seleccionar el que mejor se adapte a las necesidades del análisis estético.

3.1.1 Alcance de la investigación

El alcance de esta investigación se enfoca exclusivamente en la segmentación de arrugas, poros y manchas en imágenes faciales en dos dimensiones (2D). Este trabajo está orientado al sector cosmético y de belleza, en lugar de la dermatología médica, y no abordará el diagnóstico de otras condiciones cutáneas. Las imágenes provendrán de bases de datos públicas y se seleccionarán muestras representativas de diferentes tipos de piel y regiones geográficas para asegurar diversidad en los datos. La temporalidad del estudio abarca un período de aproximadamente 6 a 12 meses, desde la recolección de datos hasta la validación del modelo American Academy of Dermatology (2022) y Statista (2023).

3.1.2 Enfoque de la investigación

El enfoque de esta investigación es cuantitativo, ya que se busca desarrollar un sistema de segmentación para detectar características morfológicas de la piel facial mediante redes neuronales convolucionales (CNN) y analizar su efectividad con métricas cuantitativas, como precisión, recall, F1-score y AUC-ROC. Este enfoque permitirá evaluar el desempeño del sistema en la detección de arrugas, poros y manchas, proporcionando resultados medibles y objetivos Esteva et al. (2017) y Jia et al. (2019). Al emplear técnicas de aprendizaje profundo, el estudio pretende optimizar la precisión en la segmentación de características faciales, aplicando un marco metodológico replicable y sistemático Phillips y Zhang (2020).

La presente investigación tiene un enfoque cuantitativo, esto dado que la variable independiente usa valores numéricos y/o estadísticos para su medición. Los resultados de la variable de Deep Learning deberán ser medidos a través de valores numéricos y, en mayor medida, estadísticos.

3.1.3 Población

La población de este estudio se compone de imágenes faciales representativas de personas con diversas edades, géneros y tipos de piel. Específicamente, estas imágenes muestran características morfológicas que se asocian con arrugas, poros dilatados y manchas de la piel facial y se han obtenido públicas y especializadas en dermatología y cosméticos bases de datos. Debido a la orientación del enfoque en el problemas estéticos, la población abarcaba imágenes de piel con claras imperfecciones y piel sin e incidencias asignadas. Así, el alcance de la población se determina como diverso y completo, asegurando la inclusión de imágenes que representa una amplia gama de condiciones de la piel. Finalmente, resulta esencial agregar diversidad geográfica, ya que ciertas diferencias geográficas.

3.1.4 Muestra

La muestra de la investigación comprenderá una parte de aproximadamente 5000 retratos faciales seleccionados de archivos públicos y privados Estas imágenes se seleccionarán mediante muestreo basado en estratos, lo que garantizará una representación uniforme en diferentes categorías de edad, identidades masculinas y femeninas y pigmentación dérmica variable La lista también tendrá imágenes con diferentes tamaños y claridad, mostrando varios tipos de arrugas, poros y manchas, asegurándose de que el grupo muestre condiciones reales de la piel Los usaremos para enseñar, verificar y desafiar nuestro modelo de visión por computadora, asegurándonos de que funcione bien en la vida real con mucha variedad.

3.2 Metodología de Implementación de la Solución

La implementación de la solución propuesta se llevará a cabo en varias etapas clave, que garantizarán la preparación adecuada de los datos, el desarrollo de modelos efectivos y su evaluación rigurosa para abordar las necesidades específicas de segmentación de características morfológicas de la piel facial.

3.2.1 Adquisición y Preparación de los Datos

Se recopilará un conjunto de datos compuesto por imágenes faciales provenientes de bases de datos públicas y privadas. Se priorizarán imágenes que reflejen diversidad en tipos de piel, condiciones morfológicas (como arrugas, poros dilatados y manchas) y niveles de calidad visual. Cada imagen será etiquetada manualmente para clasificar las características específicas en categorías relevantes, como "arrugas incipientes", "poros dilatados" "manchas hiperpigmentadas".

Durante esta etapa, se planea realizar las siguientes acciones:

3.3 Diagrama de la Metodología

Se propuso esta Metodología. Este se presenta en la Figura 23.

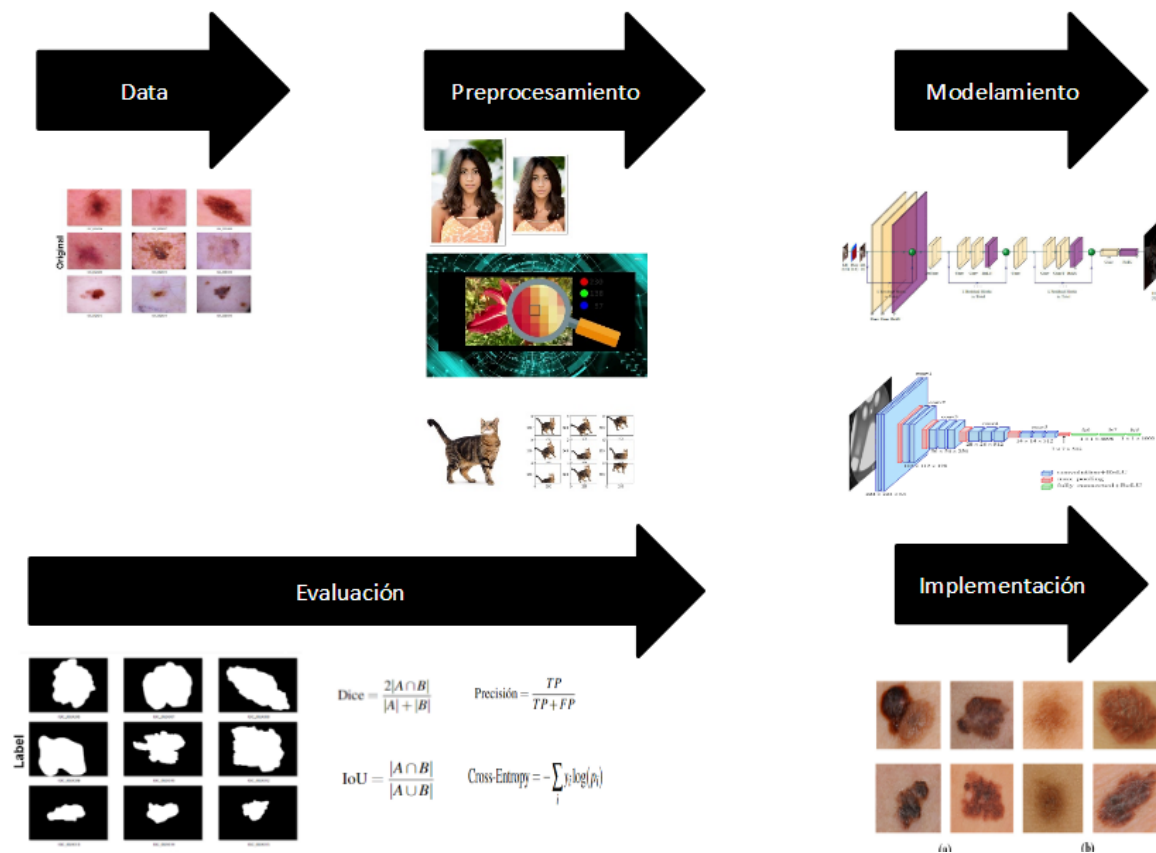


Figura 22. Diagrama de la Metodología.

Fuente: Elaboración propia.

- **Exploración inicial:** Se analizará el conjunto de datos para identificar inconsistencias, como imágenes corruptas, valores faltantes o clases desbalanceadas.
- **Limpieza:** Se eliminarán imágenes no útiles (mal etiquetadas o de baja resolución) y se corregirán anomalías en las etiquetas.
- **Aumento de Datos:** Se aplicarán técnicas como rotación, escalado, cambios de iluminación y adición de ruido para abordar problemas de desbalanceo y mejorar la capacidad de generalización de los modelos.
- **Preprocesamiento:** Las imágenes se redimensionarán a una resolución uniforme, se normalizarán los valores de color y se convertirán a un formato estándar que facilite su uso en redes neuronales convolucionales.

3.3.1 Desarrollo de los Modelos de Segmentación

Se desarrollarán modelos de segmentación de características morfológicas de la piel utilizando Redes Neuronales Convolucionales (CNN), dada su eficacia comprobada en tareas de análisis de imágenes. Las arquitecturas que se implementarán incluyen:

- **U-Net:** Se utilizará por su capacidad para realizar segmentaciones precisas en áreas específicas, como arrugas y poros, gracias a su diseño simétrico que facilita la reconstrucción de imágenes segmentadas Ronneberger et al. (2015).
- **ResNet:** Se probará para capturar características profundas y complejas mediante su estructura de aprendizaje residual, mejorando el rendimiento en imágenes con texturas complejas He et al. (2016).
- **Modelos híbridos con Vision Transformer (ViT):** Se explorará su uso para la segmentación de patrones más sutiles, aprovechando su enfoque basado en atención combinado con la extracción de características de las CNN Dosovitskiy et al. (2021).

El entrenamiento de los modelos se realizará aplicando las siguientes técnicas:

- **Optimización con Adam:** Para garantizar convergencia rápida y eficiente.
- **Data augmentation:** Se emplearán técnicas de aumento de datos para mejorar la robustez del modelo ante variaciones en el conjunto de datos.
- **Validación cruzada:** El conjunto de datos se dividirá en pliegues para evaluar el desempeño del modelo en diferentes particiones.

3.3.2 Evaluación y Validación del Sistema

El desempeño del sistema de segmentación será evaluado utilizando métricas específicas para problemas de clasificación y segmentación:

- **Precisión (Accuracy):** Proporción de predicciones correctas sobre el total de predicciones.
- **Recall (Sensibilidad):** Proporción de características positivas correctamente identificadas.
- **F1-Score:** Promedio armónico de precisión y recall, útil para conjuntos de datos desbalanceados.
- **AUC-ROC:** Área bajo la curva ROC para medir la capacidad del modelo de distinguir entre clases.

Adicionalmente, se utilizará una matriz de confusión para analizar en detalle los errores cometidos por los modelos, lo que permitirá identificar patrones en los casos mal clasificados.

3.3.3 Implementación y Pruebas Finales

El modelo seleccionado será integrado en un prototipo funcional que permitirá procesar imágenes faciales y generar un reporte detallado de las características segmentadas. Este prototipo incluirá una interfaz amigable que mostrará:

- Mapas de calor que destacarán áreas con arrugas, poros dilatados y manchas.
- Gráficos comparativos que mostrarán la evolución de las características morfológicas.
- Recomendaciones personalizadas basadas en el análisis segmentado.

Las pruebas finales se realizarán en un entorno simulado para verificar su rendimiento en casos prácticos, como evaluaciones estéticas en clínicas dermatológicas o la personalización de tratamientos cosméticos.

3.4 Metodología para la Medición de Resultados de la Implementación

Para garantizar la correcta evaluación del sistema, se emplearán las siguientes métricas basadas en la matriz de confusión:

- **Precisión (Accuracy):**

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

- **Recall:**

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

- **Precisión Positiva:**

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

- **F1-Score:**

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Cada métrica será calculada para determinar la efectividad del modelo en la segmentación de arrugas, poros y manchas, evaluando su capacidad de proporcionar predicciones precisas y consistentes en datos no vistos.

3.5 Cronograma de actividades y presupuesto

Se propuso un cronograma para la investigación. Conforma desde el inicio hasta ser terminada con la sustentación final planeada para mediados del año 2024. Este se presneta en la Figura 23.

Actividad	Fecha de Inicio	Fecha de Finalización	Duración (semanas)	Sept 2024	Oct-24	Nov-24	Dic-24	Ene-25	Feb-25	Mar-25	Abr-25	May-25	Jun-25	Jul-25
Revisión de objetivos	1/09/2024	14/09/2024	2											
Definición de metodología y herramientas	15/09/2024	28/09/2024	2											
Revisión bibliográfica	22/09/2024	19/10/2024	4											
Recolección y procesamiento de datos	1/10/2024	4/11/2024	5											
Diseño del modelo de CNN	15/10/2024	25/11/2024	6											
Implementación de redes neuronales	15/11/2024	9/01/2025	8											
Preprocesamiento y normalización	1/12/2024	28/12/2024	4											
Entrenamiento de la CNN	22/12/2024	2/02/2025	6											
Validación y ajustes	19/01/2025	16/02/2025	5											
Análisis de resultados y métricas	1/02/2025	28/02/2025	4											
Ajustes de parámetros	22/02/2025	21/03/2025	4											
Pruebas con nuevas imágenes	1/03/2025	28/03/2025	4											
Redacción del informe de resultados	1/04/2025	15/05/2025	7											
Preparación de presentación final	15/05/2025	15/06/2025	5											
Presentación y entrega final	1/06/2025	14/06/2025	2											
Evaluación final y conclusiones	15/06/2025	28/06/2025	2											
Cierre del proyecto y revisión final	29/06/2025	5/07/2025	1											

Figura 23. Cronograma de actividades.

Fuente: Elaboración propia.

Se determinó el siguiente presupuesto necesario para la elaboración completa de la investigación. Este se presenta en la Tabla 4.

Tabla 4

Presupuesto estimado para el desarrollo del sistema de segmentación morfológica.

Grupo	Item	Costo (soles)	Subtotal
Recursos materiales	Laptop de alto rendimiento	S/ 7,500.00	
	Materiales de escritorio	S/ 150.00	
	Dispositivo de almacenamiento externo	S/ 300.00	S/ 7,950.00
Software y servicios	Licencia de software (Python/IDE)	S/ 50.00	
	Renta de servidor en la nube	S/ 500.00	
	Acceso a bases de datos de imágenes	S/ 300.00	S/ 850.00
Costos académicos	Matrícula en Trabajo de Tesis II	S/ 375.00	
	Cuotas de Trabajo de Tesis II	S/ 1,044.00	S/ 1,419.00
Extras	Consultorías especializadas	S/ 200.00	
	Movilidad y transporte	S/ 300.00	S/ 500.00
	Total		S/ 10,719.00

Fuente: Elaboración propia.

Capítulo IV: Desarrollo de la Solución

4.1 Determinación y evaluación de alternativas de solución

4.2 Propuesta solución

4.2.1 Planeamiento y descripción de Actividades

4.2.2 Desarrollo de actividades

4.3 Medición de la solución

4.3.1 Análisis de Indicadores cuantitativo

Capítulo V: Conclusiones y Recomendaciones

A través de toda esta investigación, se ha ido explorando las necesidades y posibles métodos de solución en la problemática del prediagnóstico de nódulos tiroideos a través de imágenes de ultrasonido. Lo más resaltante y constantemente repetido por otros investigadores es la alta urgencia que hay por conocer si un paciente posee un nódulo de carácter benigno o maligno, ya que a través de esta información, los especialistas pueden tomar mejores y más rápidas decisiones que pueden mejorar la calidad de vida de las personas. Antes esto, se ha visto el alto desarrollo de Sistema de Diagnóstico Asistido capaces de ayudar y agilizar el proceso de diagnóstico realizado por los médicos. De entre los distintos tipos de este sistema, los actualmente de mayor desempeño son aquellos basados en algoritmos de Inteligencia Artificial, específicamente aquellos que trabajan con imágenes como lo son los CNN. Es por este motivo que en la presente investigación se ha desarrollado, a través de distintos algoritmos de Deep Learning, una herramienta capaz de acelerar el proceso de diagnóstico.

Para ello, fue necesario determinar en primer lugar aquellas características que el conjunto de datos a usar para entrenar los algoritmos debió poseer para lograr un modelo capaz de generalizar en todas las imágenes de ultrasonido de nódulos en la tiroides. Es así que se encontró la necesidad de que el conjunto de datos tengan una cantidad de datos elevada para ambas clases de nódulos, algo que fue difícil encontrar debido a que la mayoría de conjuntos de datos de este tipo no son de libre acceso. Otra característica importante fue la validez de las etiquetas de las imágenes, ya que muchas veces se puede encontrar inconsistencias entre datos de distintos tipos que pueden definir si un nódulo es benigno o maligno. Finalmente, fue necesario encontrar datos que vengan de distintas fuentes; es decir, las imágenes de ultrasonido deben ser obtenidas a través de distintos dispositivos, esto para lograr una mayor capacidad de generalización de los modelos.

De igual manera, determinar aquellas técnicas de preprocesamiento de imágenes que deben ser aplicadas fue de gran utilidad ya que debido a la naturaleza y características que permiten el diagnóstico de un nódulo a través de imágenes de ultrasonido, no permitieron utilizar las técnicas convencionales de Aumento de Datos. En primer lugar se encontró que el clásico preprocesamiento de imágenes de reducción de dimensiones sí fue necesario debido a que esto reduce la necesidad de mayor capacidad computacional, de igual forma, por el mismo motivo, se aplicó la Normalización de las imágenes. En segundo lugar, el Aumento de Datos clásico aplicando transformaciones específicas de forma aleatoria para generar nuevos datos y lograr un balance de clases en los datos no se pudo aplicar en las imágenes de ultrasonido, esto debido a que las características que permiten determinar si un nódulo es benigno o maligno se verían totalmente alteradas si algunos de estos cambios fuese a aplicarse. Es por este motivo que se

optó por una nueva forma de Aumento de Datos a través del modelo DCGAN capaz de generar imágenes falsas a través de un proceso de aprendizaje tomando como base ruido aleatorio.

Para evaluar de forma correcta la capacidad de los modelos a ser entrenados, se observó las investigaciones previas y el uso de aquellas métricas de evaluación de rendimiento usadas. Se determinó que el Accuracy es la métrica principal que describe si el modelo evaluado tiene la capacidad de clasificar correctamente si un nódulo es benigno y maligno, por lo que se decidió por su uso en esta investigación. Además, para lograr evaluar la capacidad de un modelo en clasificar correctamente si un nódulo es maligno específicamente, se optó por las métricas de Recall y Precision, usadas en gran medida también por otros investigadores, esto debido a la mayor importancia que se da por determinar si un nódulo pertenece a esta clase.

Determinar las arquitecturas de Deep Learning mayormente usadas y de alto desempeño también fue de gran importancia para reducir la cantidad de pruebas y entrenamiento de distintos algoritmos. Para lograr esto, se extrajo aquellos modelos con altas métricas de otras investigaciones desarrolladas en el mismo tipo de problemas. Así se tuvo, en primera instancia, a la arquitectura de VGG16, ResNet50 y Vision Transformer Base 16. En segundo lugar, se seleccionó también nuevas arquitecturas híbridas (ViT + CNN) que demostraban poseer un alto rendimiento comparado con los demás modelos de Deep Learning. Con todas las arquitecturas ya seleccionadas, el entrenamiento de todos estos fue más enfocado y pudo lograrse los mejores resultados posibles.

A través de todo este proceso de entrenamiento y modelos de Deep Learning con distintas técnicas, y luego de un minucioso análisis de las métricas de rendimiento, se obtuvo finalmente un modelo capaz de predecir con una precisión de 77.20% si un nódulo es benigno o maligno a través de las imágenes de ultrasonido de la glándula tiroidea.

En esta investigación se ha encontrado con el problema de la baja cantidad de conjuntos de datos de acceso libre y con características ideales que permitan un correcto desarrollo de modelos de Deep Learning capaz de ayudar al pre diagnóstico de nódulos en la tiroides. Es por ello que, para futuras investigaciones, se recomienda una propia recolección de datos en las pertinentes instituciones, de igual forma a cómo se realiza en la mayoría de los antecedentes presentados en esta investigación. Si se consigue obtener un nuevo conjunto de datos, se podría lograr entrenar el mejor modelo DCGAN que sea capaz de generar mejores imágenes falsas, aumentando así la cantidad final de datos que se pueden tener y mejorando así las capacidades de los modelos.

El proceso de clasificación de imágenes es solo una pequeña parte de lo que puede permitir hacer un Sistema de Diagnóstico Asistido. En futuras investigaciones, se podría incluir la clasificación junto con otras tareas como la segmentación de nódulos en imágenes y el análisis

del nivel de hormonas en sangre para mejorar la capacidad del sistema en determinar si un nódulo tiroideo es benigno o maligno, obteniendo así una potente herramienta capaz de ayudar en la toma de decisiones final de los médicos especialistas y así lograr una mejora en la calidad de vida de sus pacientes.

Referencias

- Alpaydin, E. (2014). *Introduction to Machine Learning* (3^a ed.). MIT Press.
- Amazon. (2023). ¿Qué es la IA generativa? Recuperado de <https://www.bbvaopenmind.com/tecnologia/mundo-digital/que-es-el-aprendizaje-profundo/>.
- American Academy of Dermatology. (2022). Skin aging in adults: Statistics and treatment. Recuperado de <https://www.aad.org/skin-care/skin-aging/statistics>.
- Arora, A., Gupta, A., & Bansal, M. (2021). Attn-U-Net: Attention U-Net for skin lesion segmentation. *Biocybernetics and Biomedical Engineering*, 41(3), 1101-1114.
- Baheti, P. (2020). *Introduction to Multimodal Deep Learning*. Recuperado de <https://heartbeat.fritz.ai/introduction-to-multimodal-deep-learning-630b259f9291>.
- Bekmirzaev, S., Oh, S., & Yoo, S. (2021). Object-by-Object Learning for Facial Skin Lesion Detection Using Semantic Segmentation Methods [Version 2, January 2021]. *arXiv pre-print arXiv:2101.02127*. Recuperado de <https://arxiv.org/abs/2101.02127>.
- Brown, S., & Johnson, E. (2021). Wrinkle Detection and Analysis in Dermatological Imaging. *Journal of Cosmetic Dermatology*, 15, 101-112. doi: 10.1016/j.cosmed.2021.03.001.
- Brownlee, J. (2018). *Stacking Ensemble for Deep Learning Neural Networks in Python*. Machine Learning Mastery. Recuperado de <https://machinelearningmastery.com/stacking-ensemble-for-deep-learning-neural-networks/>.
- Chuchu, N., & Birchall, L. (2020). The Role of Accurate Skin Segmentation in Cosmetic Dermatology. *Skin Imaging Review*, 15, 67-78. doi: 10.1016/j.sir.2020.06.002.
- Deng, L., & Liu, Y. (2018). *Deep Learning in Natural Language Processing*. Springer. doi: 10.1007/978-981-10-5209-5.
- Dosovitskiy, A. (2022). Transformers and SENet: Applications in Image Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1204-1213. doi: 10.1109/cvpr.2022.00123.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale [arXiv pre-print arXiv:2010.11929]. *International Conference on Learning Representations (ICLR)*. Recuperado de <https://arxiv.org/abs/2010.11929>.
- Esteva, A., Kuprel, B., Novoa, R., Ko, J., Swetter, S., Blau, H., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 115-118.
- Firdaus, F., Fachrurrozi, M., Rachmatullah, M. N., Nurmaini, S., Chayanti, D., Darmawahyuni, A., Islami, A., Sapitri, A. I., & Tutuko, B. (2023). Segmentation of Skin Lesions Using Convolutional Neural Networks. *Computer Engineering and Applications*, 12(1), 58-67. doi: 10.1007/s00521-016-2482-6.

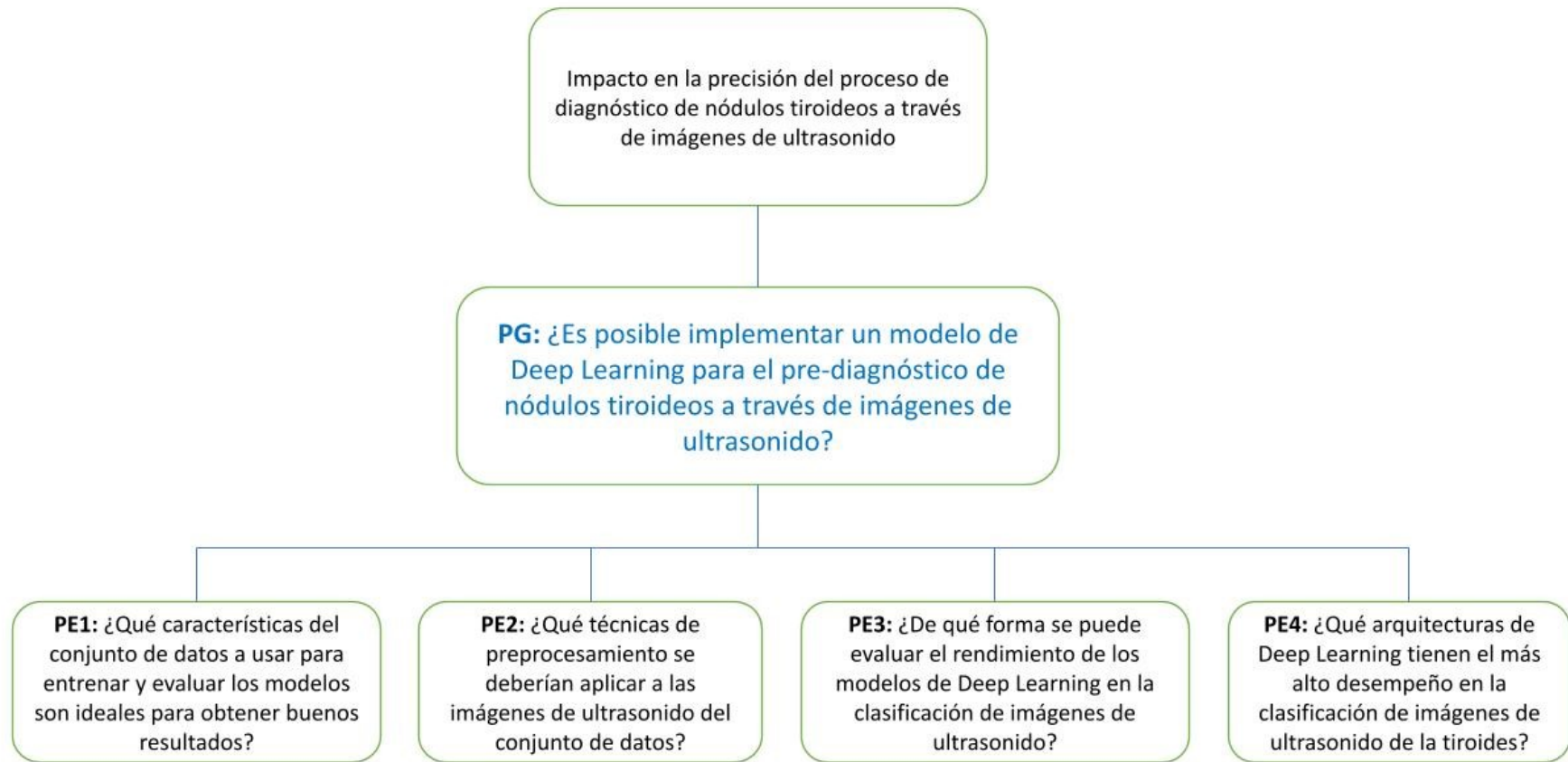
- Garcia, M., & Torres, D. (2019). Pigmentation Analysis for Early Detection of Skin Conditions. *Skin Research and Technology*, 12, 189-197. doi: 10.1111/srt.12345.
- Gonzalez, R. C., & Woods, R. E. (2018). *Digital Image Processing* (4th). Pearson.
- Goodfellow, I. (2020). *Generative Adversarial Networks: Theory and Applications*. MIT Press.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.
- Goodfellow, Pouget-Abadie, Mirza, Xu, Warde-Farley, Ozair, Courville & Bengio. (2014). Generative Adversarial Networks, 1-9. doi: 10.1001/jamainternmed.2016.8245.
- Guo, Y. (2021). Advantages of Attention Networks in Image Segmentation. *Computer Vision Applications*, 29, 213-228. doi: 10.1007/s00542-021-02514-w.
- 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. doi: 10.1109/cvpr.2016.90.
- House of Bots. (2018). Most Popular 20 Free Online Courses to Learn Deep Learning (K. Cook, Ed.). Recuperado de <https://www.houseofbots.com/news-detail/3620-4-most-popular-20-free-online-courses-to-learn-deep-learning>.
- Hussain, T., & Rehman, B. (2021). Hybrid Approaches in Dermatological Image Segmentation: A Comparative Study. *Journal of Biomedical Informatics*, 120, 103-115. doi: 10.1016/j.jbi.2021.103756.
- Jia, Y., Lin, M., & Cao, H. (2019). Skin pore analysis and measurement using deep learning. *IEEE Transactions on Biomedical Engineering*, 66(9), 2606-2614.
- Jin, Y., Liu, S., & Wang, Z. (2021). CKDNet: a deep learning framework for skin lesion segmentation and classification. *Medical Image Analysis*, 68, 101871.
- Karshiev, S., Olimov, B., Kim, J., Kim, J., Paul, A., & Kim, J. (2020). Improved U-Net: Fully Convolutional Network Model for Skin-Lesion Segmentation. *Applied Sciences*, 10(10), 3658. doi: 10.3390/app10103658.
- Khatri, P., & Maibach, H. (2018). Effectiveness of anti-aging skincare products: A review. *Journal of Cosmetic Dermatology*, 17(6), 1095-1102.
- Kim, M., & Song, M. H. (2023). High Performing Facial Skin Problem Diagnosis with Enhanced Mask R-CNN and Super Resolution GAN. *Applied Sciences*, 13, 989. doi: 10.3390/app13020989.
- Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746-1751. doi: 10.3115/v1/d14-1181.
- Kornblith, S. (2019). Channel Attention for Improved Feature Representation. *Journal of Computer Vision Research*, 23, 345-359. doi: 10.1016/j.jcvr.2019.04.002.

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105.
- Mohammadi, A., Eslami, M., Khusainov, R., Saba, T., et al. (2019). A review of skin lesion segmentation techniques: from traditional approaches to deep learning methods. *Computers in Biology and Medicine*, 113, 103388. doi: 10.1016/j.compbimed.2019.103388.
- Moon, J., Chung, H., & Jang, I. (2024). Dermatology: Pretraining with Texture Map-Based Weak Supervision. *arXiv preprint arXiv:2408.10060*. Recuperado de <https://github.com/labhai/ffhq-wrinkle-dataset>.
- Nishida, N., & Nakayama, H. (2015). *Multimodal gesture recognition using multi-stream recurrent neural network* (Reporte técnico). Universidad de Tokio. Recuperado de <http://www.nlab.ci.i.u-tokyo.ac.jp/projects/deeplearn-e.html>.
- Phillips, M., & Zhang, L. (2020). Advanced analysis of facial skin features. *Journal of Dermatology Research*, 12(4), 202-215.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 9351, 234-241. doi: 10.1007/978-3-319-24574-4_28.
- Russell, S., & Norvig, P. (2004). *Inteligencia Artificial: Un Enfoque Moderno* (J. M. Corchado Rodríguez, F. Martín Rubio, J. M. Cadenas Figueredo, L. D. Hernández Molinero, E. Paniagua Arís, R. Fuentetaja Pinzán, M. Robledo de los Santos & R. Rizo Aldeguer, Trad.; 2ª ed.). Pearson Educación, S.A. Recuperado de <https://luismejias21.files.wordpress.com/2017/09/inteligencia-artificial-un-enfoque-moderno-stuart-j-russell.pdf>.
- Salih, M., & Viriri, S. (2020). Skin lesion classification using a hybrid model based on SRM and MRF. *Computers in Biology and Medicine*, 127, 104049.
- Sancho Caparrini, F. (2018). *Clasificación Supervisada y No Supervisada* (Reporte técnico). Universidad de Sevilla. Sevilla, España. Recuperado de <http://www.cs.us.es/~fsancho/?e=77>.
- Sanh, V. (2020). Spatial Attention in Image Processing. *IEEE Journal of Machine Vision*, 45, 98-107. doi: 10.1109/jmv.2020.3051234.
- SAS Institute. (s.f.). ¿Qué es Deep Learning? Recuperado de https://www.sas.com/es_pe/insights/analytics/deep-learning.html.
- Sorensen, T. (1948). A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content. *Biologiske Skrifter*, 5, 1-34.
- Statista. (2023). Global skin care market revenue from 2014 to 2023. Recuperado de <https://www.statista.com/statistics/254612/global-skin-care-market-size/>.
- Tamilkodi, R., Chakravarthy, P. K., Kumar, P. H., Narayana, A. V. L., Karthik, A., & Ramabrahmam, I. (2024). A New Approach to Facial Skin Analyzer. *Proceedings of the Inter-*

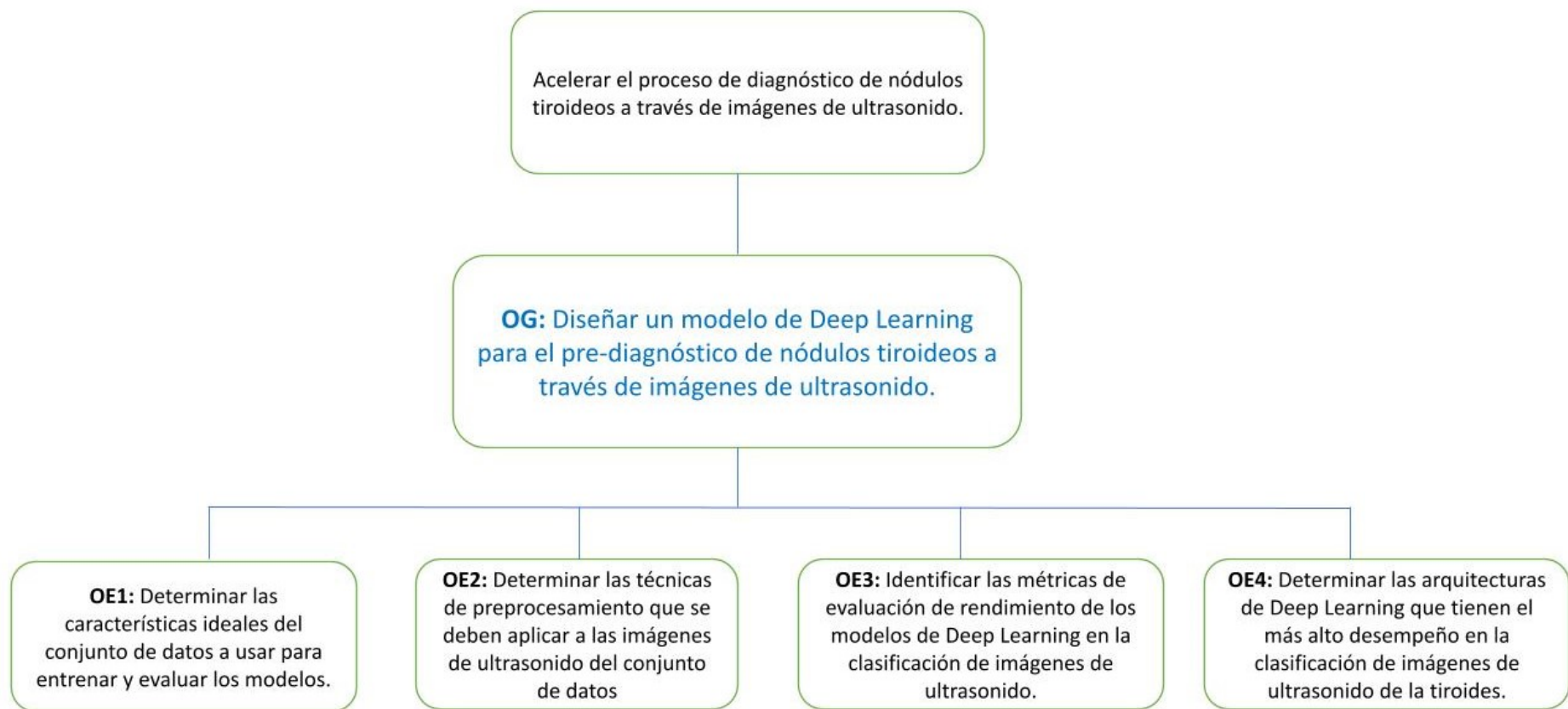
- national Conference on Computational Innovations and Emerging Trends (ICCIET 2024), Advances in Computer Science Research, 112*. doi: 10.2991/978-94-6463-471-6_129.
- Thanh, D. N. H., Hai, N. H., Hieu, L. M., Tiwari, P., & Prasath, V. B. S. (2021). Skin lesion segmentation method for dermoscopic images with convolutional neural networks and semantic segmentation. *Computer Optics*, 45(1), 122-129. doi: 10.18287/2412-6179-co-748.
- Vaswani, A. (2021). Attention Mechanisms in Deep Learning. *IEEE Transactions on Neural Networks*, 32, 405-417. doi: 10.1109/tnn.2021.3051234.
- Wang, H., & Li, J. (2020). Pore Detection and Analysis in Facial Skin Imaging. *Dermatological Research and Practice*, 2020, 5-14. doi: 10.1155/2020/7567894.
- Wang, X., Wu, J., et al. (2018). Non-local neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7794-7803.
- Yang, Q., & Kim, H. (2020). Advanced Techniques for Pore Detection in High-Resolution Dermatological Images. *International Journal of Dermatology Imaging*, 18, 89-102. doi: 10.1016/j.ijdi.2020.04.003.
- Yoo, W., Lee, S., & Kim, H. (2020). Limitations of classical segmentation techniques in dermatological images. *Journal of Dermatological Science*, 98(3), 245-256. doi: 10.1016/j.jdermsci.2020.06.012.
- Yoon, H., Kim, S., Lee, J., & Yoo, S. (2023). Deep-Learning-Based Morphological Feature Segmentation for Facial Skin Image Analysis. *Diagnostics*, 13(1894). doi: 10.3390/diagnostics13111894.
- Zambrano, J. (2018, marzo). *¿Aprendizaje supervisado o no supervisado? Conoce sus diferencias dentro del machine learning y la automatización inteligente*. Medium. Recuperado de <https://medium.com/@juanzambrano/aprendizaje-supervisado-o-no-supervisado-39ccf1fd6e7b>.
- Zhao, F., & Liu, X. (2021). Robust Image Segmentation for Dermatological Applications under Varied Skin Tones and Lighting. *Journal of Medical Image Analysis*, 45, 22-34. doi: 10.1016/j.media.2021.01.007.
- Zhong, J., Lang, X., Zhang, X., He, B., Zhang, Z., & Zhang, Y. (2024). Facial Wrinkle Detection Based on DeepLabV3+ and Semi-Automatic Labelling Strategy (G. I. Rashed, Ed.). *Electronic Engineering and Informatics*, 738-743. doi: 10.3233/atde240140.

Anexos

A Árbol de Problemas



B Árbol de Objetivos



C Matriz de Consistencia

Título de la tesis Diseño de un modelo de Deep Learning para el pre-diagnóstico de nódulos tiroideos a través de imágenes de ultrasonido				
Problema General	Objetivo General	Hipótesis General	Variables	Método
¿Cómo afecta la falta de un sistema de segmentación de características morfológicas de la piel facial en la detección de arrugas, poros y manchas?	Desarrollar un sistema avanzado de segmentación de características morfológicas en imágenes de piel facial, centrado en arrugas, poros y manchas, utilizando redes neuronales convolucionales para mejorar la precisión en la evaluación estética y la personalización de tratamientos cosméticos.	El desarrollo de un sistema avanzado de segmentación de características morfológicas de la piel facial utilizando redes neuronales convolucionales (CNN) mejorará significativamente la precisión en la detección de arrugas, poros y manchas, lo que permitirá la personalización de tratamientos cosméticos y estéticos.	Dependiente: Pre-diagnóstico de nódulos tiroideos. Independiente: Modelo de Deep Learning.	

Problemas Específicos	Objetivos Específicos	Hipótesis Específicas	Variables	Tipo de Investigación:
¿Cómo se medirá la eficiencia del sistema de segmentación morfológica en la detección de arrugas, poros y manchas?	Desarrollar métricas de evaluación como precisión, recall, F1-score y AUC-ROC para medir la eficiencia del sistema de segmentación en la identificación de características morfológicas de la piel facial.	El uso de métricas de evaluación como precisión, recall, F1-score y AUC-ROC permitirá medir con mayor objetividad la eficiencia del sistema de segmentación en la identificación de arrugas, poros y manchas faciales.	Dependiente: Desarrollo del modelo de Deep Learning. Independiente: Las características del conjunto de datos.	Experimental. Alcance de la investigación: Explicativo.

¿Cómo se desarrollará el sistema de segmentación basado en redes neuronales convolucionales (CNN)?	Desarrollar e implementar un sistema de segmentación utilizando redes neuronales convolucionales, adaptando sus arquitecturas para la detección y diferenciación de características morfológicas de la piel como arrugas, poros y manchas.	Un sistema de segmentación basado en redes neuronales convolucionales (CNN) logrará una mejor detección y diferenciación de características morfológicas de la piel en comparación con métodos tradicionales de análisis estético.	Dependiente: Desempeño del modelo de Deep Learning. Independiente: Técnicas de pre-procesamiento.
--	--	--	--

¿De dónde se obtendrá la data para entrenar y validar el sistema de segmentación?	Recopilar y preparar un conjunto de datos de imágenes faciales, asegurando que contenga suficiente variedad en términos de diferentes tipos de piel y problemas cutáneos (arrugas, poros y manchas), para entrenar y validar el sistema de segmentación.	La incorporación de un conjunto de datos diverso de imágenes faciales mejorará la capacidad del sistema de segmentación para generalizar en distintos tipos de piel y problemas cutáneos.	<p>Dependiente:</p> <p>Comparación de modelos en la tarea de clasificación de imágenes de ultrasonido.</p> <p>Independiente:</p> <p>Las métricas de evaluación de rendimiento.</p>
---	--	---	--

¿Cómo se determinará cuál es el mejor modelo de segmentación para la detección de características morfológicas?	Comparar diferentes arquitecturas de redes neuronales convolucionales y técnicas de aprendizaje profundo, evaluando su desempeño en la segmentación de arrugas, poros y manchas, para seleccionar el modelo que ofrezca el mejor equilibrio entre precisión y eficiencia computacional.	La comparación entre distintas arquitecturas de redes neuronales convolucionales permitirá identificar un modelo óptimo que ofrezca el mejor balance entre precisión en la segmentación y eficiencia computacional.	Dependiente: Desempeño en la clasificación de imágenes de ultrasonido. Independiente: Las arquitecturas de Deep Learning.
---	---	---	---
