



UNIVERSIDAD DE GUANAJUATO

CAMPUS IRAPUATO-SALAMANCA

DIVISIÓN DE INGENIERÍAS

**Aplicación de Modelos basados en el Comportamiento
del Sistema Visual Humano y DCNN al Procesamiento
de Imágenes Médicas**

TESIS

QUE PARA OBTENER EL GRADO DE:

DOCTOR EN INGENIERÍA ELÉCTRICA

PRESENTA:

M. en I. José Luis LÓPEZ RAMÍREZ

DIRECTORES DE TESIS:

Dr. José RUIZ-PINALES

Dr. Juan Gabriel AVIÑA-CERVANTES

SALAMANCA, GUANAJUATO

Enero, 2026

Resumen

En los últimos años, se ha logrado un éxito significativo con la implementación de algoritmos de aprendizaje automático en tareas de reconocimiento, clasificación y segmentación de imágenes en el ámbito médico. Actualmente, los hospitales y centros de salud disponen de un número considerable de registros y expedientes médicos electrónicos que contienen imágenes, datos y señales extraídas de diversos tipos de análisis realizados a pacientes con diferentes patologías.

En la actualidad, existe una amplia variedad de modalidades de imágenes médicas, y su uso está en constante aumento. Entre estas modalidades destacan el ultrasonido (*Ultrasound*, US), los Rayos-X, la Tomografía Computarizada (*Computed Tomography*, CT) y la Imagen por Resonancia Magnética (Magnetic Resonance Imaging, MRI), así como los escáneres de Tomografía por Emisión de Positrones (*Positron Emission Tomography*, PET), la fotografía retiniana, las diapositivas de histología y las imágenes de dermatoscopia. Algunas de estas modalidades están diseñadas para examinar múltiples órganos, como el CT y el MRI, mientras que otras se centran en órganos específicos, como la fotografía retiniana y la dermatoscopia. La cantidad de datos generados por cada estudio puede variar significativamente, desde unos pocos megabytes para una sola toma, hasta varios cientos de megabytes para una sola imagen de MRI. Estas diferencias tienen implicaciones técnicas en el procesamiento de los datos y en el diseño de la arquitectura de los algoritmos, considerando las limitaciones de los recursos computacionales disponibles para procesar el conjunto de imágenes.

Este trabajo introduce los algoritmos de aprendizaje automático aplicados al análisis de imágenes médicas, centrándose en el uso de modelos biológicamente inspirados en la percepción del sistema visual humano y en las Redes Neuronales Convolucionales Profundas (*Deep Convolutional Neural Networks*, DCNN). Se enfatizan y se abordan soluciones para tres problemas fundamentales en medicina. La primera es la segmentación de imágenes de fundus, donde se identifican y segmentan las venas presentes en el ojo humano. El segundo es la segmentación de tumores cerebrales a partir de imágenes obtenidas mediante ultrasonidos preoperatorios. Por último, se aborda la identificación automática de puntos cefalométricos en imágenes de radiografías, las cuales son comúnmente utilizadas en el diagnóstico maxilofacial por odontólogos expertos en el área.

La investigación en aprendizaje automático distingue entre la localización y la detección. Mientras que la segmentación traza contornos alrededor de los bordes de los objetos de interés y los etiqueta, el registro se refiere a ajustar una imagen sobre otra. Estas tareas tradicionalmente han sido realizadas por médicos especialistas en diversas áreas médicas. La llegada de algoritmos computacionales facilita la interpretación médica y reduce el tiempo necesario para diagnosticar diferentes enfermedades.

Salamanca, Gto., a 8 de enero del 2026.

MTRO. JUAN SIGFRIDO LÓPEZ CUSTODIO
COORDINADOR DE ASUNTOS ESCOLARES
P R E S E N T E.-

Por medio de la presente, se otorga autorización para proceder a los trámites de impresión, empastado de tesis y titulación al alumno **José Luis López Ramírez** del **Programa de Doctorado en Ingeniería Eléctrica** y cuyo número de **NUA** es: 143376 del cual soy director. El título de la tesis es:

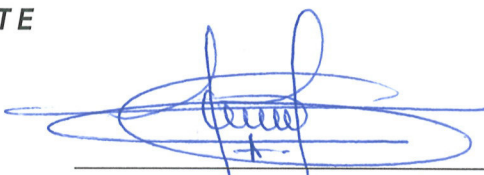
“Aplicación de Modelos basados en el Comportamiento del Sistema Visual Humano y DCNN al
Procesamiento de Imágenes Médicas”

Hago constar que he revisado dicho trabajo y he tenido comunicación con los sinodales asignados para la revisión de la tesis, por lo que no hay impedimento alguno para fijar la fecha de examen de titulación.

A T E N T A M E N T E



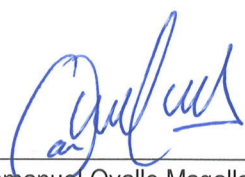
Dr. José Ruiz Pinales
DIRECTOR DE TESIS
SECRETARIO



Dr. Juan Gabriel Aviña Cervantes
DIRECTOR DE TESIS



Dr. Carlos Hugo García Capulín
PRESIDENTE



Dr. Emmanuel Ovalle Magallanes
VOCAL 1



Dr. Joel Quintanilla Domínguez
VOCAL 2



Dr. David Granados Lieberman
VOCAL 3

Dedicatoria

A mi esposa, **Perla**, por su apoyo, amor y compañía en nuestro hogar.

A mis hijos, **Diego Isaac** y **Perla Sofía**, desde su llegada al mundo, le han dado un nuevo sentido a mi existir.

A mis padres, **María Esther** y **José Luis**, por su apoyo incondicional y por todo lo que he recibido desde que nací.

A mi hermana **Sara**, por ser mi mejor amiga y estar conmigo en todo momento.

Siempre tendrán en mí, a alguien que hará todo lo posible por ayudarlos cuando lo necesiten y procurar que siempre estén bien.

José Luis López-Ramírez.

Agradecimientos

Agradecimientos Institucionales

- **Universidad de Guanajuato:** Por permitirme ser parte del posgrado en el programa del Doctorado en Ingeniería Eléctrica bajo en NUA: 143376, por la utilización de sus instalaciones, el acceso a becas, así como los apoyos y facilidades brindadas para la asistencia a congresos, estancias académicas y el pago por la publicación de artículos científicos.
- **Dirección de la División de Ingenierías del Campus Irapuato-Salamanca:** Por su apoyo para la asistencia de eventos académicos.
- **Departamento de Ingeniería Electrónica, División de Ingenierías Campus Irapuato-Salamanca:** Por el apoyo a la realización de la estancia doctoral en el año 2018-2019, para la asistencia al congreso RIIAA 2.0 en agosto de 2019, presentación en el congreso ROPEC en noviembre de 2020, presentación en el congreso COMIA 2020.
- **Universidad de Leipzig:** Por el apoyo y facilidades brindadas durante el periodo de mayo 2018 a enero 2019 mientras realizaba la estancia en sus instalaciones y por permitirme ser parte de un proyecto de investigación.
- **Centro de Innovación en Cirugía Asistida por Computadora (ICCAS):** Por su aceptación, apoyo, asesoría y facilidades brindadas durante el periodo de la estancia, mientras realizaba un proyecto de investigación bajo la supervisión de la **Dra. Claire Chalopin**, a quien se le agradece todas las gestiones para que fuera posible.
- **Secretaría de Ciencias, Humanidades, Tecnología e Innovación (SECIHTI):** Agradezco por el apoyo otorgado a través del CV: 495556 y SGN: 4555203. Mediante la beca mixta que permitió la realización de la estancia académica en Alemania, así como por la beca otorgada durante mis estudios de posgrado. Su respaldo fue fundamental para el desarrollo y conclusión de este trabajo.
- **Instituto Tecnológico Superior de Purísima del Rincón (ITSPR):** Agradezco por la oportunidad laboral que se me ha brindado y por el apoyo otorgado a través de las convocatorias que se emiten desde el TecNM. Por darme crecimiento Profesional.

Agradecimientos Personales

- Al **Dr. José Ruiz-Pinales**, por todo el apoyo y acompañamiento recibido durante la carrera, maestría y sobre todo en el Doctorado, así como en el desarrollo de la tesis, sus consejos y confianza depositada en mí. Por ayudarme a exigirme cada día más con la finalidad de desarrollar nuevas habilidades.
- Al **Dr. Juan Gabriel Aviña-Cervantes**, por todo el apoyo recibido durante mi estancia en la Universidad. Especialmente por las asesorías y confianza brindada en diversos proyectos y por motivarme a buscar la excelencia en todo lo que haga. Adicionalmente por el pago del artículo JCR publicado.
- Al **M.I. Fernando Daniel Hernández-Gutiérrez**, por todo el apoyo recibido en la redacción y experimentación del artículo de investigación. Especialmente por la confianza brindada y por motivarme a concluir este último ciclo de estudios.
- Al **Dr. Joel Quintanilla-Dominguez**, por todo el apoyo recibido en la lectura y revisiones del presente trabajo. por todos los consejos, especialmente por la confianza brindada y por motivarme.
- Al **Dr. David Granados-Lieberman**, por todo el apoyo recibido en las revisiones, consejos y comentarios que seguramente me ayudaran. Especialmente por la confianza brindada y por motivarme a concluir este último ciclo de estudios.
- Al **Dr. Carlos Hugo García-Capulin**, por las revisiones al presente manuscrito.
- Al **Dr. Emmanuel Ovalle-Magallanes**, por las valiosas y acertadas revisiones al presente manuscrito.
- Al **Dr. Mario Alberto Ibarra-Manzano**, por todo el apoyo recibido durante la carrera, la asesoría y confianza brindada en diversos proyectos, por incentivar mi desarrollo integral.
- Al **Dr. Rogelio Castro-Sánchez**, por el apoyo recibido, los consejos y comentarios hacia mí. Por su pasión por la enseñanza y la investigación y por motivarme a dar siempre mi máximo esfuerzo.
- Al **Dra. Ma. de Guadalupe García-Hernández**, por el apoyo recibido, los consejos y comentarios hacia mí. Principalmente por motivarme a dar el siguiente paso en mi carrera profesional y llegar hasta este momento.
- A mi Familia, Tios: **Toño y Agustín**; Tias: **Rita, Susana y Virginia**; Prim@s: **Luisa, Karla, Alejandro, Melina, Pancho**; Mis suegros: **Don Rogelio y Doña Chuy**; Mi Cuñado: **Rogelio** y Concuño: **Daniel Sustaita**, por todos los consejos y platicas.

- A los amigos que hice durante mi estancia en UG y aún persisten: **Pedro López-Rodríguez, José de Jesús Colín-Robles, Hugo Aguirre-Ramos, Elisee Ilunga-Mbuyamba, Alejandro López-Ramírez, Raúl Pérez-Mosqueda, Dr. Iván Hernández-Romano y José Cruz (Chepe).**
- A mis compañeros Docentes, **Juan de Anda, Germán Pérez, Fernando Villanueva, Rosa Jazmín, Guadalupe Ordaz, Alejandro Jerusalen, Daniel Jiménez, Miguel Ángel, José de Jesús y Oscar Omar Cuin** por animarme a seguir adelante en mi formación personal y profesional dentro del ITSPR.
- A todos **mis compañeros con los que compartí clases**, por las buenas anécdotas y el conocimiento adquirido.
- A **mis profesores**, por ayudarme en mi formación personal y profesional.

Gracias de corazón por influir en mi crecimiento personal y profesional, quiero que sepan que son personas que admiro y respeto.

Índice general

Resumen	I
Dedicatoria	II
Agradecimientos	III
1. Introducción	1
1.1. Hipótesis General	5
1.2. Justificación	5
1.3. Objetivos	7
1.3.1. Objetivo General	7
1.3.2. Objetivos Específicos	7
1.4. Descripción del Proyecto	8
1.5. Organización de la Tesis	9
1.6. Conclusiones de Capítulo	10
2. Análisis Cefalométrico	11
2.1. Introducción	11
2.2. Historia de la Cefalometría	11
2.3. Proceso de Captura	12
2.4. Proceso de Trazado y Etiquetado	14
2.4.1. Etiquetas Cefalométricas	15
2.5. Estado del Arte	18
2.5.1. Tipos de Imagen Utilizadas para el Análisis Cefalométrico	21
2.5.2. Técnicas de Detección de Puntos de Referencia Cefalométricos Empleados en la Literatura	22
2.5.3. Enfoques Basados en el Conocimiento	23
Método de Emparejamiento de Plantillas	24
Técnicas de Detección de Bordes	28

Métodos de Detección de Patrones	29
Modelo de Apariencia Activa (MAA)	29
Enfoques Basados en Algoritmos Genéticos	30
Modelos Híbridos	30
2.5.4. Modelos Basados en Inteligencia Artificial	32
Metodologías Basadas en Aprendizaje (<i>Machine-Learning</i>)	32
Arquitecturas Basadas en Aprendizaje Profundo (<i>Deep-Learning</i>)	37
2.6. Conclusiones del Capítulo	50
3. Redes Neuronales	52
3.1. Introducción	52
3.1.1. Aprendizaje de Máquina	53
3.2. La Neurona Biológica	54
3.3. Antecedentes Históricos	56
3.4. La Neurona Artificial	57
3.4.1. Funciones de Activación	58
3.4.2. Aprendizaje	59
3.5. Perceptrón Multicapa	60
3.5.1. Tipos de Redes Neuronales	64
3.5.2. Aprendizaje en Redes Neuronales de Más de Una Neurona	65
3.5.3. Backpropagation	67
3.5.4. Optimizadores	67
3.5.5. Validación del Modelo	68
3.5.6. Sobreajuste y Regularización	69
3.6. Aprendizaje Profundo	71
3.7. Conclusiones del Capítulo	72
4. Redes Neuronales Convolucionales	74
4.1. Introducción	74
4.2. Similitud Biológica	75
4.2.1. Procesamiento de la Luz	75
4.2.2. Procesamiento Visual a Alto Nivel	76
4.3. Antecedentes e Historia Reciente	76
4.4. La Operación de Convolución	77
4.4.1. Convolución en Imágenes	78
4.5. Capas Utilizadas en Redes Neuronales Convolucionales	78

4.5.1. Capa Convolutacional	80
4.5.2. Funciones Pooling	81
4.6. Aprendizaje en Redes Convolucionales	82
4.6.1. Ajuste de Hiperparámetros	83
4.7. Aplicaciones de las Redes Neuronales Convolucionales	86
4.7.1. Aplicaciones de las Redes Convolucionales al Área Médica	86
4.8. Conclusiones del Capítulo	88
5. Metodología	89
5.1. Introducción	89
5.2. Aproximación al Problema	90
5.3. Descripción de la Base de Datos Utilizada	90
5.3.1. Conjunto de Datos del Desafío IEEE ISBI 2015	91
5.3.2. Métricas de Evaluación del Rendimiento para la Detección Automática de Puntos Cefalométricos	93
5.3.3. Error Radial Medio (MRE)	93
5.3.4. Desviación Estándar	95
5.3.5. Tasa de Detección Exitosa (SDR)	95
5.3.6. Interpretación Clínica	96
5.4. Conjunto de Datos	96
5.4.1. Creación del Conjunto de Entrenamiento y Prueba	97
5.4.2. Intra-Observabilidad Entre Expertos	99
5.4.3. Cálculo del Error Interobservador	100
5.4.4. Cálculo del Error Interobservador para la Base de Datos ISBI 2015	101
5.4.5. Visualización del Error Interobservador	102
5.5. Diseño de la Arquitectura de la Red Neuronal Convolutacional	104
5.5.1. Hardware Utilizado Durante el Entrenamiento	105
5.5.2. Software Utilizado	105
TensorFlow	105
Keras	107
5.5.3. Diseño de la Arquitectura de la Red	109
Diagrama de la Arquitectura del Modelo de Red Convolutacional	110
5.5.4. Optimizador Nadam	111
5.5.5. Ajuste Adaptativo de la Tasa de Aprendizaje	114
5.5.6. Conclusiones del Capítulo	116

6. Resultados y Conclusiones	118
6.1. Arquitectura de Red Final	119
6.2. Análisis de la Función de Pérdida	120
6.2.1. Unidades de la predicción y del MRE	121
6.2.2. Conversión de píxeles a milímetros	122
6.2.3. Interpretación de la curva de aprendizaje	122
6.2.4. Hiperparámetros y parámetros del modelo	122
6.2.5. 6.2.4. Hiperparámetros y parámetros del modelo	123
6.3. Análisis de Resultados de Localización por cada Imagen de los Subconjuntos de Prueba	125
6.3.1. Resultados por Imagen en el Conjunto Test1	128
Análisis Cualitativo del Desempeño en el Conjunto Test1	128
6.3.2. Resultados por Imagen en el Conjunto Test2	130
Análisis Cualitativo del Desempeño en el Conjunto Test2	131
6.4. Análisis de Resultados de Localización por Etiqueta Cefalométrica	133
6.4.1. Análisis del Error Radial Medio por Landmark Cefalométrico	136
6.5. Análisis de Resultados Generales considerando la Tasa de Detección Exitosa (SDR) . .	138
6.6. Comparación de Resultados	139
6.6.1. Desarrollo de la Interfaz Gráfica de Usuario para el etiquetado cefalométrico .	141
6.7. Conclusiones Generales	146
6.8. Trabajo a Futuro	147
6.9. Introducción	148
6.10. Criterios de Categorización	148
6.11. Categoría 1: Inteligencia Artificial y Visión por Computadora en Medicina	149
6.11.1. Descripción de la línea	149
6.11.2. Productos académicos asociados	149
6.11.3. Aportaciones clave	149
6.12. Categoría 2: Señales Biomédicas y Reconocimiento de Patrones	149
6.12.1. Descripción de la línea	149
6.12.2. Productos académicos asociados	150
6.12.3. Aportaciones clave	150
6.13. Categoría 3: Inteligencia Artificial de las Cosas (AIoT), Robótica y Visión Artificial . .	150
6.13.1. Descripción de la línea	150
6.13.2. Productos académicos asociados	150
6.13.3. Aportaciones clave	150
6.14. Categoría 4: Baterías, Electromovilidad y Sistemas Energéticos Inteligentes	151

6.14.1. Descripción de la línea	151
6.14.2. Productos académicos asociados	151
6.14.3. Aportaciones clave	151
6.15. Categoría 5: Energía Renovable, Semiconductores e Inteligencia Artificial	151
6.15.1. Descripción de la línea	151
6.15.2. Productos académicos asociados	151
6.15.3. Aportaciones clave	152

Índice de figuras

1.1. Diferencia entre sistemas de Diagnóstico Asistido por Computadora y sistemas de Diagnóstico Automatizado por Computadora.	2
1.2. Proceso de diagnóstico utilizando análisis cefalométrico.	5
1.3. Número de publicaciones que involucran el término <i>Computer-Aided Diagnosis</i> desde el año 2010 hasta el año 2025.	6
2.1. Radiografía Lateral de Cráneo.	13
2.2. Medidas estándar requeridas para la captura de una radiografía lateral de cráneo. . . .	14
2.3. Radiografía con el trazado de marcas de referencia y algunas estructuras de tejido blando y duro.	15
2.4. Ubicación de las 19 etiquetas cefalométricas a utilizar.	16
2.5. Distribución anual de los artículos revisados basados en conocimiento, aprendizaje automático y aprendizaje profundo.	20
2.6. Esquema de evolución metodológica para la detección automática de puntos cefalométricos.	24
2.7. Enfoques Basados en Conocimiento Utilizados en la Detección Automática de Puntos Cefalométricos.	26
2.8. Línea de tiempo de los Modelos Híbridos aplicados a la identificación automática de etiquetas cefalométricas.	32
2.9. Métodos basados en Aprendizaje Automático para la Detección de Puntos Cefalométricos	33
2.10. Línea de Tiempo de las Técnicas basadas en Aprendizaje Automático aplicadas a la Localización de Etiquetas Cefalométricas.	33
2.11. Línea de tiempo compacta de arquitecturas CNN aplicadas a la detección automática de puntos cefalométricos.	50
3.1. Clasificación y distintos algoritmos de aprendizaje de máquina.	54
3.2. Neurona biológica.	56
3.3. Modelo de neurona artificial.	59
3.4. Función Booleana XOR.	62

3.5. Diagrama de un perceptrón multicapa.	63
3.6. MLP para la solución de XOR.	64
3.7. Descenso del gradiente para una función unidimensional.	66
3.8. Diferencia entre subajuste, buen ajuste y sobreajuste de un modelo.	69
3.9. Curvas de pérdida en entrenamiento y en validación y su relación con el sobreajuste.	70
3.10. Efecto del dropout sobre un perceptrón multicapa.	71
3.11. Relación de pertenencia entre Inteligencia Artificial, Aprendizaje de Máquina y Aprendizaje Profundo.	72
4.1. Convolución para imágenes, realizada de la misma manera que las bibliotecas de aprendizaje de máquina más comunes.	79
4.2. Efecto del relleno aplicado a imágenes.	80
4.3. Efecto en la variación del paso aplicado a imágenes.	81
4.4. Variantes de la operación de agrupación.	83
4.5. Filtros aprendidos dentro de las primeras capas de la arquitectura VGG16.	84
4.6. Mapas de características en diversas capas de la arquitectura VGG16.	85
5.1. Esquema de un modelo de red neuronal convolucional utilizado para realizar regresión.	91
5.2. Máquina Cranex,	92
5.3. Comparación de las anotaciones realizadas por los expertos: Senior y Junior sobre la imagen 007.	93
5.4. Tasas de éxitos para la etiqueta señalada por el experto 1.	96
5.5. Error promedio por landmark entre expertos. La línea punteada indica el error global promedio.	102
5.6. Archivo 007 de la base de datos ISBI 2015, correspondiente al error máximo entre observadores. El punto marcado en rojo corresponde al experto senior y el azul al experto junior. La distancia entre ambas marcas es de 22.01 mm, representando la discrepancia más alta registrada para la etiqueta llamada "Pos".	103
5.7. Mapa de calor de errores interobservador por archivo y etiqueta cefalométrica (en mm). Los tonos más claros indican errores mayores, observándose mayor variabilidad en los puntos asociados a P (Porion) y Pos (Pogonion de tejido blando).	104
5.8. Arquitectura "Encoder", propuesta como Solución.	110
5.9. Plataforma NVIDIA Jetson Nano B01 utilizada durante la etapa de pruebas. Esta tarjeta de desarrollo cuenta con GPU Maxwell de 128 núcleos y está optimizada para la ejecución de modelos de inteligencia artificial en el borde.	115

6.1. Red Neuronal Convolutacional Propuesta Como Solución en la Identificación Automática de Etiquetas Cefalométricas.	118
6.2. Curva de pérdida durante el entrenamiento del modelo propuesto.	121
6.3. Distribución del Error Radial Medio (MRE) por imagen para el conjunto de prueba <i>Test1</i> . La línea discontinua representa el MRE promedio global del conjunto.	128
6.4. Imágenes del conjunto <i>Test1</i> con los seis valores más bajos de Error Radial Medio (MRE).129	
6.5. Imágenes del conjunto <i>Test1</i> con los seis valores más altos de Error Radial Medio (MRE).130	
6.6. Distribución del Error Radial Medio (MRE) por imagen para el conjunto de prueba <i>Test2</i> . La línea discontinua representa el MRE promedio global del conjunto.	131
6.7. Imágenes del conjunto <i>Test2</i> con los seis valores más bajos de Error Radial Medio (MRE).132	
6.8. Imágenes del conjunto <i>Test2</i> con los seis valores más altos de Error Radial Medio (MRE).133	
6.9. Error Radial Medio (MRE) promedio por punto cefalométrico con desviación estándar para el conjunto <i>Test1</i>	135
6.10. Error Radial Medio (MRE) promedio por punto cefalométrico con desviación estándar para el conjunto <i>Test2</i>	135
6.11. Distribución del Error Radial Medio (MRE) para cada punto cefalométrico en las 150 imágenes de prueba del conjunto de datos <i>Test1</i>	136
6.12. Distribución del Error Radial Medio (MRE) para cada punto cefalométrico en las 100 imágenes de prueba del conjunto de datos <i>Test2</i>	137
6.13. Arquitectura adversarial Encoder–Decoder de Dai <i>et al.</i> (2019), empleada como referencia representativa de redes profundas para localización cefalométrica.	140
6.14. Tool Action flow.	144
6.15. Medical Graphical User Interface for cephalometric labeling.	144
6.16. CLR labeled using the proposed tool.	145

Índice de tablas

2.1. Descripción de planos cefalométricos que se pueden trazar a partir de las etiquetas cefalométricas seleccionadas.	17
2.2. Descripción de las 19 etiquetas cefalométricas utilizadas.	18
2.3. Comparación de modalidades de imagen cefalométrica: ventajas y desventajas.	23
2.4. Métodos propuestos para la Detección Automática de Etiquetas Cefalométricas Basadas en Conocimiento.	25
2.5. Métodos de Aprendizaje Automático (ML) para la Detección Automática de Etiquetas Cefalométricas.	37
3.1. Sucesos que definieron el desarrollo de las redes neuronales.	58
3.2. Funciones de activación comunes.	61
5.1. Descripción del conjunto de datos.	91
5.2. Descripción anatómica y clasificación por tipo de tejido de los puntos de referencia cefalométricos utilizados en el estudio.	94
5.3. Error medio por etiqueta cefalométrica entre los expertos.	101
5.4. Descripción del hardware utilizado.	105
5.5. Parámetros del optimizador <i>Nadam</i> empleados durante el entrenamiento de la red neuronal.	113
5.6. Parámetros del programador de tasa de aprendizaje <i>ReduceLROnPlateau</i>	114
5.7. Especificaciones técnicas de la NVIDIA Jetson Nano B01 empleada durante las pruebas.	116
6.1. Detalles de las Capas de la Arquitectura de Red Neuronal Propuesta.	120
6.2. Resumen del modelo obtenido de Keras.	124
6.3. Error radial medio (MRE) por imagen para el conjunto <i>Test1</i> de 150 muestras.	126
6.4. Error radial medio (MRE) por imagen para el conjunto <i>Test2</i> de 100 muestras.	127
6.5. MRE promedio obtenido para cada etiqueta cefalométrica.	134
6.6. Resultados del error radial medio (MRE), desviación estándar y porcentaje de detecciones dentro de distintos umbrales SDR (<i>Test1</i>).	138

6.7. Resultados del error radial medio (MRE), desviación estándar y porcentaje de detecciones dentro de distintos umbrales SDR (Test2).	139
6.8. Rango de SDR para los conjuntos de datos Test1 y Test2.	139
6.9. Detalles de las capas de la arquitectura adversarial Encoder–Decoder propuesta.	142
6.10. Comparación del modelo propuesto con métodos del estado del arte en el conjunto IEEE ISBI 2015 (Test1)	143
6.11. Comparación del modelo propuesto con métodos del estado del arte en el conjunto IEEE ISBI 2015 (Test2)	143

Capítulo 1

Introducción

El crecimiento sostenido de la población humana durante los últimos años (el crecimiento poblacional dentro de los próximos 30 años se estima en 2000 millones de personas más), sumado al aumento en la esperanza de vida mundial [1], han ocasionado un incremento en la demanda en todos los servicios, desde alimentación, educación, energía y telecomunicaciones y demás áreas como la medicina, la cual requiere cada vez más de herramientas, métodos y equipos que agilicen los procesos de diagnóstico, tratamiento y monitoreo de procesos médicos.

Como consecuencia del crecimiento exponencial en el poder de cómputo de los dispositivos electrónicos [2], diversas industrias migraron sus servicios a una versión digital, como sucedió con algunos sensores, medios de comunicación, entre otros. La medicina no fue la excepción, con el surgimiento de sensores e instrumentos de captura digitales, como las máquinas de rayos X, con las que inició la radiografía digital, que permitió la generación de grandes cantidades de imágenes de diversos estudios en formato digital.

Posterior a la digitalización de la industria de la radiología, durante la década de 1980 [3], el número de radiografías digitales disponibles para su estudio creció de una manera considerable, sin embargo no fue hasta los inicios del tercer *boom* de la inteligencia artificial en el año 2020 [4], gracias al cómputo paralelo y demás mejoras en el procesamiento de datos en los equipos de cómputo que se aprovechó la cantidad de datos generada con la aplicación de técnicas de aprendizaje automático a imágenes médicas con la finalidad de crear sistemas que permitan automatizar ciertas tareas en el proceso de diagnóstico médico, basados en diagnóstico asistido por computadora.

Derivado de las mejoras mencionadas, surgen los sistemas de diagnóstico asistido por computadora (CAD, por sus siglas en inglés: *Computer Aided Diagnosis*), los cuales permiten la detección temprana y medición del avance de alguna enfermedad o patología reconocida en la imagenología a la par que reduce los tiempos de lectura de las imágenes. La misión de los sistemas CAD es brindar una “segunda

opinión” al profesional de la salud, una salida que sea complementaria y permita al profesional de la salud reforzar o refutar su hipótesis respecto al estudio en turno.

Es importante hacer la distinción entre diagnóstico asistido por computadora y diagnóstico automatizado por computadora.

Figura 1.1 muestra de manera gráfica el flujo de acción en los dos tipos de sistemas. Figura 1.1a muestra el sistema de diagnóstico asistido: el sistema CAD captura los síntomas del paciente y emite un dictamen; posteriormente, el especialista analiza, con base en su experiencia, el dictamen del sistema y, finalmente, emite un diagnóstico al paciente. En caso contrario, en el esquema de diagnóstico automatizado, el dictamen final del sistema se transfiere directamente al diagnóstico del paciente, como se muestra en la Figura 1.1b.

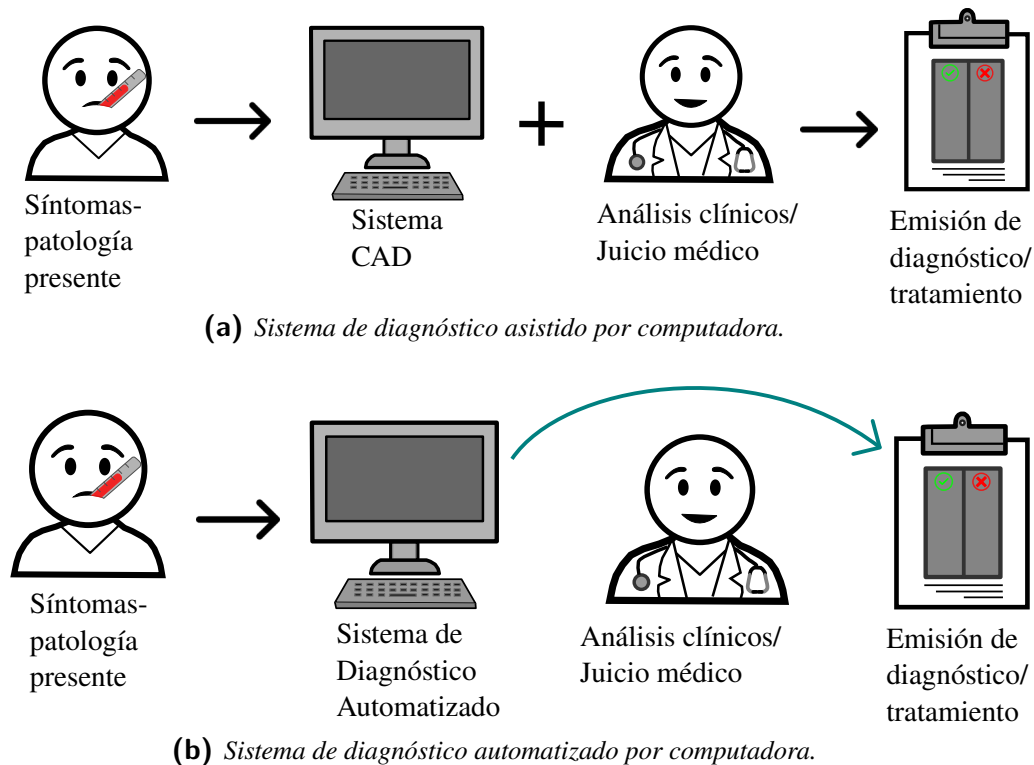


Figura 1.1: Diferencia entre sistemas de Diagnóstico Asistido por Computadora y sistemas de Diagnóstico Automatizado por Computadora.

Existen 3 tipos de sistemas CAD clasificados según su propósito de especialización:

- **CADe:** Enfocado a la detección, utiliza la salida de la computadora (ubicación de regiones sospechosas), para que el profesional realice la caracterización, diagnóstico y manejo del paciente [5].

- **CADx:** Enfocado al diagnóstico, busca ampliar el análisis computacional para caracterizar una región sospechosa o una lesión previamente identificada por un experto o un sistema CAD [5].
- **CADt:** Enfocado al triaje o selección, los cuales buscan establecer prioridades de acción en el tratamiento a realizar, utilizado en su mayoría en lesiones ocurridas en desastres o accidentes para alertar el nivel de urgencia de cierto tratamiento [4].

A pesar de que los orígenes de los sistemas CAD se remontan a la década de 1960, la investigación de sistemas CAD se volvió sistemática en la década de 1980 en la que investigadores de los “Laboratorios Kurt Rossmann para la investigación de imágenes radiológicas” de la Universidad de Chicago iniciaron el desarrollo de sistemas enfocados a 3 proyectos de investigación [6]:

- Detección de grupos de microcalcificaciones en mamografías [7].
- Localización de nódulos pulmonares en radiografías de pecho [8, 9].
- Detección de lesiones que involucran imagenología vascular [10, 11] .

Después de un largo proceso de pruebas y evaluación, el sistema para la detección de lesiones en mamografías se patentó en 1987 y fue el primer sistema aprobado por la Administración de Alimentos y Medicamentos (*Food and Drug Administration*, FDA), que es una agencia gubernamental de los Estados Unidos [12]. Para 2008 ya existían más de 1500 sistemas CAD para el diagnóstico temprano de cáncer de mama [13].

El desarrollo de sistemas CAD aumentó debido a que, si se usan como segunda opinión al momento del diagnóstico, pueden ayudar a evitar errores tanto de detección como de interpretación ocasionados por limitaciones en el sistema humano ojo-cerebro, fatiga del lector y la presencia de estructuras que se superponen y ocultan características importantes para el estudio [5], esto sin olvidar el dominio requerido por parte del lector de la imagen.

Dado que la mayoría de los sistemas CAD funcionan con algoritmos de aprendizaje de máquina es de suma importancia la generación de las bases de datos utilizadas con propósitos de entrenamiento, para las cuales es tan importante la cantidad como la calidad de los datos. Ciertos estudios de imagenología no han obtenido su sistema CAD debido a la falta de bases de datos, ya sea por falta de estandarización de las imágenes o por asuntos de cesión de derechos de las imágenes médicas.

Actualmente los sistemas CAD son cada día más utilizados dentro del proceso de diagnóstico de diversas patologías, funcionando siempre como un segundo espectador. El impacto de los sistemas CAD es tal que en 2016 cerca de 92 % de las lecturas de mamografías involucraban el uso de algún sistema CAD [4].

Debido a los resultados mostrados por sistemas como los mencionados anteriormente, el diagnóstico asistido por computadora se ha extendido a diversos tipos de estudios y modalidades de captura como imagen por resonancia magnética (MRI, por sus siglas en inglés: *Magnetic Resonance Imaging*,) o tomografía por emisión de positrones (PET, por sus siglas en inglés: *Positron Emission Tomography*). Entre las patologías que se beneficiaron de la utilización de sistemas CAD destacan [5]:

- Aneurismas de placa vascular.
- Cáncer de colon.
- Detección de nódulos pulmonares.
- Identificación de tumores en mamografías.
- Osteoporosis.
- Patologías identificadas mediante análisis cefalométrico.

El análisis cefalométrico es una herramienta utilizada en odontología, principalmente para análisis relacionados a ortodoncia. Para la realización de este tipo de análisis es necesario trazar líneas a partir de landmarks o etiquetas cefalométricas que son identificadas en radiografías del cráneo.

El análisis cefalométrico se lleva a cabo con la finalidad de realizar mediciones lineales y angulares que permitan comparar dichos resultados con mediciones que se toman como estándar y emitir un diagnóstico a partir de las diferencias existentes entre las mediciones obtenidas y la plantilla estándar.

Debido a que actualmente los procesos de captura están totalmente estandarizados además de las herramientas con las que cuentan las máquinas digitales de captura, el número de errores en el proceso de captura se ha reducido significativamente, no así en el caso de la identificación de las etiquetas, esto se debe principalmente a tres razones [14]:

- La calidad de la imagen radiográfica (ruido intrínseco al capturar, mal ajuste de contraste, nitidez).
- La precisión con la que están definidas algunas etiquetas (temas de ambigüedad al describir donde se encuentran las etiquetas cefalométricas) así como su reproducibilidad.
- Errores respecto al procedimiento de etiquetaje, fatiga o falta de dominio por parte del etiquetador.

La Figura 1.2 muestra, a manera de introducción, el proceso a seguir para obtener un diagnóstico utilizando análisis cefalométrico. Véase el Capítulo 2, Sección 2.3 para una explicación más detallada acerca de cada una de las etapas que conforman el proceso.

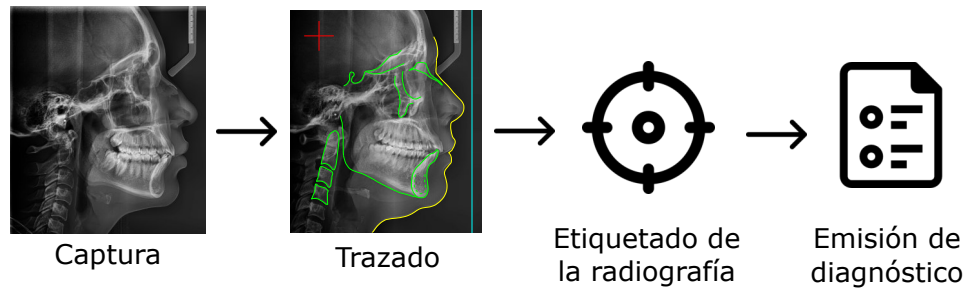


Figura 1.2: Proceso de diagnóstico utilizando análisis cefalométrico.

1.1. Hipótesis General

El uso de redes neuronales convolucionales profundas (DCNN, por sus siglas en inglés: *Deep Convolutional Neural Networks*), entrenadas con la base de datos pública ISBI 2015 (por sus siglas en inglés: *International Symposium on Biomedical Imaging 2015*) y optimizadas para ejecutarse en plataformas de bajo costo como la NVIDIA Jetson Nano, permite la localización automática de etiquetas cefalométricas en radiografías laterales de cráneo con un error radial medio cercano a 3.0 mm, alcanzando un desempeño óptimo con respecto al de métodos reportados en el estado del arte donde se utilizan computadoras de alto desempeño, lo que posibilita su aplicación en sistemas de diagnóstico asistido por computadora (CAD) de bajo costo y alta eficiencia.

1.2. Justificación

Un buen tratamiento médico depende de la velocidad y la precisión con la que se detecta la patología a tratar, por lo que el tiempo que ahorran los sistemas CAD, aún sirviendo como una segunda opinión para el profesional de la salud, es de suma importancia para evitar el avance de ciertas enfermedades degenerativas o que puedan desencadenar otros padecimientos.

Es necesario tratar de aprovechar los avances tecnológicos con los que contamos hoy en día para enfocarlos en problemas que tengan impacto a gran escala, como lo es el problema de cobertura de salud, que requiere de sistemas CAD para el apoyo de los expertos de la salud, que tienen que realizar una gran cantidad de diagnósticos por día, lo que podría reducir la precisión y consistencia en la manera en la que se lleva a cabo el etiquetado de radiografías. También es importante resaltar que la imagenología continúa siendo uno de los pilares actuales de la medicina, de ahí la importancia de generar diagnósticos certeros.

Los sistemas CAD basados en técnicas de aprendizaje profundo han logrado obtener valores equiparables a los de los expertos de la salud para algunas tareas en cuanto a precisión y eficiencia, incluso algunos han alcanzado la aprobación por la FDA debido a su desempeño [15]. Derivado de los

buenos resultados obtenidos y la cantidad masiva de datos para analizar, gobiernos e iniciativa privada han incrementado el apoyo a la implementación de Inteligencia Artificial en el área médica, resultando en un crecimiento de 40 % y un valor estimado de mercado de 6,600 millones de dólares para 2021 [16].

Las mejoras realizadas en el hardware actual, incluyendo los procesadores, unidades procesadoras de gráficos (GPUs, por sus siglas en inglés: *Graphics Processing Units*) y unidades procesadoras de tensores (TPUs, por sus siglas en inglés: *Tensor Processing Units*), así como en la manera en la que se procesan los datos nos permiten ahora utilizar técnicas de aprendizaje profundo para extraer la mayor cantidad de características aún cuando la cantidad de datos disponible no sea la más adecuada para la aplicación, como es el caso del procesamiento de imágenes médicas. La Figura 1.3 muestra el número de publicaciones que involucran el término Computer-Aided Diagnosis desde 2010, que indica una tendencia al alza en los últimos años.

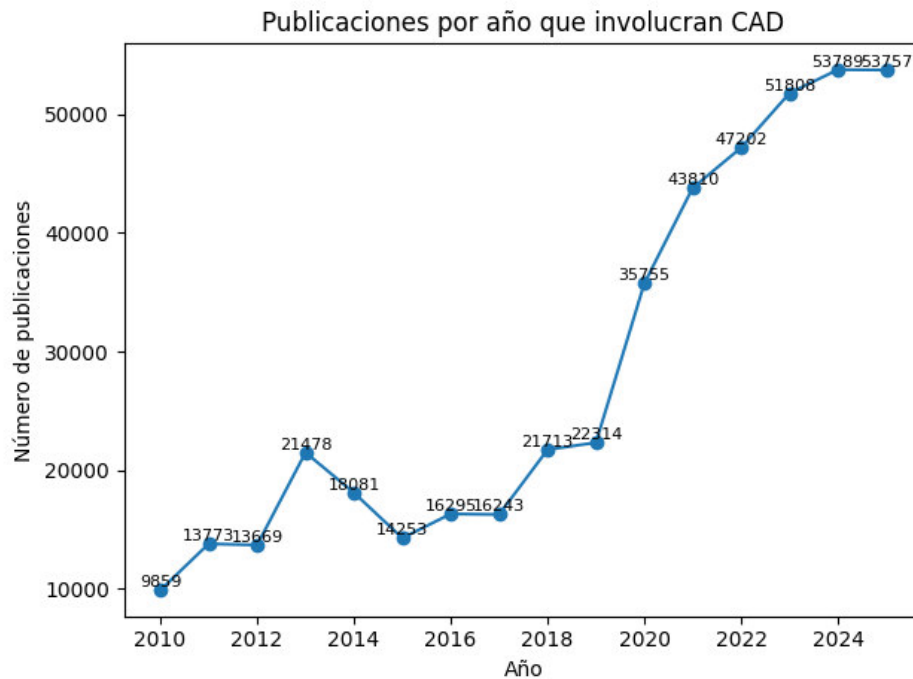


Figura 1.3: Publicaciones realizadas por año desde 2010-2025 que involucran el término Computer-Aided Diagnosis [17].

Es importante recalcar que a pesar de los esfuerzos que se han realizado para crear sistemas CAD, aún con las herramientas y modelos del estado del arte utilizados, todavía no se alcanza la precisión requerida para decir que la tarea del análisis cefalométrico se ha resuelto, por lo que es importante redoblar esfuerzos para crear mejores sistemas a partir de lo que se conoce que funciona y a partir de ahí buscar mejorar.

1.3. Objetivos

1.3.1. Objetivo General

Diseñar e implementar un sistema de localización automática de marcas cefalométricas en radiografías digitales laterales de cráneo mediante técnicas de aprendizaje profundo o también conocidas como redes neuronales convolucionales profundas DCNN, en Python, usando la base de datos ISBI 2015 como conjunto de entrenamiento y validación, y optimizar el modelo para su despliegue en una tarjeta embebida de bajo costo (NVIDIA Jetson Nano), con el fin de generar un prediagnóstico cuya precisión cumpla los límites de error aceptables definidos por la literatura y por las métricas del concurso ISBI 2015.

1.3.2. Objetivos Específicos

- Comprender los fundamentos teóricos de las redes neuronales profundas, sus orígenes, arquitecturas y principales áreas de aplicación, con énfasis en su uso en el campo de la salud.
- Aplicar las redes neuronales convolucionales profundas (DCNN, por sus siglas en inglés: *Deep Convolutional Neural Networks*) al análisis cefalométrico automatizado, contribuyendo al desarrollo de herramientas de apoyo al diagnóstico médico.
- Descargar, analizar y preparar la base de datos pública ISBI 2015, realizando procesos de normalización, aumento de datos, alineación y anotación, con el fin de obtener un conjunto adecuado para el entrenamiento, validación y prueba del modelo.
- Diseñar, entrenar y comparar diferentes arquitecturas de redes DCNN, para la detección y localización automática de puntos cefalométricos en radiografías laterales de cráneo.
- Evaluar el desempeño de los modelos desarrollados mediante métricas estándar como el error radial medio (MRE, por sus siglas en inglés: *Mean Radial Error*), la desviación media, y la tasa de detección exitosa (SDR, por sus siglas en inglés: *Successful Detection Rate*) en umbrales de 2, 2.5, 3 y 4 mm, seleccionando el modelo más preciso y robusto para su uso en una Computadora de Placa Única (SBC, por sus siglas en inglés: *Single-Board Computer*).
- Optimizar el modelo seleccionado para su ejecución en la plataforma embebida "NVIDIA Jetson Nano", mediante técnicas de reducción de parámetros, cuantización, poda y aceleración con TensorRT, buscando alcanzar un desempeño en tiempo real con recursos limitados.
- Implementar un pipeline funcional en la Jetson Nano que permita procesar radiografías, localizar automáticamente las etiquetas cefalométricas y generar un informe de prediagnóstico con la estimación del error asociado.

- Documentar de manera detallada el código, el procedimiento experimental, los resultados obtenidos y las pruebas de robustez realizadas (ruido, variaciones de exposición, resolución), comparando el desempeño frente a resultados de referencia del *ISBI 2015*.
- Sentar las bases para el desarrollo futuro de proyectos académicos enfocados en la aplicación de técnicas de aprendizaje profundo en el área médica, fomentando la continuidad de líneas de investigación en inteligencia artificial aplicada a la salud.
- Desarrollar una interfaz gráfica de usuario médica que permita visualizar los resultados del modelo y sirva como base para la creación futura de una base de datos pública de imágenes y etiquetas cefalométricas.

1.4. Descripción del Proyecto

Las fases esenciales para el desarrollo de un proyecto de aprendizaje automático se seleccionaron de acuerdo con lo propuesto por [15, 18]. Estas fases sirvieron como guía para estructurar el plan de trabajo y definir el alcance del presente proyecto. A continuación, se describen las etapas consideradas:

1. **Identificación y comprensión del problema:** En esta fase se analizó la relevancia del análisis cefalométrico en el diagnóstico ortodóntico, así como la necesidad de sistemas automáticos que permitan la detección y localización de puntos de referencia de manera precisa y eficiente. Las ventajas clínicas y tecnológicas de este tipo de sistemas se discuten con mayor detalle en el Capítulo 2 en la Sección 2.5, donde se realizó un estudio extenso acerca de las soluciones propuestas para la identificación automática de etiquetas cefalométricas por científicos desde el año 1999.
2. **Selección de la solución potencial:** En esta etapa, se revisaron los enfoques existentes empleados para resolver problemáticas similares, tales como la detección automática de landmarks faciales y la estimación de pose. En los últimos años, las redes neuronales convolucionales (CNN, por sus siglas en inglés: *Convolutional Neural Networks*) se han consolidado como la herramienta más eficaz para abordar tareas relacionadas con el procesamiento de imágenes. Los fundamentos teóricos de este enfoque se abordan en el Capítulo 4.
3. **Recolección de datos:** Se realizó la búsqueda e investigación de bases de datos públicas relacionadas con el análisis cefalométrico, considerando su disponibilidad, volumen de imágenes, anotaciones y formato. El conjunto de datos seleccionado, proveniente del *ISBI Challenge 2015*, se describe en detalle en la Sección 5.4, donde se presentan sus principales características, así como el procedimiento de generación y liberación de dicha base.

4. **Preprocesamiento de los datos:** De acuerdo con la naturaleza de las imágenes y los requisitos del modelo, se aplicaron diversas transformaciones para facilitar su uso en el sistema de aprendizaje profundo. Entre las operaciones realizadas se incluyen la normalización, recorte, reducción de ruido, aumento de datos (*data augmentation*) y ajustes de brillo o contraste, con el objetivo de mejorar la robustez del modelo y reducir el riesgo de sobreajuste.
5. **Desarrollo del modelo:** Con base en la hipótesis de solución planteada, se diseñó y entrenó una arquitectura de red neuronal convolucional profunda para la localización automática de etiquetas cefalométricas. Para ello, se realizaron la separación de los conjuntos de entrenamiento, validación y prueba, así como la selección de las métricas de evaluación más adecuadas. El proceso de desarrollo se detalla en el Capítulo 5.
6. **Evaluación del modelo:** Una vez entrenada la red, se evaluó su desempeño utilizando métricas cuantitativas estándar, tales como el error radial medio y la tasa de detección exitosa en distintos umbrales. En esta etapa se identificaron posibles áreas de mejora, como la ampliación del conjunto de datos o la optimización de hiperparámetros. Los resultados obtenidos se presentan en el Capítulo 6.
7. **Despliegue del modelo final:** Finalmente, se preparó el modelo para su implementación en una plataforma embebida de bajo costo, específicamente la *NVIDIA Jetson Nano*, con el objetivo de generar un sistema de asistencia al diagnóstico (CAD) funcional. Se exploraron posibles optimizaciones para mejorar el desempeño del sistema y se discutieron futuras líneas de desarrollo en el Capítulo 6.

1.5. Organización de la Tesis

Este trabajo está organizado de la siguiente manera:

- **Capítulo 1. Introducción:** Se presenta una descripción general del trabajo desarrollado, el enfoque seguido y una breve introducción a los Sistemas de Diagnóstico Asistido por Computadora. Asimismo, se exponen los objetivos y la justificación del proyecto.
- **Capítulo 2. Análisis Cefalométrico:** Este capítulo aborda los fundamentos teóricos del análisis cefalométrico, su evolución histórica, el proceso de captura y trazado, así como las principales etiquetas cefalométricas empleadas. Además, se presenta el estado del arte y las técnicas más relevantes aplicadas en este campo.
- **Capítulo 3. Redes Neuronales:** Se introduce el concepto de redes neuronales artificiales, sus antecedentes y principios de funcionamiento. Se explican las bases del aprendizaje en redes

neuronales, las principales funciones de activación, los métodos de entrenamiento y optimización, así como los problemas de sobreajuste y regularización. El capítulo concluye con una descripción de las arquitecturas más representativas.

- **Capítulo 4. Redes Neuronales Convolucionales:** Se presentan la inspiración biológica, los fundamentos teóricos y matemáticos de las redes neuronales convolucionales (CNN), describiendo la operación de convolución y las capas que conforman este tipo de modelos. También se discuten sus principales aplicaciones, con énfasis en el ámbito médico.
- **Capítulo 5. Metodología:** Este capítulo describe el proceso de diseño e implementación de la arquitectura convolucional propuesta para la detección automática de puntos cefalométricos. Se detalla la base de datos empleada, las métricas de evaluación, el entorno de desarrollo, así como la arquitectura de red utilizada y los parámetros utilizados durante el entrenamiento y validación del modelo.
- **Capítulo 6. Resultados y Conclusiones:** Se presentan los resultados obtenidos al evaluar la arquitectura propuesta, comparando el desempeño del modelo con el de otros enfoques reportados. Finalmente, se exponen las conclusiones generales del trabajo y se plantean posibles líneas de investigación futura orientadas al mejoramiento del sistema CAD desarrollado.

1.6. Conclusiones de Capítulo

En este primer capítulo se establecieron los fundamentos que motivan el desarrollo de un sistema de localización automática de etiquetas cefalométricas en radiografías digitales laterales de cráneo mediante el uso de redes neuronales convolucionales profundas. Se expuso la relevancia de automatizar este proceso, considerando que el análisis cefalométrico manual requiere de tiempo, precisión y experiencia clínica, además de estar sujeto a variabilidad inter e intraobservador.

Asimismo, se presentó la justificación que sustenta la aplicación de técnicas de aprendizaje profundo en el campo de la ortodoncia digital, destacando su potencial para mejorar la exactitud y reproducibilidad del diagnóstico. El planteamiento de los objetivos generales y específicos permitió delimitar el alcance del trabajo, orientando los esfuerzos hacia el diseño, implementación y validación de un modelo eficiente capaz de ejecutarse en una plataforma de bajo costo como la NVIDIA Jetson Nano.

Por último, se describió la estructura general de la tesis, la cual integra tanto los fundamentos teóricos del análisis cefalométrico y las redes neuronales, como la metodología empleada para el entrenamiento, validación y evaluación del modelo propuesto. Con ello, se sientan las bases conceptuales y técnicas que servirán de guía para los capítulos posteriores, donde se abordará en detalle el desarrollo y los resultados del sistema de detección automática.

Capítulo 2

Análisis Cefalométrico

2.1. Introducción

Como se mencionó anteriormente, el Análisis Cefalométrico constituye una herramienta indispensable para tratamientos relacionados con ortodoncia. Dicha técnica utiliza radiografías del cráneo para la medición de líneas rectas y mediciones angulares, y de esta manera, estudiar la anatomía y proporciones de la cabeza humana. Posteriormente dicho estudio será interpretado para tomar decisiones clínicas respecto al tratamiento del paciente.

El objetivo del análisis cefalométrico es comparar las medidas obtenidas del paciente contra valores estándar [19]. Este capítulo trata de explicar los aspectos principales del análisis cefalométrico, desde su historia y aplicaciones hasta los métodos actuales utilizados para llevarlo a cabo. De igual manera se presenta la revisión bibliográfica referente a los sistemas CAD aplicados al análisis cefalométrico.

2.2. Historia de la Cefalometría

La cefalometría tiene sus orígenes en la craneometría, que es la medición del cráneo y se remontan al siglo IV a.C. Dada la metodología y logros que alcanzó la craneometría, algunas de las etiquetas, planos y ángulos usados en cefalometría son los mismos que los que se usaban anteriormente en craneometría [20].

Un hecho que potencializó el desarrollo del análisis cefalométrico fue el descubrimiento de los rayos X por Wilhem Conrad Roentgen en 1895. Después de este hecho, surgieron nuevas técnicas para obtener radiografías del cuerpo humano, sin embargo, no fue hasta 1922 que Pacini y Carrera tomaron las primeras radiografías del cráneo en vista lateral de una forma estandarizada [14], dicha posición era alcanzada utilizando un vendaje de gasa que fijaba la cabeza contra la película receptora.

A pesar de los esfuerzos por tomar radiografías de una manera estandarizada, los métodos utilizados hasta ese entonces no eran tan reproducibles (Percy Brown en 1921 y A.J. Pacini en 1922) lo que impedía que los análisis para comparar el crecimiento no prosperaran. Fue hasta 1931 que B.H. Broadbent publicó en Estados Unidos los resultados de un aparato desarrollado un año antes. Al mismo tiempo pero en Alemania, H. Hofrat llegó al mismo resultado: Un aparato llamado cefalostato que permitía estandarizar la posición de la cabeza al momento de tomar radiografías laterales [19, 20], lo cual favoreció el desarrollo de comparaciones de estudios de la misma cabeza a lo largo del tiempo.

El cefalostato de Broadbent-Hofrat se compone de dos barras o extensiones que se introducen en el meato auditivo externo además de una guía para ubicar el punto anatómico conocido como “Nasion”. Posterior a la invención y utilización del cefalostato, se resolvió el problema de estandarización en cuanto a la distancia a la que se tomaban las radiografías para compararlas sin importar el lugar o equipo con el que se tomaron.

Una vez que se estandarizó el proceso de captura de las radiografías, surgieron diversos estudios que no fueron exitosos por diversos motivos. Fue hasta el año 1948 que Willian B. Downs publicó un análisis extenso con diversas mediciones realizadas para caracterizar la cara [21]. A través de los años se fueron incorporando nuevos métodos de análisis haciendo uso de herramientas como el cefalostato y mediciones para realizar análisis comparativos [22].

Después de algunos años de mejorar los procesos de medición lineal y angular para hacer análisis comparando radiografías, se enfatizó la necesidad de una definición clara y reproducible de las etiquetas cefalométricas para realizar un mejor análisis de crecimiento (análisis para una sola persona) y para la comparación de valores de etiquetas en diversas personas siguiendo el mismo tratamiento.

Dada la diversidad de estudios dentro del análisis cefalométrico (utilizando radiografías laterales o posteroanteriores, cada una con sus variaciones), muchas de las veces es necesario utilizar un tipo específico de radiografía. Para el presente trabajo se utilizan Radiografías Laterales de Cráneo (*LCR*, por sus siglas en inglés *Lateral Cranes Radiography*), también conocidas como cefalogramas, que es una imagen 2D del cráneo en vista lateral y permite establecer relaciones entre dientes, huesos craneales, tejido suave entre otros. La Figura 2.1 muestra un ejemplo de radiografía lateral de cráneo.

2.3. Proceso de Captura

Una parte importante para que el análisis cefalométrico sea exitoso, es una radiografía de calidad, que se adapte a los estándares para ser etiquetada y que permita una interpretación correcta y, por ende, un tratamiento correcto. La máquina de captura deberá de calibrarse para obtener una imagen de buena calidad y nitidez, es necesario recordar que al tratar de proyectar un objeto 3D como lo es la cabeza

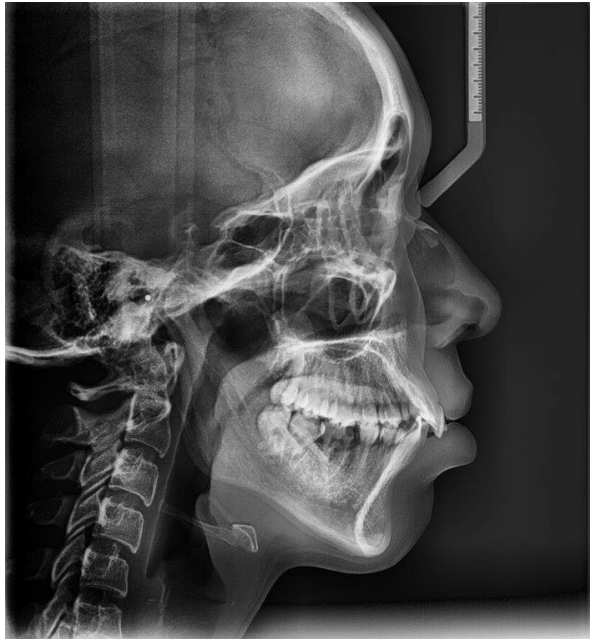


Figura 2.1: Ejemplo del tipo de radiografía utilizada para el desarrollo de este trabajo.

humana sobre una imagen 2D existirá un poco de distorsión debido a la manera en la que actúan los rayos X sobre la película receptora.

A continuación se presenta la manera en la que se captura la radiografía lateral de cráneo de acuerdo con [20]:

1. La captura se puede realizar con el paciente de pie o sentado.
2. La cabeza del paciente se posiciona con el lado izquierdo de la cara cercano al receptor de la imagen. La distancia entre el receptor y el plano sagital medio debe de ser de 15 cm.
3. La cabeza se coloca en la posición estándar utilizando el cefalostato.
4. Se ajusta la fuente emisora de radiación tal que la distancia entre el aparato y el plano sagital medio sea de 152 cm (aproximadamente 60 in).
5. Se mantiene la posición alcanzada mientras se captura la radiografía (el tiempo dependerá del tipo de captura que se haga, analógica o digital). La Figura 2.2 muestra una vista simplificada con las distancias necesarias para capturar una radiografía lateral de cráneo.

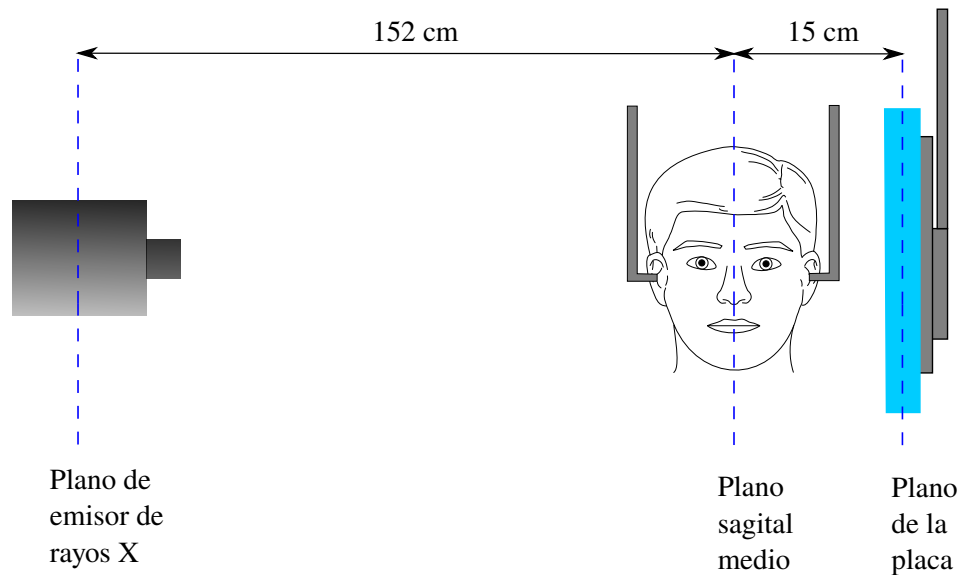


Figura 2.2: Medidas estándar al momento de la captura de una radiografía lateral de cráneo.

2.4. Proceso de Trazado y Etiquetado

Antes de comenzar a identificar las etiquetas cefalométricas necesarias para el análisis a realizar, es necesario cerciorarse que la radiografía está bien tomada y que el paciente no se movió durante el proceso de captura, esto evitará problemas al momento de identificar las etiquetas.

El primer paso en el proceso de trazado es colocar la imagen capturada en una película de celuloide sobre una caja de luz, también conocida como “negatoscopio” que permita apreciar mejor la radiografía o en la computadora si el proceso de trazado será digital. El procedimiento manual se realiza sobre una lámina de acetato o papel vegetal para no dañar la radiografía de cara a estudios posteriores.

Algunas veces se trazan una o dos marcas de referencia (+ o X) tanto en la radiografía original como en el acetato, esto con la finalidad de servir como guías para alinear la radiografía y el acetato, si es que se requiere. También se traza una línea vertical paralela al borde derecho del cefalograma para asegurarse que el soporte de la radiografía es completamente perpendicular al suelo.

Posterior a ello, se realiza el trazado del tejido blando comenzando por parte alta de la frente y las estructuras duras empezando por la frente o las cervicales. Para una identificación más rápida de algunas estructuras del tejido blando y duro, existen reglas de Ricketts y plantillas comerciales que ayudan a resolver esta tarea. La Figura 2.3 muestra una radiografía trazada con el perfil de tejido blando, así como algunas estructuras esqueléticas importantes.

Una vez que el cefalograma está trazado con las estructuras esqueléticas y de tejido blando necesarias para ejecutar el análisis cefalométrico seleccionado, se procede con el etiquetaje de la radiografía ya sea

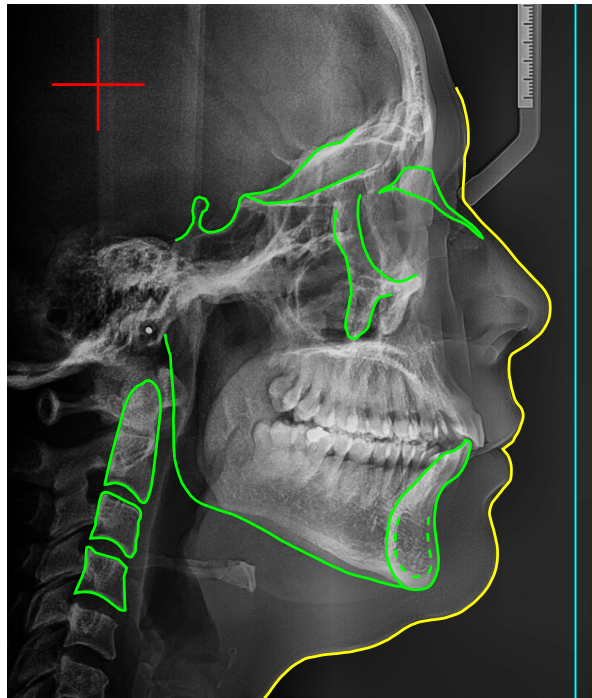


Figura 2.3: Trazado de marcas de referencia y algunas estructuras de tejido blando y duro. El número y tipo de estructuras a etiquetar dependerá del tipo de análisis a realizar.

de manera manual o digital. En algunas ocasiones se escanea el cefalograma tomado de manera manual en formato DICOM (*Digital Imaging and Communications in Medicine*, por sus siglas en inglés) para evitar pérdida de información.

2.4.1. Etiquetas Cefalométricas

Las etiquetas cefalométricas, también conocidas como *landmarks*, son el componente principal del análisis cefalométrico, puesto que permiten establecer relaciones cuantitativas entre etiquetas cefalométricas para evaluar avances de ciertas patologías, así como la planeación de tratamiento maxilofacial.

Una etiqueta cefalométrica es un punto anatómico bien definido que se ubica sobre la radiografía previamente trazada y que ayuda a establecer relaciones medibles entre estructuras dentales y esqueléticas para realizar el análisis pertinente. A partir de la unión de etiquetas cefalométricas se crean las líneas que permiten la medición lineal y angular con propósitos de comparación con las medidas consideradas como estándar.

Las etiquetas cefalométricas son tan variadas como el número de análisis cefalométricos; es posible añadir o crear nuevas etiquetas siempre y cuando cumpla con dos condiciones [23]: Fácil de ubicarla, con

esto se busca que factores como la calidad de la radiografía o contornos anatómicos que se traslapan en la imagen tengan la menor influencia posible, y que la ubicación de la etiqueta se vea lo menos comprometida por aspectos de sexo, raza o edad.

Es importante hacer la distinción entre los tres tipos de etiquetas cefalométricas que existen:

1. **Etiquetas anatómicas:** Se ubican sobre estructuras esqueléticas. Normalmente este tipo de etiquetas se ubican en estructuras óseas puntiagudas o con una estructura curva.
2. **Etiquetas antropológicas:** Muchas de estas etiquetas se mantienen desde la craneometría, se ubican dentro de estructuras esqueléticas.
3. **Etiquetas radiológicas:** También conocidas como puntos construidos, se encuentran en la intersección de líneas trazadas previamente en la radiografía.

Una vez que se conoce la definición y tipos de etiquetas cefalométricas se mostrarán las 19 etiquetas a utilizar en este trabajo. La Figura 2.4 y las Tablas 2.1 y 2.2 muestran la ubicación dentro de una LCR, así como el nombre y definición de las etiquetas a utilizar y planos que surgen a raíz de las etiquetas mencionadas.

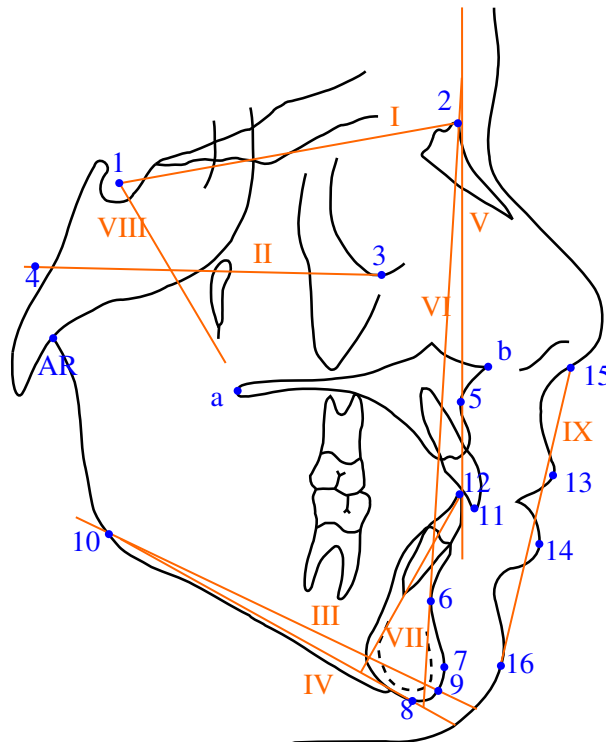


Figura 2.4: Ubicación en una LCR de las etiquetas cefalométricas a utilizar, así como algunos planos que se relacionan con las mismas, adaptada de [24].

Tabla 2.1: Descripción de planos cefalométricos que se pueden trazar a partir de las etiquetas cefalométricas de la Figura 2.4, adaptada de [25, 20, 26].

Identificador	Plano	Descripción
I	Línea Sella-Nasion	Línea que une la etiqueta sella turcica con nasion. Utilizada para representar la base craneal anterior, con la que se pueden relacionar los dientes con la mandíbula.
II	Plano horizontal de Frankfurt	Línea que une las etiquetas Porion con Orbitale. Decretado como posición anatómica del cráneo humano.
III	Plano mandibular de Steiner	Línea que une las etiquetas Gonion y Gnathion.
IV	Plano mandibular de Tweed	Línea que une las etiquetas Gonion y Menton. Por definición es tangente al borde inferior de la mandíbula.
V	Línea A Nasion	Línea que une la etiqueta Nasion con la etiqueta Subspinale (punto A).
VI	Línea B Nasion	Línea que une la etiqueta Nasion con la etiqueta Supramentale (punto B).
VII	Eje incisivo inferior central	Línea que va desde el incisivo inferior hasta el plano mandibular de Tweed.
VIII	Línea Sella turcica-gnathion	También conocida como eje Y, une el centro de la etiqueta sella turcica con gnathion y forma un ángulo con el plano de Frankfurt.
IX	Línea estética de Steiner	Une la etiqueta Point PM (punta de la nariz) con Soft Tissue Pogonion.

Es importante recordar que la calidad del análisis cefalométrico depende directamente de la precisión con la que se establecen las etiquetas cefalométricas, ya que los resultados obtenidos constituyen la base para futuros estudios o tratamientos. Por ello, resulta fundamental contar con sistemas de diagnóstico asistido por computadora (CAD) que posean un alto grado de reproducibilidad.

Asimismo, es necesario destacar la relevancia del aval médico en la definición de estas etiquetas, puesto que los sistemas desarrollados mediante métodos supervisados o semisupervisados dependen, en gran medida, de las anotaciones realizadas por especialistas durante la fase de entrenamiento.

A continuación, se presenta el estado del arte, en el cual se analizan los principales avances, enfoques metodológicos y modelos propuestos en la literatura especializada, orientados a la detección automática de puntos cefalométricos mediante el uso de técnicas de aprendizaje profundo y visión por computadora.

Tabla 2.2: Listado y descripción de las 19 etiquetas cefalométricas utilizadas para el desarrollo del presente trabajo, adaptada de [14, 19, 25, 27, 28].

Identificador	Etiqueta	Descripción
1	Sella turcica	Etiqueta que representa el punto medio de la fosa pituitaria.
2	Nasion (N)	Se ubica en la unión de los huesos nasal y frontal en el plano medio. N se ubica en el punto de máxima convexidad entre la nariz y la frente.
3	Orbitale (Or)	Punto más inferior sobre el margen orbital en la radiografía.
4	Porion (P)	El punto más alto en el margen del meato auditorio externo.
5	Subspinale	También conocido como punto A, se ubica en la concavidad más profunda sobre el perfil anterior de la mandíbula.
6	Supramentale	También conocido como punto B, es el punto más adelantado en la concavidad sobre la superficie anterior de la mandíbula en el plano medio.
7	Pogonion (Pog)	El punto más frontal en el hueso de la barbilla.
8	Menton (Me)	Ubicado en el punto más bajo de la mandíbula.
9	Gnathion (Gn)	Ubicado en el plano medio, es el punto más inferior y frontal en la sínfisis mandibular.
10	Gonion (Go)	El punto trasero más inferior en el ángulo de la mandíbula.
11	Lower incisal incision	Punta de la corona del incisivo mandibular central más frontal.
12	Upper incisal incision	Punta de la corona del incisivo mandibular central más frontal.
13	Upper lip	Etiqueta de tejido blando, indica la posición del labio superior.
14	Lower lip	Etiqueta de tejido blando, indica la posición del labio inferior.
15	Subnasale	Punto ubicado donde la curvatura del borde frontal de la sínfisis cambia de cóncavo a convexo.
16	Soft tissue pogonion	Punto más frontal de la piel de la barbilla. Ubicado en el plano medio.
a	Posterior nasal spine	Ubicado en la intersección entre la continuación del muro frontal de la fosa pterigopalatina y el fondo de la nariz.
b	Anterior nasal spine	Ubicado en el plano medio, es el punto más frontal de la piel de la barbilla en el plano medio.
AR	Articulate	La intersección del margen posterior de la rama ascendente y el margen exterior de la base craneal.

2.5. Estado del Arte

La detección automática de puntos de referencia cefalométricos se ha convertido en un área fundamental de investigación que combina la imagenología médica, la visión artificial y la ortodoncia. La identificación de puntos de referencia cefalométricos es fundamental en el campo de la ortodoncia, ya que contribuye significativamente al proceso de diagnóstico y planificación de tratamientos, así como a la investigación sobre aspectos craneofaciales. Esta práctica tiene el potencial de mejorar la toma de decisiones clínicas y, en última instancia, mejorar los resultados para los pacientes.

La detección de puntos de referencia cefalométricos es un aspecto fundamental del tratamiento de ortodoncia y la cirugía maxilofacial, y se utiliza para evaluar la posición y el crecimiento de los maxilares

y los dientes. Los puntos de referencia cefalométricos se refieren a puntos anatómicos distintivos del cráneo que sirven como indicadores para evaluar las relaciones esqueléticas y dentales, así como los patrones de crecimiento. Para descartar anomalías esqueléticas o dentales y evaluar las características del crecimiento craneofacial, se evalúan la distancia y las desviaciones angulares entre los puntos de referencia y sus valores normativos.

La detección precisa y fiable de puntos de referencia a partir de cefalogramas es crucial para la planificación del tratamiento, el diagnóstico y la predicción de resultados. Las variaciones anatómicas, los problemas de calidad de la imagen o la superposición de estructuras pueden dificultar la identificación precisa de ciertos puntos de referencia, lo que podría comprometer la fiabilidad de las mediciones y los análisis cefalométricos [29].

La variabilidad interobservador e intraobservador en la identificación manual de puntos de referencia puede generar inconsistencias en la identificación de puntos de referencia entre profesionales y para el mismo profesional a lo largo del tiempo, lo que podría socavar la fiabilidad de las mediciones y los análisis derivados de las radiografías cefalométricas.

Esta variabilidad refleja la magnitud de los errores en la identificación de puntos de referencia y subraya la importancia de comprender la distribución espacial de los errores en relación con cada punto de referencia específico. Los errores en la identificación manual de puntos de referencia pueden deberse a factores humanos como la fatiga, la distracción y sesgos inherentes, lo que puede resultar en imprecisiones en los análisis cefalométricos y la posterior planificación del tratamiento [30].

Esta tarea, que requiere mucho tiempo, especialmente con numerosas radiografías, puede aumentar la carga de trabajo clínico y causar retrasos en la planificación del tratamiento y la toma de decisiones. La aplicación de procedimientos de rastreo digital está sustituyendo cada vez más a los métodos de rastreo manual como medio para mejorar la eficiencia y agilizar la operación. El uso de la radiografía digital y la conversión de películas cefalométricas tradicionales a formato digital presenta diversas ventajas, como una mayor usabilidad, un mejor rendimiento gracias a la capacidad de realizar múltiples análisis simultáneos, predicciones automatizadas de tratamientos y un almacenamiento eficiente [31]. El uso de cefalogramas digitales para la captura directa de imágenes ofrece varios beneficios a los pacientes, como la reducción de la exposición a la radiación, la reducción del uso de productos químicos y la mitigación de los problemas ambientales asociados [32].

El uso de algoritmos de Aprendizaje Profundo (DL, por sus siglas en inglés: *Deep Learning*) y Aprendizaje Automático o Aprendizaje Máquina (ML, por sus siglas en inglés: *Machine Learning*) ha facilitado la aparición de sistemas automatizados capaces de detectar y localizar eficazmente puntos de referencia cefalométricos. Este avance tecnológico ha resultado en una reducción del error humano, así

como en una mejora significativa de la eficiencia y la velocidad del proceso. Sin embargo, la detección automática de puntos de referencia aún presenta algunos problemas, como la necesidad de conjuntos de datos extensos y diversos para entrenar eficazmente los modelos de aprendizaje automático, así como la necesidad de que los algoritmos muestren resiliencia ante las variaciones en la calidad de la imagen.

Los sistemas informáticos ahora son capaces de detectar y localizar puntos de referencia cefalométricos con precisión y rapidez, gracias al uso de métodos de IA como las redes neuronales recurrentes (RNN, por sus siglas en inglés; *Recurrent Neural Networks*) y las redes neuronales convolucionales [33]. Estos algoritmos son capaces de aprender de grandes conjuntos de datos de imágenes cefalométricas, lo que les permite identificar puntos de referencia en tiempo real con una precisión excepcional. Además, las técnicas de Inteligencia Artificial (IA), como el aprendizaje por transferencia [34], el aumento de datos y los métodos de ensembles, también han contribuido a mejorar la detección de puntos de referencia cefalométricos.

La adopción de estos procedimientos ofrece un potencial considerable para mejorar la precisión diagnóstica y la planificación del tratamiento en ortodoncia, mejorando así la precisión y la eficiencia del tratamiento, lo que supone un avance notable en la atención odontológica con el potencial de elevar los resultados de los pacientes y los estándares de atención médica.

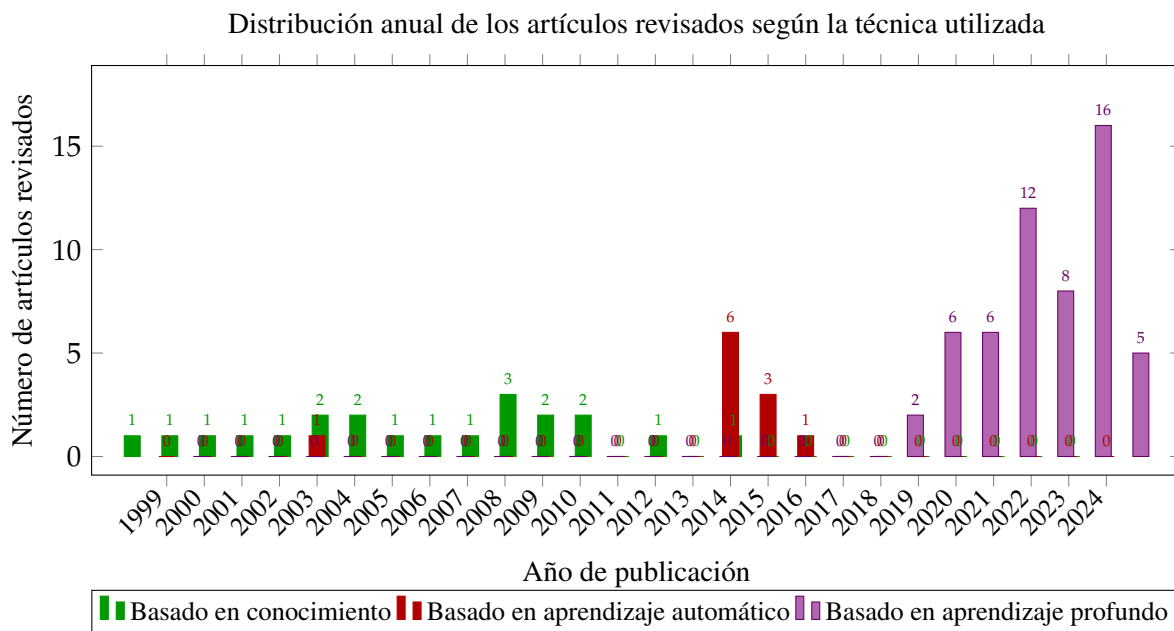


Figura 2.5: Distribución anual de los artículos revisados basados en conocimiento, aprendizaje automático y aprendizaje profundo.

2.5.1. Tipos de Imagen Utilizadas para el Análisis Cefalométrico

La detección automática de puntos de referencia cefalométricos utiliza diversas modalidades de imagen, que abarcan tanto la radiografía bidimensional (2D) estándar como las técnicas avanzadas de imagen tridimensional (3D). Las radiografías ofrecen una representación bidimensional de las estructuras craneofaciales, lo que puede dificultar la evaluación precisa de la profundidad y las relaciones espaciales entre los puntos de referencia anatómicos.

La superposición de estructuras de diferentes planos en las radiografías 2D dificulta la distinción precisa de los puntos de referencia anatómicos individuales, y la visualización limitada de tejidos blandos como músculos, cartílagos y paladar blando dificulta un análisis cefalométrico completo, especialmente en ortodoncia y cirugía ortognática. Además, la distorsión y la magnificación en las radiografías, a menudo debidas a una posición incorrecta del sujeto o a variaciones en la geometría del haz, pueden comprometer la precisión de la medición. A pesar de las bajas dosis de radiación, la exposición repetida y el uso de radiación ionizante presentan riesgos, especialmente para los pacientes pediátricos [35].

Las modalidades de imagen 2D más utilizadas en el análisis cefalométrico son las radiografías de cráneo laterales y posteroanteriores (PA). Las radiografías laterales proporcionan una vista lateral de la cabeza y se utilizan para evaluar las relaciones esqueléticas sagitales, mientras que las cefalogramas PA proporcionan una vista frontal de la cabeza y se utilizan para evaluar las relaciones esqueléticas transversales. Las modalidades de imagen 3D proporcionan datos volumétricos que pueden utilizarse para identificar puntos de referencia cefalométricos con mayor precisión. La Tomografía Computarizada (CT, por sus siglas en inglés: *Computer Tomography*), la resonancia magnética (MRI, por sus siglas en inglés: *Magnetic resonance imaging*) y la tomografía computarizada de haz cónico (CBCT, por sus siglas en inglés: *Cone Beam Computed Tomography*) son técnicas de imagen tridimensional ampliamente utilizadas en el campo de la ortodoncia.

La Academia Americana de Radiología Oral y Maxilofacial (AAOMR, por sus siglas en inglés, *American Academy of Oral & Maxillofacial Radiology*) ha considerado el uso de la CBCT en odontología como una modalidad de imagen prometedora [36].

La CBCT emplea una configuración cónica de haz de rayos X para obtener representaciones visuales tridimensionales de la región craneal del paciente, ofreciendo así una resolución espacial mejorada y al mismo tiempo, mitigando la exposición a la radiación, a diferencia de las técnicas de CT convencionales. La CT emplea una secuencia de Rayos-X para generar una representación tridimensional de la región craneal del paciente. En consecuencia, la cefalografía 3D CBCT surgió como la aplicación más destacada de la tecnología CBCT [37].

La CT se correlaciona con una mayor exposición a la radiación en comparación con la CBCT y se utiliza comúnmente en situaciones que requieren altos niveles de precisión. El uso de las Imágenes de Resonancia Magnética (MRI, por sus siglas en inglés) en odontología ofrece un método seguro para obtener imágenes tridimensionales sin el uso de radiación ionizante. Sin embargo, su aplicación se ve limitada por el elevado coste de la RM (Resonancia Magnética), lo que restringe su uso a casos específicos donde se requiere explícitamente para un diagnóstico preciso [38].

Este método de imagen destaca por su excelente resolución espacial, su alta resolución de contraste en tejidos blandos, su especificidad y su excelente penetración en profundidad [39]. La obtención de imágenes por MRI de tejidos duros dentales y hueso es difícil debido a su baja concentración de protones y a sus cortas duraciones de relajación transversal (T2) [40]. Debido a la necesidad de una duración de examen concisa y un enfoque bien tolerado, el estudio sobre la aplicación de la RM en la cefalometría 3D ha sido limitado [41]. Posiblemente una exposición excesiva a la radiación pueda conducir al desarrollo de cáncer [42].

La mayoría de los experimentos analizados en este trabajo utilizan la modalidad de imágenes por rayos X, mientras que un subconjunto más pequeño emplea las modalidades de imágenes CBCT y CT. La Tabla 2.3 proporciona una descripción detallada de los diversos métodos de imágenes cefalométricas, incluidas sus respectivas ventajas y desventajas.

2.5.2. Técnicas de Detección de Puntos de Referencia Cefalométricos Empleados en la Literatura

Se han desarrollado diversas técnicas para la detección automática de puntos de referencia cefalométricos.

Estas técnicas incluyen la coincidencia de plantillas, la detección de bordes, métodos de aprendizaje automático o también conocido como Aprendizaje Máquina, (ML, por sus siglas en inglés), modelos estadísticos de forma, modelos deformables, métodos de aprendizaje profundo (Deep-Learning ,DL), métodos de ensembles y enfoques híbridos [46].

Cada técnica presenta sus propias ventajas y limitaciones, y la selección de una metodología depende de diversos aspectos, como la complejidad de la tarea, la accesibilidad de los datos y los recursos computacionales disponibles. Sin embargo, aún existen desafíos, como la variabilidad de la imagen, las oclusiones y las configuraciones inusuales de puntos de referencia, que requieren investigación y desarrollo continuos de algoritmos robustos y conjuntos de datos de entrenamiento. La clasificación básica de las técnicas de detección automática de puntos de referencia cefalométricos se muestra en la Figura 2.6.

Tabla 2.3: Comparación de modalidades de imagen cefalométrica: ventajas y desventajas.

Modalidad de imagen	Ventajas	Desventajas
Radiografías de cráneo en 2D	■ Bajo costo y facilidad de uso ■ Exposición mínima a radiación ■ Adecuada para planificación ortodóntica rutinaria ■ Simplifica el análisis cefalométrico	■ Capacidad limitada para mostrar detalles anatómicos ■ Errores por la naturaleza 2D de las imágenes ■ Dificultad para ubicar puntos de referencia en profundidad ■ La posición del paciente afecta la precisión
TC Cefalométrica 3D [43]	■ Visualización 3D de estructuras craneofaciales ■ Útil en casos complejos o asimétricos ■ Imágenes de alta resolución mejoran la precisión ■ Permite predecir resultados y mejorar la planificación	■ Mayor exposición a radiación ■ Más costosa que otras modalidades ■ Tamaño de archivo grande requiere más almacenamiento ■ Artefactos por metales pueden afectar referencias ■ Requiere habilidades especializadas
CBCT Cefalométrica 3D [44]	■ Proporciona detalles volumétricos y espaciales ■ Menor distorsión y mejor detección de referencias ■ Más segura que la TC ■ Eficiente para casos complejos/asimétricos ■ Permite detección automática para análisis rápido	■ Costosa y no siempre accesible ■ Archivos grandes requieren almacenamiento adicional ■ Artefactos metálicos pueden afectar mediciones ■ Requiere conocimientos especializados
Imagen Cefalométrica por RM [45]	■ Visualización precisa de tejidos blandos ■ Imágenes multiplanares permiten localización precisa ■ Segura por no usar radiación ionizante ■ Alta resolución de contraste ■ Proporciona datos cuantitativos de tejidos	■ Alto costo y accesibilidad limitada ■ Menor detalle óseo comparado con TC/CBCT ■ Requiere conocimientos avanzados ■ Imágenes grandes requieren gestión de datos compleja ■ Sensible a artefactos por movimiento o metales

2.5.3. Enfoques Basados en el Conocimiento

Un sistema basado en el conocimiento aprovecha la información actual de un sistema dado para resolver problemas complejos mediante el desarrollo de una base de conocimiento. Los métodos de coincidencia (o emparejamiento) de plantillas comparan plantillas predefinidas de puntos de referencia con regiones de la imagen para encontrar la mejor coincidencia basándose en medidas de similitud.

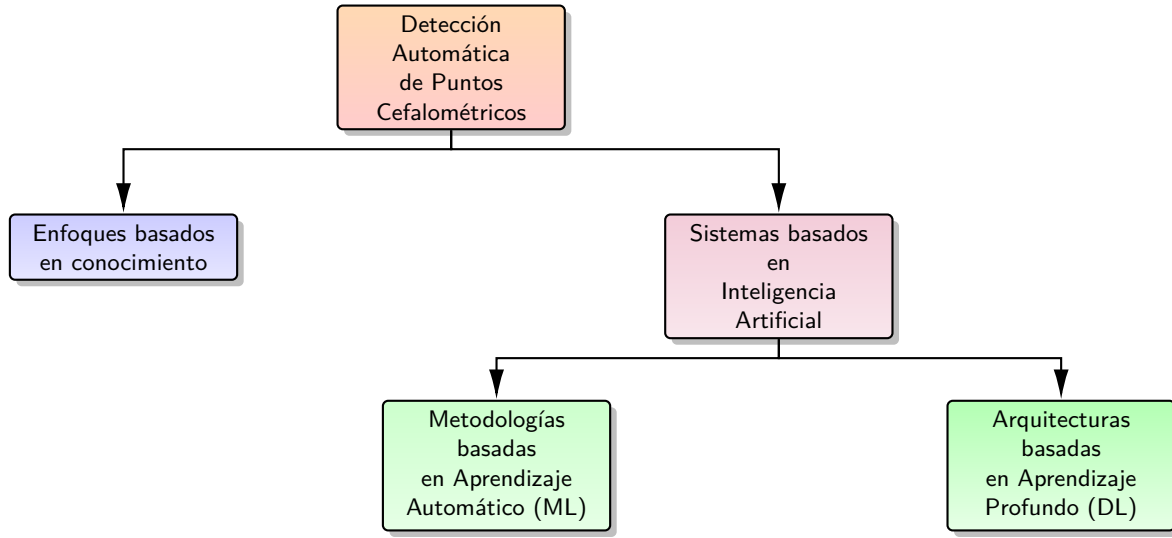


Figura 2.6: Metodologías Utilizadas para la Detección Automática de Puntos Cefalométricos.

Los métodos de detección de bordes se centran en identificar límites o bordes en la imagen y localizar puntos de referencia según cambios de intensidad o gradiente. Los modelos estadísticos de forma y los modelos deformables capturan variaciones de forma y apariencia para alinear plantillas o contornos con la imagen, refinando la ubicación de los puntos de referencia. Los enfoques híbridos combinan múltiples técnicas para mejorar la precisión y la eficiencia de la predicción.

En la Figura 2.7 se muestran varios enfoques basados en el conocimiento para la identificación de puntos de referencia cefalométricos. La Tabla 2.4 muestra un análisis comparativo de diferentes enfoques basados en el conocimiento para el análisis cefalométrico.

Método de Emparejamiento de Plantillas

- **Kaur et al.** [59], examinaron la importancia de las características globales basadas en los *momentos de Zernike* en el cálculo de pequeñas ventanas de expectativa para puntos de referencia individuales y la estimación inicial de estos. Las características basadas en el momento de Zernike, Z_{or} , con orden o y repetición r , se extrajeron de la imagen mediante (2.1).

$$Z_{or} = \frac{o+1}{\pi} \sum_{j=0}^{M-1} \sum_{k=0}^{M-1} f_n(x_j, y_k) V_{or}^*(x_j, y_k) dx dy, \quad x_j^2 + y_k^2 \leq 1, \quad (2.1)$$

donde

$$x_j = \frac{2j+1-M}{M/\sqrt{2}}, \quad y_k = \frac{2k+1-M}{M/\sqrt{2}}, \quad j, k = 0, 1, \dots, M-1, \quad dx = dy = \frac{\sqrt{2}}{M/\sqrt{2}},$$

Tabla 2.4: Métodos propuestos para la Detección Automática de Etiquetas Cefalométricas Basadas en Conocimiento.

Autor y Año	Método	Rad.	Land.	Resultados (SDR % / MRE $\pm$ SD mm)	Inferencias
Liu <i>et al.</i> [47]	Método basado en bordes	10	13	—	Aceptable sólo para un número limitado de etiquetas
Grau <i>et al.</i> [48]	Algoritmo de detección de patrones	20	17	88.6 % en 2 mm; MRE=1.1	Sensible al ruido
Ciesielski <i>et al.</i> [49]	Programación genética	36	4	85 % en 2 mm	Requiere tiempos de ejecución elevados
Rueda <i>et al.</i> [50]	AAM + morfología matemática	96	43	50.04–91.44 % (2–5 mm); MRE=2.48 $\pm$ 1.66	Múltiples contornos mejoran los resultados
Mohseni <i>et al.</i> [51]	Matriz afín + bordes	50	8	93 % en 2 mm	Precisión limitada por el número de etiquetas
Kafieh <i>et al.</i> [52]	ASM modificado	63	16	24–93 % (1–3 mm)	Requiere optimización de parámetros
Mosleh <i>et al.</i> [53]	Fusión de datos + preprocesamiento	20	11	Err. máx. áng=1.15 mm; lin=0.16	Precisión sensible a variaciones musculares
Pouyan <i>et al.</i> [54]	Descriptores HOG	63	16	85 % en 4 px	Tasa en 2 mm menor que en otros métodos
Vucinic <i>et al.</i> [55]	ASM	60	10	61–95 % (2–5 mm); Med=1.68	Mejora reduciendo ruido y oclusión
Shahidi <i>et al.</i> [56]	Template matching + bordes	40	16	12.5–93.75 % (1–4 mm); MRE=2.59	Baja tasa en 2 mm
Majd <i>et al.</i> [57]	Edge tracing + análisis + matching	40	16	30.83–87.34 % (1–4 mm); MRE<1.7	Complejidad por relaciones inter-etiquetas
Montufar <i>et al.</i> [58]	ASM	24	18	SD=2.3° err. angular	Difícil identificar Porion y Sella
Kaur <i>et al.</i> [59]	Momentos de Zernike + matching	163	18	89.5 % en 2 mm; MRE=1.84 $\pm$ 1.24	Puede mejorar con plantillas paramétricas
Sari <i>et al.</i> [60]	Transformada afín + matching	106	1	73.6–92.5 % (2–4 mm); MRE=2.04 $\pm$ 3.37	Uso de Menton mejora precisión anatómica
Favaedi <i>et al.</i> [61]	Relajación probabilística	4	21	30.95–71.4 % (1–2 mm)	Deforma plantilla para ajuste completo
Vasamsetti <i>et al.</i> [62]	Template matching optimizado	37	24	66.6–100 % (2–3 mm); MRE=1.44 $\pm$ 0.98	Plantillas multiescales mejoran rendimiento
Neelapu <i>et al.</i> [63]	Simetría + template matching	30	20	64.16–93.60 % (2–4 mm); MRE=1.88 $\pm$ 1.10	Robusto ante diversidad y deformidades
Vasamsetti <i>et al.</i> [62]	Wavelets	20	15	MRE=1.61; 8 soft, 7 hard en 2 mm	Robusto ante variaciones anatómicas

Aquí, $f_n(x_j, y_k)$ es la función imagen definida sobre un dominio discreto cuadrado de tamaño $M \times M$, y $V_{or}(x_j, y_k)$ son los polinomios complejos de Zernike. $V_{or}^*(x_j, y_k)$ representa el conjugado complejo

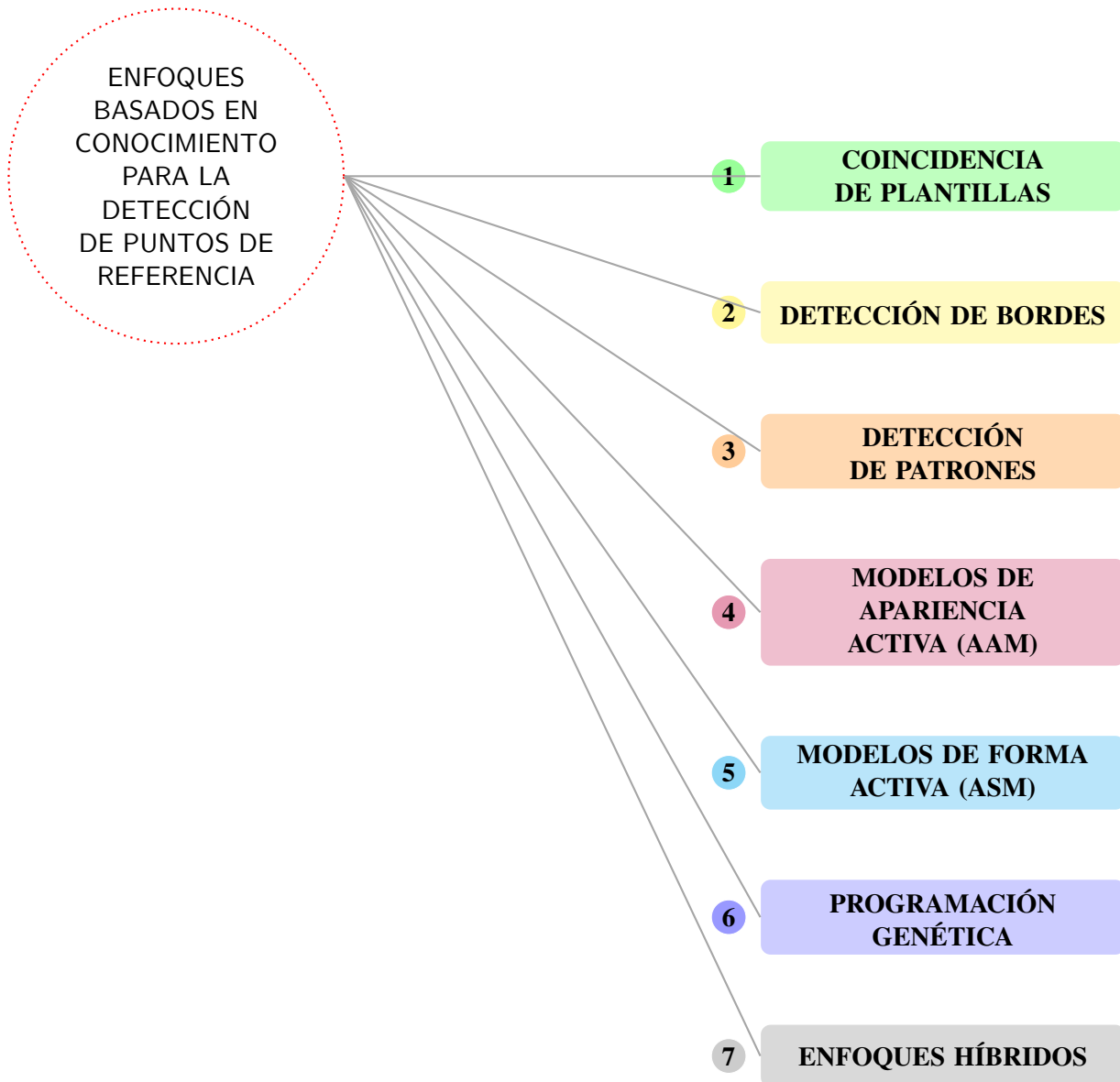


Figura 2.7: Enfoques Basados en Conocimiento Utilizados en la Detección Automática de Puntos Cefalométricos.

de $V_{or}(x_j, y_k)$.

Utilizando un conjunto de datos de 85 cefalogramas seleccionados aleatoriamente, el sistema se evaluó en 18 ubicaciones de referencia. Una ventana de 2 mm contiene el 89.5 % de las localizaciones de los 18 puntos de referencia seleccionados. El error medio para las 18 referencias fue de 1.84 mm, con una desviación estándar promedio de 1.24 mm.

- **Sari et al.** [60], exploraron la aplicación de técnicas de *transformación afín* y *coincidencia de plantillas* para detectar tres puntos de referencia esenciales: el mentón y dos proyecciones del cefalostato. En el primer paso, se detectan las proyecciones mediante un método de coincidencia

de plantillas. Se obtiene una matriz de transformación afín ($AFMatrix$) de dimensiones 5×2 , comparando los puntos de referencia x_r, y_r y x_o, y_o obtenidos en el paso anterior, como se muestra en (2.2),

$$AF_{matrix} = f(x_r, y_r, x_o, y_o), \quad (2.2)$$

El conjunto de entrenamiento incluye las coordenadas reales del terreno del punto *Menton*, indicadas como x_t, y_t . La matriz de transformación afín se utiliza posteriormente para calcular las coordenadas iniciales del mentón en una imagen no observada m , como se describe en (2.3).

$$(x_m, y_m) = AF_{matrix} \cdot (x_t, y_t), \quad (2.3)$$

La tasa de detección obtenida dentro de una ventana de precisión de 4 mm fue del 92.5 %.

- **Favaedi et al.** [61], desarrollaron un sistema basado en *relajación probabilística* que integra información relacional —la posición relativa entre puntos de referencia— con información contextual local derivada de la morfología craneal global.
- **Vasamsetti et al.** [62], introdujeron un método de *coincidencia optimizada de plantillas* (OTM) para la identificación automatizada de puntos anatómicos de tejidos duros y blandos en imágenes cefalométricas laterales. El *coeficiente de correlación* (CC) entre la plantilla de referencia generada, que incluye la región de interés (ROI), y la plantilla obtenida de la imagen de prueba, se calcula mediante (2.4).

$$CC = \frac{\sum_{j=0}^p \sum_{k=0}^q (RT_{jk} - \overline{RT})(TT_{jk} - \overline{TT})}{\sqrt{\left(\sum_{j=0}^p \sum_{k=0}^q (RT_{jk} - \overline{RT})^2\right) \left(\sum_{j=0}^p \sum_{k=0}^q (TT_{jk} - \overline{TT})^2\right)}} \quad (2.4)$$

donde RT_{jk} y $\overline{RT}$ representan la plantilla de referencia generada y su valor de intensidad promedio, respectivamente; TT_{jk} y $\overline{TT}$ corresponden a la plantilla obtenida de la imagen de prueba y su intensidad promedio. Los índices p y q indican las filas y columnas de la plantilla generada.

El estudio analizó 37 cefalogramas laterales, identificando 24 puntos de referencia dentro de un rango de error de 3.0 mm, 23 puntos dentro de 2.5 mm y 16 dentro de 2.0 mm.

- **Neelapu et al.** [63], desarrollaron un método para la *localización automática de puntos cefalométricos en 3D* empleando datos de *CBCT*. De un total de 20 puntos de referencia seleccionados, doce se ubicaron en el plano sagital medio. El error medio global fue de 1.88 mm con una desviación estándar de 1.10 mm.
- **Vasamsetti et al.** [62], propusieron una técnica de extracción de características basada en *wavelets* para analizar 20 cefalogramas y detectar automáticamente puntos anatómicos. Los resultados mostraron la detección correcta de 7 puntos óseos y 8 puntos de tejidos blandos con una precisión de 2 mm.

Técnicas de Detección de Bordes

- **Liu et al.** [47], proponen un sistema computarizado para la identificación automática de puntos de referencia, el cual se evaluó mediante características basadas en bordes. La evaluación se realizó sobre 10 cefalogramas, analizando 13 puntos de referencia específicos. En cinco de estos trece puntos —sella, nasion, porion, orbital y gnathion— no se encontró una diferencia estadísticamente significativa en la inexactitud entre la identificación manual y la computarizada.
- **Mohseni et al.** [51], propusieron un método eficiente, rápido y automatizado para la localización de puntos de referencia en el plano medio cefalométrico en radiografías digitales. La imagen se divide en múltiples regiones, cada una conteniendo tres puntos de control primarios: la punta de la nariz ($NP1$), el punto más bajo de la cadena ($CLP2$) y el cóndilo ($CP3$). Con estos puntos se calcula una matriz de transformación afín (AT_{Matrix}), como se muestra en (2.5):

$$AT_{Matrix} = \begin{bmatrix} x_{NPRI} & y_{NPRI} & 1 \\ x_{CLPR2} & y_{CLPR2} & 1 \\ x_{CPR3} & y_{CPR3} & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} x_{NP1} & y_{NP1} & 1 \\ x_{CLP2} & y_{CLP2} & 1 \\ x_{CP3} & y_{CP3} & 1 \end{bmatrix}, \quad (2.5)$$

donde $NP1$, $CLP2$ y $CP3$ son los puntos de control identificados, y $NPRI$, $CLPR2$ y $CPR3$ son los puntos de referencia. La localización de algunos puntos en la imagen objetivo se estimó a partir de la matriz de transformación afín utilizando la Ecuación (2.6):

$$\begin{bmatrix} x_{NPi} \\ y_{NPi} \end{bmatrix} = [AT_{Matrix}] \cdot \begin{bmatrix} x_{NPRI} \\ y_{NPRI} \end{bmatrix}, \quad (2.6)$$

Posteriormente, cada punto individual se alinea con un punto correspondiente en la imagen de referencia para determinar la ubicación apropiada de otros puntos desde la imagen de referencia hacia la imagen objetivo. El método permite determinar estimaciones iniciales precisas, restringe las regiones de búsqueda y logra una localización precisa de puntos de referencia con una exactitud superior al 90 %.

- **Mosleh et al.** [53], Con el objetivo de optimizar el proceso manual intensivo del análisis cefalométrico, desarrollaron un sistema novedoso compuesto por tres módulos que facilitan el análisis y las mediciones cefalométricas. El sistema detectó exitosamente 11 puntos de referencia en 20 radiografías, demostrando una mejora significativa en eficiencia comparado con el método manual, con una velocidad de procesamiento 10 veces mayor. El error angular máximo observado fue de aproximadamente 1.15 grados, mientras que el error lineal máximo fue de alrededor de 0.16 mm.

Métodos de Detección de Patrones

- **Grau et al.** [48], introdujeron un paradigma compuesto por dos etapas: un módulo de detección de líneas y un módulo de detección de puntos de referencia. Los investigadores analizaron la espina nasal y la línea mandibular con el fin de identificar líneas prominentes y claramente distinguibles. Para evaluar la eficacia del paradigma en la detección de 17 características anatómicas distintas, se utilizaron 20 imágenes, obteniéndose una tasa de detección del 90.3 %.
- **Pouyan et al.** [54], evaluaron el desempeño de las rejillas del descriptor (*Histogram of Oriented Gradients*, HOG), como herramienta para la identificación de puntos de referencia. El método propuesto demostró una exactitud del 85 % en la identificación de la ventana del punto anatómico, considerando una desviación máxima de cuatro píxeles respecto a su posición real. El estudio destacó la eficacia de emplear variables HOG normalizadas localmente dentro de una rejilla altamente superpuesta, lo cual permitió una identificación precisa de las ventanas que contienen los puntos de referencia.

Modelo de Apariencia Activa (MAA)

- **Rueda et al.** [50], utilizando Modelos de Apariencia Activa (MAA), propusieron un sistema automatizado que emplea modelos de apariencia activa. Se construyó un conjunto de entrenamiento coherente aplicando una transformación top-hat para ajustar la inhomogeneidad en la intensidad de las radiografías. Las estructuras brillantes se extrajeron computando la transformación top-hat $\rho(f_s)$. Al encontrar la diferencia entre la imagen f_s y su apertura $\gamma(f_s)$, se obtiene la siguiente expresión:

$$\rho(f_s) = f_s - \gamma(f_s), \quad (2.7)$$

Este modelo analizó 96 imágenes anotadas manualmente, obteniendo resultados de segmentación del 50.04 %, 72.62 % y 91.44 % en umbrales de 2 mm, 3 mm y 5 mm, respectivamente.

- **Vucinić et al.** [64], propusieron un modelo estadístico que abarca variaciones en la forma y textura, derivadas mediante la aplicación de Análisis de Componentes Principales (PCA, por sus siglas en inglés) sobre una colección de vectores que describen las formas y texturas dentro del conjunto de entrenamiento. La representación de la forma de un objeto puede denotarse mediante el vector $\mathbf{s}$, mientras que la textura, es decir, los niveles de gris, se denota mediante el vector $\mathbf{t}$. El modelo de apariencia se determina por parámetros denotados como am , los cuales combinan tanto la forma como las componentes de textura, como se indica en (2.8),

$$\mathbf{s} = \bar{\mathbf{s}} + V_s \cdot am; \quad \mathbf{t} = \bar{\mathbf{t}} + V_t \cdot am \quad (2.8)$$

donde $\bar{s}$ representa la forma media, $\bar{t}$ representa la textura media, y V_s y V_t representan las matrices que describen los modos de variación obtenidos del conjunto de entrenamiento. Esta técnica logró una exactitud del 61 % y 95 % en los umbrales de 2 mm y 5 mm, respectivamente. En promedio, el error radial medio (MRE) del método de detección basado en modelos fue de 1.68 mm.

Enfoques Basados en Algoritmos Genéticos

- **Ciesielski et al.** [49], describieron la evolución de programas de detección de objetos basados en programación genética aplicados a radiografías digitales para identificar características craneofaciales. El proceso de entrenamiento se basa en una función de aptitud (fitness function) que combina la tasa de detección y la tasa de falsos positivos, aplicando ponderaciones adecuadas a cada métrica.

Para el análisis se emplearon un total de 73 radiografías en el conjunto de entrenamiento y 36 radiografías en el conjunto de prueba, considerando cuatro puntos de referencia específicos. Los resultados mostraron una tasa de detección del 100 % en la nariz, 79 % en la línea media del labio superior, 96.3 % en el incisivo, y 63.9 % en el sella.

Modelos Híbridos

- **Majid et al.** [57], categorizaron los puntos de referencia utilizando una combinación de técnicas de coincidencia ponderada de plantillas, estimación basada en análisis y estimación basada en bordes, considerando las propiedades anatómicas y su relevancia en los análisis cefalométricos establecidos. El proceso de detección identifica aproximadamente 16 etiquetas, mostrando una tasa promedio de error inferior a 1.7 mm para todos los puntos identificados.
- **Kafieh et al. (2007)** [52], propusieron una versión modificada del modelo ASM (por sus siglas en inglés, *Active Shape Model*) para la predicción automática de etiquetas, donde el 24 % de las 16 etiquetas cefalométricas se encontraron dentro de un margen de 1 mm respecto a las coordenadas reales, el 61 % dentro de un margen de 2 mm, y un impresionante 93 % dentro de un margen de 5 mm. Además, se empleó una red neuronal de cuantización vectorial de aprendizaje (LVQ, *Learning Vector Quantization*) con el propósito de categorizar imágenes en función de sus características geométricas.
- **Jetwansi et al.** [65], eliminaron la subjetividad del operador utilizando métodos de segmentación y detección de bordes mediante wavelets difusos (*fuzzy wavelet edge detection*).

Dada una imagen de entrada I , y dos imágenes GD_x y GD_y que representan las aproximaciones derivativas horizontal y vertical en cada punto, se calculan de acuerdo con (2.9) y (2.10).

$$GD_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * I \quad (2.9)$$

$$GD_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * I \quad (2.10)$$

La magnitud del gradiente (GD) puede obtenerse combinando las aproximaciones de gradiente en cada punto de la imagen, como se muestra en (2.11),

$$GD = \sqrt{(GD_x)^2 + (GD_y)^2} \quad (2.11)$$

El cálculo del gradiente normalizado de Sobel (SG_N) se realiza utilizando (2.12), donde GD_{Max} representa el valor máximo de la magnitud del gradiente dentro de la imagen de entrada I ,

$$SG_N(x, y) = \frac{GD(x, y)}{GD_{Max}}. \quad (2.12)$$

- **Sanei et al.** [66], presentaron un enfoque iterativo de análisis de datos autoorganizado que combina el reconocimiento estadístico de patrones con la toma de decisiones basada en lógica difusa.
- **Shahidi** [56], se seleccionaron aleatoriamente 40 radiografías cefalométricas digitales para diseñar un software capaz de localizar etiquetas cefalométricas y evaluar su precisión en dicha tarea. Para facilitar la comparación de resultados con otros estudios y garantizar la consistencia, se determinaron los equivalentes en grados de los umbrales correspondientes a los extremos de las líneas con longitudes de 1, 2 y 4 mm.

Sea ($L1 - L2$) la longitud del segmento de línea y $0.5 * l_{mm}$ el umbral conocido, es posible determinar el umbral angular (β_t) utilizando la siguiente Ecuación matemática, mostrada en (2.13).

$$\beta_t = \arctan \left(\frac{0.5 * l}{0.5 * |L1 - L2|} \right), \quad (2.13)$$

donde $arctg$ es la función tangente inversa. El error promedio general, expresado en milímetros, calculado a partir del error medio de 2.59 mm en la detección de etiquetas y de 1.27 mm en la detección de líneas, se determinó en 2.18 mm. Un total del 12.5 % de las etiquetas analizadas presentaron errores medios inferiores a 1 mm, mientras que el 43.75 % se encontró por debajo del umbral de 2 mm.

- **Rakhshan et al.** [67], desarrollaron una técnica para la detección de 15 etiquetas cefalométricos laterales utilizando algoritmos basados en bordes, morfología y conocimiento previo. El sistema

identificó con exactitud 5 etiquetas dentro del rango clínicamente aceptable de 2 mm.

La línea de tiempo de los artículos publicados basados en modelos híbridos se muestra en la Figura 2.8.

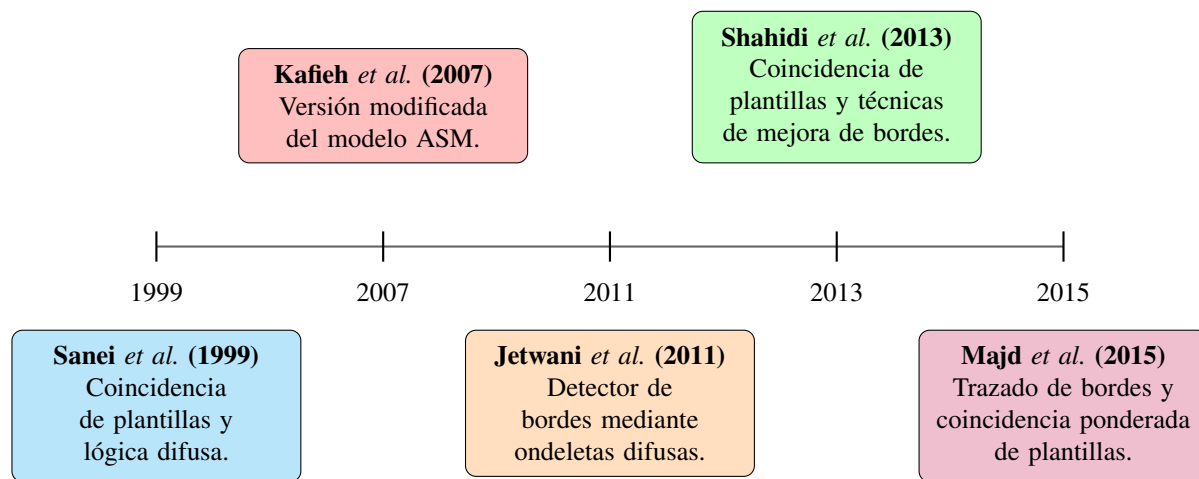


Figura 2.8: Línea de tiempo de los Modelos Híbridos aplicados a la identificación automática de etiquetas cefalométricas.

2.5.4. Modelos Basados en Inteligencia Artificial

La Inteligencia Artificial (IA) ha ganado relevancia en los campos de la ortodoncia y el análisis craneofacial como un mecanismo altamente efectivo para automatizar la identificación de puntos cefalométricos mediante la utilización de modelos computacionales.

Metodologías Basadas en Aprendizaje (*Machine-Learning*)

El campo de la detección automática de puntos cefalométricos ha experimentado avances significativos como resultado de las técnicas de aprendizaje automático (ML, por sus siglas en inglés). Dado que estos enfoques emplean algoritmos para aprender patrones y relaciones a partir de grandes conjuntos de datos, son capaces de detectar con exactitud los puntos cefalométricos basados en la información disponible. Estos algoritmos incluyen Máquinas de Soporte Vectorial (MSV), *Random Forest*, (RF), métodos de Ensamble, Regresión y modelos PCA lineales, como se ilustra en la Figura 2.9.

Un análisis detallado de las técnicas basadas en ML se presenta en la línea del tiempo representada en la Figura 2.10.

- **Ibragimov *et al.*** [68], utilizó características de tipo Haar y un algoritmo de ML para construir un detector de posibles ubicaciones candidatas a puntos de referencia.



Figura 2.9: *Métodos basados en Aprendizaje Automático para la Detección de Puntos Cefalométricos*

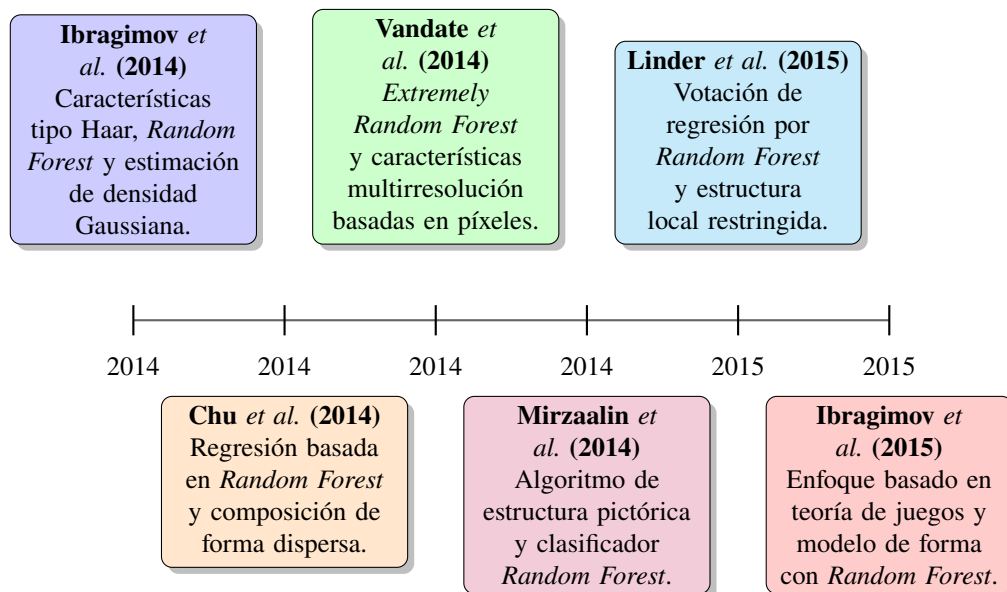


Figura 2.10: *Línea de Tiempo de las Técnicas basadas en Aprendizaje Automático aplicadas a la Localización de Etiquetas Cefalométricas.*

Tras la validación con un conjunto de 100 imágenes de prueba, se determinó una tasa de detección del 72.7 %, 78.8 %, 82.7 % y 87.7 % dentro de rangos de precisión de 2 mm, 2.5 mm, 3 mm y 4 mm, respectivamente.

- **Chu et al.** [69], implementaron regresiones mediante *Random Forest*) y composiciones de forma dispersa para detectar puntos cefalométricos utilizando 80 radiografías cefalométricas para entrenamiento y 20 para prueba, obteniendo una tasa de detección del 77.79 % de los puntos dentro de un error menor a 4.0 mm.
- **Vandale et al.** [70], emplearon *Random Forest* Extremadamente Aleatorizados en conjunto con

características multirresolución basadas en píxeles. Se utilizó una distribución normal multivariada, definida en (2.14), para generar 16 observaciones aleatorias considerando la varianza en las direcciones X e Y, representadas como $\text{Var}(X)$ y $\text{Var}(Y)$.

$$ND(POS_{train}, \text{Covar}(POS_{train})), \quad (2.14)$$

donde POS_{train} denota las posiciones de los puntos en el conjunto de entrenamiento. X_{POS} y Y_{POS} representan las posiciones en los ejes X y Y, respectivamente. Como resultado de una validación cruzada de diez particiones, fue posible detectar algunos puntos con una exactitud del 96 % dentro de un rango de 2.5 mm.

- **Mirzaalian et al.** [71], introdujeron un método de estructura pictórica orientado a la identificación automática de puntos cefalométricos a partir de radiografías. Este sistema utiliza radiografías cefalométricas para el proceso de reconocimiento. El desafío de localizar los 19 puntos de referencia ($LM \in \{lm_1, lm_2, \dots, lm_{19}\}$) se formuló como un problema de optimización mostrado en (2.15),

$$\hat{LM}(I) = \arg \min_{LM} \left(\sum_{jk} \theta^\phi(LM_j, LM_k) + \sum_{j=1}^{19} \theta^\psi(LM_j, I) \right) \quad (2.15)$$

donde el término de regularización por pares se representa como θ^ϕ del vector $LM_{jk} = LM_j - LM_k$. $\theta^\phi(LM_j, LM_k)$ se calcula mediante la distancia de Mahalanobis, según la Ecuación 2.16, entre LM_{jk} y el valor medio de los bordes correspondientes:

$$\theta^\phi(LM_j, LM_k) = (LM_j, LM_k)^T \text{CoVar}_{jk}^{-1} (LM_{jk} - \overline{LM}_{jk}) \quad (2.16)$$

donde CoVar_{jk}^{-1} es la matriz de covarianza inversa y $\overline{LM}_{jk}$ representa el vector medio. Este enfoque alcanzó una exactitud de 62.32 %, 70.42 %, 75.68 % y 84.05 % dentro de rangos de 2 mm, 2.5 mm, 3 mm y 4 mm, respectivamente.

- **Linder et al.** [72], utilizando un método de votación por regresión basado en *Random Forest*, analizaron 150 imágenes y lograron un error promedio de 1.6 mm–1.7 mm, con tasas de detección que oscilaron entre 75 % y 85 % dentro de un rango de precisión de 2 mm. Este método genera un modelo lineal que captura las variaciones en la forma representando la información posicional de cada punto lm_i , usando (2.17).

La posición estimada de la etiqueta se expresa como:

$$S_{lm,n} = GT_{lm}(S_{lm} + V_{lm,n} + d_{lm,n}) \quad (2.17)$$

donde, la posición media de la etiqueta se denota como S_{lm} . Los parámetros del modelo de forma y

el conjunto de variaciones se representan como $V_{lm,n}$ y $d_{lm,n}$, siendo las desviaciones modales y las transformaciones globales GT_{lm} .

La predicción mediante *Random Forest* (RF) se calcula agregando todas las predicciones individuales, resultando en histogramas bidimensionales de votos $V_{T,lm}$ para cada etiqueta lm . El objetivo de combinar los histogramas mediante restricciones de forma se obtiene maximizando (2.18),

$$C(\theta, \phi) = \sum_{lm=1}^L V_{T,lm} \left(GT_{lm}(S_{lm} + V_{lm,n} + d_{lm,n}) \right) \quad (2.18)$$

- **Linder et al.** [73], llevaron a cabo un estudio con el objetivo de crear y verificar un sistema automático de localización de etiquetas anatómicas faciales, denominado *FALA system*. El sistema fue diseñado para identificar y clasificar anomalías esqueléticas faciales utilizando 19 etiquetas cefalométricas en imágenes radiográficas. El sistema FALA emplea modelos de regresión multivariada y *Random Forest* (RF).

Para cada etiqueta, el error puntual específico (*landmark-specific point-to-point error*, PPE_{lm}) se calcula mediante la siguiente relación:

$$PPE_{lm} = \sqrt{\frac{1}{N} \sum_{i=1}^N (MP_{lm,i} - AP_{lm,i})^2} \quad (2.19)$$

donde $MP_{lm,i}$ y $AP_{lm,i}$ representan las coordenadas predichas y las reales del etiquetas lm en la imagen i , respectivamente. El error promedio punto a punto (PPE) se obtiene promediando el error sobre todas las etiquetas, como se muestra a continuación,

$$PPE = \frac{1}{L} \sum_{lm=1}^L PPE_{lm} \quad (2.20)$$

donde N y L representan el número de imágenes y de etiquetas, respectivamente. El error promedio total punto a punto (*total point-to-point error*, TPPE) se calcula mediante la Ecuación (2.21):

$$TPPE = \sqrt{\frac{1}{N} \sum_{lm=1}^L PPE_{lm}^2} \quad (2.21)$$

El estudio utilizó 400 radiografías cefalométricas digitales para la detección de 19 etiquetas cefalométricas empleando el sistema FALA. La precisión media obtenida fue de 0.42 mm.

- **Wang et al.** [24], en el contexto del *IEEE International Symposium on Biomedical Imaging* (2015), investigaron 300 radiografías cefalométricas capturadas de individuos con edades entre 9 y 66 años,

provenientes del departamento de ortodoncia del Hospital General de Tianjin. Se reportó un rango de precisión promedio de 4 mm.

- **Hwang et al.** [74], propusieron un enfoque basado en aprendizaje profundo con una precisión media de 1.34 ± 1.09 mm, y detección exitosa dentro de un rango de ± 2 mm del 71.7 %.
- **Chakrabarty et al.** [75], propusieron una máquina de soporte vectorial (*Support Vector Machine*, SVM) para distinguir las etiquetas faciales, obteniendo una exactitud de más del 95 %.

Coefficiente de correlación craneal

- **Romaniuk et al.** [76], propusieron el uso de tres metodologías para la localización estadística de las etiquetas faciales.

El Coeficiente de Correlación Craneal (CC_{ij}) se define como:

$$CC_{ij} = \frac{1}{N} \sum_{k=1}^N C_{ik} C_{jk} \quad (2.22)$$

donde C_{ik} y C_{jk} son los puntos centrales de referencia y N es el número de etiquetas analizadas. Para cada eje $\alpha \in \{x, y, z\}$, la media se expresa como:

$$\mu_\alpha = \frac{1}{N} \sum_{i=1}^N (v_{i,\alpha} - v_{0,\alpha}) \quad (2.23)$$

Sea $S_k^l = (x_k^l, y_k^l)$ la k -ésima etiqueta predicha de la imagen l ($l = 1, 2, \dots, n$), y sea $S_{k,GT}^l$ la posición real (*ground truth*). Entonces, el error cuadrático medio (RMSE) por etiqueta cefalométrica se calcula como:

$$S_{RMSE} = \sqrt{\frac{1}{nL} \sum_{l=1}^n \sum_{k=1}^L [(x_k^l - x_{k,GT}^l)^2 + (y_k^l - y_{k,GT}^l)^2]}. \quad (2.24)$$

El error cuadrático medio (RMSE) se redujo a 1.86 mm. El modelo incluyó una función de pérdida geométrica (GD) y un parámetro de gradiente (GP), definidos respectivamente como:

$$GD_{lm} = |P_{lm}^{GT} - P_{lm}^P| \quad (2.25)$$

$$GP_{lm} = \arctan \frac{GD_x}{GD_y} \quad (2.26)$$

donde P_{lm}^{GT} representa la posición real y P_{lm}^P la predicha para la etiqueta cefalométrica lm .

La Tabla 2.5 muestra un análisis comparativo de diferentes metodologías de Aprendizaje Automático aplicadas al análisis cefalométrico.

Tabla 2.5: Métodos de Aprendizaje Automático (ML) para la Detección Automática de Etiquetas Cefalométricas.

Autor y Año	Método	N. Rad.	Etq.	Resultados (SDR % / MRE ± SD mm)	Inferencias
Ibragimov <i>et al.</i> [68]	<i>Random Forest</i> (RF), características Haar-like y estimación de densidad basada en Gaussianas	300	19	SDR en 2, 2.5, 3 y 4 mm: Método robusto con buena 72.75–97.68 %; MRE=1.62 mm; SD=1.12 mm	Precisión promedio
Chu <i>et al.</i> [77]	<i>Regresión Random Forest</i> (RFR) junto con composiciones de forma sparse	300	19	SDR en 2–4 mm: Precisión dependiente de la 39.74–94.26 %; MRE=2.21 mm	composición morfológica
Vandelaer <i>et al.</i> [70]	Árboles de decisión y <i>Random Forest</i> en combinación con registros no rígidos	300	19	SDR=37.45–84.97 % (2–4 mm); MRE=1.85±1.19 mm	Mejora con registro morfológico adaptativo
Mirzaalian <i>et al.</i> [71]	Estructura probabilística y clasificador RF	300	19	SDR=58.21–91.27 % (2–4 mm); MRE=2.06±1.15 mm	Reduce error medio mediante modelado jerárquico
Chen <i>et al.</i> [78]	Optimización covariada y selección de plantillas mediante muestreo aleatorio	450	19	SDR=44.82–90.45 % (2–4 mm); MRE=1.84±1.22 mm	Utiliza covarianza para mejorar estabilidad espacial
Lindner <i>et al.</i> [72]	<i>Regresión Random Forest</i> con Votación	400	19	SDR=85.84–98.46 % (2–4 mm); MRE=1.61 mm; SD=1.47 mm	Precisión superior gracias al voto ponderado
Ibragimov <i>et al.</i> [79]	Modelo basado en teoría de juegos y RF	400	19	SDR=71.41–91.81 % (2–4 mm); SD=1.76 mm	Modelo cooperativo mejora predicción colectiva
Lindner <i>et al.</i> [73]	Regresión RF y Modelo Local Restringido (CLM)	400	19	SDR=84.70–96.30 % (2–4 mm); MRE=1.20±0.60 mm	Precisión alta, pero dependiente del alineamiento local
Wang <i>et al.</i> [24]	Árboles de decisión multirresolución con regresión	465	19	SDR=73.77–90.67 % (2–4 mm); MRE=1.69±1.39 mm	Mantiene desempeño propio: estable en bases propias y externas
Romaniuk <i>et al.</i> [76]	PCA lineal y no lineal con modelos afines	227	14	Error medio: eje-x=3.4 mm, eje-y=3.5 mm	Precisión baja, depende de relación espacial rígida
Chakrabarty <i>et al.</i> [75]	Máquina de Vectores de Soporte (SVM) con índice Gini	40	8	96.12 % dentro de 1.5 mm; 93.87 % en 1 mm	Alta precisión en conjuntos pequeños

Arquitecturas Basadas en Aprendizaje Profundo (*Deep-Learning*)

Ha habido un aumento notable en la utilización de metodologías de aprendizaje profundo (DL) para la detección automatizada de puntos cefalométricos, como se muestra en la Fig. 10. La cefalometría

tiene una importancia significativa en los campos de la ortodoncia y la cirugía maxilofacial, ya que sirve como una herramienta crucial para el diagnóstico, la planificación de terapias y la evaluación de la progresión de la estructura y función facial de un paciente. Las radiografías cefalométricas han demostrado una eficacia notable en la identificación confiable de puntos anatómicos mediante la utilización de modelos de aprendizaje profundo, específicamente redes neuronales convolucionales (CNN) y redes neuronales recurrentes (RNN). Una línea temporal detallada de los estudios relacionados con técnicas de aprendizaje profundo basadas en CNN se presenta en la Figura 2.11.

- **Arik et al.** [33], emplearon redes neuronales convolucionales (CNN) para la detección de puntos de referencia mediante el análisis de la mandíbula y la base del cráneo, obteniendo una tasa de éxito ligeramente superior del 1–2 % para la detección de estructuras anatómicas dentro de un rango de 2 mm y una exactitud de clasificación promedio del 76 %.
- **Lee et al.** [80], introdujeron una técnica de regresión por coordenadas basada en CNN para reconocer 19 puntos cefalométricos en radiografías dentales. Las operaciones de convolución en las capas CNN se realizan utilizando (2.27) y (2.28):

$$fn(conv) = (im * knl)[conv] = \sum_{k=1}^N im[k] knl[k + conv] \quad (2.27)$$

donde la imagen se denota como im , la cual se representa como una matriz bidimensional de píxeles, y knl es el kernel. El proceso matemático produce resultados a partir de la característica de mapa fn .

$$O(q, r) = \frac{(IN * FLR)_{(q,r)}}{\sum_{SD}(m, n)G(m)G(n)}(r - n) \quad (2.28)$$

La matriz de entrada, denominada como IN , se convoluciona con un filtro bidimensional FLR , donde (q, r) representa las coordenadas de salida. El resultado de la convolución se representa como $O(q, r)$. El símbolo $IN * G$ denota el proceso de convolución.

- **Kang et al.** [81], desarrollaron una CNN tridimensional utilizando datos cefalométricos 3D de una muestra de 27 individuos sanos, de los cuales 18 se usaron para el entrenamiento y los 9 restantes como conjunto de prueba. El objetivo fue predecir las coordenadas de 12 puntos anatómicos distintos. Este modelo mostró errores promedio de 3.26, 3.18 y 4.81 mm a lo largo de los tres ejes, y 7.61 mm en el espacio tridimensional.
- **Gupta et al.** [82], propusieron un enfoque basado en conocimiento para identificar y localizar 20 puntos cefalométricos utilizando imágenes CBCT de 30 sujetos mediante programación en MATLAB. El método propuesto arrojó un error promedio de 2.01 mm con una desviación estándar de 1.23 mm en todos los puntos anatómicos. La exactitud promedio en la detección de puntos fue de 64.67 %, 82.67 % y 90.33 % al considerar márgenes de error de 2, 3 y 4 mm, respectivamente.

- **Qian et al.** [83], usando Fast R-CNNs (CephaNet) propusieron un enfoque en dos etapas para detectar puntos cefalométricos mediante una estrategia de pérdida multi-tarea, mitigando así la variabilidad intra-clase. Este estudio propone una técnica de regresión jerárquica en dos etapas dirigida a resolver la presencia de puntos anatómicos identificados incorrectamente, incluidos los innecesarios y los no anotados. La regresión bidimensional no lineal y la transformación Laplaciana están diseñadas para eliminar valores atípicos, logrando un rendimiento superior en el conjunto de datos 2015 ISBI, con una mejora del 6 % para un rango de error de 2 mm.
- **Nishimoto et al.** [84], analizaron 219 imágenes cefalométricas laterales, obteniendo un error promedio de 1.702 píxeles, mientras que el error de predicción mediano fue de 1.622 píxeles. Kunz et al. [85] desarrollaron un enfoque de atención piramidal utilizando 192 radiografías cefalométricas para identificar 18 puntos.
- **Lee et al.** [86], introdujeron redes neuronales de atención, específicamente atención en parches, entrenadas con 935 radiografías cefalométricas. En total, se localizaron 19 puntos, y 11 de ellos estaban relacionados con tejido blando. La función de pérdida para la localización de los puntos estimados incluyó la norma del error de Mahalanobis, expresada en (2.29):

$$\beta = CSD^2 \left(\frac{x_i - \hat{x}_i}{\sigma_x} \right)^2 - 2CoVar \left(\frac{x_i - \hat{x}_i}{\sigma_x} \right) \left(\frac{y_i - \hat{y}_i}{\sigma_y} \right) + \left(\frac{y_i - \hat{y}_i}{\sigma_y} \right)^2 \quad (2.29)$$

donde β es el límite de confianza. CSD^2 representa la función que retorna el valor β de la distribución ji-cuadrado con dos grados de libertad. (x_i, y_i) son los valores de coordenadas con desviaciones estándar (σ_x, σ_y) , y $CoVar$ denota la covarianza entre las coordenadas. El modelo propuesto para la detección de puntos anatómicos en tejido duro alcanzó una tasa de éxito promedio del 96.4 % dentro de un rango de error de 1.32 a 3.5 mm. De manera similar, el modelo logró una tasa de éxito promedio del 75.2 % para los puntos de tejido blando dentro de un rango de error de 1.16 a 4.37 mm.

- **Song et al.** [87], propusieron un algoritmo de dos etapas para la identificación automatizada de puntos cefalométricos en radiografías esqueléticas. Para evaluar la efectividad del algoritmo propuesto, se emplearon dos conjuntos de datos: el ISBI 2015 Grand Challenge en Imágenes de Radiografía Lateral Cefalométrica y el conjunto de datos proporcionado por la Universidad de Shandong. Los resultados experimentales demostraron que el método propuesto logró un SDR promedio del 86.4 %, 74 % para los conjuntos de datos ISBI 2015 Test1 y Test2, y 62 % en su propio conjunto de datos dentro de una precisión de 2 mm. El error radial medio (MRE) obtenido fue de 1.077 mm, 1.542 mm y 2.1 mm.
- **Qian et al.** (2019) analizaron 1028 imágenes cefalométricas usando las técnicas de detección de objetos *You-Only-Look-Once* versión 3 (YOLOv3) y *Single Shot Multibox Detector* (SSD).
- **Lee et al.** (2020) utilizaron una Red Neuronal Convolutiva Bayesiana (*Bayesian Convolutional Based Neural Network*, BCNN) para analizar el conjunto de datos ISBI 2015. El marco propuesto

incluyó dos subdivisiones, a saber, muestreo de baja resolución (LRS) y muestreo de alta resolución (HRS). El componente LRS es responsable de extraer parches de referencia relevantes de las radiografías. Por otro lado, el componente HRS determina la localización adecuada de los puntos de referencia dentro de los parches generados por el LRS.

Sea $(\overline{c}_i^L, \overline{r}_i^L)$ el centro del punto de referencia i^{th} dentro de la ROI del LRS. Entre el número total de imágenes verdaderas n_{TR} , la ubicación categorizada como verdadera del punto i^{th} del LRS puede calcularse usando la Ecuación (2.30):

$$(\overline{c}_i^L, \overline{r}_i^L) = \frac{1}{n_{TR}} \sum_{p=1}^{n_{TR}} P_i^{p,L} \quad (2.30)$$

Un preprocesamiento de ponderación de puntuación SW_i^p se introduce como en la Ecuación (2.31), y usando la puntuación determinada para cada píxel ROI, la ubicación final del punto de referencia $(\overline{c}_i^H, \overline{r}_i^H)$ se calcula como en la Ecuación (2.32). Cada píxel ROI P_i^R fue posteriormente elegido usando una metodología similar a la del LRS, y se denota como $IN(P_i^R)$, para usarse como entrada para el HRS.

$$SW_i^p = e^{-\frac{(y_i^p - \mu_i)^2}{2 \tan(h)^2}} \left(\frac{\sum_{x,y} I_{xy}^p}{\sum_{x,y}} \right) \quad (2.31)$$

donde μ_i es la media, y h está asociado con los valores suavizados $\tan(h)$ de cada entrada.

$$(\overline{c}_i^H, \overline{r}_i^H) = \frac{\sum_p (SW_i^p \times \vec{P}_i^p)}{\sum_p SW_i^p} \quad (2.32)$$

Los resultados del marco propuesto demostraron un SDR del 82.11 %, 92.28 % y 95.95 %, con un rango de precisión de 2 mm, 3 mm y 4 mm, respectivamente.

- **Qian et al.** [88], crearon una red neuronal única denominada CephaNN, que incorpora un mecanismo de atención multi-cabeza para la identificación precisa de los rasgos cefalométricos. El mecanismo de atención propuesto aborda el problema del desequilibrio de clases, mejorando la efectividad de regiones específicas en los mapas de calor regresados mediante la implementación de una función de pérdida de realce regional (*Region Enhancing Loss*, RE loss). De acuerdo con el conjunto de datos de referencia (*benchmark dataset*), CephaNN alcanzó una exactitud del 87.6 % y un SDE de 2.0 mm, lo que indica un nivel muy alto de exactitud analizado sobre 75 puntos de referencia.
- **Kim et al.** [89], desarrollaron un programa basado en la web con la capacidad de evaluar automáticamente radiografías cefalométricas laterales utilizando la arquitectura apilamiento de relojes de arena. Esta aplicación empleó valores anotados para un total de 23 puntos de referencia, los cuales fueron extraídos de un conjunto de datos que consta de 2075 radiografías cefalométricas. Se estimó que el 81.35 % de los puntos de referencia fueron identificados correctamente por este

sistema dentro de un rango de 2 mm.

- **Zeng et al.** [90], desarrollaron una red convolucional en cascada compuesta por tres fases (*Align*, *Proposal* y *Refine Nets*) para la detección automatizada de puntos de referencia en radiografías craneales. En la evaluación del método propuesto utilizando el conjunto de datos *Test1*, los resultados indican valores de SDR de 81.37 %, 89.09 %, 93.79 % y 97.86 % para los puntos de referencia ubicados dentro de los rangos de precisión de 2 mm, 2.5 mm, 3 mm y 4 mm, respectivamente. La tasa de error promedio obtenida fue de 1.34 mm. Para el conjunto de datos *Test2* el 70.58 %, 79.53 %, 86.05 % y 93.32 % de los puntos de referencia fueron identificados correctamente, con una tasa de error promedio de 1.64 mm.
- **He et al.** [91], desarrollaron un marco de dos etapas que integró la regresión de mapa de calor gaussiano y mapa de desplazamiento (*offset map regression*). Una red de refinamiento local fue entrenada utilizando estos parches de imagen como entrada, con el fin de mejorar la precisión de la localización de los puntos de referencia mediante la intensificación de las características locales en imágenes cefalométricas de alta resolución. El modelo propuesto mostró un rendimiento consistente tanto en el conjunto de datos público ACXLDC como en el conjunto de datos privado Huaxi-Analysis. El modelo entrenado en el conjunto de datos ACXLDC se utilizó para identificar los 19 puntos de referencia frecuentes presentes en el conjunto Huaxi-Analysis. El Error Radial Medio (MRE) de las ubicaciones de los puntos de referencia respecto al valor real fue de 1.77 ± 0.82 mm y 1.76 ± 0.92 mm para los conjuntos de validación y prueba, respectivamente. En el conjunto de datos ACXLDC, los resultados de MRE para validación y prueba fueron de 1.92 ± 1.36 mm y 2.22 ± 1.94 mm, respectivamente.
- **Yueyuan et al.** [92], introdujeron un novedoso modelo basado en *transformers* denominado Swin-CE, el cual integra el codificador *Swin Transformer* con un codificador convolucional para el reconocimiento de puntos cefalométricos. Los valores de SDR a 2 mm, 2.5 mm, 3 mm y 4 mm fueron de 87.31 %, 92.93 %, 95.87 %, 98.74 % para *Test1*, y de 76.23 %, 83.36 %, 89.15 %, 95.24 % para el conjunto *Test2* del desafío IEEE ISBI 2015. El MRE para los 19 puntos cefalométricos fue de 1.13 mm y 1.44 mm, respectivamente.
- **Neeraja et al.** [93], presentan un análisis y comparación exhaustivos de diferentes técnicas de inteligencia artificial para la automatización del reconocimiento de puntos cefalométricos en radiografías craneales.
- **Oh et al.** [94], propusieron un aprendizaje profundo basado en características contextuales anatómicas mediante redes neuronales totalmente convolucionales (*fully CNNs*), que construye un *Local Feature Perturbator* y una función de pérdida de configuración anatómica (*Anatomical Configuration Loss*) para extraer características contextuales.

La imagen se define como $IN \in \mathbb{R}^{H_t \times W}$, donde $LL_i \in \mathbb{R}^{J \times \dots}$ representa el conjunto de coordenadas

de los puntos de referencia, J indica el número total de puntos de referencia, y $HP(y; LL_i, SD_{LL})$ es el mapa de calor correspondiente a un punto de referencia específico. El mapa de calor se define matemáticamente como la función de Laplace, como se muestra en la Ecuación (2.33).

$$HP(y; LL_i, SD) = \frac{1}{2SD} \exp \left(-\frac{|y - LL_i|}{2SD^2} \right) \quad (2.33)$$

Se detectaron correctamente los puntos de referencia con porcentajes de 86.2 %, 91.2 %, 94.4 %, 97.7 % y 75.9 %, 83.4 %, 89.3 %, 94.7 % para los conjuntos de datos *Test1* y *Test2* dentro de los rangos clínicamente aceptados de precisión de 2 mm, 2.5 mm y 3 mm, respectivamente, con valores de MRE que oscilaron entre 11.77 y 14.55 píxeles.

- **Kim et al.** [89], generaron cefalogramas laterales mediante la utilización de una Red Generativa Antagónica de Crecimiento Progresivo (*Progressive Growing Generative Adversarial Network*, PGGAN) basada en técnicas de aprendizaje profundo (*Deep Learning*, DL). El trazado cefalométrico, que involucró la detección de 42 puntos de referencia, se llevó a cabo tanto en imágenes reales como sintetizadas por dos ortodoncistas. Los resultados del análisis indicaron un alto nivel de consistencia, con una diferencia media de $2,08 \pm 1,02$ mm.
- **Bulatova et al.** [95], evaluaron la exactitud del reconocimiento de puntos cefalométricos comparando el desempeño de dos métodos: el uso de CNN empleando la técnica You Only Look Once, Versión 3 (YOLOv3), junto con el equipo que anotó manualmente utilizando 110 radiografías. No se encontró una diferencia estadísticamente significativa en 12 de los 16 puntos de referencia al comparar las diferencias absolutas entre los grupos de anotación manual y de IA.
- **Kim et al.** [89], proponen un proyecto, cuyo objetivo es determinar la exactitud de la detección automatizada de puntos cefalométricos mediante la implementación de CNN en cascada en cefalogramas laterales. Se recopiló un conjunto de datos compuesto por 3150 cefalogramas laterales de 10 hospitales universitarios ubicados en Corea del Sur, específicamente con el propósito de entrenamiento. La precisión del modelo generado se evaluó mediante 100 cefalogramas laterales como un conjunto de datos de validación externa. Se encontró que la diferencia promedio entre examinadores fue de $1,31 \pm 1,13$ mm. El error general en la detección automatizada fue de $1,36 \pm 0,98$ mm.
- **Wang et al.** [96], también analizaron el conjunto de datos público IEEE ISBI 2015 para la detección cefalométrica utilizando DCNN, que incluye regiones de predicción inicial para varios puntos de referencia. La metodología propuesta obtuvo valores SDR de 87.51 %, 91.83 %, 94.74 % y 98.01 % en los rangos de precisión de 2, 2.5, 3 y 4 mm, respectivamente.
- **Reddy et al.** [97], propusieron una arquitectura de atención para abordar eficazmente tanto la apariencia local como el contexto global. Este enfoque permite el entrenamiento estable incluso cuando la cantidad de datos disponibles es baja. La interacción entre estas dos subredes se expresa

mediante una operación de multiplicación punto a punto, como se muestra en la Ecuación (2.34).

$$HMP_i(y) = HMP_i^{LAN}(y) \times HMP_i^{GCN}(y) \quad (2.34)$$

donde la salida de las redes de apariencia local y contexto global son $HMP_i^{LAN}(y)$ y $HMP_i^{GCN}(y)$. Para el punto de referencia L_k^i , con $i \in \{1, 2, \dots, M\}$, donde M representa el total de puntos de referencia y HMP_i es el mapa de calor generado.

Los resultados de la implementación demostraron tasas de detección de 86.28 %, 91.24 %, 94.81 % y 97.58 % para *Test1* y de 75.21 %, 82.63 %, 97.95 % y 94.89 % para *Test2* en el conjunto de datos ISBI IEEE 2015 dentro del rango de error de 2, 2.5, 3 y 4 mm. En el mismo conjunto de datos, la metodología logró 39.2 %, 48.35 %, 73.9 % y 81.23 % dentro de los rangos de precisión clínicamente aceptados.

- **Schwendicke et al. [98]**, realizaron una revisión sistemática y un metaanálisis con el fin de evaluar la exactitud y la evidencia de respaldo del DL en la detección de puntos de referencia cefalométricos tanto en radiografías 2D como 3D. Dos revisores independientes detectaron y extrajeron investigaciones sobre exactitud diagnóstica publicadas entre 2015 y 2020 en las bases de datos Medline, Embase, IEEE y arXiv. Estos estudios utilizaron técnicas de DL con el propósito de identificar puntos de referencia cefalométricos. Los investigadores realizaron un metaanálisis de efectos aleatorios, análisis de subgrupos y metarregresión. Además, se evaluó la calidad de los estudios incluidos utilizando QUADAS-2.
- **He et al. [91]**, introdujeron un novedoso modelo de cascada-refinamiento que aborda las limitaciones del número y la profundidad en redes neuronales mediante el uso de redes neuronales conectadas en cascada (CCNNs) y la transferencia de parámetros entre las estructuras de red apiladas. En este estudio, las CCNNs se consideran aproximaciones discretas de ecuaciones diferenciales ordinarias.
- **Lu et al. [99]**, propusieron un modelo cascade-refine que se entrenó utilizando el conjunto de datos IEEE 2015, se basaron en un enfoque novedoso para la localización de puntos de referencia en el análisis cefalométrico, empleando redes neuronales convolucionales gráficas basadas en parches multiescala. Las relaciones espaciales entre los puntos de referencia se establecen mediante redes neuronales convolucionales de grafos (Spatial Graph Convolutional Networks, GCN) de tres capas. La utilización de GCN se implementó dentro del marco de aprendizaje profundo y se dividió en tres componentes distintos. Los parámetros aprendibles se inicializan con matrices de peso $L^{CNN} \in \mathbb{R}^{F_{IN} \times F_{LW}}$, donde F_{LW} representa los puntos de referencia aprendidos. Los mapas de características finales se representan por $N^F = R^{F_{LW}} L W^F$, donde F_{LV} es el número de vectores de características de entrada. Los mapas de características aprendibles se denotan como $LW^{LIM1}, LW^{LIM2} \in \mathbb{R}^{F_{IN} \times OFV}$, donde OFV representa la dimensión de los mapas de características de salida. La capa de subsiguiente del grafo es caracterizada en la Ecuación (2.35)

$$N^F = R^F (LW_M^{LIM1} + F_{LM2}^{LW}) \quad (2.35)$$

Para actualizar simultáneamente muchos puntos de referencia y acercarlos más a la proximidad de sus respectivos objetivos, se propuso un enfoque en cascada de tipo burdo a fino. El método logró que el 87.93 %, 92.42 %, 86.59 %, 89.32 % y 86.11 %; 83.68 %, 68.67 %, 95.21 % de los puntos de referencia fueran detectados con éxito para Test1 y Test2 dentro de rangos de precisión clínicamente aceptados de 2 mm, 2.5 mm, 3 mm y 4 mm.

- **Uğurlu et al.** [100], involucraron la adquisición de 1620 radiografías cefalométricas laterales, que abarcan un total de 21 puntos de referencia distintos. De estos, 1360 radiografías fueron utilizadas como conjunto de entrenamiento, 140 radiografías se asignaron para el conjunto de validación y 180 radiografías fueron designadas para el conjunto de prueba. Los resultados experimentales indican que el método propuesto obtuvo un SDR promedio de 76.82 %, 83.54 %, 88.22 % y 93.4 % en rangos de 2, 2.5, 3 y 4 mm para el conjunto de prueba con MRE de 3.4 mm y SD de 1.57 mm.
- **Moreno et al.** [101] realizaron un estudio que tuvo como objetivo evaluar la precisión y exactitud de dos programas de software, WebCeph™ y DentalIQ®, en la detección de puntos de referencia anatómicos. Un total de 68 radiografías cefalométricas laterales fueron elegidas para el análisis de 24 puntos de referencia anatómicos. Las disparidades entre los puntos de referencia identificados manualmente y los detectados por software fueron evaluadas. Los dos programas de software, WebCeph™ y DentalIQ®, presentaron una exactitud promedio de 96.57 % y 66.5 %.
- **Lee et al.** [102] propusieron un nuevo enfoque para el reconocimiento efectivo de 19 puntos de referencia utilizando una CNN de paso único. La red mostró resultados iniciales de 96.42 %, 92 %, 95.68 %, 98.46 % y 74.58 %, 81.79 %, 87.53 %, respectivamente, en el conjunto de datos público ISBI IEEE 2015, desafiando los conjuntos de datos de prueba Test1 y Test2.
- **Filho et al.** [103], examinaron la consistencia y confiabilidad del programa Cefbot-AI en la identificación precisa de puntos de referencia en radiografías cefalométricas laterales. Se adquirieron treinta imágenes utilizando Orthophos XG5/Ceph para analizar 19 puntos de referencia. Los puntos de referencia fueron anotados utilizando Cefbot-AI, calibrado (ECal) y examinador de control (ECont). El análisis del rendimiento de ECont y ECal demostró diferencias estadísticamente significativas en algunas ubicaciones cefalométricas a lo largo del eje x, específicamente S, Po y ENA.
- **Yao et al.** [104], utilizaron una CNN para construir un sistema automatizado para identificar y localizar 37 puntos de referencia utilizando 512 radiografías cefalométricas laterales. Esta estimación fue posteriormente refinada utilizando el módulo modificado localmente para mejorar la precisión de los resultados. Los resultados experimentales indican que el método propuesto produjo un SDR promedio de 45.95 %, 89.19 %, 97.30 %, 97.30 %, 97.30 % y 54.05 %, 91.89 %, 97.30 %, 100 %, 100 % y 100 % para los conjuntos de datos de validación y prueba en los rangos de precisión de 1,

1.5, 2, 2.5, 3 y 4 mm. Los valores de MRE con SD fueron 1.127 ± 1.028 mm y 1.038 ± 0.893 mm para los conjuntos de validación y prueba.

- **Innes et al.** [105], propusieron como objetivo principal, el evaluar la viabilidad de utilizar redes neuronales basadas en pulsos acoplados (PCNN) con el propósito de identificar ciertas características craneofaciales en radiografías digitales. Las PCNN se componen de dos compartimentos principales: un compartimento de alimentación y un compartimento de enlace. Las operaciones realizadas en las PCNN, incluyendo la salida final, se muestran en (2.36) a (2.40).

$$FC_{jk}[n_e] = e^{\alpha_{FC}\delta n_e}, F_{jk}[n_e - 1] + SI_{jk} + C_{FC} \sum_{pq} W_{Fjkpq} O_{pq}[n_e - 1] \quad (2.36)$$

$$LC_{jk}[n_e] = e^{\alpha_{LC}\delta n_e}, L_{jk}[n_e - 1] + C_{LC} \sum_{pq} W_{Ljkpq} O_{pq}[n_e - 1] \quad (2.37)$$

$$NI_{jk}[n_e] = FC_{jk}[n]1 + \gamma, LC_{jk}[n_e] \quad (2.38)$$

$$O_{jk}[n_e] = \begin{cases} 1 & \text{si } NI_{jk}[n_e] > DT_{jk}[n - 1] \\ 0 & \text{en otro caso} \end{cases} \quad (2.39)$$

$$DT_{jk}[n_e] = e^{\alpha_{DT}\delta n_e}, DT_{jk}[n - 1] + C_D, TO_{pq}[n_e - 1] \quad (2.40)$$

donde las señales de los compartimentos de alimentación y enlace se denotan como FC_{jk} y LC_{jk} . NI_{jk} es el estado interno de la neurona. O_{jk} está representada como la salida. DT_{jk} y SI son el umbral dinámico y el estímulo de entrada. W_{Fjkpq} y W_{Ljkpq} denotan los pesos sinápticos de los compartimentos de alimentación y enlace respectivamente. Las constantes de la red son C , δ y α . La PCNN tuvo una tasa de detección exitosa de 93.6 % y 88.1 % para segmentar con precisión regiones que contienen tejidos blandos y estructuras esqueléticas, respectivamente, en un conjunto de prueba que incluía 109 imágenes. Durante la extracción de características alrededor del punto Sella, la PCNN pudo lograr una tasa de éxito muy baja de 36.7 %.

- **Ristau et al.** [106], tuvieron como objetivo principal de su investigación el examinar y evaluar la precisión y la fiabilidad en la identificación de puntos de referencia cefalométricos al comparar un software de trazado automatizado que emplea redes neuronales convolucionales (CNN) con trazadores humanos. Dos ortodoncistas marcaron manualmente un total de sesenta cefalogramas, mientras que se utilizó un software de inteligencia artificial denominado *AudaxCeph*® para trazar y reconocer trece puntos de referencia. No se encontraron diferencias estadísticamente significativas entre los ortodoncistas y el programa de trazado automático *AudaxCeph*®, con la excepción de las

coordenadas X y Y del Porion y la coordenada Y del ápice del incisivo inferior (L1 apex).

- **Zhao et al.** [107], introdujeron un enfoque denominado *YOLOV3 basado en muestreo multiescala* (*Multi-Scale based YOLOV3*), que incorpora algoritmos de muestreo multiescala para capturar características superficiales y profundas en diferentes resoluciones. La evaluación del modelo involucró la utilización de dos conjuntos de datos distintos: cefalogramas laterales disponibles públicamente y cefalogramas anteroposteriores (AP) no divulgados. El algoritmo propuesto MS-YOLOV3 demostró un rendimiento mejorado, alcanzando tasas de detección exitosa (SDR) del 80.84 %, 93.75 % y 98.14 % dentro de los rangos de error de 2, 3 y 4 mm respectivamente, para los cefalogramas laterales. De manera similar, para los cefalogramas AP, el algoritmo alcanzó tasas SDR del 85.75 % en 2 mm, 92.87 % en 3 mm y 96.66 % en 4 mm de rango de error.
- **Feghi et al.** [103], propusieron un enfoque para la localización de puntos de referencia cefalométricos. Para determinar la ubicación exacta de un punto de referencia, se realizan dos pasos en el proceso de localización. Se emplea un sistema neurodifuso (NFS, *Neuro-Fuzzy System*) para generar una ventana de expectativa reducida para cada punto de referencia. El modelo fue entrenado con 565 imágenes, identificando 20 puntos de referencia y alcanzando una tasa de reconocimiento del 90 %.
- **Leonardi et al.** [28], propusieron como objetivo evaluar la precisión de las redes neuronales basadas en celdas para identificar de forma automática los puntos de referencia cefalométricos a partir de cefalogramas digitales. Una celda en una matriz $A \times B$ se representa como $CL(i, j)$ y el vecindario de celdas interactivas $NE_r(i, j)$ se define como en la Ecuación (2.41)

$$NE_r(i, j) = \{CL(t, u) : \max(|t - j|, |v - q|) \leq r\} \quad (1 \leq s \leq A; 1 \leq t \leq B) \quad (2.41)$$

41 radiografías cefalométricas laterales fueron analizadas, adquiridas con el equipo Siemens OrthoPhos DS Ceph, que se sometieron a un proceso automatizado para identificar 10 puntos de referencia con un error medio de 0.59 mm en comparación con los procedimientos de marcado manual.

- **Hwang et al.** [74], propusieron un algoritmo modificado de YOLOv3, el cual fue utilizado para entrenar 1028 cefalogramas digitales donde se incluyeron 86 imágenes de prueba, 34 puntos de referencia de tejidos duros y 2 puntos de referencia de tejidos blandos verticales. El estudio observó que, cuando varios examinadores manualmente detectaban los puntos de referencia múltiples veces, existía un alto grado de variabilidad, con un error de detección de 0.97 ± 1.03 mm. En contraste, el sistema de IA identificó de manera confiable las posiciones exactas de los puntos de referencia en pruebas repetidas. La diferencia promedio de detección entre IA y examinadores humanos se encontró entre 1.46 ± 2.97 mm. Entre los examinadores humanos, se observó una diferencia promedio de 1.50 ± 1.48 mm. El error promedio de detección entre la IA y los humanos fue inferior a 0.9 mm.

- **Dai et al.** [108], proponen un mapeo cefalométrico utilizando el enfoque de Redes Generativas Antagónicas (GAN) . El modelo GAN se entrenó en la base de datos de la competencia ISBI 2014 y generó un mapa de distancia para cada punto de referencia. Durante la fase de prueba, se observó que una proporción significativa de los puntos de referencia fueron identificados correctamente dentro de un rango de error, con un SDR de 80 %.
- **Hong et al.** [109], utilizaron la aplicación de Deep Q-network (DQN) y técnicas de doble Deep Q-network (DDQN) en el contexto de la predicción automatizada de puntos de referencia cefalométricos utilizando el conjunto de datos de la competencia ISBI 2015 con un promedio de SDR de 67.33 % y 66.04 % respectivamente para el rango de precisión de 2 mm.
- **Weingart et al.** [110], proponen evaluar la precisión de las arquitecturas de redes neuronales profundas (DNNs), como un marco de aprendizaje profundo basado en DL, en la detección automatizada de puntos de referencia cefalométricos mediante el uso de 60 puntos de referencia en 30 imágenes de TC. El promedio del error de detección obtenido en el estudio fue de 1.45 mm.
- **Ye et al.** [111], se basaron en un conjunto de 43 radiografías cefalométricas laterales para detectar los puntos X e Y de 32 coordenadas marcadas por 3 ortodoncistas experimentados, y el promedio de la diferencia fue de 1.32 mm. La red neuronal entrenada obtuvo una exactitud del 96.4 % dentro de un rango de error de 1.32 a 3.5 mm, y una tasa de éxito promedio del 85 % para puntos de referencia de tejidos blandos dentro de un rango de error de 1.16 a 4.37 mm.
- **Chang et al.** [112], desarrollaron un modelo de detección de puntos de referencia cefalométricos mediante la utilización de un modelo de red de alta resolución usando 2000 cefalogramas laterales para identificar 51 puntos de referencia. El SDR de 2, 2.5, 3 y 4 mm fue de 89.00 %, 94.00 %, 96.33 % y 98.67 % con una MRE de 1.08 mm y una SD de 0.87 mm.
- **Duran et al.** [113], valoraron la precisión comparativa de dos programas de análisis cefalométrico totalmente automatizados, como OrthoDx™ y WebCeph, que utilizan algoritmos de IA con el marcado manual de puntos de referencia humanos utilizando el software informático no automatizado Dolphin 11.8 con radiografías de cráneos de pacientes de entre 12 y 20 años. Con base en los resultados de la prueba de consistencia, se observó un grado de consistencia estadísticamente significativo entre las mediciones obtenidas de Dolphin y OrthoDx™, específicamente en relación con las mediciones angulares.
- **Bao et al.** [114], evaluaron la precisión de la localización y medición automática de puntos de referencia cefalométricos mediante el uso de análisis cefalométrico a través de IA en comparación con el análisis manual asistido por computadora. Se seleccionó un total de 85 pacientes en los cuales se utilizaron cefalogramas laterales reconstruidos (RLCs) obtenidos a partir de CBCT. El estudio incluyó análisis manual asistido por computadora, específicamente utilizando Dolphin Imaging 11.9, así como análisis automático basado en IA, empleando específicamente Planmeca Romexis 6.2, para

identificar y localizar con precisión 19 puntos anatómicos de referencia y obtener 23 mediciones distintas. Los valores de SDR dentro de 1 mm, 2 mm, 2.5 mm, 3 mm y 4 mm fueron de 18.82 %, 58.58 %, 71.70 %, 82.04 % y 91.39 %. La MRE para los 19 puntos de referencia cefalométricos se encontró en 2.07 ± 1.35 mm utilizando el software automático.

- **Neeraja et al.** [115] introdujeron la arquitectura CephXNet, en la cual un modelo CNN personalizado incorporó un bloque de atención Squeeze-and-Excitation (SEB) para la clasificación automática y la predicción de las coordenadas XY de 19 puntos de referencia utilizando el conjunto de datos de cefalogramas laterales IEEE ISBI 2015 grand challenge. El modelo CephXNet alcanzó un SDR de 88.06 %, 94.64 %, 96.87 % y 98.68 % en el conjunto de prueba Test1 y 78.72 %, 83.54 %, 91.26 %, 96.36 % en el conjunto Test2, y 73.94 %, 80.76 %, 86.88 % y 91.97 % para los rangos de 2, 2.5, 3 y 4 mm. Los valores de MRE con SD obtenidos para Test1, Test2 y el conjunto de datos privado fueron $0.92 \text{ mm} \pm 0.76 \text{ mm}$, $1.41 \pm 0.91 \text{ mm}$ y $1.58 \text{ mm} \pm 1.08 \text{ mm}$ respectivamente.
- **Lee et al.** [116], desarrollaron un modelo automatizado novedoso y completo para identificar 19 puntos de referencia cefalométricos en 1032 radiografías posteroanteriores (PA) utilizando técnicas de DL y obtuvieron un SDR de 34.7 %, 67.5 % y 91.5 % dentro de los rangos de error de 1 mm, 2 mm y 4 mm.
- **Zhu et al.** [117], emplearon un Coordinated Transformer (CoorTransformer) para la detección automática de puntos de referencia cefalométricos y lograron un SDR de 79.28 % en 2 mm utilizando 250 imágenes de prueba del conjunto de datos ISBI 2015.
- **Feng et al.** [118], generaron y evaluaron un modelo DL que puede segmentar y cuantificar automáticamente la silla turca en radiografías cefalométricas. Se utilizaron un total de 1129 imágenes con fines de entrenamiento, mientras que un conjunto adicional de 50 imágenes fue empleado para evaluar las capacidades de generalización de la red. La red de segmentación alcanzó un coeficiente de Dice de 92.84 %.
- **Chen et al.** [119], emplearon método de regresión de mapa de calor utilizando una CNN en forma de U para predecir automáticamente puntos de referencia a partir de cefalogramas y alcanzó tasas SDR de 79.65 % para Test1 y 80.05 % para Test2 del conjunto de datos ISBI 2015.
- **Wu et al.** [120], propusieron un codificador híbrido que integra CNN y Swin Transformer para la extracción de características fusionadas de múltiples escalas, junto con capas convolucionales 1D agrupadas para una incrustación eficiente de características. Este enfoque demostró alta precisión y eficiencia en la detección de puntos de referencia cefalométricos, incluso aquellos con texturas sutiles.
- **Juneja et al.** [121], examinaron y evaluaron varios enfoques empleados en la identificación de puntos de referencia cefalométricos, proporcionando un análisis crítico de su efectividad.

- **Londono et al.** [122], tuvieron como objetivo realizar un metaanálisis y una revisión sistemática para evaluar la precisión del software de ML en la identificación y predicción de puntos anatómicos de referencia en imágenes cefalométricas laterales 2D. El meta-análisis se adhirió a los requisitos establecidos por los Elementos de Reporte Preferidos para Revisiones Sistemáticas y Meta-análisis (PRISMA) con el fin de seleccionar y examinar artículos de investigación. Las tasas combinadas de detección efectiva para los rangos de 1 mm, 2 mm, 2.5 mm, 3 mm y 4 mm fueron 65 %, 81 %, 86 %, 91 % y 96 %, respectivamente.
- **Cao et al.** [123], realizaron una búsqueda electrónica exhaustiva en cuatro bases de datos, incluyendo MEDLINE, Embase, IEEE y arXiv, con el objetivo de identificar estudios de precisión diagnóstica que se publicaron en idioma inglés durante el período de enero de 2015 a diciembre de 2020. El enfoque de búsqueda utilizó términos relacionados con la identificación de puntos de referencia, cefalometría, ortodoncia, así como metodologías relevantes que incluyen IA, DL y CNNs. Además, se realizó una búsqueda manual para identificar otras fuentes académicas examinando las listas de referencias.
- **Lin et al.** [124], propusieron como objetivo del estudio evaluar el nivel de conocimiento, experiencia y actitud de ortodoncistas y estudiantes de ortodoncia hacia las tecnologías cefalométricas asistidas por IA. Se realizó una encuesta transversal en línea utilizando una plataforma profesional (www.wjx.cn) durante el período del 11 al 17 de octubre de 2022. Entre los encuestados, el 68.8 % expresó su acuerdo con el posible reemplazo de los procedimientos manuales y semiautomáticos por aplicaciones cefalométricas basadas en IA.
- **Khalid et al.** [125], propusieron un marco de regresión de dos etapas para la detección de puntos de referencia cefalométricos. La etapa inicial predice las ubicaciones de los puntos de referencia identificando regiones potenciales utilizando una sola red neuronal, y un bloque de fusión semántica se incorpora en la segunda etapa para refinar estas predicciones. El modelo obtuvo un SDR de 87.17 % y 75.79 % para los conjuntos de datos Test1 y Test2 en el rango de 2 mm.
- **Strunga et al.** [126], compararon el rendimiento del trazado cefalométrico automatizado por IA con el trazado digital operado por humanos utilizando el software Invivo 7.1.2 en 120 imágenes CBCT, anotadas por cuatro expertos humanos y una IA a través de 14 variables cefalométricas. Se utilizó la prueba de Wilcoxon para evaluar las diferencias en los tiempos de medición entre humanos e IA, revelando diferencias estadísticamente significativas en 6 de las 14 variables.
- **Khan et al.** [127], presentaron una arquitectura CNN en cascada utilizando un nuevo conjunto de datos denominado “Aariz” para la identificación automática de puntos de referencia cefalométricos. El modelo obtuvo un SDR de 78.44 % y 94.44 % en 2 mm y 4 mm respectivamente.
- **Wu et al.** [120] propusieron el enfoque llamado CeLDA incorporando una red prototípica para detectar automáticamente puntos de referencia cefalométricos entre varios grupos de edad utilizando

el conjunto de datos CephAdoAdu. La red obtuvo un SDR promedio de 89.13 % para el grupo combinado Adulto y Adolescente, 88.33 % para Adulto y 89.93 % para Adolescente en un rango de 2 mm.

- **Wang et al. [128]** introdujeron un nuevo conjunto de datos “WebCeph2k” para la detección de puntos de referencia cefalométricos integrando transformación de profundidad espacial y enfoques de decodificación de mapas de calor. El marco alcanzó un SDR de 77.45 % y 79.80 % para los conjuntos de datos de prueba y validación WebCeph2k. La red obtuvo tasas SDR de 88.21 % y 76.26 % en 2 mm para el conjunto de datos ISBI 2015.

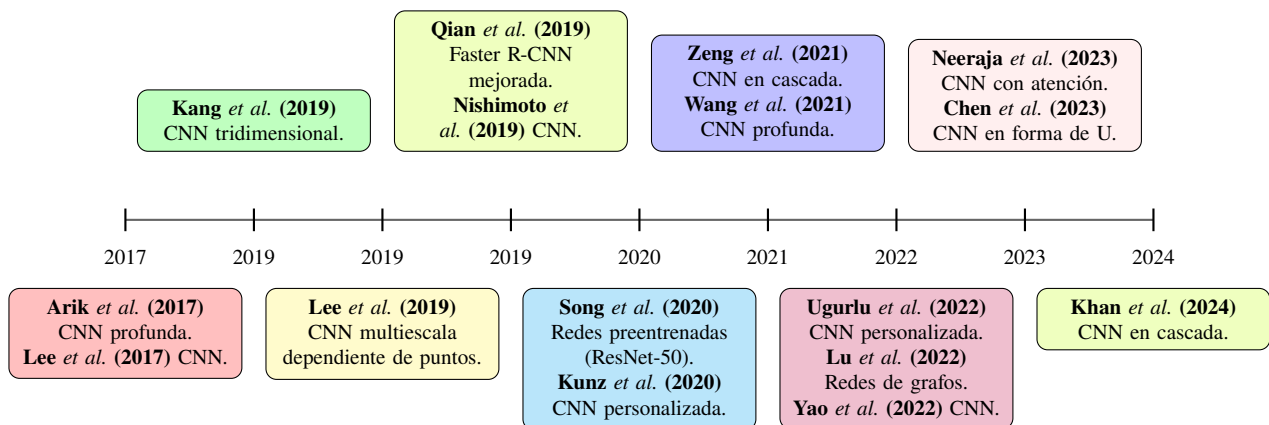


Figura 2.11: Línea de tiempo compacta de arquitecturas CNN aplicadas a la detección automática de puntos cefalométricos.

2.6. Conclusiones del Capítulo

En este capítulo se abordaron los fundamentos teóricos y metodológicos del análisis cefalométrico, destacando su relevancia dentro del diagnóstico ortodóncico y craneofacial. Se revisó el proceso histórico que condujo a la estandarización de la cefalometría radiográfica, así como las etapas involucradas en la captura, trazado y etiquetado de puntos anatómicos de referencia. Dicho procedimiento constituye la base para el estudio de las relaciones morfológicas y funcionales del complejo cráneo-maxilofacial.

Asimismo, se presentaron las distintas estrategias empleadas en la literatura científica para la detección de puntos cefalométricos, clasificadas en enfoques basados en el conocimiento y enfoques basados en inteligencia artificial. Los primeros incluyen métodos de emparejamiento de plantillas, detección de bordes, modelos de apariencia activa y técnicas evolutivas, los cuales, aunque han demostrado cierto grado de precisión, presentan limitaciones importantes ante la variabilidad anatómica y las condiciones heterogéneas de las imágenes radiográficas.

Por otro lado, los métodos modernos sustentados en aprendizaje automático y, particularmente, en aprendizaje profundo, han mostrado una mejora sustancial en la automatización y exactitud del proceso de localización de puntos cefalométricos. Dichos modelos aprenden de grandes volúmenes de datos las características discriminantes necesarias para identificar estructuras anatómicas con una intervención mínima del especialista, lo que representa un avance significativo hacia sistemas de diagnóstico asistido por computadora.

El análisis cefalométrico ha evolucionado desde procedimientos manuales basados en la experiencia clínica hasta metodologías computacionales avanzadas que integran inteligencia artificial. Esta evolución marca la transición hacia una nueva etapa de precisión y eficiencia en el diagnóstico ortodóncico. Con base en esta revisión, el siguiente capítulo se centrará en el estudio detallado de las redes neuronales y sus principios de funcionamiento, con el propósito de establecer los fundamentos para el diseño de modelos convolucionales aplicados a la detección automática de puntos cefalométricos.

Capítulo 3

Redes Neuronales

Desde épocas remotas, el ser humano ha tratado de entender la manera en que funciona el cerebro para descubrir cómo pensamos. Esto ha dado lugar a una gran cantidad de avances en la neurociencia, a tal grado que 10 % de los premios Nobel de Medicina otorgados entre 1901 y 1991 se dieron a personas que contribuyeron a mejorar el entendimiento de la mente humana y los procesos involucrados [129].

Como parte de estos esfuerzos por entender cómo pensamos, se han utilizado diversos enfoques para automatizar tareas “humanas” mediante el uso de las herramientas computacionales para crear algoritmos capaces de resolver dichas tareas.

El objetivo de este capítulo es explicar cada uno de los bloques sobre los que se sustenta la teoría del aprendizaje profundo, el enfoque elegido para la solución del problema presentado en este trabajo, partiendo desde la neurona artificial, el perceptrón multicapa y su funcionamiento para terminar con la aplicación que tienen éstos modelos en la solución de problemas de diversas áreas. Esta introducción se presenta para facilitar la comprensión del Capítulo 4, que es parte fundamental de este trabajo.

3.1. Introducción

La Inteligencia Artificial es una rama de las ciencias de la computación en la que se involucran más áreas del conocimiento, entre las que destacan [130]: Neurociencias, física, matemáticas, estadística, entre otras. La inteligencia artificial busca “automatizar tareas intelectuales normalmente realizadas por humanos” [131].

La inteligencia artificial surge de manera formal en 1956 durante un congreso organizado por John McCarthy en la Universidad de Dartmouth, donde los 10 investigadores asistentes a dicho congreso presentaron diversas ideas y trabajos previos con la idea de replicar tareas humanas. Después de esto han existido diversos episodios en la historia de la inteligencia artificial con periodos de gran expectación,

así como los conocidos “inviernos” de la inteligencia artificial, donde se redujo considerablemente la cantidad de investigación realizada en esta área.

Actualmente es común encontrar aplicaciones de la inteligencia artificial al realizar tareas cotidianas como utilizar el teléfono inteligente o utilizar redes sociales, dichas aplicaciones automatizan tareas utilizando algoritmos de aprendizaje de máquina, esto gracias al incremento considerable en el poder de cómputo de los dispositivos actuales así como la gran cantidad de datos existente para entrenar dichos algoritmos.

3.1.1. Aprendizaje de Máquina

El Aprendizaje de Máquina (*Machine Learning*, ML) es una rama de la inteligencia artificial, el cual busca “aprender” a realizar una tarea mediante el uso de un modelo estadístico construido a partir de un conjunto de datos. Otra definición para el aprendizaje de máquina de acuerdo con [132] es: “Una forma de estadística aplicada con especial énfasis en el uso de computadoras para estimar estadísticamente funciones complicadas”, el modelo estadístico estará basado en los datos de la problemática a resolver.

Dentro de los tipos de aprendizaje de máquina encontramos:

- **Aprendizaje supervisado:** Consiste en la utilización de un conjunto de datos para entrenar el algoritmo con la finalidad de que sea capaz de asociar la entrada x con la salida y . La mayoría de los algoritmos de aprendizaje supervisado tienen como finalidad estimar una densidad de probabilidad $p(y|x)$, a partir de la cual se pueda predecir el valor de la salida dada una entrada arbitraria.

La solución del algoritmo de aprendizaje supervisado se logra encontrando la función:

$$f(X_m) = Y_m \quad (3.1)$$

Que nos servirá para mapear futuras observaciones X_s a Y . En la Ecuación (3.1) X_m corresponde al vector (puede ser vector, matriz o tensor dependiendo del problema a resolver) de características, mientras que Y_m corresponde al vector de etiquetas objetivo o salidas, que actuará como señal de supervisión.

El aprendizaje supervisado se da mediante el ajuste de los parámetros del modelo estadístico utilizando el conjunto de datos de entrenamiento que regularmente se presentan de la siguiente manera:

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1N} \rightarrow Y_1 \\ x_{21} & x_{22} & x_{23} & \dots & x_{2N} \rightarrow Y_2 \\ & & \vdots & & \\ x_{M1} & x_{M2} & x_{M3} & \dots & x_{MN} \rightarrow Y_M \end{bmatrix} \quad (3.2)$$

El enfoque de aprendizaje supervisado es especialmente útil al momento de resolver problemas de regresión y de clasificación.

- **Aprendizaje no supervisado:** A diferencia del aprendizaje supervisado, en este enfoque el conjunto de datos de entrenamiento sólo cuenta con el vector de características $\{x_i\}_{i=1}^M$ a partir del cual se obtendrá un modelo que permita transformar dicho vector en otro o en un valor que pueda utilizarse para resolver el problema en cuestión. Regularmente el aprendizaje no supervisado se utiliza para resolver tareas de clustering, reducción de dimensionalidad o detección de valores atípicos [133].
- **Aprendizaje por refuerzo:** Este tipo de algoritmos interactúa con el ambiente para obtener una retroalimentación del mismo que le permitirá aprender de las experiencias del estado del ambiente, para las cuales obtendrá una recompensa o castigo. En este caso la salida es una política bajo la cual actuará el algoritmo para resolver la tarea.

La Figura 3.1 muestra un desglose de los principales algoritmos de aprendizaje de máquina.

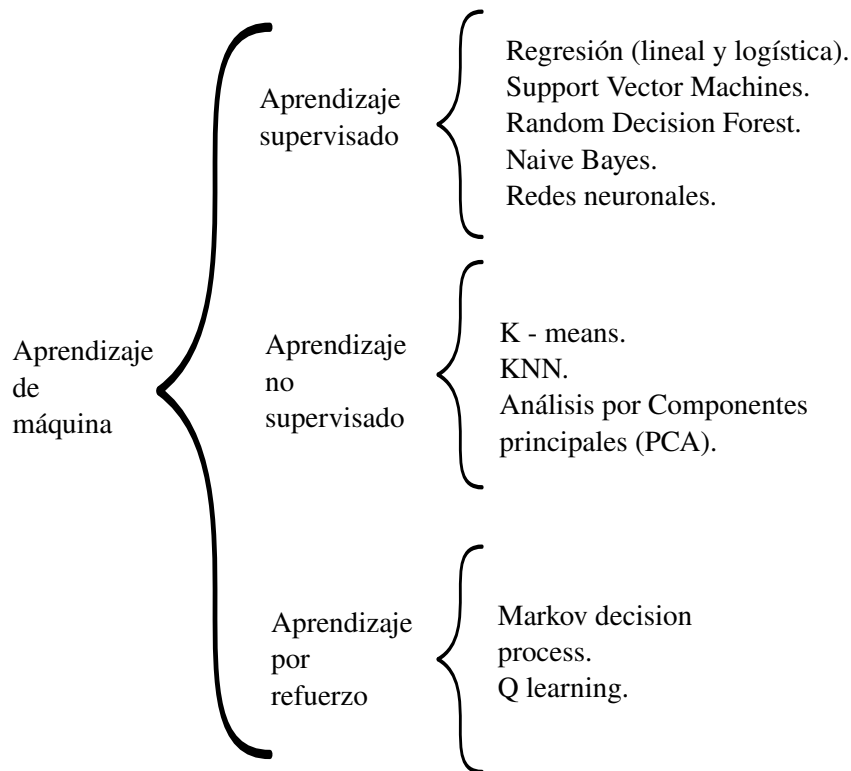


Figura 3.1: Tipos de aprendizaje de máquina, así como algunos de los principales algoritmos utilizados.

3.2. La Neurona Biológica

La neurociencia comprende el estudio del sistema nervioso, haciendo énfasis particular en el cerebro. A pesar de que el interés del ser humano por comprender los procesos racionales se remonta al

menos a tiempos de Aristóteles, ha sido en siglos recientes que se han visto los avances más importantes respecto a esta área a partir de 1873 cuando Camillo Golgi desarrolló un método de tinción que permitió la observación de neuronas individuales [134]. Dicho método fue utilizado y perfeccionado por Santiago Ramón y Cajal para la realización de diagramas de estructuras neuronales a partir de los cuales el profesor Wilhem Waldeyer acuñó el término neurona (del griego *νεῦρον*, nervio) [129].

La neurona es el bloque fundamental del sistema nervioso. El cerebro humano cuenta con aproximadamente 86,000 millones de neuronas y existen alrededor de 10,000 tipos diferentes de neuronas [135]. La invención del microscopio electrónico ayudó a descubrir que todas las neuronas se pueden construir partiendo de los mismos componentes básicos:

- **Soma o cuerpo:** Aquí se ubica el núcleo de la neurona, la mayor parte del citoplasma. También se realiza la síntesis y transporte de proteínas para el funcionamiento de la célula.
- **Dendritas:** Estructuras en forma de rama que parten del cuerpo de la neurona. Su función es recibir señales entrantes de neuronas adyacentes.
- **Axón:** Estructura que sale del cuerpo de la neurona (en específico del cono axónico) y se divide al final en pequeñas ramas. Su función es la transmisión de las señales eléctricas de la neurona a otras células nerviosas o bien a fibras musculares. El axón se encuentra cubierto por un tejido graso que lo aísla y permite que la señal recorra el axón sin pérdidas [136].

La Figura 3.2 muestra las partes fundamentales de la neurona.

La comunicación entre neuronas sucede en la sinapsis. La sinapsis es la unión de una rama axónica con una neurona, ésta unión no se da de manera física puesto que las neuronas se encuentran separadas por una distancia aproximada de 200 nm [137], en lo que se conoce como hendidura sináptica. Vista como un bloque, la sinapsis incluye el final del axón, la dendrita de la neurona adyacente y la hendidura sináptica. La neurona que envía la señal es conocida como neurona presináptica mientras que la que recibe la señal es denominada neurona postsináptica.

El procesamiento de información entre neuronas (transmisión de señales nerviosas) se lleva a cabo a través de fenómenos eléctricos y químicos. Cuando la neurona se encuentra inactiva existe un potencial de reposo dentro de la célula de alrededor de -70 mV [137].

En el interior de la neurona ocurren los fenómenos eléctricos, es decir las descargas eléctricas que viajan desde el cuerpo de la célula hasta el axón. El cuerpo de la neurona funciona como un dispositivo para “sumar” el efecto despolarizador de cada entrada. Si la magnitud de la suma de dichas señales supera el umbral crítico de la membrana (≈ 10 mV), se abre una serie de canales iónicos con los cuales la neurona pierde su impermeabilidad contra iones de Na^+ , lo que libera un impulso eléctrico llamado *potencial de acción*, que viaja desde el cono axónico hasta el axón [135].

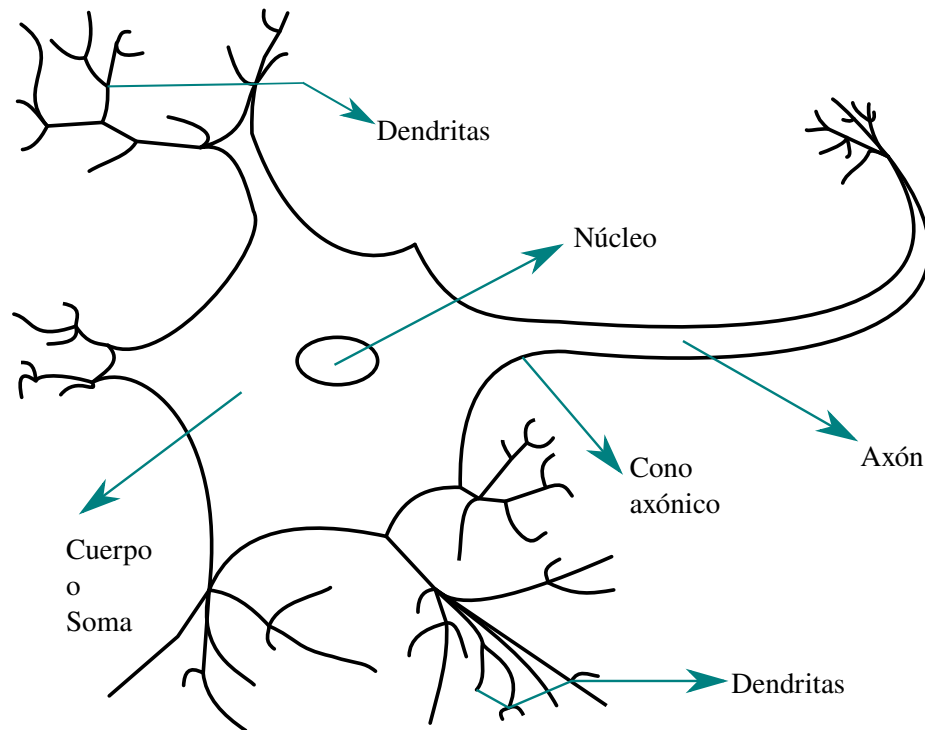


Figura 3.2: Partes fundamentales de una neurona biológica, adaptada de [130, 137, 135].

Con la finalidad de comunicarse (enviar las señales eléctricas), el potencial de acción se transforma en un mensaje químico que cruzará la sinapsis. Éstas señales químicas son conocidas como neurotransmisores y existen del tipo excitatorio e inhibitorio, el neurotransmisor excitatorio más común en el cerebro humano es el glutamato, mientras que el neurotransmisor inhibitorio más común es el ácido gamma-aminobutírico (GABA).

Se estima que cada neurona en el cerebro humano recibe conexiones de aproximadamente 10^4 sinapsis [130]. En el contexto de la neurona biológica, el aprendizaje se puede entender como la creación de nuevas conexiones entre neuronas así como la modificación de las existentes, que es posible gracias a la plasticidad sináptica.

3.3. Antecedentes Históricos

El campo del aprendizaje de máquina y en especial la rama de las redes neuronales, al igual que muchas ramas de la ciencia, han tenido un desarrollo no lineal, con periodos de crecimiento exponencial y expectativas muy altas, así como periodos de escaso o nulo interés en su desarrollo, conocidos como inviernos.

El avance reciente en el campo de las redes neuronales se debe principalmente a dos razones: El aumento en la cantidad de bases de datos suficientemente grandes para entrenar modelos de aprendizaje profundo, a tal grado que sólo en el ámbito médico se espera una generación de datos de 2,314 Exabytes en el 2020 [138], sumado a las mejoras en hardware para el procesamiento paralelo de datos, creciendo a una razón de 1000 veces cada década desde el año 2000, lo cual permitió utilizar a gran escala los algoritmos desarrollados en años anteriores.

Una vez que se tuvo cierto grado de conocimiento de cómo funcionaba la neurona biológica en su forma más fundamental, surgieron intentos de modelar su comportamiento, realizando muchas suposiciones para limitar la complejidad del modelo. A continuación se presenta una línea del tiempo con algunos de los acontecimientos que permitieron los avances actuales.

3.4. La Neurona Artificial

La neurona artificial es la unidad básica de procesamiento o cómputo dentro de una red neuronal. A pesar de que existen varias similitudes entre la neurona biológica y la artificial es importante aclarar que la neurona artificial para nada iguala la complejidad de una neurona biológica, la cual es comparable con un microprocesador actual [130]. La Figura 3.3, muestra el esquema de una neurona artificial, cuya salida está dada por la expresión:

$$\hat{y} = g \left(\sum_{i=1}^n w_i x_i + b \right), \quad (3.3)$$

o expresado de manera vectorial:

$$\hat{y} = g(w^T x + b), \quad (3.4)$$

donde:

$\hat{y}$ = Salida computada por la neurona. Depende de la aplicación, puede ser una clase o un valor continuo, aplicada a clasificación o regresión lineal respectivamente.

w = Vector de pesos sinápticos.

x = Vector de entrada de la neurona.

b = Umbral, también conocido como *offset* o *bias*, es un escalar.

g = Función de activación, determina si la neurona se “activará”.

Es fácil notar las similitudes de funcionamiento entre este modelo simplificado con las partes de la neurona biológica presentada anteriormente: Las terminales o nodos de entrada asemejan las dendritas, los pesos sinápticos a las sinapsis entre neuronas y el sumador al cuerpo de la neurona biológica.

Tabla 3.1: *Sucesos que definieron el desarrollo de las redes neuronales.*

1920	●	Durante esta década, Nicolás Rashevsky intenta modelar la neurona con ecuaciones diferenciales para aplicarla al estudio del sistema nervioso [139].
1943	●	Warren McCulloch y Walter Pitts proponen una teoría de procesamiento de información utilizando modelos de neuronas binarias [140], en el que el estado siguiente de la neurona será dictado por la influencia de otras neuronas, expresando este estado como una combinación lineal de sus valores de entrada [137].
1949	●	Donald Hebb publica su libro “The organization of behavior”, donde demuestra una regla para modificar conexiones entre neuronas [136]. Actualmente esto se conoce como aprendizaje Hebbiano.
1958	●	Frank Rosenblatt, conocido como padre de la neurocomputación, presenta el perceptrón, en el que introduce pesos numéricos en el modelo de la neurona [141].
1969	●	Marvin Minsky y Seymour Papert publican el libro “Perceptrons” [142], donde muestran las limitaciones del perceptrón, lo que ocasiona que se pierda el interés en el desarrollo de nuevos modelos.
1986	●	Rumelhart et al. [143] muestran la aplicación del algoritmo de backpropagation para redes con neuronas como unidades de cómputo.
1987	●	Se realiza la primera conferencia de redes neuronales [136].
1988	●	Surge la Sociedad Internacional de Redes Neuronales (INNS por sus siglas en inglés).
1989	●	Yann LeCun presenta la aplicación exitosa de redes neuronales convolucionales a reconocimiento óptico de caracteres [144].
1990	●	Se publica por primera vez el <i>IEEE Transactions on Neural Networks</i> [136].
1997	●	Hochreiter et al. presentan las redes <i>Long-Short Term Memory</i> , como una extensión de las redes neuronales recurrentes [145]
1998	●	LeCun et al. presentaron una red neuronal convolucional más eficiente y eficaz llamada LeNet-5 cuyo éxito preparó el camino para el surgimiento de las redes neuronales profundas del siglo XXI.
2006	●	NVIDIA lanza CUDA (Arquitectura Unificada de Dispositivos de Cómputo), que permitió mayor paralelización de cálculos en las GPUs.
2012	●	Krizhevsky et al. ganan la competencia de ImageNet 2012 utilizando una red neuronal convolucional para identificar las 1000 clases presentes en el dataset de la competencia [146].
2014	●	Goodfellow et al. presentan las redes neuronales generativas adversariales [147].
2015	●	Google hace su biblioteca Tensorflow software <i>open source</i> .

3.4.1. Funciones de Activación

La función de activación $g(z)$, también conocida como función de transferencia permite conocer si la neurona se activará o disparará, similar al potencial de acción necesario para superar el umbral en

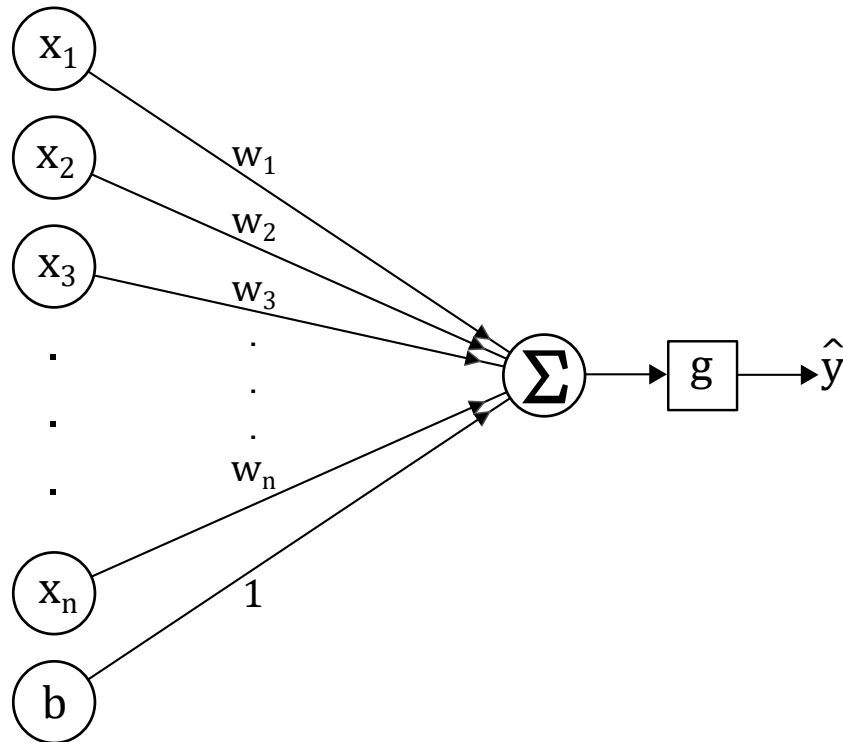


Figura 3.3: Neurona artificial y sus elementos.

las neuronas biológicas. La Tabla 3.2 muestra algunas de las funciones de activación más utilizadas en el contexto de redes neuronales artificiales, así como su derivada que será importante en la etapa del ajuste de los pesos sinápticos de la neurona.

Es importante mencionar la función de activación debe de ser preferentemente diferenciable ya que esto hará que la función calculada por la neurona, y por lo tanto por la red neuronal en arquitecturas con más de una neurona, sea diferenciable de cara a la aplicación de un algoritmo de optimización de primer grado para minimizar el error presente en nuestra red.

3.4.2. Aprendizaje

Dentro de la neurona artificial, el aprendizaje consiste en ajustar los valores del vector de pesos sinápticos w de manera que la salida del algoritmo difiera lo menor posible de la salida esperada. Para cumplir esta tarea es necesario definir una regla de aprendizaje, con la cual podamos modificar los pesos de manera iterativa y con cada iteración, reducir el error a la salida.

Para ilustrar el concepto de regla de aprendizaje se tomará el ejemplo del “perceptrón”, una red con una neurona, utilizada para realizar tareas de clasificación binaria, es decir, establecer una línea que separe en dos grupos los datos de entrada. En ésta arquitectura se define una métrica para el error e en

cada iteración como la siguiente:

$$e = y - \hat{y} \quad (3.5)$$

Una vez que se ha definido el error, podemos definir las reglas de actualización de acuerdo con los 3 posibles valores del error [148]:

- Si $e = 1$, entonces $w_{i+1} = w_i + x_i$
- Si $e = -1$, entonces $w_{i+1} = w_i - x_i$
- Si $e = 0$, entonces $w_{i+1} = w_i$

Es posible tomar las tres consideraciones anteriores para formar expresiones de la regla de aprendizaje:

$$w_{i+1} = w_i + ex_i = w_i + (y - \hat{y})x_i \quad (3.6)$$

Y para el caso del umbral:

$$b_{i+1} = b_i + e \quad (3.7)$$

Una clara desventaja de las redes con una neurona es que están limitadas a resolver problemas con datos de entrada linealmente separables, debido a la naturaleza de las operaciones vectoriales que realizan, en la que la frontera de separación de los datos es una línea recta ortogonal al vector w definida por la expresión:

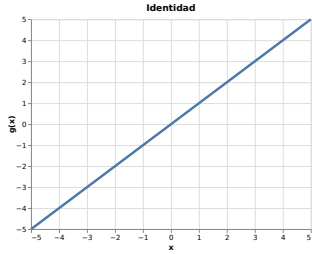
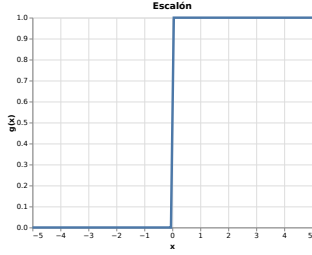
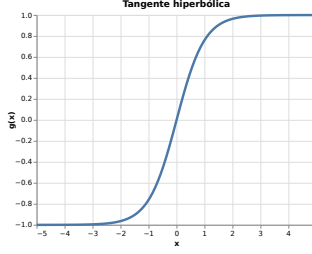
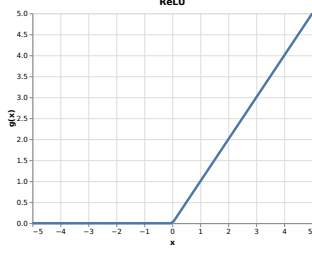
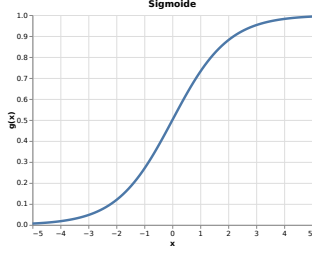
$$w^T x + b = 0 \quad (3.8)$$

A pesar de que este tipo de redes se pueden aplicar a una gran variedad de problemas, la gran mayoría de problemas involucran datos de entrada multidimensionales, un ejemplo con datos no separables linealmente es la función computada por la compuerta lógica XOR (or exclusiva), de la misma manera la Figura 3.4 muestra la gráfica de la función, donde se observa que es imposible separar el espacio de entradas con una línea recta.

3.5. Perceptrón Multicapa

El problema del cómputo de la compuerta lógica XOR, así como cualquier función Booleana (entendiendo función Booleana como el mapeo de $\{0, 1\}^n$ a $\{0, 1\}$), se puede resolver utilizando un perceptrón que contenga una capa oculta y la arquitectura adecuada, dando como resultado un perceptrón multicapa.

Tabla 3.2: Funciones de activación comunes utilizadas en el ámbito de redes neuronales.

Función de activación	Ecuación	Derivada	Gráfica
Identidad	$g(x) = x$	$g'(x) = 1$	
Escalón o <i>hardlimit</i>	$g(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$	$g'(x) = \begin{cases} 0 & \text{si } x \neq 0 \\ \text{ind. COC} & \end{cases}$	
Tangente hiperbólica	$g(x) = \frac{2}{1+e^{-2x}} - 1$	$g'(x) = 1 - g(x)^2$	
Unidad lineal rectificadora (ReLU)	$g(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases}$	$g'(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$	
Softmax	$g(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$	—	—
Sigmoide	$g(x) = \frac{1}{1+e^{-x}}$	$g'(x) = g(x)(1 - g(x))$	

Un perceptrón multicapa (MLP por sus siglas en inglés), también conocido como red neuronal *feedforward* es una estructura de procesamiento de información distribuida en paralelo representado

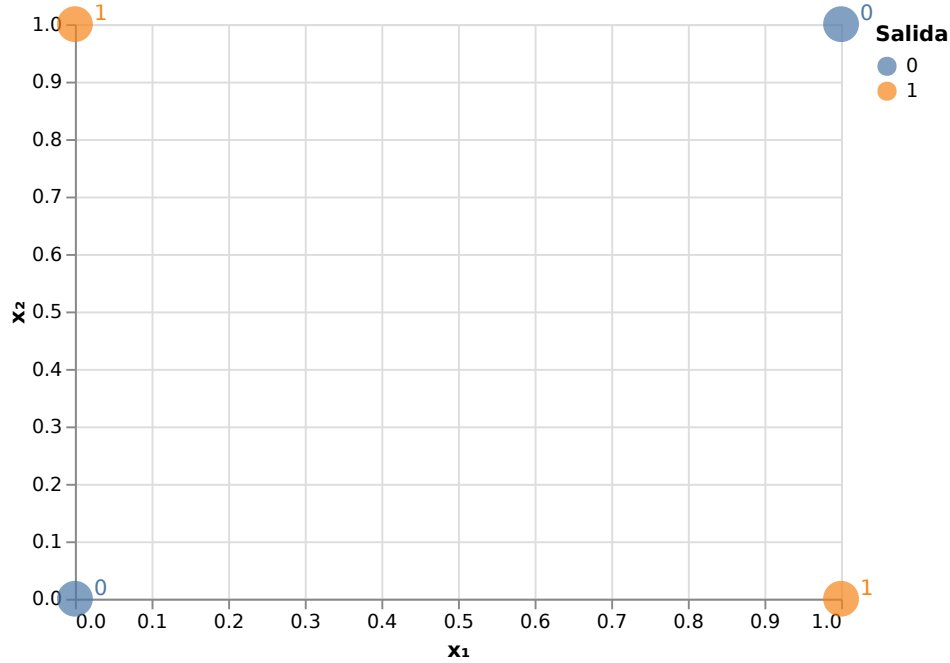


Figura 3.4: *Función Booleana XOR.*

por un grafo acíclico que se construye a partir de las neuronas presentadas al inicio de la Sección 3.4, realizando arreglos verticales de neuronas conocidos como capas, las cuales se interconectan para lograr transformaciones cada vez más complejas mediante el uso de funciones de activación que apliquen transformaciones no lineales a los datos de entrada. La primera capa de izquierda a derecha es conocida como la capa de entrada, mientras que la capa final es conocida como capa de salida.

La Figura 3.5 muestra un diagrama de un perceptrón multicapa con una capa oculta.

El cómputo a realizar dentro del perceptrón multicapa de la Figura 3.5 está definido por:

$$\mathbf{h}^{(1)} = g^{(1)}\left(\mathbf{w}^{(1)\top}\mathbf{x} + b^{(1)}\right) \quad (3.9)$$

$$\mathbf{y} = g^{(2)}\left(\mathbf{w}^{(2)\top}\mathbf{h}^{(1)} + b^{(2)}\right) \quad (3.10)$$

Podemos extender las capacidades de la red añadiendo más capas ocultas, donde la entrada de dicha capa será la salida de la capa anterior. De esta manera podemos considerar a la red neuronal como una máquina que realiza composiciones de funciones para aproximar la función que resuelva el problema a tratar con los datos de entrada que tiene, pues la salida para una red con tres capas estará dada por:

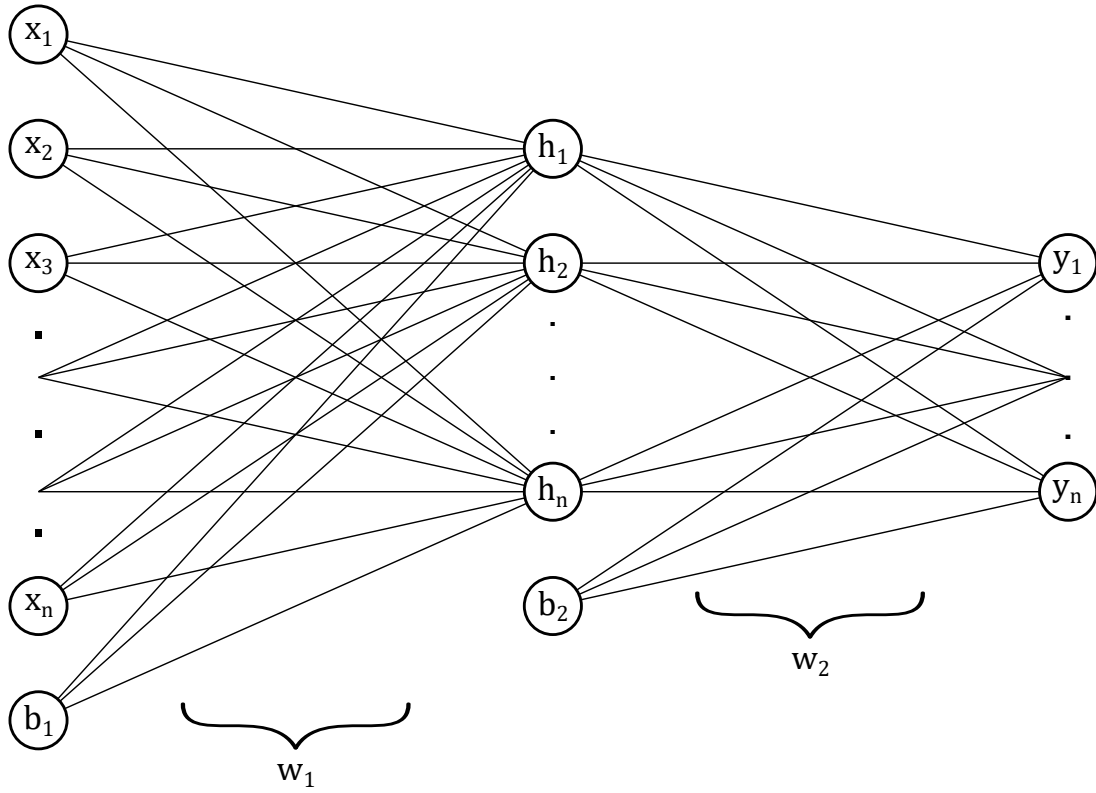


Figura 3.5: Diagrama de un perceptrón multicapa.

$$f(x) = f^{(3)}\left(f^{(2)}\left(f^{(1)}(x)\right)\right) \quad (3.11)$$

donde $f^{(1)}$ es la salida o transformación computada por la capa uno, $f^{(2)}$ la salida de la capa dos y así sucesivamente. La longitud general de la cadena de funciones determinará la profundidad del modelo, mientras que la dimensionalidad de las capas ocultas determinará el ancho de este [132].

La Figura 3.6 muestra un perceptrón con una capa oculta, capaz de resolver el problema de la compuesta lógica XOR, definida como: $y = x_1 \oplus x_2 = (x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge x_2)$ [129]; la primera neurona de la capa oculta calcula $x_1 \wedge \neg x_2$, la segunda neurona $\neg x_1 \wedge x_2$ y la neurona final realiza el cálculo de la función OR. Las tres neuronas utilizan la función de activación:

$$g(x) = \left\{1 \text{ si } \sum_{i=1}^n w_i x_i \geq 0.5; 0 \text{ si } \sum_{i=1}^n w_i x_i < 0.5\right\} \quad (3.12)$$

Donde:

- La operación lógica XOR se define como $x_1 \oplus x_2$, que es verdadera si exactamente uno de los dos valores es verdadero.

- La forma equivalente en términos de AND ($\wedge$), OR ($\vee$) y negación ($\neg$) es:

$$y = (x_1 \wedge \neg x_2) \vee (\neg x_1 \wedge x_2).$$

Esto significa que y es verdadero si x_1 es verdadero y x_2 es falso, o si x_1 es falso y x_2 es verdadero.

- La función de activación $g(x)$ se define como:

$$g(x) = \begin{cases} 1 & \text{si } \sum_{i=1}^n w_i x_i \geq 0.5 \\ 0 & \text{si } \sum_{i=1}^n w_i x_i < 0.5 \end{cases}$$

Es decir, la neurona final devuelve 1 si la suma ponderada de las entradas supera 0.5, y devuelve 0 en caso contrario.

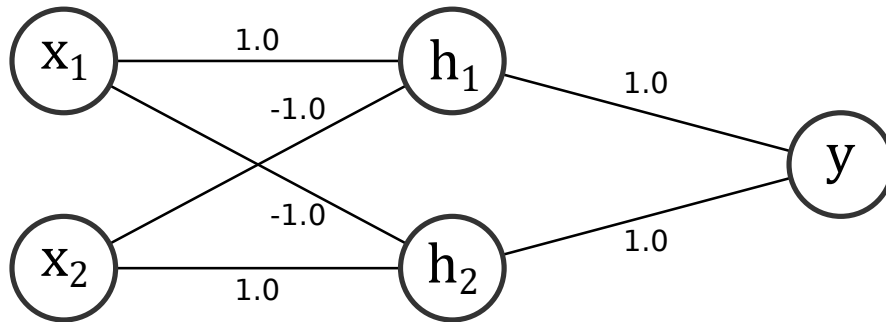


Figura 3.6: MLP para la solución de XOR, adaptado de [129].

3.5.1. Tipos de Redes Neuronales

Como se ilustró con la Figura 3.6, los cálculos a realizar dentro de la red están determinados por la manera en la que se conectan las neuronas, lo que da origen a diversas maneras de ubicar las neuronas, así como crear distintos tipos de capas que se adaptan mejor a algunos tipos de datos específicos como imágenes o texto.

De acuerdo con [149], podemos clasificar a las redes neuronales en cinco grupos:

- Feedforward Neural Network: Este tipo de redes no tiene ningún tipo de ciclos dentro de su arquitectura. La información fluye de izquierda a derecha dentro del grafo acíclico que representa la organización de las neuronas en el modelo.
- Redes Neuronales Recurrentes (RNN): Este tipo de redes presenta ciclos de retroalimentación dentro de su arquitectura, en los que la salida de esa neurona es la entrada a la misma unidad de cómputo, con lo que se crea una memoria acerca de los estados anteriores. Regularmente las RNN se aplican a problemas que involucran datos secuenciales como voz y texto.

- **Redes Neuronales de Base Radial:** La salida en este tipo de redes depende de la distancia euclidiana de la entrada a un centro c , este cálculo se implementa en la que se conoce como función radial y se encuentra en la capa oculta de la red. Este tipo de redes se utilizan para propósitos de clasificación y constan de tres capas.
- **Mapas Autoorganizados de Kohonen:** Es un modelo de red neuronal que realiza aprendizaje no supervisado. Cuentan sólo con la capa de entrada y capa de salida, también conocida como mapa topológico o mapa de características. Se utilizan para propósitos de agrupamiento y representación de datos.
- **Redes Neuronales Modulares:** Consisten en un conjunto de redes neuronales que forman módulos independientes que realizan tareas específicas que se combinan para obtener la salida de la red.

3.5.2. Aprendizaje en Redes Neuronales de Más de Una Neurona

La manera en que sucederá la transformación de los datos de entrada, también conocida como representación, depende del valor de los pesos sinápticos en cada una de las capas de la red neuronal, por lo que es conveniente que la transformación que realiza la capa esté parametrizada por sus pesos, con lo que se obtiene:

$$\phi(x; \theta) \tag{3.13}$$

donde:

ϕ = Transformación, regularmente no lineal, que se aplicará a los datos para poder aproximar la función f del problema a resolver.

x = Datos de entrada del modelo.

θ = Pesos sinápticos (w), también conocidos como parámetros del modelo.

La finalidad de entrenar el algoritmo es aprender ϕ a través del uso de algoritmos de optimización que permitan encontrar los valores de θ , tal que se obtenga una buena representación [132, 131]. Para lograr esto es necesario definir una función de costo $J(\theta)$, también conocida como función objetivo o función de pérdida, que cuantificará el error de predicción de nuestro modelo con el valor actual de θ en cada iteración basado en una métrica previamente establecida dependiendo de la naturaleza del problema a tratar. Algunas de las funciones de costo comunes son entropía cruzada para problemas de clasificación o error cuadrático medio (MSE, por sus siglas en inglés) y error medio absoluto (MAE, por sus siglas en inglés) para tareas de regresión.

Una vez que tenemos definida la función que será minimizada, es necesario definir el algoritmo de optimización. Gracias a que todas nuestras funciones computadas son diferenciables (incluyendo la función de transferencia) podemos utilizar un enfoque basado en el descenso del gradiente. La Figura 3.7

muestra de manera gráfica el descenso del gradiente aplicado a una función de costo dependiente de un solo parámetro.

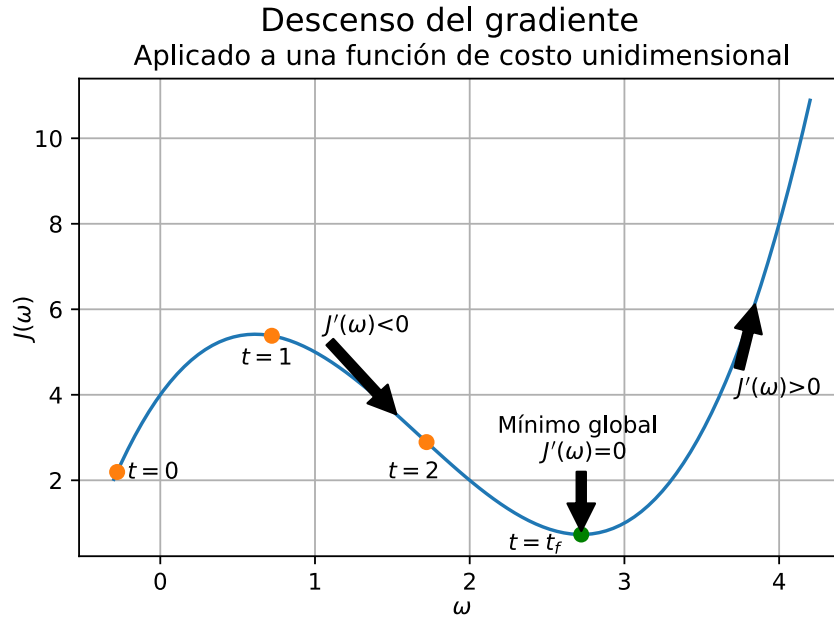


Figura 3.7: Descenso del gradiente para una función unidimensional.

Del cálculo diferencial se conoce que es posible encontrar el mínimo de una función de manera analítica utilizando la derivada, para funciones multivariables como el caso de las redes neuronales, es necesario recurrir al concepto de gradiente, el cual se calcula de la siguiente manera:

$$\nabla_x y = \left(\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \dots, \frac{\partial y}{\partial x_i} \right)^T, \quad (3.14)$$

donde:

$\nabla_x y$ = Gradiente de y , calculado respecto al tensor x .

$\frac{\partial y}{\partial x_i}$ = Derivada parcial de y respecto del i -ésimo término del tensor x .

Ahora es posible calcular el gradiente de nuestra función de costo respecto de los vectores de pesos sinápticos presentes en las l capas del modelo:

$$\nabla_w J = \left(\frac{\partial J}{\partial w_1}, \frac{\partial J}{\partial w_2}, \dots, \frac{\partial J}{\partial w_l} \right) \quad (3.15)$$

Utilizando la Figura 3.7 como referencia se puede observar que si la derivada (o gradiente en su defecto) de la función es negativa es necesario moverse hacia la derecha con la finalidad de reducir el

valor de $J(\omega)$ y para el caso donde se tiene una derivada positiva es necesario desplazarse hacia la izquierda para el mismo fin.

3.5.3. Backpropagation

La retropropagación o *backpropagation*, es un algoritmo que permite calcular el gradiente a través del uso de la regla de la cadena para las derivadas, haciéndolo computacionalmente eficiente. Fue aplicado de manera exitosa en redes neuronales por primera vez en 1986 por Rumelhart et al. [143].

El gradiente se calcula, a grandes rasgos, en dos etapas:

- **Propagación hacia adelante:** Aquí la información fluye de izquierda a derecha en el grafo, es decir, de la capa de entrada a la capa de salida, pasando por las capas ocultas. La salida de esta etapa es la predicción del modelo con el valor de los parámetros en ese momento, es decir, se realiza la evaluación de la composición de funciones realizada por el modelo. Se compara dicha salida con la esperada para calcular la métrica de error.
- **Propagación hacia atrás:** Inicia a partir de la capa de salida hacia la capa de entrada, propagando el error calculado entre cada una de las capas. Cada neurona recibe una parte del error proporcional a su influencia en el cálculo de la salida que originó ese error. El gradiente se calcula aplicando la regla de la cadena de manera recursiva con técnicas de memorización, donde se almacenan los resultados para evitar recalcular la misma función, hasta llegar a la expresión deseada.

3.5.4. Optimizadores

Una vez que se calcula el gradiente de la función de coste con respecto a los parámetros del modelo, es posible realizar la minimización de esta utilizando el método de optimización que más se acople al problema a resolver. Un ejemplo de esto es el algoritmo del descenso del gradiente por lotes para realizar la actualización de manera iterativa en cada época de los valores de los pesos sinápticos del modelo, de acuerdo con la siguiente ecuación:

$$w_{i+1} = w_i + \Delta w_i = w_i - \alpha \frac{\partial J}{\partial w_i} \quad (3.16)$$

Que es equivalente a aplicar:

$$\theta = \theta - \alpha \cdot \nabla_{\theta} J(\theta) \quad (3.17)$$

Donde α es conocido como tasa de aprendizaje, se recomienda utilizar un valor pequeño, ya que controla el tamaño de la actualización. Con la ecuación anterior, los valores de los pesos sinápticos serán actualizados de acuerdo a la magnitud del error que tienen en la salida del modelo.

Algunos de los optimizadores más utilizados actualmente son mejoras realizadas al descenso del gradiente, entre los que destacan:

- Descenso del gradiente estocástico (SGD): Computa el gradiente en mini-lotes de los datos de entrenamiento en vez de calcular el gradiente completo sobre todos los datos disponibles, con lo que se evitan algunos problemas del descenso del gradiente por lote como cálculos redundantes para conjuntos de datos extensos con ejemplos similares [150].
- Propagación del error cuadrático medio (RMSProp): Optimizador de tasa de aprendizaje variable propuesto por Geoffrey Hinton, es una versión en mini-lotes de rprop, que busca utilizar el signo del gradiente para adaptar el tamaño del paso de manera separada para cada peso sináptico [151].
- Adam: Optimizador de tasa de aprendizaje variable, el cual combina los beneficios de RMSProp y del descenso del gradiente estocástico con momentum para aplicarse a problemas con gradientes muy ruidosos [152].
- AdaGrad: Utiliza el algoritmo de gradiente adaptativo para generar la tasa de aprendizaje variable o adaptativa [153], regularmente converge más rápido que el SGD.

3.5.5. Validación del Modelo

Dentro del modelo a desarrollar, existen parámetros e hiperparámetros. Los parámetros son el conjunto de valores w que se aprenden durante la etapa de entrenamiento. Los hiperparámetros son aquellos valores se pueden modificar al momento de compilar el modelo, algunos de los hiperparámetros son:

- Número de capas ocultas.
- Número de neuronas en cada capa.
- Funciones de activación.
- Función de pérdida.
- Tipo de optimizador y tamaño de lote.
- Tasa de aprendizaje α .
- Número de épocas de entrenamiento.

Los hiperparámetros del modelo deben de ser ajustados con la finalidad de obtener un mejor resultado en el modelo a desarrollar. El ajuste de los hiperparámetros se realiza utilizando un conjunto de datos de validación o desarrollo, datos que el modelo no ha visto con los cuales se mide la métrica de éxito con los valores actuales del modelo. El tamaño del conjunto de datos de validación dependerá del

tamaño del conjunto de datos disponible. Se recomienda separar el conjunto de datos en tres porciones, una para entrenamiento, una para validación y una para prueba. El porcentaje que tendrá cada porción estará determinado por la cantidad de datos.

Para conjuntos de datos pequeños se utilizan métodos de validación cruzada como “Leave One Out Cross-Validation”, donde se utilizan como datos de entrenamiento todos los datos disponibles excepto uno, que será utilizado como conjunto de validación, este proceso se realiza tantas veces como se tengan muestras y el error es el promedio de los errores calculados en cada iteración. Otro método de validación cruzada que no tiene un costo computacional tan alto es el “K-Fold Cross Validation”, donde se separan k grupos y se entrena en $k - 1$ grupos y se valida en el grupo restante.

3.5.6. Sobreajuste y Regularización

A medida que se avanza en el proceso de entrenamiento es importante graficar a la par la pérdida en entrenamiento y la pérdida de validación, con la finalidad de detectar sobreajuste. El sobreajuste u *overfitting* se refiere al momento cuando el modelo comienza a ajustarse demasiado bien a los datos de entrenamiento, a tal grado de modelar parte del ruido existente en los datos o características que no son relevantes para la solución del problema. La Figura 3.8 muestra la diferencia entre un modelo con menor complejidad de la necesaria (subajuste, también conocido como *underfitting*), con un buen ajuste, y con sobreajuste.

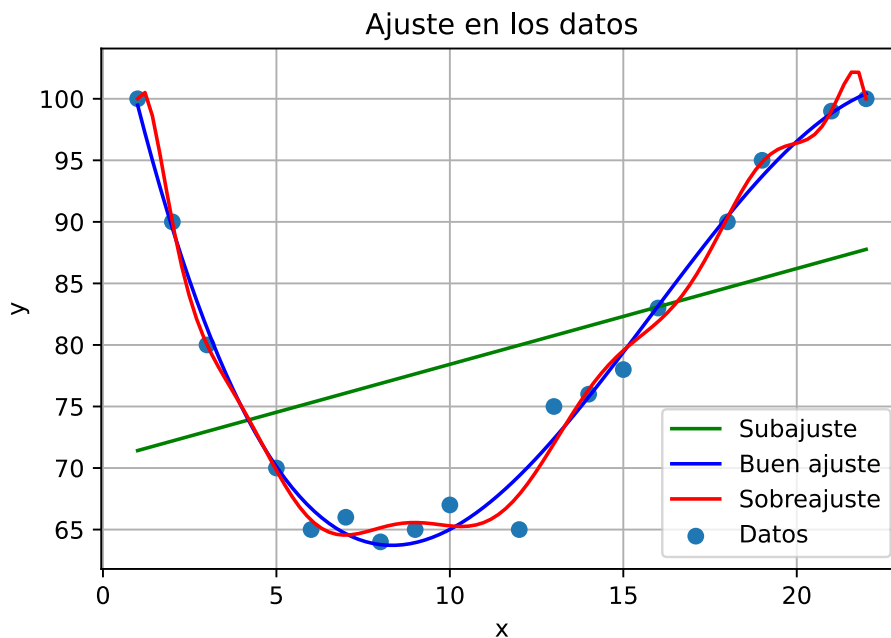


Figura 3.8: Diferencia entre subajuste, buen ajuste y sobreajuste de un modelo.

A pesar de que la línea roja de la Figura 3.8 parece ajustarse mejor a los datos dados, se tiene sobreajuste, pues la complejidad del modelo es mayor de la necesaria para modelar los datos y el

modelo termina por memorizar los datos, con lo que se reduce la capacidad del modelo para generalizar ante datos que nunca se han visto.

Una de las maneras más sencillas para detectar cuando el modelo desarrollado presenta sobreajuste, es graficar las pérdidas de entrenamiento y validación contra las épocas de entrenamiento. El sobreajuste se presenta cuando las dos curvas dejan de tener una forma similar. La Figura 3.9 muestra dos curvas para ilustrar el momento en el que se presenta el sobreajuste, al igual que cuando se tiene subajuste.

Con la finalidad de reducir o evitar el sobreajuste, además de la reducción de la complejidad del modelo o detener prematuramente el entrenamiento al detectar los primeros signos de sobreajuste, se desarrollaron técnicas de regularización, que permiten penalizar mediante un término regularizador $\Omega(\theta)$ a la función de costo J a optimizar, en el caso de las redes neuronales se penaliza a los pesos sinápticos del modelo.

De acuerdo con [132], regularización es “cualquier modificación realizada a un algoritmo de aprendizaje dirigido a reducir el error de generalización, pero no el de entrenamiento”. La función de costo regularizada tendrá la siguiente forma:

$$\tilde{J}(\theta; X; y) = J(\theta; X; y) + \alpha\Omega(\theta) \quad (3.18)$$

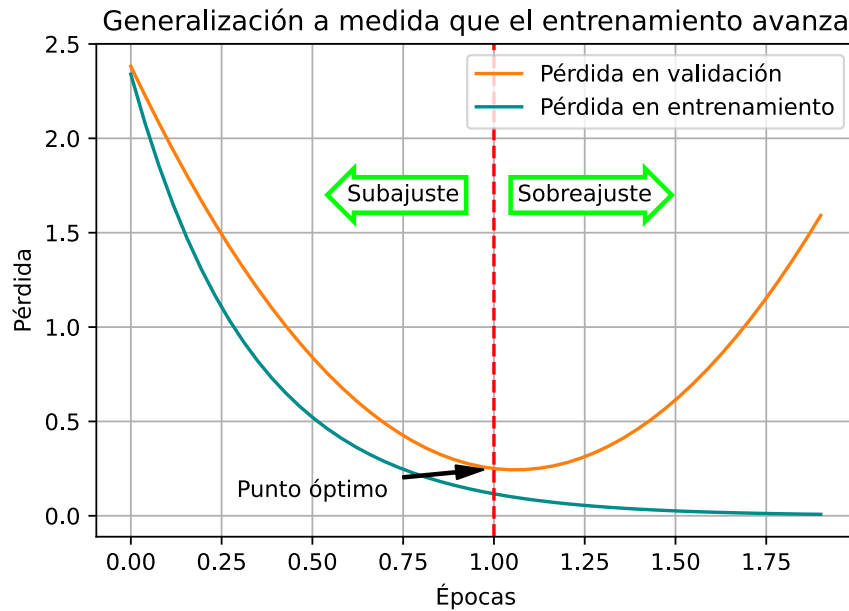


Figura 3.9: Curvas de pérdida en entrenamiento y en validación y su relación con el sobreajuste.

Una de las técnicas de regularización más populares en el área de las redes neuronales es el *dropout*, propuesto por Srivastava *et al.* [154] en el que de manera aleatoria se eliminan neuronas que

no pertenezcan a la capa de salida, así como las conexiones que la involucran, con lo que se reduce la complejidad del modelo de una manera computacionalmente eficiente.

La Figura 3.10 muestra el efecto del *dropout* sobre un perceptrón multicapa.

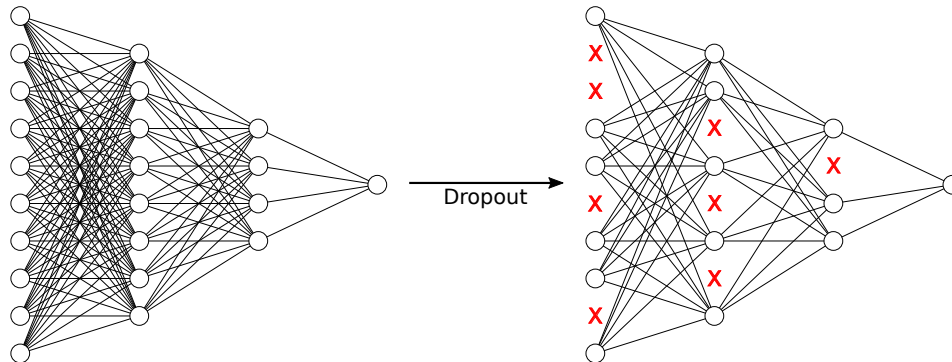


Figura 3.10: Efecto del dropout sobre un perceptrón multicapa.

3.6. Aprendizaje Profundo

Gracias a los avances de los últimos años en hardware de procesamiento paralelo, la cantidad masiva de datos etiquetados con los que contamos hoy en día, así como el desarrollo de nuevos algoritmos permitieron el avance y aplicación de técnicas de aprendizaje profundo a diversos problemas. El aprendizaje profundo es una rama del aprendizaje de máquina que hace uso de múltiples capas de procesamiento computacional para aprender representaciones de los datos con un nivel mayor de abstracción [155]. La Figura 3.11 muestra la relación existente entre inteligencia artificial, aprendizaje de máquina y aprendizaje profundo. La mayor ventaja del aprendizaje profundo en comparación con los otros enfoques existentes en el aprendizaje de máquina es que elimina, de manera práctica, la necesidad de una etapa de extracción manual de características al realizar el modelo, pues las capas generan niveles incrementales de representación de los datos, que les permiten extraer de manera automática las características relevantes para resolver la tarea en turno.

El hecho de que el aprendizaje profundo elimine la extracción manual de características permite que los modelos de aprendizaje profundo sean aplicados a diversos tipos de datos como imágenes, audio o series de tiempo, que permiten resolver problemas de áreas muy diversas.

Dado que el aprendizaje profundo ha tenido un gran impacto en diversas áreas de nuestra vida, desde medicina hasta marketing o finanzas, la Asociación para la Maquinaria Computacional (ACM por sus siglas en inglés), entregó el Premio Turing del 2018 a los investigadores Yoshua Bengio, Geoffrey Hinton y Yann LeCun por los avances que permitieron considerar al aprendizaje automático como un componente crítico en el área del cómputo, pues en años recientes técnicas de aprendizaje profundo

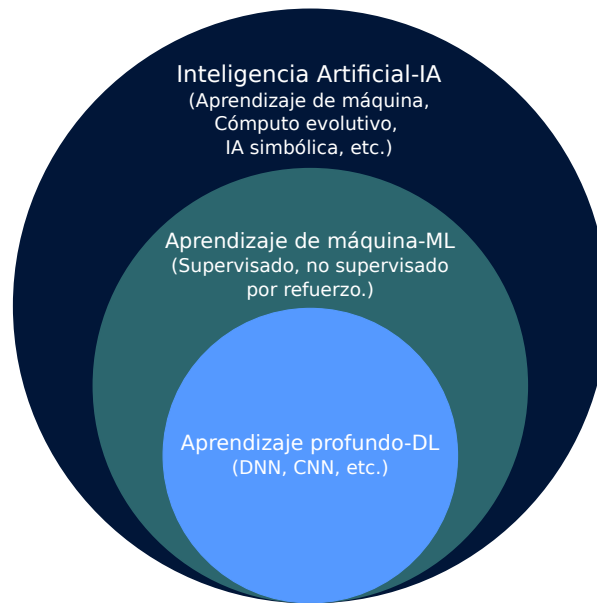


Figura 3.11: *Relación de pertenencia entre Inteligencia Artificial, Aprendizaje de Máquina y Aprendizaje Profundo.*

han obtenido resultados espectaculares en aplicaciones tan diversas como generación de texto con GPT3 [156], detección de objetos con YOLOv4 [157] o plegamiento de proteínas con Alphafold [158], entre muchos otros modelos.

Dentro de las redes neuronales *feedforward* más utilizadas se encuentran las redes neuronales convolucionales (CNN, por sus siglas en inglés), que utilizan convoluciones sobre imágenes para aprender filtros que pueden “descomponer” a la imagen en sus características más importantes, de manera similar a lo que realiza una transformada de Fourier para generar un banco de funciones sinusoidales [131].

El diseño y la inspiración biológica que tienen las redes neuronales convolucionales les permiten capturar algunas características importantes del córtex visual primario de los mamíferos por lo que son ideales para trabajar con imágenes. Una revisión más extensa se realizará a lo largo del Capítulo 4, puesto que es la arquitectura seleccionada para atacar el problema presentado en este trabajo.

3.7. Conclusiones del Capítulo

En este capítulo se expusieron los fundamentos teóricos que sustentan el funcionamiento de las redes neuronales artificiales, destacando su origen inspirado en la estructura biológica del sistema nervioso humano. Se analizó la evolución histórica de los modelos neuronales, desde los primeros

perceptrones hasta las arquitecturas multicapa, las cuales introdujeron el concepto de aprendizaje supervisado mediante el algoritmo de retropropagación del error.

Se revisaron también los elementos esenciales que definen a una red neuronal, tales como las funciones de activación, los mecanismos de aprendizaje y los algoritmos de optimización, cuya correcta selección influye de manera decisiva en el rendimiento del modelo. Asimismo, se discutieron las principales estrategias para evitar el sobreajuste, incluyendo la regularización y la validación cruzada, con el fin de mejorar la capacidad de generalización de los sistemas de aprendizaje automático.

El capítulo permitió comprender que las redes neuronales constituyen una herramienta fundamental para la modelación de relaciones no lineales en datos complejos, siendo capaces de aproximar funciones arbitrarias mediante procesos iterativos de ajuste de pesos. No obstante, también se reconoció que las arquitecturas tradicionales presentan limitaciones al enfrentarse a datos de alta dimensionalidad, como imágenes, debido a la pérdida de información espacial durante el procesamiento.

Finalmente, se introdujo el concepto de aprendizaje profundo (*Deep-Learning*), el cual extiende los principios de las redes neuronales convencionales mediante el uso de múltiples capas jerárquicas que permiten extraer representaciones progresivamente más abstractas de los datos. Este enfoque constituye la base conceptual para el estudio de las redes neuronales convolucionales, que serán analizadas en el siguiente capítulo como un paradigma especializado en el procesamiento y reconocimiento de patrones en imágenes.

Capítulo 4

Redes Neuronales Convolucionales

Debido a la gran capacidad de las redes neuronales para reconocer patrones en cantidades masivas de datos, la investigación en esta rama del aprendizaje automático creció mucho, lo que propició la creación de nuevas arquitecturas optimizadas para algún tipo de dato en específico, como las imágenes. Podemos tratar a las imágenes como tensores de dimensión $(w \times h \times d)$, donde w representa el ancho, h la altura, y d los canales de la imagen, para una imagen a escala de grises $d = 1$ mientras que para una imagen a color en RGB $d = 3$.

En este capítulo se presentan los conceptos fundamentales para entender las redes neuronales convolucionales, su fundamento matemático, su relación con el sistema visual humano, así como las aplicaciones que tienen este tipo de redes.

4.1. Introducción

En los perceptrones multicapa, cada entrada es una característica diferente del conjunto de datos del problema, lo que lo hace propicio para trabajar con datos sin estructura definida. Para el caso de las imágenes cada píxel será una entrada a la red neuronal, con la que se tendría una complejidad de ejecución de $\mathcal{O}(w \times h \times d)$. Este enfoque hace que la cantidad de parámetros del modelo crezca muy rápido a medida que la resolución de la imagen de entrada aumenta, por ejemplo, para una imagen de 1920×1080 píxeles en escala de grises tendríamos 2,073,600 parámetros tan solo en la capa de entrada y por cada capa l que se quiera añadir, el modelo incrementará $(tam_{l-1} + 1) \cdot tam_l$ parámetros.

Además del problema que implica trabajar con tantos parámetros a la entrada, regularmente los datos de entrada se vectorizan para facilitar el manejo de los datos, este enfoque hace que se pierdan las relaciones espaciales existentes en las imágenes de entrada, con lo que se dificulta la tarea. Las redes neuronales convolucionales resuelven este problema, al aprender patrones locales de la entrada.

4.2. Similitud Biológica

El sistema visual humano involucra alrededor del 30 % del córtex cerebral humano [135], puesto que no solo incluye al ojo, sino también partes del sistema nervioso central. De la misma manera que con el cerebro, se han realizado diversos estudios para comprender el sistema visual humano y cómo sucede todo el proceso, desde la transformación de la luz visible en imágenes, hasta su interpretación.

4.2.1. Procesamiento de la Luz

Los estudios más importantes para entender el proceso de transformación de la luz en señales eléctricas se realizaron en ratones. En el sistema visual humano, este proceso inicia con la córnea haciendo el enfoque inicial y recibiendo la luz, que pasará a la pupila y posteriormente al iris que regulará la luz que entra. La transformación en señales eléctricas se da en la retina, una continuación del sistema visual que se encuentra organizada en capas de neuronas, principalmente de tres tipos: Fotorreceptores, neuronas integradoras y las células ganglionares.

Después de pasar por la córnea y el cristalino, la luz incide en los fotorreceptores que la transforman en pulsos eléctricos que pasan a través de las células ganglionares y las neuronas integradoras, que, aunque no responden directamente a estímulos de luz, permiten el procesamiento y transmisión de información. Las neuronas integradoras o de asociación permiten la conexión entre distintas partes del cerebro. Los fotorreceptores son células sensibles a la luz que convierten la luz en señales eléctricas que estimulan diversos procesos fisiológicos y cada retina humana tiene aproximadamente 150 millones de fotorreceptores [159].

Los conos y bastones son fotorreceptores encargados de darle a la percepción de la imagen un gran nivel de detalle. Los conos permiten ver colores debido a los tres tipos existentes en el ojo humano, que reaccionan a un rango distinto de colores y su traslape nos otorga el espectro visual conocido. Los bastones permiten la percepción de luz, con ello es posible ajustar nuestra percepción del brillo y mejorar nuestra visión nocturna y periférica, además de que constituyen el 95 % de los fotorreceptores en los seres humanos.

Una vez que la entrada al sistema visual es procesada por las diversas capas de la retina se mapea como una imagen bidimensional invertida, tal como ocurre en las cámaras oscuras. La imagen de salida será enviada al cerebro mediante el nervio óptico, formado por los axones de las células ganglionares y que conecta directamente la parte posterior del ojo con el cerebro. Dicha señal es recibida en el núcleo geniculado lateral para posteriormente ser enviada al córtex visual para su procesamiento.

4.2.2. Procesamiento Visual a Alto Nivel

El córtex visual primario, también conocido como V1 o córtex estriado, es una hoja delgada de tejido neuronal en la que se realiza procesamiento de estímulos visuales y que contiene células que tienen campos receptivos más complejos y son capaces de reaccionar a estímulos en forma de líneas en distintas orientaciones o bordes [135]. La información generada en ésta región será combinada en otras áreas para crear campos receptivos cada vez más selectivos que sean capaces de distinguir patrones más complejos.

El procesamiento a más alto nivel se estudió principalmente en simios y gatos. David Hubel y Torsten Wiesel publicaron en 1962 [160] un estudio realizado en el córtex visual de los gatos, donde descubrieron las células simples, que responden a patrones de barras, y las células complejas, que responden a características similares a las de las simples pero son más robustas ante desplazamientos espaciales. Posteriormente uno de los estudios más importantes para conocer el funcionamiento del sistema visual de los primates también fue realizado por Hubel y Wiesel en 1968 [161].

A partir del conocimiento de modelado y funcionamiento del córtex visual de algunos animales, se formularon algunas de las funciones existentes en las redes neuronales convolucionales como las capas convolucionales y las capas de *pooling* están fuertemente inspiradas en el funcionamiento de las células simples y células complejas [155], respectivamente.

4.3. Antecedentes e Historia Reciente

El cognitrón y neocognitrón son los antecedentes inmediatos de las redes neuronales convolucionales. Éstos modelos fueron propuestos por Kunihiko Fukushima en 1975 y 1980 respectivamente. El cognitrón [162], propone una red neuronal multicapa que funciona bajo dos condiciones principales para la modificación de sus pesos sinápticos, para imitar células sensibles a distintas orientaciones presentadas anteriormente en modelos del córtex visual.

El neocognitrón [163], una extensión del cognitrón, tiene una estructura similar al modelo de Hubel y Wiesel. El neocognitrón presenta una arquitectura piramidal que reduce la resolución de la imagen a medida que se acerca a la salida. El neocognitrón busca imitar el funcionamiento de las células simples para detección de características locales y células complejas que hagan al modelo robusto ante desplazamientos y distorsiones locales, trabajando en conjunto para tareas de reconocimiento de patrones.

La primera aplicación exitosa de las redes neuronales convolucionales la realizó Yann LeCun en 1989 en los laboratorios Bell de AT&T, utilizando una red para el reconocimiento de caracteres en códigos postales del servicio postal estadounidense a partir de la imagen normalizada de 16×16

píxeles como entrada a la red. El modelo utilizó un conjunto de datos de 9298 imágenes etiquetadas, 7291 separadas para el entrenamiento y 2007 imágenes para la validación del mismo [144].

Fue hasta 2012 que ocurrió el resurgimiento de las redes neuronales convolucionales, con el triunfo de la arquitectura AlexNet presentada por **Alex Krizhevsky et al.** [146], en el desafío a gran escala de reconocimiento visual (ILSVRC, por sus siglas en inglés) utilizando el conjunto de datos de ImageNet que incluía 1.3 millones de imágenes repartidas en 1000 clases diferentes. AlexNet ganó el desafío por un margen muy por encima de los resultados ganadores en ediciones anteriores.

A partir del hito que representó el triunfo de la arquitectura AlexNet se incrementó la cantidad de investigación relacionada a las redes neuronales convolucionales y aplicaciones de las mismas. La Sección 4.7 muestra algunas de las aplicaciones actuales de este tipo de modelos.

4.4. La Operación de Convolución

La convolución es una operación que se realiza entre 2 funciones $x, k : \mathbb{R}^d \rightarrow \mathbb{R}$ y se define de la siguiente manera:

$$s(t) = (x * k)(t) = \int x(z)k(t - z)dz, \quad (4.1)$$

donde:

x = Función de entrada.

k = Función kernel.

s = Salida de la convolución.

La operación de convolución $(*)$, tiene las siguientes propiedades [164]:

- Conmutatividad: $x(t) * k(t) = k(t) * x(t)$
- Distributiva: $x(t) * [k_1(t) + k_2(t)] = x(t) * k_1(t) + x(t) * k_2(t)$
- Asociativa: $[x(t) * k_1(t)] * k_2(t) = x(t) * [k_1(t) * k_2(t)]$

También es posible definir la convolución discreta para x y k discretas:

$$s(t) = (x * k)[t] = \sum_{z=-\infty}^{\infty} x[z]k[t - z] \quad (4.2)$$

4.4.1. Convolución en Imágenes

Para realizar la convolución sobre una imagen $I(i, j)$ bidimensional de al menos un canal, es necesario definir un *kernel* o filtro $K(i, j)$ que será deslizado sobre la entrada para obtener la salida o mapa de características. Al ser K e I funciones discretas podemos expresar la convolución como:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n) K(i - m, j - n) \quad (4.3)$$

Aplicando la propiedad conmutativa de la convolución se tiene:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (4.4)$$

Al escribir la Ecuación 4.4 sin invertir el kernel y manipulando los índices, se obtiene la operación de correlación cruzada, la cual es la operación que está implementada en la mayoría de las bibliotecas de aprendizaje automático [132], y que es equivalente a la convolución para propósitos de aprendizaje de máquina.

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n) K(m, n) \quad (4.5)$$

La Figura 4.1 ilustra la manera en que se efectúa la convolución 2D en algunas bibliotecas de aprendizaje automático como Keras [165] con la configuración por defecto, en la que únicamente se consideran las posiciones “válidas” del filtro dentro de la imagen, con lo cual se obtiene un resultado de tamaño menor al de la entrada.

Para el caso particular de la Figura 4.1 se realiza la convolución entre una imagen de 5×5 y un filtro de 3×3 (matriz verde), que para este caso particular es un filtro de Sobel para calcular bordes en dirección x , así como el proceso para obtener los elementos que conforman la salida. Las matrices en azul son conocidas como campo receptivo, por su similitud con el campo receptivo biológico, que es la porción del espacio visual donde un estímulo luminoso puede provocar una respuesta en una célula ganglionar retiniana.

4.5. Capas Utilizadas en Redes Neuronales Convolucionales

Las redes neuronales convolucionales presentan algunas características importantes por las que destacan al momento de trabajar con imágenes, entre las que se encuentran [132, 131, 166]:

0	1	2	3	4
5	6	7	8	9
10	11	12	1	2
3	4	5	6	7
8	9	10	11	12

 $\ast$

-1	0	1
-2	0	2
-1	0	1

 $=$

8	-4	-4
8	-16	-16
8	-4	-4

Donde cada elemento se calcula de la siguiente manera:

0	1	2	3	4
5	6	7	8	9
10	11	12	1	2
3	4	5	6	7
8	9	10	11	12

 $\ast$

-1	0	1
-2	0	2
-1	0	1

 $=$

8		

$$(0)(-1) + (1)(0) + (2)(1) + (5)(-2) + (6)(0) + (7)(2) + (10)(-1) + (11)(0) + (12)(1)$$

0	1	2	3	4
5	6	7	8	9
10	11	12	1	2
3	4	5	6	7
8	9	10	11	12

 $\ast$

-1	0	1
-2	0	2
-1	0	1

 $=$

8	-4	

$$(1)(-1) + (2)(0) + (3)(1) + (6)(-2) + (7)(0) + (8)(2) + (11)(-1) + (12)(0) + (1)(1)$$

0	1	2	3	4
5	6	7	8	9
10	11	12	1	2
3	4	5	6	7
8	9	10	11	12

 $\ast$

-1	0	1
-2	0	2
-1	0	1

 $=$

8	-4	-4

$$(2)(-1) + (3)(0) + (4)(1) + (7)(-2) + (8)(0) + (9)(2) + (12)(-1) + (1)(0) + (2)(1)$$

0	1	2	3	4
5	6	7	8	9
10	11	12	1	2
3	4	5	6	7
8	9	10	11	12

 $\ast$

-1	0	1
-2	0	2
-1	0	1

 $=$

8	-4	-4
8		

$$(5)(-1) + (6)(0) + (7)(1) + (10)(-2) + (11)(0) + (12)(2) + (3)(-1) + (4)(0) + (5)(1)$$

0	1	2	3	4
5	6	7	8	9
10	11	12	1	2
3	4	5	6	7
8	9	10	11	12

 $\ast$

-1	0	1
-2	0	2
-1	0	1

 $=$

8	-4	-4
8	-16	-16
8	-4	-4

0	1	2	3	4
5	6	7	8	9
10	11	12	1	2
3	4	5	6	7
8	9	10	11	12

 $\ast$

-1	0	1
-2	0	2
-1	0	1

 $=$

8	-4	-4
8	-16	-16
8	-4	-4

Figura 4.1: Convolución para imágenes, realizada de la misma manera que las bibliotecas de aprendizaje de máquina más comunes.

- **Invariancia a la traslación:** Puesto que los filtros se aplican de manera idéntica en cada imagen, los patrones aprendidos serán detectados sin importar la ubicación donde aparezcan.
- **Principio de localidad:** Las capas se enfocan en detectar patrones en regiones locales, sin importar

el contenido en las demás zonas de la imagen. Dichos patrones serán utilizados para construir representaciones más abstractas a medida que la profundidad del modelo se incrementa.

4.5.1. Capa Convolutiva

La capa convolutiva es la base de las redes neuronales convolucionales y permite trabajar con datos en forma de arreglos multidimensionales como las imágenes. Dicha capa utiliza operaciones lineales y no lineales para extraer las relaciones espaciales existentes en los datos de entrada, así como en las capas anteriores. Dentro de los hiperparámetros de la capa convolutiva encontramos el padding y el paso, los cuales se describen a continuación:

- **Padding:** El padding, también llamado relleno en español, se refiere a los píxeles adicionales que se añaden al borde de la imagen antes de realizar la convolución. El relleno se utiliza para ampliar el tamaño de la salida y de esta manera obtener una salida con las mismas dimensiones espaciales que la entrada. Aún cuando el relleno se realiza utilizando valores de cero en los píxeles añadidos (conocido como *zero-padding*), la imagen resultante es útil para manipularse en capas subsecuentes. La Figura 4.2 muestra el efecto de distintos valores de relleno para una imagen original de 3×3 con un filtro de 2×2 y 3×3 respectivamente.

0	0	0	0	0
0	0	1	2	0
0	3	4	5	0
0	6	7	8	0
0	0	0	0	0

 $\ast$

0	1
2	3

 $=$

0	3	8	4
9	19	25	10
21	37	43	16
6	7	8	0

(a) Relleno de 1 píxel.

0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	1	2	0	0
0	0	3	4	5	0	0
0	0	6	7	8	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

 $\ast$

0	1	2
3	4	5
6	7	8

 $=$

0	8	23	20	12
24	58	100	70	36
63	132	204	132	63
36	70	100	58	24
12	20	23	8	0

(b) Relleno de 2 píxeles.

Figura 4.2: Efecto del relleno aplicado a imágenes.

- **Paso:** El paso, también conocido como *stride*, en el contexto de convolución define la zancada entre cada convolución, es decir la distancia que existe entre cada campo receptivo. Un paso pequeño

evitará saltar información que pueda ser importante, pero se verá reflejado en una mayor cantidad de operaciones. La Figura 4.3 muestra el efecto de distintos valores del paso para una imagen de 8×8 .

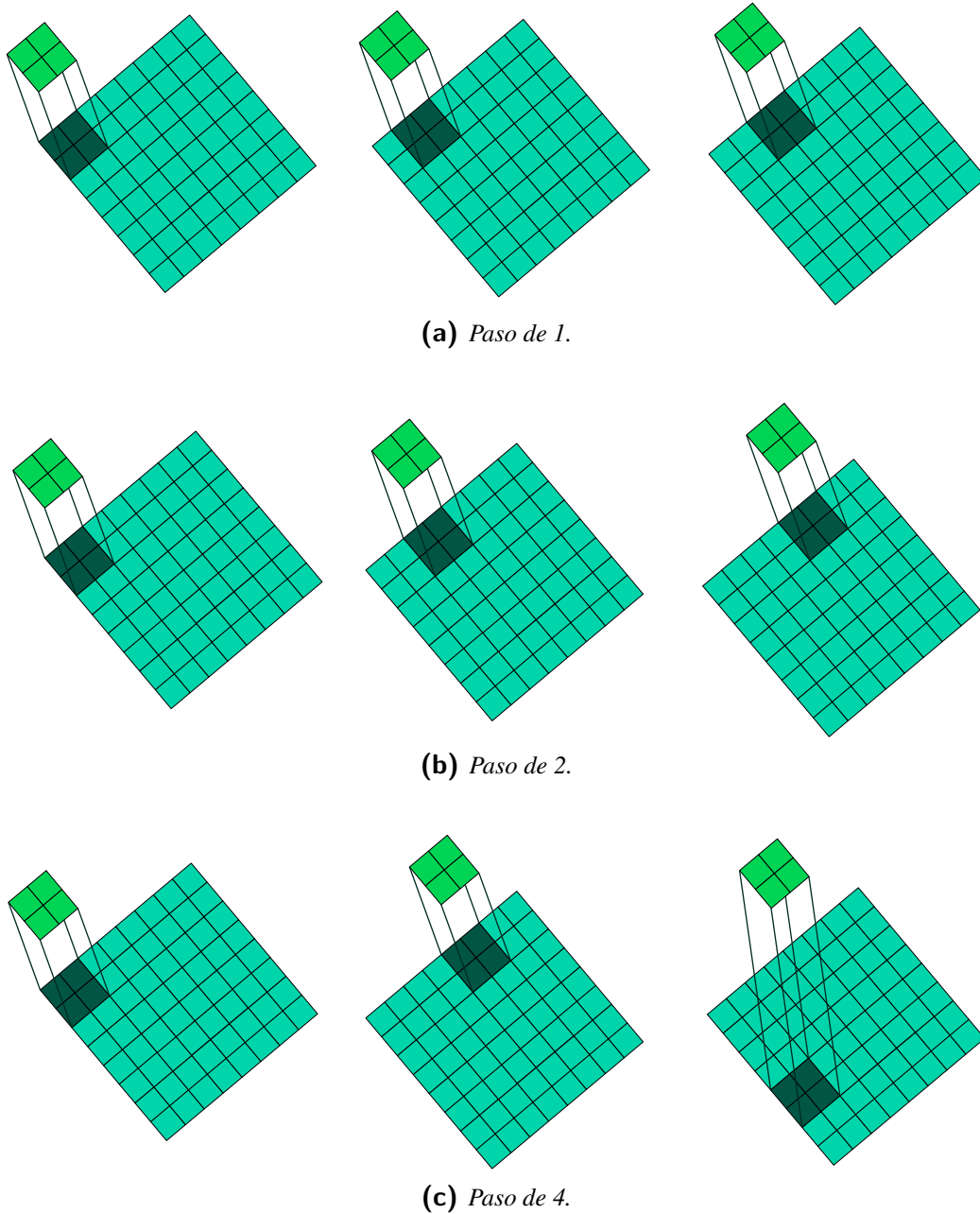


Figura 4.3: Efecto en la variación del paso aplicado a imágenes.

4.5.2. Funciones Pooling

Para ayudar a reducir la dimensión espacial del mapa de activación (la salida de una capa convolucional), se utiliza una capa llamada *pooling* (agrupación), la cual reduce el número de parámetros

del modelo, con la particularidad de que no hay mucha pérdida de información, puesto que se están uniendo características similares en una. La capa de agrupación realiza una operación predefinida sobre todo el campo receptivo (región de donde recibe información cada neurona) y entrega un único valor de salida. La reducción de parámetros del modelo al aplicar pooling, ayuda a mitigar el problema de sobreajuste al reducir la sensibilidad del modelo a ubicaciones y también a incrementar la velocidad de entrenamiento.

Algunas operaciones de pooling existentes son:

- Agrupación máxima: Conocida como *max pooling*, entrega el valor máximo de la entrada.
- Agrupación promedio: Su salida es el promedio de los valores de entrada recibidos.
- Agrupación estocástica: Propuesta por **Zeiler et al.** [167], la selección del valor en la ventana deslizante será seleccionada mediante un procedimiento estocástico con lo que se vuelve una operación que no requiere hiperparámetros. Es posible utilizar esta operación como regularizador para combatir el sobreajuste.

La Figura 4.4 muestra cómo se realizan las operaciones de pooling mostradas anteriormente para una imagen de 4×4 con un filtro de 2×2 y un paso de 2.

4.6. Aprendizaje en Redes Convolucionales

Durante la etapa de entrenamiento, se aprenden los valores de los pesos sinápticos (incluyendo los umbrales) de cada uno de los filtros presentes en nuestra arquitectura, con la finalidad realizar la convolución entre el filtro y el mapa de características precedente de las capas anteriores y así extraer características relevantes de los datos que permitan resolver la tarea en turno. La Figura 4.5 muestra seis de los 64 filtros aprendidos para cada canal en la segunda capa dentro de la arquitectura VGG16 [168], una arquitectura de red neuronal convolucional profunda de 16 capas del estado del arte para tareas de clasificación.

El proceso de entrenamiento de las redes neuronales convolucionales permite encontrar los valores de los parámetros del modelo. Dicho proceso se lleva a cabo bajo el mismo enfoque de descenso del gradiente para la optimización de la función de costo y retropropagación para el cálculo del gradiente. Algunos de los optimizadores utilizados en redes neuronales convolucionales son: Propagación del error cuadrático medio (RMSProp) [151], AdaGrad [153] y Adam [152] junto con sus variaciones como Nadam [169].

Las redes neuronales convolucionales aprenden cada vez patrones más específicos a medida que se hace más profundo el modelo, es decir, durante las primeras capas del modelo se extraerán características de bajo nivel, tales como líneas y bordes (cambios de intensidad), mientras que en las

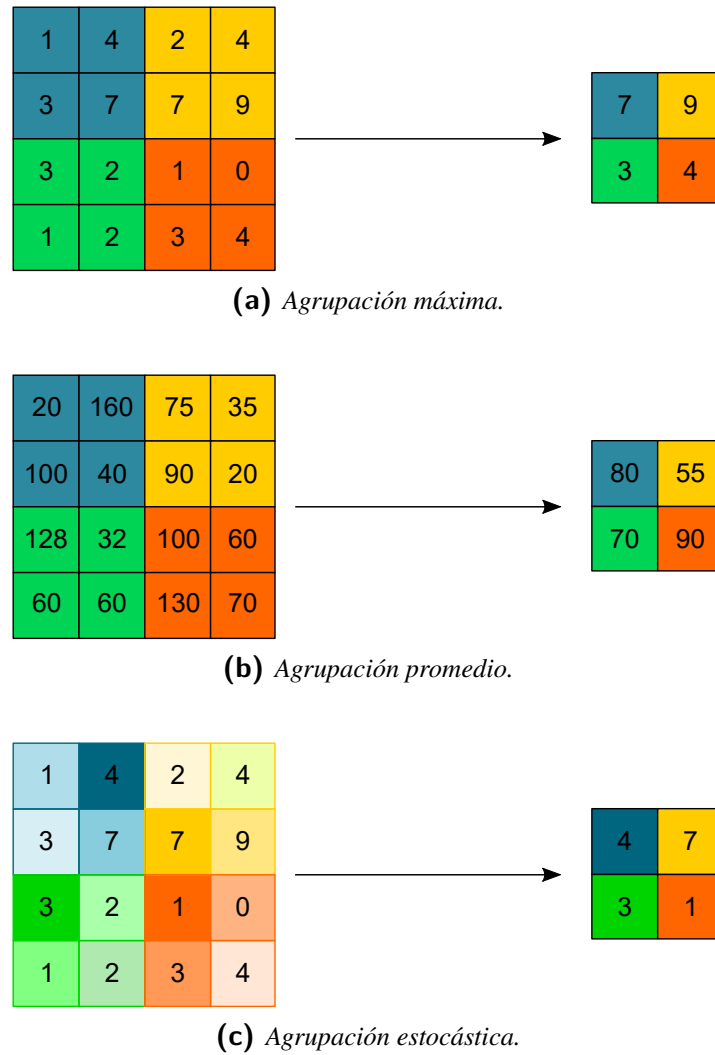


Figura 4.4: Variantes de la operación de agrupación.

capas más profundas se tendrán características de alto nivel como ojos o cara, que están compuestas de características de bajo nivel. La Figura 4.6 muestra los mapas de características o filtros resultantes en algunas de las capas de la misma arquitectura, con la que se pueden entender las características de la entrada que son detectadas como relevantes.

4.6.1. Ajuste de Hiperparámetros

El aprendizaje, y por lo tanto el desempeño de la red neuronal convolucional son altamente dependientes de una buena selección de hiperparámetros. En las redes neuronales convolucionales existen ciertos hiperparámetros que tienen gran influencia a la hora de diseñar la arquitectura. Algunos de los hiperparámetros más importantes a tomar en cuenta son [170]:

- Número de capas convolucionales.

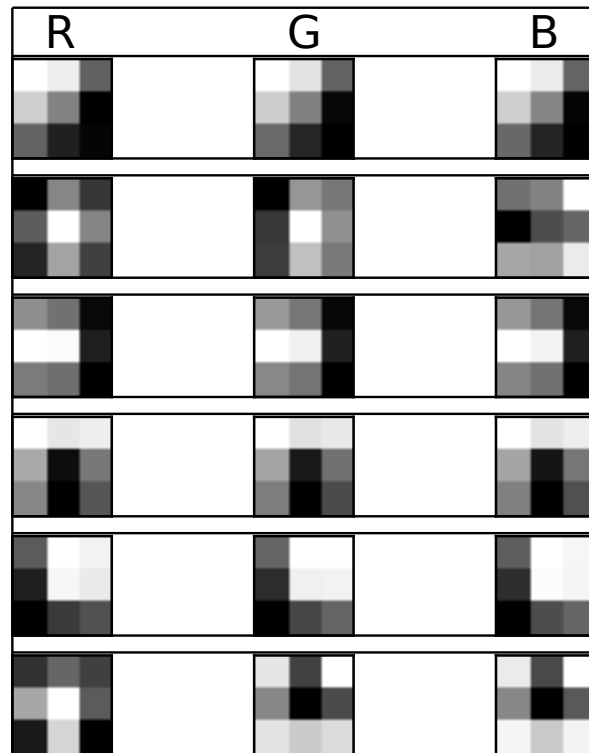


Figura 4.5: Filtros aprendidos dentro de las primeras capas de la arquitectura VGG16 [168].

- Número de filtros en cada capa convolucional.
- Tamaño de los filtros de cada capa.
- Tamaño del paso (*stride*).
- Relleno (*padding*).
- Tipo y tamaño de agrupación (*pooling*).
- Función de activación.
- Número de capas totalmente conectadas (también conocidas como *fully connected*, son capas similares a las existentes en el perceptrón multicapa).
- Número de neuronas en las capas totalmente conectadas.
- Optimizador.
- Tasa de aprendizaje.

Debido a que el número de parámetros a afinar es grande, se siguen diversas metodologías para la selección de los mismos. Una de ellas es la búsqueda de rejilla (también conocida como *grid search*) en la que se ubican en una tabla, los hiperparámetros a afinar y su valor para esa prueba, así como el valor obtenido de las métricas de éxito con esa arquitectura, los valores de los hiperparámetros serán variados dentro de un rango definido previamente para seleccionar la combinación que entregue los

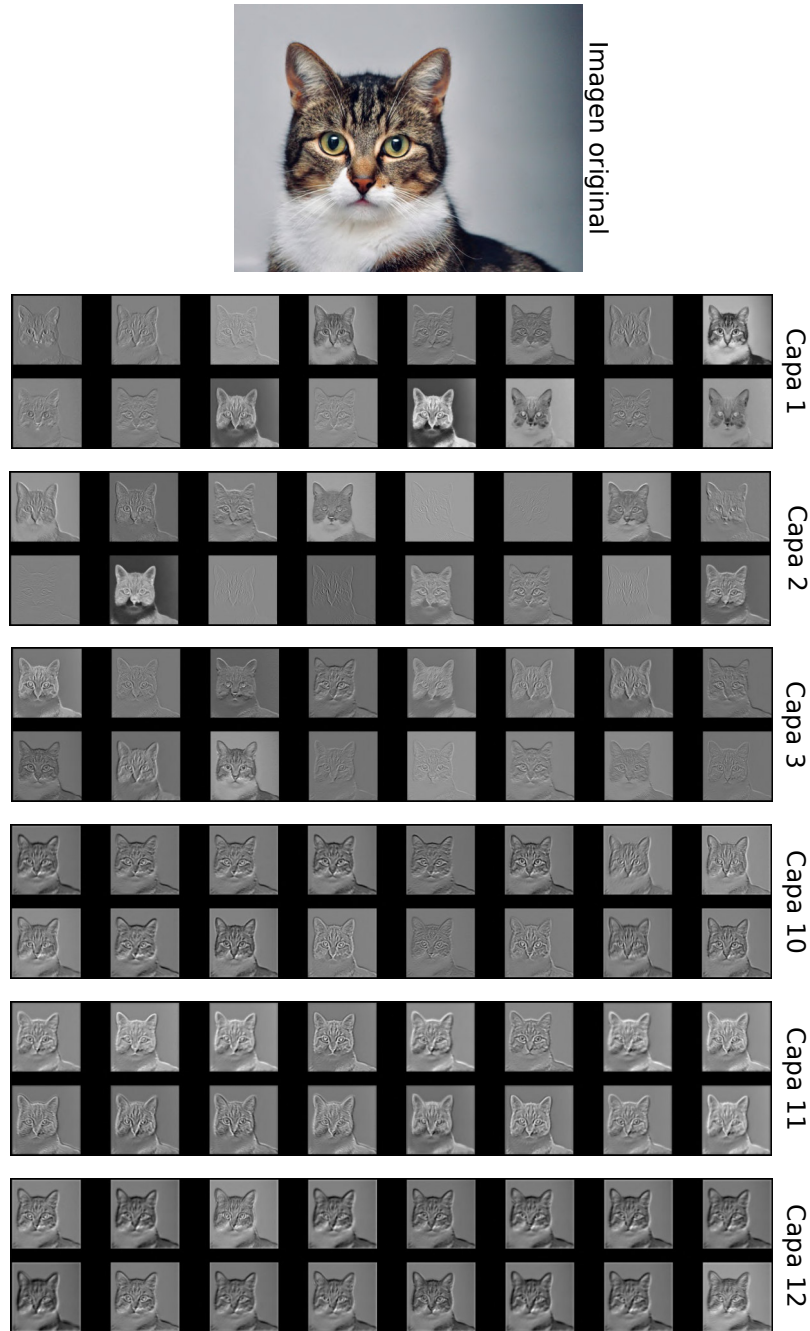


Figura 4.6: Mapas de características en diversas capas de la arquitectura VGG16 [168].

mejores resultados.

Una de las desventajas de la búsqueda de rejilla es que es demandante en términos de tiempo y cómputo, de acuerdo con [132]: “Si se tienen m hiperparámetros que tomarán n valores, entonces se requerirán $\mathcal{O}(n^m)$ pruebas de entrenamiento y validación”. Debido a lo tardado que puede resultar utilizar búsqueda de rejilla, existen otros enfoques para definir el valor de los hiperparámetros del

modelo, como la búsqueda aleatoria [171], o analizar el problema de afinación de hiperparámetros como una tarea de optimización.

Además de la manera de seleccionar los hiperparámetros, existen patrones de diseño para arquitecturas de aprendizaje profundo basados en la experiencia de diseñar diversas arquitecturas para distintas tareas. Smith y Topin presentan en [172], 14 patrones de diseño, entre los que destacan: buscar la simplicidad, modelos con estructura piramidal y capas tan anchas como lo permitan los recursos de cómputo disponibles.

4.7. Aplicaciones de las Redes Neuronales Convolucionales

Las características presentadas al inicio de la sección 4.5 así como la gran cantidad de desarrollo teórico en el área de redes neuronales convolucionales ha ocasionado que sean el enfoque por defecto al momento de trabajar con imágenes además de aplicarse a algunos tipos de datos diferentes como texto o señales biológicas. A continuación, se describen algunas de las tareas que se resuelven utilizando redes neuronales convoluciones:

- **Clasificación:** En esta tarea se busca asignar una etiqueta a la imagen que permita distinguir entre las distintas clases de la tarea. Los algoritmos que resuelven éstas tareas tienen a la salida la clase a la que pertenece la imagen o la probabilidad de pertenecer a cada una de las clases preestablecidas. Algunas de las bases de datos más comunes para realizar tareas de clasificación de imágenes son: ImageNet [173], con 1000 clases y poco más de 14,000,000 de imágenes etiquetadas; CIFAR-10 y CIFAR-100 [174] con 60,000 muestras repartidas en 10 y 100 clases respectivamente.
- **Segmentación:** Aquí se debe de dividir la imagen en segmentos o regiones discretas que no se traslapen. Esto se hace con la finalidad de definir bordes que permitan identificar y separar los objetos presentes en la imagen. De una manera más general, la segmentación consiste en clasificar cada uno de los píxeles de la imagen analizada. Existen dos tipos de segmentación: Segmentación semántica y segmentación por instancia.
- **Regresión:** Se utiliza para resolver problemas que involucren realizar predicciones de valores continuos a partir de la entrada. En el caso de las imágenes implica predecir las coordenadas de algunos puntos que permitan resolver la tarea.

4.7.1. Aplicaciones de las Redes Convolucionales al Área Médica

A continuación, se mencionan algunas aplicaciones específicas que se les dan a las redes convolucionales en el contexto de medicina:

- **Clasificación:**

- **Esteva et al.** [175] presentan un modelo de clasificación de lesiones en la piel. La red se entrenó con una base de datos de 129,450 imágenes con 2,032 enfermedades distintas y se probó a la par con 21 dermatólogos expertos, alcanzando niveles similares en la ejecución del diagnóstico.
 - **Anthimopoulos et al.** [176] presentan un modelo para la clasificación de patrones de enfermedad intersticial pulmonar (ILD, por sus siglas en inglés), utilizando una base de datos de 120 Tomografías Computarizadas. La salida del modelo indica la pertenencia de la entrada a una de las siete clases predefinidas.
 - **Farooq et al.** [177] presentan un modelo que puede clasificar casos de Alzheimer dentro de cuatro clases para imagen de resonancia magnética que recibe como entrada. La red se entrenó utilizando la base de datos de la Iniciativa en Neuroimagenología para la Enfermedad Alzheimer (ADNI, por sus siglas en inglés).
 - Diagnóstico de cáncer de mama: **Sun et al.** [178] presentan una red neuronal convolucional que utiliza aprendizaje semi-supervisado y tiene 3,158 regiones de interés, con dos clases posibles de salida.
 - Diagnóstico y clasificación de retinopatía diabética: **Pratt et al.** [179] proponen un sistema basado en redes convolucionales que clasifica la entrada asignándole una etiqueta de las cinco posibles para indicar la severidad de la retinopatía. El sistema se entrenó utilizando 80,000 imágenes.
- **Segmentación:**
- Segmentación de imágenes cardíacas: **Chen et al.** [180] presentan una revisión de los métodos de aprendizaje profundo utilizados para realizar segmentación de distintos tipos de imágenes del corazón. Las redes neuronales convolucionales son utilizadas para realizar segmentación de: arteria coronaria, ventrículos del corazón, fibrilación arterial, cicatrices miocárdicas, entre otras.
 - Segmentación de tumores cerebrales: **Pereira et al.** [181] utilizan las redes neuronales convolucionales para realizar segmentación de tumores cerebrales, incluyendo gliomas (el tumor cerebral más común y agresivo) en imágenes de resonancia magnética (MRI). El modelo fue validado utilizando la base de datos BRATS 2013.
- **Regresión:**
- Control y mejora de Electromiografía (EMG): **Ameri et al.** [182] presentan un modelo de regresión basado en redes neuronales convolucionales para extraer información de control motor en señales de electromiografía. Dichas señales se pueden utilizar para controlar interfaces, prótesis o exoesqueletos.

4.8. Conclusiones del Capítulo

En este capítulo se analizaron los fundamentos teóricos y prácticos de las redes neuronales convolucionales (CNN), destacando su relevancia en el procesamiento y análisis de imágenes digitales. Se explicó la estructura jerárquica de este tipo de redes, compuesta por capas convolucionales, de agrupamiento, normalización y completamente conectadas, las cuales permiten la extracción progresiva de características espaciales y texturales a partir de los datos de entrada.

También se abordaron las funciones de activación y pérdida más comúnmente empleadas, así como los mecanismos de optimización y regularización que contribuyen al entrenamiento estable de las redes. Este análisis permitió comprender cómo el ajuste adecuado de estos componentes impacta directamente en la precisión y capacidad de generalización del modelo.

Finalmente, las redes neuronales convolucionales constituyen una herramienta fundamental para la automatización del análisis de imágenes, al permitir la extracción de características relevantes sin requerir ingeniería manual de atributos. Su robustez frente a variaciones espaciales y su capacidad de aprendizaje jerárquico las hacen especialmente adecuadas para la detección de estructuras anatómicas en radiografías cefalométricas. Con base en estos fundamentos, el siguiente capítulo se centrará en la implementación práctica del modelo propuesto, detallando su arquitectura, y parámetros de entrenamiento.

Capítulo 5

Metodología

5.1. Introducción

En la actualidad, el avance de la inteligencia artificial en el ámbito médico ha demostrado su impacto transformador en múltiples especialidades, consolidando la validez de los enfoques automáticos para el diagnóstico por imagen. Un ejemplo notable se observa en el área de la detección temprana del cáncer de mama, donde herramientas aprobadas por la (*Food and Drug Administration*, FDA), como “*Clairity Breast*”, un software que ha mostrado la capacidad de identificar patrones imperceptibles para el ojo humano en mamografías, logrando anticipar riesgos con una precisión superior a la de los métodos tradicionales. Estos sistemas, entrenados con cientos de miles de imágenes, actúan como un segundo evaluador clínico, mejorando la sensibilidad diagnóstica entre un 20 % y un 40 % y reduciendo los tiempos de interpretación sin sacrificar exactitud.

Del mismo modo, la detección automática de puntos cefalométricos persigue un propósito análogo: complementar la labor del especialista, potenciar la precisión del diagnóstico y acelerar los procesos de planificación ortodóncica mediante el análisis cuantitativo y reproducible de imágenes radiográficas. Así como la IA ha revolucionado el diagnóstico mamográfico, su incorporación al análisis cefalométrico representa un paso decisivo hacia una odontología más predictiva, personalizada y sustentada en evidencia.

En el año 2014, durante el *IEEE International Symposium on Biomedical Imaging (ISBI)* celebrado en Beijing, China, se propuso por primera vez un reto internacional orientado a la detección automática de puntos cefalométricos en radiografías laterales de cráneo. Este desafío, conocido como “*Grand Challenge ISBI 2014*”, marcó un precedente importante al reunir una base de 300 imágenes cefalométricas con la finalidad de evaluar la capacidad de los sistemas de visión por computadora para localizar de manera precisa 19 puntos anatómicos de referencia comúnmente utilizados en el análisis cefalométrico.

Los resultados obtenidos en aquella primera edición reflejaron la complejidad del problema, la localización automática de puntos cefalométricos continua siendo una tarea desafiante, especialmente por las variaciones anatómicas entre pacientes y las características inherentes a las imágenes radiográficas.

A partir de esta experiencia, en el año 2015 se propuso un nuevo reto dentro del marco de ISBI 2015, titulado Grand Challenge on Dental Radiographic Image Analysis (ISBI2015-GC03). En esta segunda edición, se amplió la base de datos original, conformando un conjunto total de 400 imágenes radiográficas cefalométricas, que incluían las 300 utilizadas en el desafío anterior. Esta expansión permitió contar con un conjunto de datos más representativo y robusto, favoreciendo la comparación objetiva entre los diferentes métodos propuestos por la comunidad científica.

El objetivo central de este reto fue nuevamente la detección automática de los 19 puntos cefalométricos de referencia, fundamentales para la evaluación de estructuras óseas, dentales y de tejidos blandos. La importancia de este desafío radica en su contribución al desarrollo de sistemas automatizados que reduzcan la dependencia del marcado manual —un proceso que en la práctica clínica resulta laborioso y sujeto a variabilidad interobservador—, promoviendo así una mayor precisión, reproducibilidad y eficiencia en el análisis cefalométrico.

5.2. Aproximación al Problema

El enfoque utilizado como solución del problema planteado en este trabajo es la utilización de Redes Neuronales Convolucionales Profundas (DCNN) para la creación de un modelo de regresión que nos permita predecir las 19 coordenadas (x, y) de las etiquetas cefalométricas de la imagen de entrada. Tomando como base de datos experimental la propuesta de la competencia ISBI 2015. En la literatura se conoce a este tipo de red neuronal como *Vanilla deep regression*, en el que se añade a la arquitectura de red convolucional una capa final que realiza la regresión. La Figura 5.1 muestra un diagrama de una arquitectura de red neuronal convolucional para la solución de una regresión.

5.3. Descripción de la Base de Datos Utilizada

El conjunto de datos comprende 400 radiografías cefalométricas de 400 pacientes, con edades comprendidas entre los 6 y los 60 años. Las radiografías de cráneo se obtuvieron con el equipo Soredex CRANEX® Excel Ceph (Tuusula, Finlandia) y se procesaron con el software Soredex SorCom (versión 3.1.5, 2.0) con una resolución de 1935×2400 píxeles. Para su evaluación, se anotaron manualmente 19 puntos de referencia en cada radiografía, que fueron revisados posteriormente por dos ortodoncistas experimentados (llamados Senior y Junior). El conjunto de datos completo se dividió de forma aleatoria

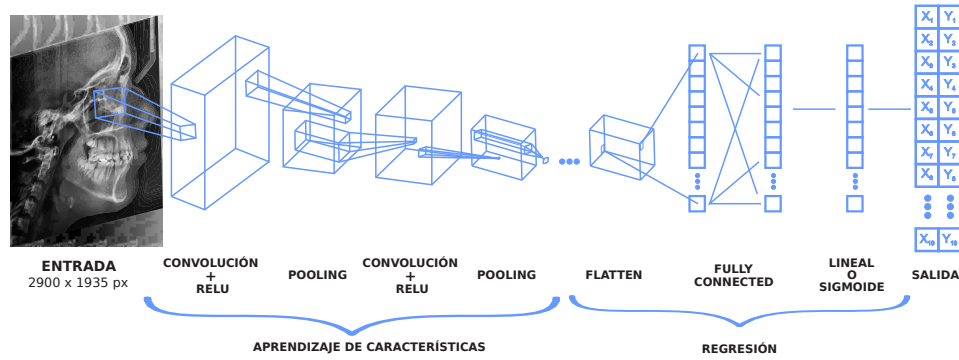


Figura 5.1: Esquema de un modelo de red neuronal convolucional utilizada para realizar regresión.

y uniforme en tres subconjuntos: datos de entrenamiento, datos de la Prueba 1 y datos de la Prueba 2 para el desafío de dos etapas. Las metodologías de aprendizaje automático propuestas por Ibragimov *et al.* [68], Mirzaalian *et al.* [71], Chu *et al.* [69], Vandaele *et al.* [70] y Chen *et al.* [78] fueron seleccionadas para la etapa 1 y participaron en la competencia de la etapa 2.

5.3.1. Conjunto de Datos del Desafío IEEE ISBI 2015

El conjunto de datos corresponde al *IEEE ISBI 2015 Challenge on Automatic Cephalometric Landmark Detection*, considerado un estándar para evaluar métodos automáticos de detección de puntos cefalométricos mediante aprendizaje profundo. La Tabla 5.1 muestra las características del hardware utilizado para la adquisición de la base de datos utilizada.

Tabla 5.1: Descripción del conjunto de datos.

Característica	Valor
Receptor de la máquina	Soredex CRANEX Excel Ceph machine.
Edad de los pacientes	Entre 6 y 60 años.
Resolución	1935×2400 píxeles.
Formato de salida	TIFF
Tamaño de píxel	$(0.1 \times 0.1)mm^2$
Imágenes de entrenamiento	Training (1-150)
Imágenes de prueba	Test1 (151-300) y Test2 (301-400)

Se anotaron manualmente 19 puntos cefalométricos ($M = 19$), en las 400 imágenes que conforman la base de datos. Estas Etiquetas fueron revisadas por dos ortodoncistas. Un ortodoncista es llamado como Experto 1 "*Senior*", y Experto 2 "*Junior*".

Para el entrenamiento y evaluación, las imágenes fueron separadas de la siguiente manera: las primeras 150 radiografías se utilizaron para el entrenamiento, mientras que las 250 restantes se dividieron en dos conjuntos de prueba:



Figura 5.2: Máquina Soredex CRANEX Excel Ceph. Utilizada para la adquisición de la base de datos del ISBI 2015.

- **TrainingData:** imágenes 001 a 150.
- **Test1Data:** imágenes 151 a 300.
- **Test2Data:** imágenes 301 a 400.

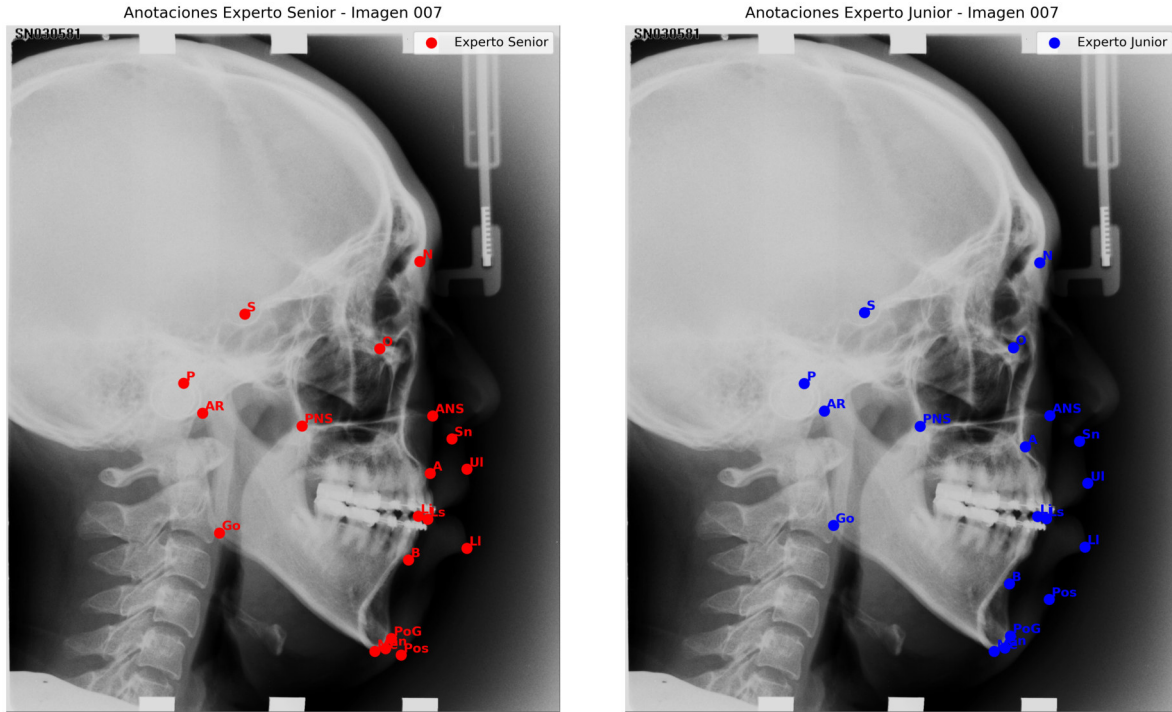
Esta división permite que todos los artículos que utilizan esta base de datos sean medidos de manera consistente y comparable.

Estas imágenes bidimensionales de cabeza y cara son comúnmente utilizadas en ortodoncia y cirugía ortognática, ya que permiten localizar con precisión distintos puntos anatómicos de referencia.

Los 19 puntos se clasifican en:

- **Tejido duro:** S, N, O, P, A, B, Pog, Me, Gn, Go, PNS, ANS, Ar.
- **Tejido blando:** Li, Ls, Ul, Ll, Sn, Pos.

La descripción anatómica detallada de estos puntos se presenta en la Tabla [5.2](#).



(a) Anotaciones del experto Senior.

(b) Anotaciones del experto Junior.

Figura 5.3: Comparación de las anotaciones realizadas por los expertos: Senior y Junior sobre la imagen 007.

5.3.2. Métricas de Evaluación del Rendimiento para la Detección Automática de Puntos Cefalométricos

Las métricas de evaluación permiten cuantificar la precisión y la fiabilidad de los algoritmos de detección automática de puntos cefalométricos en radiografías laterales del cráneo. En este estudio se emplearon dos métricas principales: el Error Radial Medio (MRE) y la Tasa de Detección Exitosa (SDR).

5.3.3. Error Radial Medio (MRE)

La métrica principal empleada para cuantificar la precisión del modelo es el Error Radial Medio (*Mean Radial Error*, MRE). El MRE mide la distancia euclidiana promedio entre las coordenadas predichas por el modelo y las coordenadas de referencia (*ground truth*) de cada punto anatómico:

$$R_{ij} = \sqrt{(x_{p,ij} - x_{o,ij})^2 + (y_{p,ij} - y_{o,ij})^2} \quad (5.1)$$

donde:

Tabla 5.2: Descripción anatómica y clasificación por tipo de tejido de los puntos de referencia cefalométricos utilizados en el estudio.

ID	Etiqueta	Tejido	Descripción
1	S — Silla Turca	Duro	Punto medio de la fosa pituitaria en el plano medio sagital.
2	N — Nasion	Duro	Punto más anterior entre los huesos frontal y nasal en la sutura frontonasal.
3	O — Orbitale	Duro	Punto más bajo del margen inferior de la órbita.
4	P — Porion	Duro	Punto más alto del contorno superior del meato auditivo externo.
5	A — Subespinal	Duro	Punto más profundo en la línea media entre ANS y la cresta alveolar de los incisivos centrales.
6	B — Supramental	Duro	Punto más profundo en la línea media entre la cresta alveolar mandibular y el proceso mentoniano.
7	Pog — Pogonion	Duro	Punto más anterior del mentón óseo en el plano medio.
8	Me — Mentón	Duro	Punto medio más inferior de la sínfisis mandibular.
9	Gn — Gnation	Duro	Punto más antero-inferior en la sínfisis del mentón.
10	Go — Gonion	Duro	Intersección del plano mandibular y el plano del ramo mandibular.
11	Li — Incision Inferius	Blando	Punto medio en el borde incisivo del incisivo central mandibular más prominente.
12	Ls — Incision Superius	Blando	Punto medio en el borde incisivo del incisivo central maxilar más prominente.
13	Ul — Labio Superior	Blando	Punto medio ubicado en el margen superior del labio superior membranoso.
14	Ll — Labio Inferior	Blando	Punto medio ubicado en el margen inferior del labio inferior membranoso.
15	Sn — Subnasale	Blando	Punto de intersección entre el margen inferior de la nariz y el contorno del labio superior.
16	Pos — Pogonion de tejido blando	Blando	Punto más prominente del contorno de tejido blando del mentón.
17	PNS — Espina Nasal Posterior	Duro	Intersección del límite anterior de la fosa pterigopalatina y el piso de la nariz.
18	ANS — Espina Nasal Anterior	Duro	Extremo anterior de la espina ósea del maxilar en el margen inferior de la apertura nasal anterior.
19	AR — Articulare	Duro	Punto de unión entre el borde posterior del ramo mandibular y el borde inferior del hueso occipital.

R_{ij} = Error radial del punto j en la imagen i .

$(x_{p,ij}, y_{p,ij})$ = Coordenadas predichas por el modelo.

$(x_{o,ij}, y_{o,ij})$ = Coordenadas de referencia anotadas manualmente.

Posteriormente, el error radial medio por cada imagen se obtiene como el promedio de los errores radiales de sus M puntos:

$$R_i = \frac{1}{M} \sum_{j=1}^M R_{ij} \quad (5.2)$$

Finalmente, el Error Radial Medio Global sobre todas las radiografías del conjunto de prueba se calcula como:

$$MRE = \frac{1}{N} \sum_{i=1}^N R_i \quad (5.3)$$

donde:

$N = 400$ = Número total de radiografías cefalométricas.

$M = 19$ = Número de puntos de referencia por radiografía.

Este valor proporciona una medida cuantitativa del desplazamiento promedio entre las predicciones y las anotaciones expertas. Un valor bajo de MRE indica mayor precisión en la localización de los puntos anatómicos.

5.3.4. Desviación Estándar

Para evaluar la consistencia del modelo, se calcula la Desviación Estándar (DE) de los errores radiales, la cual refleja la variabilidad en la precisión del sistema:

$$DE = \sqrt{\frac{1}{N \times M - 1} \sum_{i=1}^N \sum_{j=1}^M (R_{ij} - MRE)^2} \quad (5.4)$$

Una desviación estándar pequeña implica un desempeño más uniforme, mientras que una alta dispersión puede evidenciar dificultades del modelo en regiones anatómicas específicas o imágenes con ruido.

5.3.5. Tasa de Detección Exitosa (SDR)

La Tasa de Detección Exitosa (*Success Detection Rate*, SDR) mide el porcentaje de puntos localizados dentro de un margen de error clínicamente aceptable z (típicamente 2.0 mm):

$$SDR(z) = \frac{N_z}{N \times M} \times 100\% = \frac{|\{R_{ij} : R_{ij} < z\}|}{N \times M} \times 100\% \quad (5.5)$$

donde:

N_z = Número total de puntos detectados con error menor que z .

z = Umbral de distancia aceptable en milímetros (por ejemplo, 2 mm, 2.5 mm, 3 mm, 4 mm).

Esta métrica complementa al MRE, mostrando la proporción de detecciones que cumplen el criterio de precisión clínica. En estudios previos, los valores típicos de SDR con umbral de 2 mm oscilan entre 71 % y 89.5 %, siendo este rango considerado adecuado para aplicaciones clínicas reales.

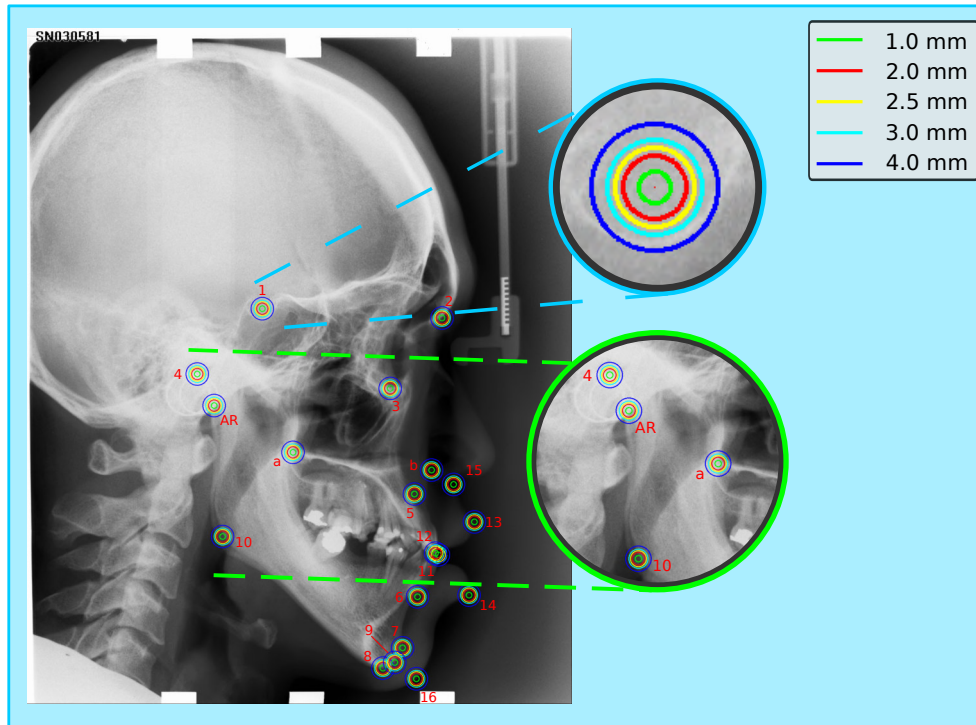


Figura 5.4: *Diversos rangos de error en comparación con la etiqueta señalada por el experto 1.*

5.3.6. Interpretación Clínica

El desafío ISBI 2015 establece que un error menor a 2 mm se considera clínicamente aceptable. Por lo tanto:

- Un $MRE < 2$ mm indica desempeño de nivel experto.
- Un $SDR(2\text{ mm}) > 85\%$ refleja fiabilidad suficiente para aplicación clínica.

Estas métricas permiten comparar objetivamente diferentes modelos de detección automática y constituyen el estándar de referencia en la literatura reciente sobre cefalometría computacional.

5.4. Conjunto de Datos

Para el entrenamiento del modelo propuesto se utilizó la base de datos pública liberada durante el Simposio Internacional de Imagenología Biomédica (ISBI 2015) del IEEE, que cuenta con 400 radiografías laterales de cráneo.

Las imágenes fueron etiquetadas en dos ocasiones por dos ortodoncistas expertos con 15 y 6 años de experiencia del hospital general de Taiwán. Para cada imagen se otorgó un archivo de texto plano (*.txt) por cada experto, dicho archivo incluye las coordenadas (x, y) de cada una de las 19 etiquetas cefalométricas.

5.4.1. Creación del Conjunto de Entrenamiento y Prueba

Con la finalidad de optimizar el manejo de los datos y reducir el tamaño de los archivos a procesar durante el entrenamiento y la evaluación del modelo, se emplearon archivos en formato *HDF5* (*.hdf5) para almacenar los conjuntos de entrenamiento y de prueba. Este formato permite la compresión interna de los datos, lo cual resulta especialmente útil para grandes volúmenes de información. Siguiendo este enfoque, el tamaño total del conjunto de datos original, de aproximadamente 5.9GB, se redujo a 986MB considerando ambos conjuntos, lo que permitió disminuir significativamente los tiempos de carga durante las fases de entrenamiento y validación del modelo.

El conjunto de datos se estructuró en dos arreglos principales: X y Y , correspondientes a las variables de entrada y salida del modelo, respectivamente. El arreglo X contiene las imágenes procesadas en formato de arreglos de *NumPy* [183] —una biblioteca de Python optimizada para el manejo de vectores, matrices y arreglos multidimensionales, que incluye una amplia gama de funciones para su manipulación—. En este caso, las imágenes se convirtieron a escala de grises y se normalizaron, sin aplicar procesos adicionales de calibración.

Por su parte, el arreglo Y almacena las coordenadas (x, y) correspondientes a los 19 puntos anatómicos por imagen, obtenidas a partir de los archivos de texto que acompañan a cada imagen en la base de datos.

Finalmente, se definió el método de compresión a utilizar (en este caso, *gzip*) y se segmentaron los datos en dos subconjuntos independientes: uno destinado al entrenamiento y otro a la prueba. Con ello, se obtuvieron los archivos finales en formato *HDF5*, los cuales son empleados posteriormente para la fase de entrenamiento del modelo (véase el Código 5.1).

```
1 # =====
2 # Creacion de los conjuntos de entrenamiento y prueba en HDF5
3 # =====
4 import numpy as np
5 import os
6 import h5py
7 from PIL import Image
8
9 DATA_PATH = "rawImage"
10 LABEL_PATH = "400_junior"
11 TRAIN_H5 = "dataset_train.hdf5"
```

```

12 TEST_H5 = "dataset_test.hdf5"
13
14 IMG_WIDTH = 1935
15 IMG_HEIGHT = 2400
16 IMG_DOWNSCALE = 0.1
17 # -----
18 # Carga de etiquetas (19 puntos por imagen)
19 # -----
20 def load_labels():
21     pts = []
22     for k in range(400):
23         data = np.genfromtxt(
24             f"{LABEL_PATH}/{k+1:03d}.txt",
25             dtype="float64,float64",
26             delimiter=",",
27             max_rows=19
28         )
29         xx = IMG_DOWNSCALE * (data["f0"] - IMG_WIDTH / 2)
30         yy = IMG_DOWNSCALE * (data["f1"] - IMG_HEIGHT / 2)
31         pts.append(np.concatenate((xx, yy), axis=0))
32     return np.array(pts)
33 # -----
34 # Carga y normalizacion de imagenes en escala de grises.
35 # -----
36 def load_images():
37     images = []
38     for folder, start, end in [
39         ("TrainingData", 1, 150),
40         ("Test1Data", 151, 300),
41         ("Test2Data", 301, 400)
42     ]:
43         for k in range(start, end + 1):
44             img_path = f"{DATA_PATH}/{folder}/{k:03d}.bmp"
45             im = Image.open(img_path)
46             if im.mode == "RGB":
47                 im = im.convert('L')
48             im = np.array(im, dtype=np.float32) / 255.0
49             images.append(im)
50     return np.expand_dims(np.array(images), -1)
51 # -----
52 # Creacion de los archivos comprimidos HDF5.
53 # Se guardan los conjuntos de entrenamiento y prueba,
54 # usando compresion gzip.
55 # -----
56 def prepare_data():
57     if os.path.isfile(TRAIN_H5) and os.path.isfile(TEST_H5):
58         with h5py.File(TRAIN_H5, 'r') as f:
59             Xtrain = np.array(f['data'])
60             Ytrain = np.array(f['label'])
61         with h5py.File(TEST_H5, 'r') as f:

```

```

62         Xtest = np.array(f['data'])
63         Ytest = np.array(f['label'])
64         print("Archivos HDF5 encontrados. Cargando conjuntos existentes...")
65     else:
66         print("Generando nuevos conjuntos HDF5...")
67         Y = load_labels()
68         X = load_images()
69
70         # Entrenamiento: primeras 150 imagenes
71         Xtrain, Ytrain = X[:150], Y[:150]
72
73         # Imagenes de Prueba Test1 151 a 300 Test1 y Test2 301 a 400
74         Xtest, Ytest = X[150:], Y[150:]
75
76         with h5py.File(TRAIN_H5, 'w') as f:
77             f.create_dataset('data', data=Xtrain, compression='gzip')
78             f.create_dataset('label', data=Ytrain, compression='gzip')
79         with h5py.File(TEST_H5, 'w') as f:
80             f.create_dataset('data', data=Xtest, compression='gzip')
81             f.create_dataset('label', data=Ytest, compression='gzip')
82
83         print("Archivos HDF5 creados y comprimidos correctamente.")
84         return (Xtrain, Ytrain), (Xtest, Ytest)
85
86 if __name__ == "__main__":
87     (Xtrain, Ytrain), (Xtest, Ytest) = prepare_data()
88     print("Conjuntos de entrenamiento y prueba listos para su uso.")

```

Código 5.1: Ejemplo de función para preparación de la base de datos.

5.4.2. Intra-Observabilidad Entre Expertos

La evaluación de la intraobservabilidad proporciona un *referente objetivo* sobre el límite de precisión que se puede esperar de los modelos automáticos y contribuye a mejorar la reproducibilidad y confiabilidad de los estudios cefalométricos.

La intraobservabilidad entre expertos (o concordancia inter-experto) es un indicador fundamental para evaluar la *consistencia y confiabilidad de las mediciones manuales* realizadas sobre un mismo conjunto de datos. En estudios de análisis cefalométrico, donde las etiquetas anatómicas son marcadas manualmente, es común que existan pequeñas variaciones entre distintos observadores debido a la interpretación subjetiva de los puntos de referencia.

Calcular la intraobservabilidad permite:

- **Cuantificar la precisión de las anotaciones humanas:** Identificar qué tan consistentes son las etiquetas cefalométricas marcadas por distintos expertos.

- **Detectar etiquetas críticas o problemáticas:** Algunos puntos anatómicos son más difíciles de identificar; el análisis de errores inter-experto permite localizar estos casos.
- **Validar modelos automáticos de detección:** Al conocer el rango de variación humana, se puede evaluar si un modelo de aprendizaje automático alcanza un desempeño comparable al de un experto.
- **Garantizar la calidad del dataset:** La reducción de la variabilidad entre anotaciones asegura que los datos de entrenamiento sean confiables y robustos.

5.4.3. Cálculo del Error Interobservador

Sea un conjunto de n imágenes, donde cada imagen contiene m etiquetas anatómicas. Denotemos las coordenadas de una etiqueta cefalométrica j en la imagen i anotadas por el Experto 1 (Senior) como

$$\mathbf{p}_{i,j}^{(1)} = (x_{i,j}^{(1)}, y_{i,j}^{(1)})$$

y las correspondientes anotaciones del Experto 2 (Junior) como

$$\mathbf{p}_{i,j}^{(2)} = (x_{i,j}^{(2)}, y_{i,j}^{(2)}).$$

La distancia euclidiana entre las coordenadas de una misma etiqueta cefalométrica j para la imagen i se define como:

$$d_{i,j} = \sqrt{(x_{i,j}^{(1)} - x_{i,j}^{(2)})^2 + (y_{i,j}^{(1)} - y_{i,j}^{(2)})^2}.$$

Dado que la resolución de la imagen se conoce en milímetros por píxel, se puede escalar cada coordenada de manera que la distancia se exprese en milímetros:

$$d_{i,j}^{mm} = \sqrt{\left((x_{i,j}^{(1)} - x_{i,j}^{(2)}) \cdot r_x\right)^2 + \left((y_{i,j}^{(1)} - y_{i,j}^{(2)}) \cdot r_y\right)^2}$$

donde r_x y r_y representan la resolución de cada píxel en milímetros en los ejes x y y , respectivamente.

El error medio por etiqueta cefalométrica se calcula como el promedio sobre todas las imágenes:

$$\bar{d}_j = \frac{1}{n} \sum_{i=1}^n d_{i,j}^{mm}, \quad j = 1, \dots, m.$$

Finalmente, el error global interobservador se obtiene como el promedio de todas las etiquetas cefalométricas y todas las imágenes:

$$\bar{D} = \frac{1}{m} \sum_{j=1}^m \bar{d}_j = \frac{1}{n \cdot m} \sum_{i=1}^n \sum_{j=1}^m d_{ij}^{mm}.$$

Este valor representa la discrepancia promedio entre las anotaciones de ambos expertos y sirve como referencia para evaluar la consistencia de las mediciones manuales y la precisión que se puede esperar de modelos automáticos de detección de etiquetas.

5.4.4. Cálculo del Error Interobservador para la Base de Datos ISBI 2015

Los errores promedio por etiqueta cefalométrica, calculados como la distancia euclidiana entre las anotaciones de los dos expertos y expresados en milímetros, se presentan en la Tabla 5.3. El error global promedio entre los expertos fue de 2.15 mm.

Tabla 5.3: Error medio por etiqueta cefalométrica entre los expertos.

Índice	Etiqueta	Error medio (mm)
1	S	0.93
2	N	1.51
3	O	3.08
4	P	3.31
5	A	2.89
6	B	3.02
7	PoG	1.24
8	Me	1.32
9	Gn	0.99
10	Go	2.86
11	Li	0.67
12	Ls	0.48
13	Ul	2.72
14	Ll	2.18
15	Sn	1.62
16	Pos	6.57
17	PNS	1.44
18	ANS	1.82
19	AR	2.13
Error global promedio:		2.15 mm

Como se puede observar, algunas etiquetas presentan errores significativamente mayores, como **Pos** con 6.57 mm, lo que indica que este punto anatómico es más difícil de identificar de manera

consistente entre los expertos. Otras etiquetas, como **Ls** y **Li**, muestran errores muy bajos (0.48 mm y 0.67 mm respectivamente), reflejando mayor consistencia en su anotación.

El error global promedio de 2.15 mm proporciona una referencia objetiva del límite de precisión alcanzable por los expertos humanos y sirve como base para evaluar la confiabilidad de modelos automáticos de detección de etiquetas.

Estos resultados se complementan con las gráficas generadas, donde se visualiza tanto el error promedio por etiqueta como un mapa de calor de los errores por archivo y etiqueta, permitiendo identificar de manera rápida los puntos con mayor discrepancia.

5.4.5. Visualización del Error Interobservador

Para complementar la Tabla 5.3, se generó una gráfica de barras que muestra el error medio por cada landmark entre los dos expertos. Además, se incluyó una línea horizontal que representa el *error global promedio* (2.15 mm) sobre todos los landmarks y todas las imágenes.

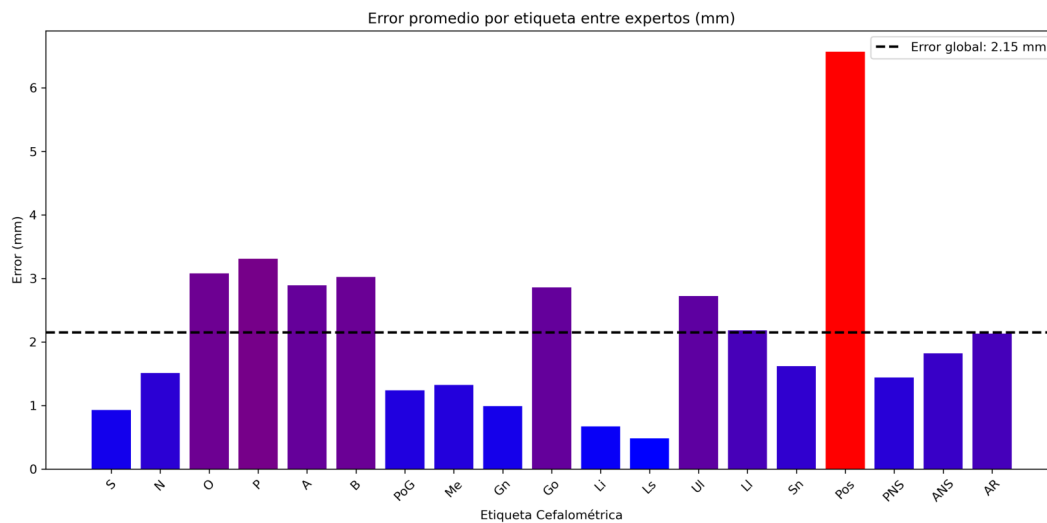


Figura 5.5: Error promedio por landmark entre expertos. La línea punteada indica el error global promedio.

Como se observa en la Figura 5.5, algunos landmarks presentan errores superiores al promedio, destacando especialmente **Pos** (Pogonion de tejido blando) con 6.57 mm, mientras que otros puntos como **S**, **N**, **PoG**, **Me**, **Gn**, **Li**, **Ls**, **Sn**, **PNS**, **ANS** y **AR** se encuentran por debajo del error global, reflejando una mayor consistencia entre los expertos. Esta representación permite identificar de manera visual los puntos anatómicos más difíciles de etiquetar y proporciona un referente claro para evaluar la precisión de futuros modelos automáticos.

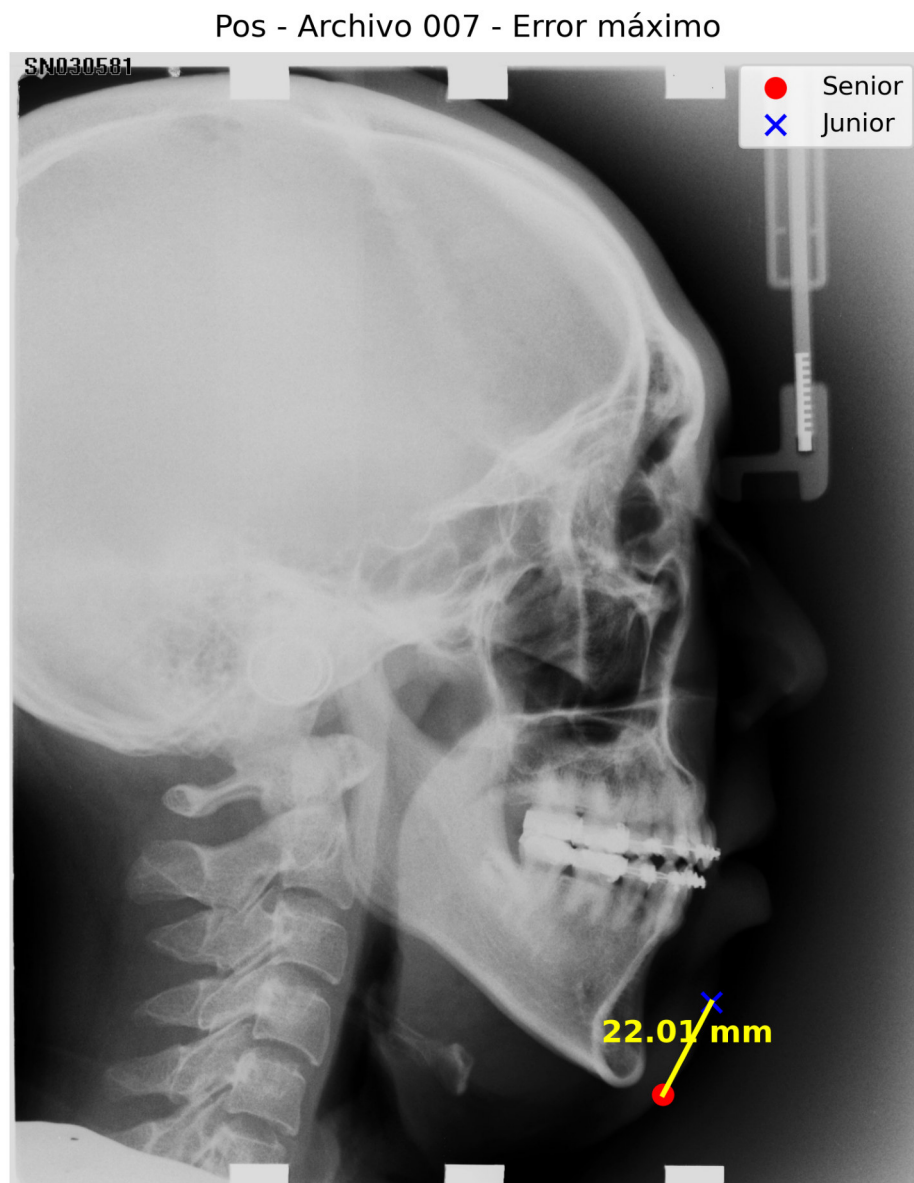


Figura 5.6: Archivo 007 de la base de datos ISBI 2015, correspondiente al error máximo entre observadores. El punto marcado en rojo corresponde al experto senior y el azul al experto junior. La distancia entre ambas marcas es de 22.01 mm, representando la discrepancia más alta registrada para la etiqueta llamada "Pos".

La Figura 5.6 ilustra la etiqueta (Pogonion de tejido blando) con mayor discrepancia entre observadores. En esta radiografía lateral se observa la diferencia en la localización de un punto anatómico específico entre un Experto Senior y el Experto Junior. Este tipo de casos permite identificar las regiones anatómicas más susceptibles a interpretaciones subjetivas, lo que puede deberse a la baja visibilidad de los contornos o a diferencias en los criterios anatómicos empleados.

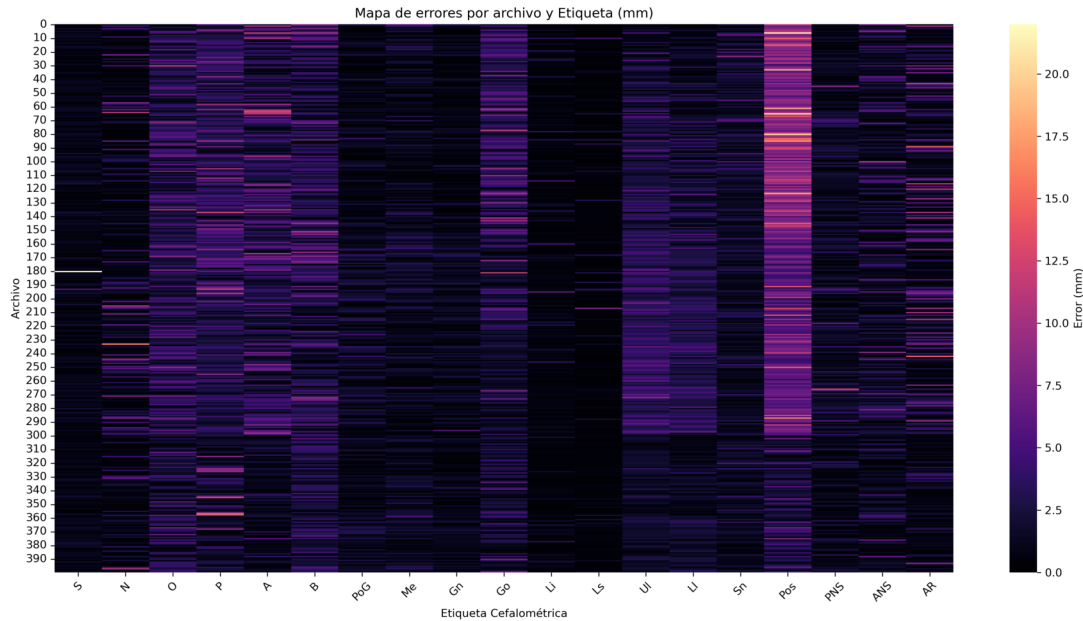


Figura 5.7: Mapa de calor de errores interobservador por archivo y etiqueta cefalométrica (en mm). Los tonos más claros indican errores mayores, observándose mayor variabilidad en los puntos asociados a **P** (Porion) y **Pos** (Pogonion de tejido blando).

La Figura 5.7 presenta un mapa de calor de los errores por archivo y etiqueta cefalométrica. Se aprecia que los errores no son uniformes entre las distintas etiquetas; en particular, los puntos relacionados con los contornos mandibulares, como **P** y **Pos**, muestran una mayor variabilidad interobservador. En contraste, puntos de referencia más definidos, como **N** (Nasion) o **S** (Sella), presentan errores considerablemente menores, reflejando una mejor consistencia en su identificación.

En conjunto, estas Figuras permiten visualizar tanto la magnitud puntual del error máximo como la distribución general de las discrepancias entre observadores, ofreciendo un panorama integral sobre la reproducibilidad de las anotaciones cefalométricas humanas.

5.5. Diseño de la Arquitectura de la Red Neuronal Convolutiva

En esta sección se describe el proceso de diseño, configuración y entrenamiento del modelo convolutiva propuesto, incluyendo los aspectos de hardware, software, hiperparámetros y la estructura final de la red.

5.5.1. Hardware Utilizado Durante el Entrenamiento

La Tabla 5.4 muestra las características técnicas del equipo empleado para el diseño, entrenamiento y pruebas del modelo.

Tabla 5.4: Descripción del hardware utilizado.

Característica	Valor
Modelo	MSI GF63
Sistema Operativo	Windows 11
Procesador	Intel Core i7, 14ª generación
Memoria RAM	96 GB DDR5
GPU	NVIDIA RTX 5090 (32 GB VRAM)
Almacenamiento	2 TB SSD

5.5.2. Software Utilizado

En el desarrollo del modelo se emplearon diversas bibliotecas de código abierto para aprendizaje profundo, preprocesamiento y visualización de datos utilizando el lenguaje de programación *python*.

TensorFlow

TensorFlow [184] es una biblioteca de software de código abierto desarrollada por Google Brain Team en 2015, diseñada para facilitar el desarrollo, entrenamiento y despliegue de modelos de aprendizaje automático y profundo. Su arquitectura flexible permite definir y ejecutar grafos computacionales que describen flujos de datos y operaciones matemáticas complejas, lo que la convierte en una herramienta fundamental para la investigación y la producción de sistemas de inteligencia artificial.

El nombre *TensorFlow* proviene de la estructura fundamental con la que opera: los *tensores*. Estos representan arreglos multidimensionales que fluyen a través de un grafo compuesto por operaciones matemáticas llamadas *nodos*. Cada nodo del grafo representa una operación, y las aristas representan los tensores que transportan datos entre operaciones. Este enfoque permite un control detallado sobre las dependencias computacionales y posibilita la ejecución paralela de tareas, lo que resulta especialmente útil para arquitecturas de alto rendimiento.

Una de las principales ventajas de TensorFlow es su capacidad para ejecutar cálculos de manera eficiente en diferentes tipos de hardware, incluyendo CPU (Computer Processing Unit), GPU (Tensor Processing Unit) y TPU (Graphics Processing Unit), aprovechando automáticamente los recursos disponibles. Además, la biblioteca implementa técnicas de *autodiferenciación*, mediante las cuales se calculan los gradientes de forma automática durante la fase de entrenamiento de los modelos.

Esta característica es esencial para el algoritmo de retropropagación del error (*backpropagation*), ampliamente utilizado en el ajuste de pesos sinápticos de redes neuronales.

TensorFlow dispone de una interfaz de programación (API) en varios lenguajes, siendo Python la más utilizada por su facilidad de uso, extensas bibliotecas complementarias y amplia comunidad de desarrolladores. En este trabajo se optó por la API de Python, ya que ofrece compatibilidad con herramientas de análisis, limpieza y visualización de datos, tales como *NumPy*, *Matplotlib* y *Pandas*. Asimismo, su integración nativa con entornos interactivos como Jupyter Notebooks permite una rápida experimentación, depuración y ajuste de modelos.

En cuanto a su estructura interna, TensorFlow cuenta con dos modos de ejecución: el modo *grafo* y el modo *inmediato* o *Eager Execution*. En el primero, las operaciones se definen previamente en un grafo estático, el cual se optimiza antes de su ejecución. Este enfoque ofrece un rendimiento superior en escenarios de producción debido a la posibilidad de aplicar optimizaciones a nivel de grafo, como la fusión de operaciones o la ejecución distribuida en múltiples dispositivos. En contraste, el modo imperativo permite ejecutar las operaciones de manera inmediata, facilitando el proceso de depuración y la comprensión del flujo de datos. TensorFlow permite alternar entre ambos modos, lo que proporciona una gran versatilidad para investigación y desarrollo.

Otra característica destacada de TensorFlow es su ecosistema modular, que incluye herramientas para distintas etapas del ciclo de vida de un modelo de aprendizaje profundo. Entre ellas destacan:

- **TensorFlow Datasets (TFDS):** para la carga y gestión de conjuntos de datos de entrenamiento y prueba.
- **TensorFlow Hub:** un repositorio de modelos preentrenados reutilizables.
- **TensorBoard:** una herramienta de visualización que permite monitorear métricas de entrenamiento como la pérdida, la precisión, los gradientes y la evolución de los pesos.
- **TensorFlow Lite:** un entorno optimizado para desplegar modelos en dispositivos móviles o sistemas embebidos.
- **TensorFlow Serving:** una plataforma de producción para la implementación de modelos en servidores y servicios web.

Durante el desarrollo de este trabajo, TensorFlow se utilizó como la plataforma principal para la construcción, entrenamiento y validación del modelo convolucional propuesto. Su capacidad para aprovechar la aceleración por GPU permitió reducir significativamente los tiempos de entrenamiento, especialmente en las fases de búsqueda de hiperparámetros mediante técnicas como *grid search*.

Además, TensorFlow facilitó la integración con Keras, utilizada como API de alto nivel para la definición de la arquitectura, lo que permitió un desarrollo más estructurado y legible del código. El

entrenamiento del modelo se llevó a cabo bajo TensorFlow versión 2.x, que integra de forma nativa Keras como su interfaz principal, eliminando las limitaciones del grafo estático presentes en versiones anteriores.

TensorFlow constituye la base sobre la cual se desarrolló el modelo de red neuronal convolucional de este proyecto. Su robustez, flexibilidad y ecosistema complementario lo convierten en una herramienta idónea tanto para la experimentación como para el despliegue de modelos en entornos de producción y dispositivos de cómputo heterogéneos.

Keras

Keras [165] es una biblioteca de alto nivel orientada al desarrollo de redes neuronales profundas, diseñada para proporcionar una interfaz intuitiva y modular que facilite la experimentación rápida con modelos de aprendizaje profundo. Fue desarrollada inicialmente por François Chollet en 2015 y, a partir de TensorFlow 2.0, se integró de manera nativa como su API principal, lo que permite aprovechar la potencia de TensorFlow con una sintaxis simplificada y altamente legible.

La filosofía de diseño de Keras se basa en la facilidad de uso, la modularidad y la extensibilidad. Esto permite construir modelos complejos mediante una estructura de bloques o capas (*layers*) que pueden combinarse de forma secuencial o funcional. Gracias a esta característica, Keras reduce significativamente la carga de programación necesaria para la definición, entrenamiento y evaluación de modelos de redes neuronales convolucionales (CNN), recurrentes (RNN) o híbridas, sin sacrificar la flexibilidad requerida para tareas de investigación avanzada.

En su arquitectura, Keras actúa como una capa de abstracción superior que traduce las operaciones de alto nivel a instrucciones ejecutables por un *backend*, siendo TensorFlow el más utilizado actualmente. Otros backends disponibles en versiones previas incluyen Theano y CNTK, aunque estos han sido discontinuados. En la versión actual, Keras 3, se incluyen los siguientes backends: Tensorflow, PyTorch, y JAX.

Keras proporciona dos principales paradigmas para la definición de modelos:

- **API Secuencial:** es el enfoque más sencillo y lineal, donde las capas se apilan de forma ordenada mediante la clase `Sequential`. Este modo resulta apropiado para arquitecturas de flujo único, como las CNN estándar utilizadas para clasificación de imágenes o detección de puntos de referencia.
- **API Funcional:** permite construir modelos con topologías más complejas, como redes con múltiples entradas o salidas, conexiones residuales, o arquitecturas de fusión multimodal. Esta API trata las capas como funciones que pueden recibir y devolver tensores, lo que brinda gran flexibilidad para representar arquitecturas no lineales.

Durante el entrenamiento de los modelos, Keras simplifica procesos clave mediante el uso de funciones integradas para la compilación, el ajuste y la evaluación del modelo. Por ejemplo, el método `compile()` permite especificar el optimizador, la función de pérdida y las métricas de evaluación, mientras que `fit()` gestiona automáticamente la partición del conjunto de datos, los lotes de entrenamiento y la actualización iterativa de los pesos. Asimismo, funciones como `evaluate()` y `predict()` permiten validar y aplicar el modelo de forma eficiente.

Entre las ventajas más relevantes de Keras se encuentran:

- Su sintaxis clara y de alto nivel, que facilita la rápida implementación de prototipos y experimentos.
- Compatibilidad con la aceleración por GPU mediante TensorFlow, lo que reduce considerablemente los tiempos de entrenamiento.
- Integración con herramientas complementarias como TensorBoard para la visualización de métricas, y ModelCheckpoint para la gestión automática de pesos durante el entrenamiento.
- Capacidad de exportar modelos en formatos estándar (.h5 o SavedModel), lo que permite su reutilización y despliegue en diferentes entornos.

En el contexto del presente trabajo, Keras se emplea como interfaz principal para la definición y configuración de la arquitectura convolucional. A través de la API funcional se estructuró el modelo encargado de procesar las imágenes y extraer características jerárquicas para la posterior localización de los puntos cefalométricos. Esta elección permitió combinar flexibilidad en el diseño de la red con eficiencia en la ejecución, aprovechando la compatibilidad completa con TensorFlow para realizar operaciones vectorizadas y cálculos en GPU.

Durante el proceso de entrenamiento, Keras facilita el control de los hiperparámetros (tasa de aprendizaje, tamaño de lote, número de épocas y función de activación), así como la implementación de técnicas de regularización, como la normalización por lotes (*Batch Normalization*) y la desactivación aleatoria de neuronas (*Dropout*). Además, mediante el uso de `callbacks`, se monitorearon en tiempo real métricas como la pérdida y la precisión, lo que permitió ajustar dinámicamente el proceso de entrenamiento y evitar sobreajuste.

Adicionalmente, Keras proporcionó un entorno de desarrollo ágil, reproducible y transparente, que permitió concentrar los esfuerzos en el diseño conceptual del modelo y no en los detalles de implementación de bajo nivel. Gracias a su integración con TensorFlow, fue posible combinar la simplicidad de uso con la potencia computacional necesaria para ejecutar entrenamientos intensivos sobre grandes volúmenes de datos de imágenes cefalométricas.

Keras representa una herramienta esencial en la etapa de desarrollo del modelo propuesto, al ofrecer una combinación óptima entre usabilidad, flexibilidad y rendimiento, cualidades que la consolidan como

una de las bibliotecas más empleadas en la investigación actual sobre aprendizaje profundo.

5.5.3. Diseño de la Arquitectura de la Red

El diseño de la arquitectura se basó en la extracción progresiva de características mediante bloques convolucionales con *stride* unitario y kernels pequeños. La metodología seguida fue la siguiente:

```

1  def createmodel(m, n, noutputs):
2      input_img = Input(shape=(1, m, n))
3      x = Convolution2D(16, 7, activation='relu', padding='same',
4      kernel_initializer='glorot_uniform')(input_img)
5      x = AveragePooling2D(pool_size=(3, 3), padding='same')(x)
6      x = Dropout(0.2)(x)
7      x = Convolution2D(32, 7, activation='relu', padding='same',
8      kernel_initializer='glorot_uniform')(x)
9      x = AveragePooling2D(pool_size=(2, 2), padding='same')(x)
10     x = Dropout(0.2)(x)
11     x = Convolution2D(32, 7, activation='relu', padding='same',
12     kernel_initializer='glorot_uniform')(x)
13     x = AveragePooling2D(pool_size=(2, 2), padding='same')(x)
14     x = Dropout(0.2)(x)
15     x = Convolution2D(32, 7, activation='relu', padding='same',
16     kernel_initializer='glorot_uniform')(x)
17     x = AveragePooling2D(pool_size=(2, 2), padding='same')(x)
18     x = Dropout(0.2)(x)
19     x = Convolution2D(32, 7, activation='relu', padding='same',
20     kernel_initializer='glorot_uniform')(x)
21     x = AveragePooling2D(pool_size=(2, 2), padding='same')(x)
22     x = Dropout(0.2)(x)
23     x = Convolution2D(32, 7, activation='relu', padding='same',
24     kernel_initializer='glorot_uniform')(x)
25     x = AveragePooling2D(pool_size=(2, 2), padding='same')(x)
26     x = Dropout(0.2)(x)
27     x = Convolution2D(32, 7, activation='relu', padding='same',
28     kernel_initializer='glorot_uniform')(x)
29     x = AveragePooling2D(pool_size=(2, 2), padding='same')(x)
30     x = Dropout(0.2)(x)
31     x = Convolution2D(32, 7, activation='relu', padding='same',
32     kernel_initializer='glorot_uniform')(x)
33     x = AveragePooling2D(pool_size=(2, 2), padding='same')(x)
34     x = Dropout(0.2)(x)
35     x = Flatten()(x)
36     x = Dense(1024, activation='relu', kernel_initializer='glorot_uniform')(x)
37     x = Dropout(0.2)(x)
38     outputs = Dense(noutputs, activation='linear',
39     kernel_initializer='glorot_uniform')(x)
40     model = Model(input_img, outputs)
41     return model

```

Código 5.2: Ejemplo de función para creación de la arquitectura.

Diagrama de la Arquitectura del Modelo de Red Convolucional

La arquitectura propuesta en la Figura 5.8, se diseñó utilizando las biblioteca de Keras sobre la base de TensorFlow, utilizando la *API funcional* para definir una secuencia jerárquica de capas convolucionales, de agrupamiento y totalmente conectadas. El modelo se orientó a la predicción de coordenadas cefalométricas, formulado como un problema de regresión multivariable con 38 salidas lineales correspondientes a las coordenadas (x, y) de 19 puntos anatómicos.

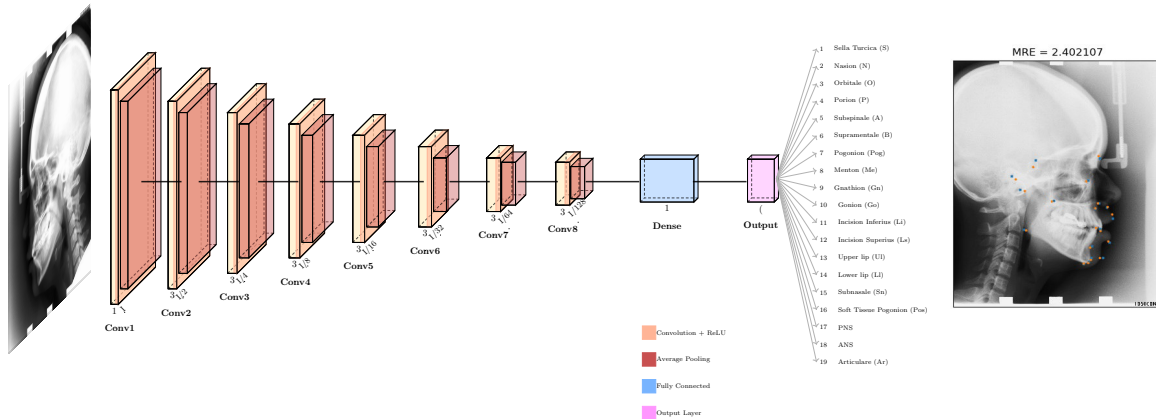


Figura 5.8: Arquitectura “Encoder”, propuesta como Solución.

- **Entrada de datos:** El modelo recibe imágenes de un solo canal de color (*grayscale*) con dimensiones de 2400×1935 píxeles, organizadas en formato *channels-first*. La capa de entrada (*InputLayer*) establece la forma de los tensores como $(1, 2400, 1935)$.
- **Bloques convolucionales jerárquicos:** La red está conformada por ocho bloques convolucionales consecutivos, cada uno integrado por:
 - Una capa Conv2D con función de activación ReLU, kernel de tamaño (7×7) , inicializador *glorot_uniform* y *padding* “same”.
 - Una capa AveragePooling2D, con kernel de (2×2) o (3×3) , que disminuye progresivamente la resolución espacial.
 - Una capa Dropout con tasa de 0.2, aplicada para reducir el sobreajuste.
 Estos bloques permiten una extracción progresiva de características, desde bordes simples hasta estructuras anatómicas complejas, conservando la estabilidad de las activaciones en cada nivel.
- **Aumento de la capacidad representacional:** A medida que se avanza en profundidad, el número de filtros convolucionales aumenta gradualmente (16, 32, 64, 128, etc.), lo que amplía la capacidad del modelo para identificar patrones más abstractos sin incrementar excesivamente la carga computacional.
- **Regularización:** Cada bloque incluye una capa Dropout para forzar la independencia de los filtros y

reducir la correlación entre activaciones, fortaleciendo la generalización del modelo frente a nuevas imágenes.

- **Transformación a vector:** Tras el último bloque convolucional, los mapas de activación bidimensionales se transforman mediante una capa Flatten, que los convierte en un vector unidimensional, facilitando la conexión con las capas totalmente conectadas.
- **Integración de características de alto nivel:** La primera capa totalmente conectada (Dense(1024)) utiliza activación ReLU y un inicializador glorot_uniform, funcionando como un nivel de integración donde se combinan las características extraídas por las capas previas. Se aplica una capa adicional de Dropout(0.2) para reforzar la regularización. Se aplica una capa adicional de Dropout(0.2) para reforzar la regularización. La selección de esta tasa de abandono se basa en la práctica reportada por Srivastava et al. (2014), quienes muestran que el valor óptimo de la probabilidad de retención p depende del tipo de capa y de la arquitectura de la red. En redes totalmente conectadas profundas, un valor cercano a $p = 0.5$ (equivalente a Dropout = 0.5) suele ser apropiado para las capas densas, mientras que para capas cercanas a la entrada y capas convolucionales se recomiendan probabilidades de retención más altas (es decir, menores tasas de dropout), típicamente en el rango $p \in [0.75, 0.9]$. Bajo este criterio, un valor de Dropout = 0.2 (equivalente a $p = 0.8$) resulta consistente con las recomendaciones para capas intermedias y convolucionales, permitiendo regularizar el modelo sin destruir excesivamente la información espacial relevante para la localización de puntos cefalométricos [185].
- **Capa de salida:** La última capa (Dense(38_outputs)) tiene activación lineal, lo que permite realizar una regresión continua de coordenadas. Cada una de las 38_outputs neuronas corresponde a una coordenada x o y que determina los 19 pares ordenados de las etiquetas cefalométricas.
- **Inicialización de pesos:** Todas las capas convolucionales y densas se inicializan con el método glorot_uniform, garantizando una distribución equilibrada de los pesos al inicio del entrenamiento y evitando la saturación de las activaciones en las primeras épocas.

A continuación se presenta el pseudo-código 1, la estructura general del modelo de Red Propuesta:

5.5.4. Optimizador Nadam

Para el proceso de entrenamiento de la red neuronal convolucional profunda (DCNN) se empleó el optimizador *Nadam* (*Nesterov-accelerated Adaptive Momentum*), el cual combina las ventajas del método *Adam* con la aceleración de gradiente propuesta por *Nesterov*. Este optimizador ajusta de manera adaptativa la tasa de aprendizaje de cada parámetro a partir de los momentos de primer y segundo orden de los gradientes, incorporando además una corrección de momento anticipado.

Algorithm 1 Pseudo-código de Red Neuronal Convolutiva Profunda Propuesta.

```

1: Inicializar entrada:  $input\_img \leftarrow Input(shape = (1, m, n))$ 
2: Primera capa convolucional:
3:  $x \leftarrow Conv2D(16, 7, activation = 'relu', padding = 'same')(input\_img)$ 
4:  $x \leftarrow AveragePooling2D(pool\_size = (3, 3), padding = 'same')(x)$ 
5:  $x \leftarrow Dropout(0.2)(x)$ 
6: for  $i = 1$  to 7 do
7:    $x \leftarrow Conv2D(32, 7, activation = 'relu', padding = 'same')(x)$ 
8:    $x \leftarrow AveragePooling2D(pool\_size = (2, 2), padding = 'same')(x)$ 
9:    $x \leftarrow Dropout(0.2)(x)$ 
10: end for
11: Aplanar características:
12:  $x \leftarrow Flatten()(x)$ 
13:  $x \leftarrow Dropout(0.2)(x)$ 
14: Capa densa totalmente conectada:
15:  $x \leftarrow Dense(1024, activation = 'relu')(x)$ 
16:  $x \leftarrow Dropout(0.2)(x)$ 
17: Capa de salida:
18:  $outputs \leftarrow Dense(noutputs, activation = 'linear')(x)$ 
19: return  $model$ 

```

El proceso de actualización de los pesos en el optimizador Nadam se define mediante las ecuaciones (5.6) a (5.10).

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (5.6)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (5.7)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad (5.8)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \quad (5.9)$$

$$\theta_{t+1} = \theta_t - \eta \left(\frac{\beta_1 \hat{m}_t + \frac{(1-\beta_1)g_t}{1-\beta_1^t}}{\sqrt{\hat{v}_t} + \epsilon} \right), \quad (5.10)$$

donde:

- m_t y v_t representan los momentos de primer y segundo orden respectivamente,
- g_t es el gradiente del error respecto a los parámetros θ_t en el instante t ,
- η es la tasa de aprendizaje,
- β_1 y β_2 son los coeficientes de decaimiento exponencial para los momentos,
- ϵ es un término pequeño que evita la división entre cero.

En este trabajo se utilizaron los parámetros mostrados en la Tabla 5.5, los cuales fueron ajustados experimentalmente para lograr una convergencia estable y minimizar el error cuadrático medio (MSE).

Tabla 5.5: *Parámetros del optimizador Nadam empleados durante el entrenamiento de la red neuronal.*

Parámetro	Símbolo	Valores Típicos	Descripción
Tasa de aprendizaje	η	2×10^{-4}	Controla la magnitud del paso en la actualización de los pesos. Un valor bajo favorece una convergencia estable.
Coefficiente de momento	β_1	0.9	Pondera el promedio exponencial de los gradientes anteriores, proporcionando inercia al descenso del gradiente.
Coefficiente del segundo momento	β_2	0.999	Controla la media cuadrática de los gradientes, estabilizando las actualizaciones ante gradientes ruidosos.
Constante de estabilidad numérica	ϵ	1×10^{-7}	Evita divisiones por cero durante la normalización de los gradientes.
Modo de actualización	—	<i>Nesterov</i>	Aplica un paso anticipado del gradiente, mejorando la velocidad de convergencia.

El optimizador Nadam tiene un equilibrio entre rapidez de convergencia y estabilidad numérica, siendo particularmente adecuado para arquitecturas profundas como la utilizada en este trabajo. Su naturaleza adaptativa permite un aprendizaje eficiente incluso en presencia de gradientes ruidosos o mal condicionados.

5.5.5. Ajuste Adaptativo de la Tasa de Aprendizaje

Durante el entrenamiento, además del optimizador principal, se implementó un mecanismo de ajuste dinámico de la tasa de aprendizaje con el fin de mejorar la convergencia del modelo y evitar oscilaciones en el error de validación. Para ello se empleó la estrategia *ReduceLROnPlateau*, disponible en la biblioteca *Keras*, la cual reduce la tasa de aprendizaje de forma automática cuando el desempeño del modelo en el conjunto de validación deja de mejorar después de un número determinado de épocas (*patience*).

El principio de funcionamiento de este programador de tasa de aprendizaje se expresa mediante la siguiente regla:

$$\eta_{t+1} = \begin{cases} \eta_t \times factor, & \text{si no hay mejora en el error de validación durante } p \text{ épocas,} \\ \eta_t, & \text{en caso contrario.} \end{cases} \quad (5.11)$$

donde:

- η_t es la tasa de aprendizaje actual,
- η_{t+1} es la nueva tasa de aprendizaje,
- *factor* es el multiplicador de reducción (por ejemplo, 0.5),
- *p* es el número de épocas de espera definido por el parámetro *patience*.

En este trabajo, el "programador" fue configurado para monitorizar la métrica de error promedio de validación (*val_myloss*), con los parámetros indicados en la Tabla 5.6.

Tabla 5.6: *Parámetros del programador de tasa de aprendizaje ReduceLROnPlateau.*

Parámetro	Símbolo / Nombre	Valor utilizado
Factor de reducción	<i>factor</i>	0.5
Paciencia	<i>patience</i>	20
Tasa mínima	$\eta_{\min}$	1×10^{-6}
Métrica monitorizada	—	<i>val_myloss</i>
Modo de reducción	—	<i>min</i>

Esto permite que el entrenamiento sea más eficiente al disminuir gradualmente la tasa de aprendizaje cuando el modelo se aproxima a un mínimo local, promoviendo una convergencia más suave y

estable. Además, evita la necesidad de definir manualmente un cronograma de aprendizaje, adaptándose automáticamente al comportamiento del error durante el entrenamiento.

Hardware Utilizado Durante las Pruebas

El hardware empleado durante la etapa de pruebas fue el descrito en la Tabla 5.7, garantizando coherencia entre las condiciones de entrenamiento y evaluación. En esta fase se utilizó la NVIDIA Jetson Nano B01, una plataforma de desarrollo optimizada para aplicaciones de inteligencia artificial de borde, que ofrece un equilibrio entre potencia de cómputo y eficiencia energética, ideal para la ejecución de modelos de visión por computadora en tiempo real.

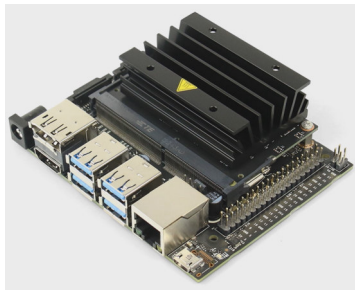


Figura 5.9: *Plataforma NVIDIA Jetson Nano B01 utilizada durante la etapa de pruebas. Esta tarjeta de desarrollo cuenta con GPU Maxwell de 128 núcleos y está optimizada para la ejecución de modelos de inteligencia artificial en el borde.*

Tabla 5.7: Especificaciones técnicas de la NVIDIA Jetson Nano B01 empleada durante las pruebas.

Especificación	Detalle
GPU	128 núcleos NVIDIA Maxwell GPU
CPU	Quad-core ARM Cortex-A57
Memoria	4 GB VRAM LPDDR4
Almacenamiento	Ranura microSD (tamaño definido por el usuario)
Conectividad	Ethernet Gigabit, Wi-Fi (mediante adaptador)
Salida de video	HDMI 2.0, DisplayPort 1.2
Interfaz de cámara	2x MIPI CSI-2 lanes
Puertos USB	4x USB 3.0 Tipo-A
GPIO	Encabezado de 40 pines (compatible con Raspberry Pi)
Entrada de energía	5V/4A (mediante conector de barril o micro-USB)
Dimensiones	100 mm × 80 mm
Sistema operativo	NVIDIA JetPack SDK basado en Ubuntu

5.5.6. Conclusiones del Capítulo

En este capítulo se presentó de forma integral la metodología seguida para el diseño, implementación y validación del modelo propuesto de detección automática de puntos cefalométricos. El desarrollo metodológico se estructuró de manera progresiva, comenzando con la comprensión del problema, la caracterización de la base de datos y la definición de las métricas de evaluación, hasta llegar a la descripción técnica de la arquitectura de la red neuronal convolucional y de las estrategias de optimización utilizadas durante el proceso de entrenamiento.

En primera instancia, se estableció la importancia de contar con un conjunto de datos de calidad y con anotaciones confiables. Para ello, se recurrió al uso de la base de datos del desafío IEEE ISBI 2015, la cual ofrece un punto de partida estandarizado y reconocido por la comunidad científica para la evaluación de algoritmos de detección cefalométrica.

El análisis de la variabilidad entre observadores fue una etapa fundamental de la metodología. A través del cálculo del error intra e interobservador, se pudo determinar el grado de consistencia en las anotaciones humanas, estableciendo un marco de referencia para la interpretación de los errores

del modelo. Esta etapa no solo permitió cuantificar la incertidumbre inherente al proceso manual de localización, sino también definir umbrales realistas de desempeño esperados para un sistema automatizado.

Posteriormente, se abordó el diseño de la arquitectura de la red neuronal convolucional, poniendo especial atención en la elección del número de capas, filtros, funciones de activación y técnicas de regularización. Estos elementos fueron seleccionados con el objetivo de lograr un equilibrio entre capacidad de representación y eficiencia computacional. La utilización de funciones de activación no lineales y capas de pooling permitió extraer características discriminantes a distintas escalas, mientras que el uso de técnicas de normalización y regularización (dropout) ayudó a reducir el sobreajuste y mejorar la capacidad de generalización.

La implementación del modelo se realizó empleando entornos de desarrollo ampliamente utilizados en la investigación actual, como TensorFlow y Keras, que facilitaron la definición modular de la arquitectura, el control del flujo de datos y la experimentación con diferentes configuraciones de hiperparámetros.

En cuanto a la fase de optimización, se seleccionó el optimizador Nadam debido a su capacidad de combinar la eficiencia adaptativa de Adam con las ventajas de la aceleración de Nesterov. Esta elección favoreció una convergencia más rápida y estable en comparación con otros métodos tradicionales de descenso de gradiente. A su vez, la incorporación de estrategias adaptativas para la modificación dinámica de la tasa de aprendizaje demostró ser un componente clave en el proceso de entrenamiento. En particular, la aplicación de la política de Cyclical Learning Rate (CLR) permitió explorar de manera más efectiva el espacio de soluciones, evitando estancamientos en mínimos locales y mejorando la capacidad de generalización del modelo en las etapas finales del aprendizaje.

La metodología desarrollada en este capítulo no solo proporciona una base sólida para el entrenamiento del modelo propuesto, sino que también establece un marco reproducible y extensible para futuras investigaciones en el campo de la cefalometría automatizada y la visión por computadora aplicada a la odontología.

Este capítulo consolida un proceso metodológico completo que permite pasar del planteamiento del problema a la implementación técnica del modelo de red neuronal convolucional. El enfoque seguido garantiza la validez y confiabilidad de los resultados que se presentan en el siguiente Capítulo 6, en el cual se analizan de manera detallada los resultados obtenidos y se discute el desempeño del modelo en comparación con los métodos existentes, tanto desde una perspectiva cuantitativa como cualitativa.

Capítulo 6

Resultados y Conclusiones

El presente capítulo presenta los resultados obtenidos a partir de la arquitectura de red neuronal convolucional propuesta para la identificación automática de etiquetas cefalométricas en radiografías laterales de cráneo, así como el análisis cuantitativo y cualitativo de su desempeño. Se describen en primer lugar los elementos que conforman la arquitectura final del modelo, seguida de la evaluación de su proceso de entrenamiento y su capacidad de generalización sobre conjuntos de datos independientes. Posteriormente, se presentan las métricas de desempeño empleadas, los resultados obtenidos para cada etiqueta cefalométrica y el análisis de los casos atípicos detectados. Finalmente, se discuten las conclusiones derivadas de los resultados experimentales.

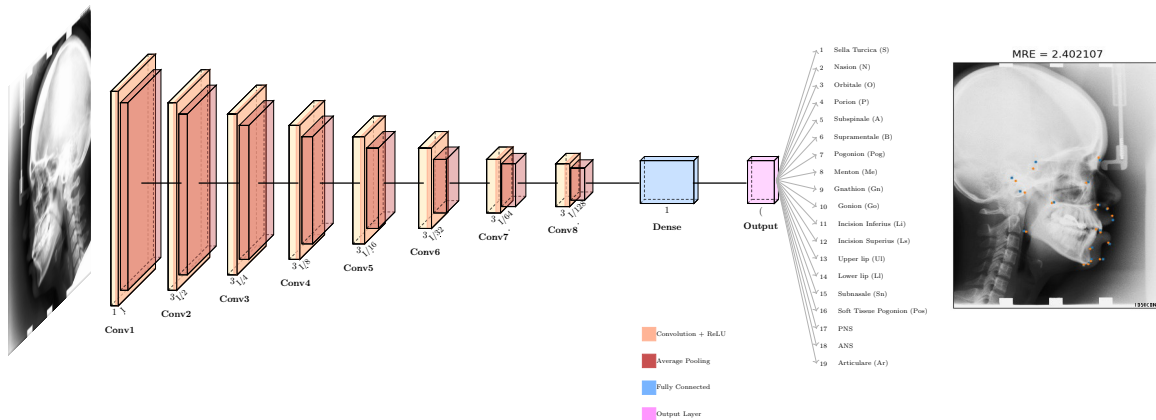


Figura 6.1: Red Neuronal Convolucional Propuesta Como Solución en la Identificación Automática de Etiquetas Cefalométricas.

En este contexto, la Figura 6.1 muestra la arquitectura de la red neuronal convolucional diseñada como solución al problema de localización automática de puntos cefalométricos. El modelo fue concebido para recibir como entrada una radiografía lateral de cráneo en escala de grises y procesarla mediante una secuencia de capas convolucionales que permiten extraer características jerárquicas

relevantes, culminando en un módulo de regresión encargado de estimar las coordenadas espaciales de cada etiqueta anatómica de interés.

6.1. Arquitectura de Red Final

En la Figura 6.1 se presenta la arquitectura de la red neuronal convolucional (CNN) propuesta para la localización automática de puntos cefalométricos en imágenes radiográficas laterales de cráneo. El modelo recibe como entrada una imagen radiográfica en escala de grises, la cual es procesada mediante una secuencia de capas convolucionales con funciones de activación no lineales. Posteriormente, se incorporan etapas de reducción espacial y, finalmente, una capa totalmente conectada encargada de realizar la regresión de las coordenadas de los puntos anatómicos de interés.

La arquitectura está compuesta por ocho bloques convolucionales principales (Conv1–Conv8), cada uno conformado por una operación de convolución seguida de una función de activación ReLU. El objetivo de estos bloques es la extracción progresiva de representaciones jerárquicas a partir de la imagen de entrada. En las primeras capas se capturan características de bajo nivel, como bordes y texturas, mientras que las capas más profundas permiten modelar estructuras anatómicas de mayor complejidad y patrones específicos asociados a los *landmarks* cefalométricos.

Después de cada bloque convolucional se incorporan etapas de *average pooling*, cuya función es reducir la dimensionalidad espacial de las características extraídas y proporcionar una invariancia traslacional parcial, al mismo tiempo que se preserva la información más relevante para la tarea de localización. La arquitectura emplea un número creciente de filtros a lo largo de las capas convolucionales, con una configuración de 16 filtros en la primera capa y 32 filtros en las capas subsecuentes, lo que permite capturar un espectro amplio de características conforme se profundiza en la red.

Las activaciones obtenidas en la última capa convolucional son posteriormente aplanadas mediante una operación *flatten* y enviadas a una capa totalmente conectada (*Fully Connected*), la cual integra la información global extraída durante el proceso convolucional. Finalmente, la red culmina en una capa de salida de regresión con 38 neuronas, correspondientes a las coordenadas bidimensionales (x, y) de los 19 puntos cefalométricos considerados en este estudio.

Los puntos anatómicos localizados por el modelo incluyen: Sella Turcica (S), Nasion (N), Orbitale (O), Porion (P), Subspinale (A), Supramentale (B), Pogonion (Pog), Menton (Me), Gnathion (Gn), Gonion (Go), Incision Inferius (Li), Incision Superius (Ls), Upper Lip (UL), Lower Lip (LL), Subnasale (Sn), Soft Tissue Pogonion (Pos), Posterior Nasal Spine (PNS), Anterior Nasal Spine (ANS) y Articulare (Ar).

Tabla 6.1: Detalles de las Capas de la Arquitectura de Red Neuronal Propuesta.

Capa	Filtros/Unidades	Kernel/Pool	Stride	Padding
Conv1	16	7x7	1	same
AvgPool1	–	3x3	1	same
Conv2	32	7x7	1	same
AvgPool2	–	2x2	1	same
Conv3	32	7x7	1	same
AvgPool3	–	2x2	1	same
Conv4	32	7x7	1	same
AvgPool4	–	2x2	1	same
Conv5	32	7x7	1	same
AvgPool5	–	2x2	1	same
Conv6	32	7x7	1	same
AvgPool6	–	2x2	1	same
Conv7	32	7x7	1	same
AvgPool7	–	2x2	1	same
Conv8	32	7x7	1	same
AvgPool8	–	2x2	1	same
Flatten	–	–	–	–
Dense	1024	–	–	–
Output	38	–	–	–

La Tabla 6.1, muestra a detalle los parámetros propuestos para cada una de las capas que conforman la arquitectura de Red Neuronal propuesta en la Figura 6.1.

6.2. Análisis de la Función de Pérdida

La Figura 6.2 muestra la evolución de la función de pérdida durante el proceso de entrenamiento del modelo propuesto, empleando el *Mean Radial Error* (MRE) como métrica principal de evaluación. El MRE cuantifica la distancia euclidiana promedio entre las coordenadas predichas por el modelo y las coordenadas reales de los puntos cefalométricos, por lo que resulta especialmente adecuado para evaluar tareas de localización automática de *landmarks* en imágenes radiográficas craneofaciales.

Durante las primeras épocas de entrenamiento se observa un descenso pronunciado del valor del MRE tanto en el conjunto de entrenamiento como en el conjunto de prueba. Este comportamiento es característico de la fase inicial de aprendizaje, en la cual el modelo ajusta rápidamente sus parámetros gracias a la elevada magnitud de los gradientes, permitiendo una corrección temprana de errores gruesos en la estimación de las coordenadas.

Posteriormente, ambas curvas entran en una fase de convergencia estable, en la cual las variaciones del MRE son progresivamente menores. Al final del entrenamiento, el modelo alcanza un valor promedio de MRE de 2.314 mm para el conjunto de entrenamiento y 3.060 mm para el conjunto de prueba. La

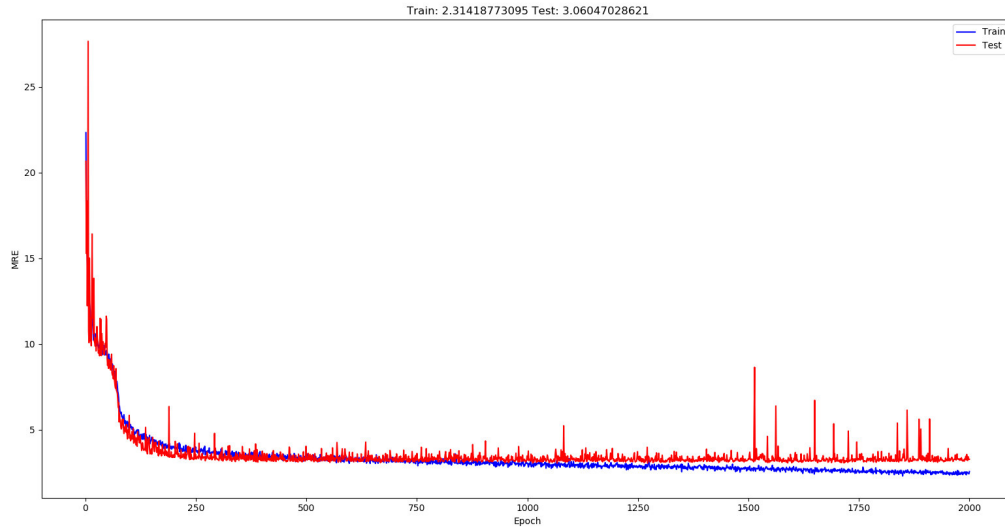


Figura 6.2: Evolución de la función de pérdida (MRE) durante el entrenamiento del modelo, utilizando como datos de referencia las anotaciones realizadas por el experto Senior.

diferencia observada entre ambas curvas ($\Delta_{\text{MRE}} \approx 0.75 \text{ mm}$) se mantiene dentro de un rango moderado, lo cual sugiere una adecuada capacidad de generalización y ausencia de sobreajuste severo.

6.2.1. Unidades de la predicción y del MRE

Es importante precisar las unidades en las que opera el modelo y en las que se reporta el error. Durante el entrenamiento, las coordenadas de los puntos cefalométricos fueron previamente normalizadas restando el centro de la imagen y aplicando un factor de escala de 0.1. En consecuencia, la red neuronal no predice directamente coordenadas en píxeles absolutos, sino en *décimas de píxel con origen en el centro de la imagen*. Matemáticamente, la transformación aplicada a las etiquetas fue:

$$\tilde{x} = 0.1 \cdot (x - x_c), \quad \tilde{y} = 0.1 \cdot (y - y_c), \quad (6.1)$$

donde (x, y) son las coordenadas originales en píxeles y (x_c, y_c) corresponde al centro de la imagen.

Dado que el modelo se entrenó con estas etiquetas normalizadas, el MRE calculado directamente a partir de la salida de la red se obtiene inicialmente en estas mismas unidades (décimas de píxel). Para recuperar el error en píxeles reales, el MRE debe multiplicarse por un factor de 10.

6.2.2. Conversión de píxeles a milímetros

Con el fin de reportar los resultados en unidades físicamente interpretables, las predicciones fueron convertidas de píxeles a milímetros mediante un factor de calibración s obtenido experimentalmente a partir de una distancia conocida en la imagen radiográfica:

$$s = \frac{D_{\text{real}} (\text{mm})}{D_{\text{imagen}} (\text{px})} \quad [\text{mm/px}] \quad (6.2)$$

Por ejemplo, si una distancia anatómica conocida de 50 mm corresponde a 250 píxeles en la imagen, entonces $s = 0.2 \text{ mm/px}$.

La conversión final de las predicciones a milímetros se realizó en dos pasos:

1. Recuperar las coordenadas en píxeles originales:

$$x = \frac{\tilde{x}}{0.1} + x_c, \quad y = \frac{\tilde{y}}{0.1} + y_c \quad (6.3)$$

2. Aplicar el factor de escala:

$$x_{\text{mm}} = s \cdot x, \quad y_{\text{mm}} = s \cdot y \quad (6.4)$$

Bajo este procedimiento, el MRE reportado en la Figura 6.2 y en los resultados finales corresponde al error radial promedio expresado en milímetros.

6.2.3. Interpretación de la curva de aprendizaje

Si bien la curva correspondiente al conjunto de prueba presenta incrementos aislados del error a lo largo del entrenamiento, estos picos pueden atribuirse a la presencia de muestras atípicas o a variaciones anatómicas particulares dentro del conjunto de validación. No obstante, dichas fluctuaciones no afectan la tendencia general de convergencia, la cual se mantiene estable a lo largo de las 2000 épocas evaluadas.

La curva mostrada en la Figura 6.2, también conocida como curva de aprendizaje, constituye una herramienta fundamental para el análisis del comportamiento del modelo durante el entrenamiento. A partir de ella es posible identificar fenómenos como el sobreajuste o el subajuste, así como evaluar el impacto de modificaciones en la arquitectura de la red y en los hiperparámetros empleados.

6.2.4. Hiperparámetros y parámetros del modelo

En este trabajo se distingue explícitamente entre *hiperparámetros* y *parámetros* del modelo. Los hiperparámetros son aquellos definidos por el diseñador y externos a la red neuronal (por ejemplo, tasa

de aprendizaje, optimizador, tamaño de lote y número de épocas). En contraste, los parámetros son los pesos y sesgos internos de la red que se ajustan automáticamente durante el entrenamiento mediante retropropagación.

6.2.5. 6.2.4. Hiperparámetros y parámetros del modelo

En este trabajo se distingue explícitamente entre *hiperparámetros* y *parámetros* del modelo. Los hiperparámetros corresponden a las decisiones de diseño fijadas por el investigador antes del entrenamiento y que no son aprendidas por la red neuronal. Por su parte, los parámetros del modelo son los pesos y sesgos internos de la red que se ajustan automáticamente durante el proceso de entrenamiento mediante el algoritmo de retropropagación y optimización.

Hiperparámetros de optimización. Para el entrenamiento del modelo se empleó el optimizador Nadam (Nesterov-accelerated Adaptive Moment Estimation), el cual combina las ventajas de Adam con el impulso de Nesterov para mejorar la estabilidad y velocidad de convergencia. La tasa de aprendizaje utilizada fue $\eta = 0.0002$, seleccionada experimentalmente para lograr un equilibrio entre velocidad de aprendizaje y estabilidad del entrenamiento.

Hiperparámetros de entrenamiento. El modelo se entrenó durante 2000 épocas con un tamaño de lote (*batch size*) igual a 1. Durante el entrenamiento se empleó mezcla aleatoria de las muestras (*shuffle=True*) y se evaluó el desempeño del modelo en cada época utilizando un conjunto de validación independiente.

Hiperparámetros de regularización. Con el fin de reducir el riesgo de sobreajuste, se incorporó regularización mediante Dropout con una tasa de 0.2 (20 %) aplicada después de cada bloque convolucional y antes y después de la capa densa principal.

Hiperparámetros de arquitectura. El modelo propuesto corresponde a una red neuronal convolucional profunda (DCNN). La arquitectura fue definida con los siguientes hiperparámetros estructurales:

- Número total de capas convolucionales: 8.
- Tamaño del kernel: 7×7 en todas las capas convolucionales.
- Tipo de *padding*: *same*.
- Función de activación en capas ocultas: ReLU.
- Inicialización de pesos: Glorot Uniform.
- Número de filtros:

- Primera capa: 16 filtros.
- Capas subsecuentes: 32 filtros cada una.
- Tipo de *pooling*: Average Pooling.
- Tamaño de ventana de pooling: (3,3) en la primera capa y (2,2) en las capas restantes.

Tabla 6.2: Resumen del modelo obtenido de Keras.

Tipo de capa	Dimensiones de salida	Número de parámetros
Capa de entrada (Input)	(1, 2400, 1935)	0
Convolutacional2D ₁	(1, 2400, 16)	1,517,056
Average Pooling ₁	(1, 800, 16)	0
Dropout ₁	(1, 800, 16)	0
Convolutacional2D ₂	(1, 800, 32)	25,120
Average Pooling ₂	(1, 400, 32)	0
Dropout ₂	(1, 400, 32)	0
Convolutacional2D ₃	(1, 400, 32)	50,208
Average Pooling ₃	(1, 200, 32)	0
Dropout ₃	(1, 200, 32)	0
Convolutacional2D ₄	(1, 200, 32)	50,208
Average Pooling ₄	(1, 100, 32)	0
Dropout ₄	(1, 100, 32)	0
Convolutacional2D ₅	(1, 100, 32)	50,208
Average Pooling ₅	(1, 50, 32)	0
Dropout ₅	(1, 50, 32)	0
Convolutacional2D ₆	(1, 50, 32)	50,208
Average Pooling ₆	(1, 25, 32)	0
Dropout ₆	(1, 25, 32)	0
Convolutacional2D ₇	(1, 25, 32)	50,208
Average Pooling ₇	(1, 13, 32)	0
Dropout ₇	(1, 13, 32)	0
Convolutacional2D ₈	(1, 13, 32)	50,208
Average Pooling ₈	(1, 7, 32)	0
Dropout ₈	(1, 7, 32)	0
Flatten	(224)	0
Dropout ₉	(224)	0
Dense ₁	(1024)	230,400
Dropout ₁₀	(1024)	0
Capa de salida (Dense)	(38)	38,950
Total de parámetros entrenables		2,112,774

Hiperparámetros de salida. La capa de salida es una capa densa con 38 neuronas, correspondientes a las coordenadas (x, y) de 19 puntos cefalométricos. Se utilizó una función de activación lineal,

adecuada para problemas de regresión continua.

Hiperparámetros de preprocesamiento. El preprocesamiento de los datos también constituye un conjunto de hiperparámetros fundamentales en el desempeño del modelo:

- Las imágenes fueron normalizadas dividiendo los valores de intensidad entre 255, de modo que los píxeles se encontraran en el rango $[0, 1]$.
- Las coordenadas de los puntos cefalométricos fueron centradas respecto al centro de la imagen y escaladas por un factor de 0.1 antes del entrenamiento.

Parámetros del modelo. Los parámetros del modelo corresponden a los pesos y sesgos aprendidos en cada capa durante el entrenamiento. De acuerdo con el resumen de la arquitectura presentado en la Tabla ??, el modelo cuenta con un total de 2,112,774 parámetros entrenables. Estos incluyen los pesos de las capas convolucionales y densas, los cuales fueron optimizados mediante el algoritmo Nadam para minimizar el error cuadrático medio durante el entrenamiento.

6.3. Análisis de Resultados de Localización por cada Imagen de los Subconjuntos de Prueba

Una vez entrenado el sistema y obtenidos los pesos óptimos de la red neuronal, se procede a la evaluación del desempeño del modelo sobre los subconjuntos de prueba *Test1* y *Test2*. Para cada imagen de estos subconjuntos se calcula el *Mean Radial Error* (MRE), el cual corresponde al promedio del error radial obtenido a partir de las 19 etiquetas cefalométricas presentes en cada imagen.

Tabla 6.3: Error radial medio (MRE) por imagen para el conjunto Test1 de 150 muestras.

Im.	MRE (mm)	Im.	MRE (mm)	Im.	MRE (mm)	Im.	MRE (mm)	Im.	MRE (mm)
1	1.86	31	1.98	61	2.36	91	2.45	121	4.08
2	3.89	32	3.16	62	2.52	92	3.25	122	3.84
3	2.97	33	2.41	63	2.84	93	2.85	123	3.61
4	2.17	34	4.01	64	2.70	94	3.10	124	3.49
5	2.84	35	2.90	65	4.45	95	4.46	125	2.68
6	3.31	36	2.08	66	3.48	96	2.61	126	1.91
7	2.36	37	3.67	67	2.26	97	3.18	127	3.13
8	2.94	38	3.29	68	1.70	98	2.85	128	2.52
9	2.60	39	2.54	69	2.23	99	4.10	129	1.60
10	2.61	40	2.58	70	3.41	100	2.99	130	2.58
11	3.15	41	2.12	71	2.93	101	3.39	131	2.09
12	1.94	42	2.47	72	3.64	102	2.87	132	2.56
13	3.39	43	2.15	73	3.42	103	2.71	133	2.09
14	3.68	44	5.78	74	2.05	104	2.71	134	3.21
15	3.58	45	2.99	75	2.88	105	2.53	135	1.98
16	2.67	46	3.47	76	2.91	106	3.78	136	3.47
17	3.93	47	5.90	77	3.09	107	2.35	137	2.11
18	3.12	48	3.38	78	2.88	108	2.34	138	3.08
19	3.10	49	2.99	79	4.08	109	3.05	139	1.39
20	3.51	50	2.60	80	2.20	110	2.98	140	2.31
21	4.35	51	2.05	81	2.96	111	3.02	141	1.92
22	3.76	52	2.13	82	10.62	112	2.08	142	2.14
23	2.24	53	1.86	83	2.81	113	2.43	143	2.44
24	3.44	54	3.56	84	4.19	114	3.73	144	3.21
25	2.82	55	4.58	85	2.94	115	2.32	145	1.53
26	3.21	56	2.96	86	2.49	116	3.02	146	1.70
27	3.75	57	2.52	87	2.04	117	5.44	147	2.22
28	4.23	58	3.77	88	2.61	118	4.09	148	4.72
29	11.04	59	2.62	89	3.21	119	3.84	149	2.70
30	4.86	60	3.61	90	4.29	120	3.61	150	2.59

Tabla 6.4: Error radial medio (MRE) por imagen para el conjunto Test2 de 100 muestras.

Im. MRE (mm)	Im. MRE (mm)	Im. MRE (mm)	Im. MRE (mm)	Im. MRE (mm)
1 2.32	21 3.01	41 2.54	61 2.47	81 18.76
2 21.39	22 2.81	42 3.10	62 16.30	82 2.23
3 4.82	23 3.04	43 2.13	63 3.13	83 2.10
4 2.20	24 2.69	44 3.44	64 3.33	84 3.52
5 1.94	25 2.62	45 3.91	65 5.92	85 2.95
6 2.74	26 2.97	46 3.94	66 2.46	86 5.11
7 2.88	27 3.06	47 2.51	67 1.89	87 3.94
8 1.91	28 3.29	48 2.50	68 3.93	88 2.89
9 2.81	29 2.87	49 4.47	69 2.40	89 5.77
10 3.34	30 3.16	50 3.35	70 2.61	90 2.76
11 2.67	31 4.14	51 1.95	71 2.37	91 3.41
12 3.30	32 2.76	52 2.26	72 2.58	92 2.63
13 3.73	33 3.13	53 2.39	73 3.19	93 3.66
14 1.89	34 3.23	54 1.94	74 2.74	94 3.55
15 5.90	35 2.09	55 3.45	75 14.90	95 2.96
16 3.72	36 2.49	56 2.06	76 4.57	96 3.11
17 2.49	37 2.66	57 2.46	77 2.39	97 2.90
18 4.50	38 4.71	58 5.67	78 2.53	98 4.26
19 3.93	39 4.64	59 6.30	79 2.08	99 3.52
20 1.48	40 2.75	60 2.95	80 12.07	100 2.00

Las Tablas 6.3 y 6.4 presenta los valores del Error Radial Medio (MRE) obtenidos para cada una de las 250 imágenes que conforman el conjunto de prueba *Test1* y *Test2*. En cada caso, el valor reportado corresponde al promedio del error de localización de las 19 etiquetas cefalométricas evaluadas en la imagen correspondiente.

Con el fin de facilitar la interpretación de los resultados y resaltar visualmente el comportamiento del modelo, se aplicó un esquema de codificación por colores en la columna correspondiente al MRE, definido de la siguiente manera:

- **Verde suave:** indica valores bajos de error, específicamente aquellos menores a 2 mm, los cuales representan estimaciones con alta precisión.
- **Rojo suave:** resalta valores elevados de error, definidos como aquellos mayores a 5 mm, que representan desviaciones significativas respecto a la referencia.
- **Rojo intenso:** identifica valores atípicos severos (*outliers*), correspondientes a errores mayores a 10 mm, los cuales evidencian fallas puntuales del modelo en imágenes específicas.

Este esquema visual permite identificar de manera inmediata tanto los casos con desempeño sobresaliente como aquellas muestras que requieren un análisis más detallado debido a su elevada desviación.

6.3.1. Resultados por Imagen en el Conjunto Test1

La Figura 6.3 muestra la distribución del Error Radial Medio (MRE) por imagen correspondiente al conjunto de prueba *Test1*, compuesto por 150 muestras.

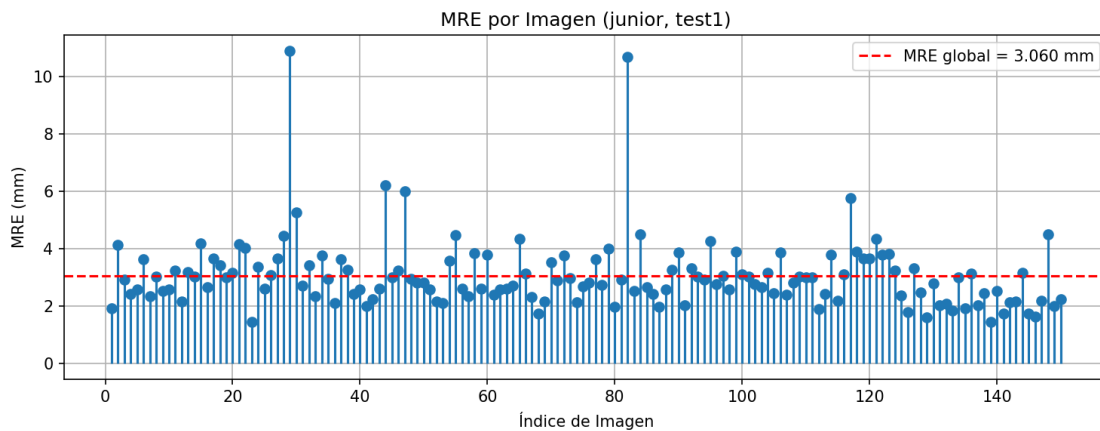


Figura 6.3: Distribución del Error Radial Medio (MRE) por imagen para el conjunto de prueba *Test1*. La línea discontinua representa el MRE promedio global del conjunto.

Cada marcador representa el MRE promedio de una imagen individual, mientras que la línea horizontal discontinua indica el valor promedio global del MRE para este conjunto, el cual es de 3.06 mm.

Se observa que la mayoría de las imágenes presentan errores inferiores a 4 mm, lo que refleja un comportamiento estable y consistente del modelo. No obstante, se identifican algunos picos aislados con valores de error significativamente mayores, los cuales corresponden a muestras atípicas. Estos casos coinciden con los valores resaltados en rojo intenso en la Tabla 6.3, y pueden atribuirse a variaciones en la calidad de la imagen, ambigüedad anatómica o discrepancias en las anotaciones de referencia.

Análisis Cualitativo del Desempeño en el Conjunto Test1

Con el fin de complementar el análisis cuantitativo del Error Radial Medio (MRE) presentado previamente para el conjunto de prueba *Test1*, se llevó a cabo un análisis cualitativo mediante la inspección visual de los casos con mejor y peor desempeño del modelo. Este análisis permite evaluar de

manera directa la correspondencia entre las predicciones generadas por la red neuronal y las anotaciones de referencia proporcionadas por el experto.

En la Figura 6.4 se presentan las seis imágenes del conjunto *Test1* que obtuvieron los valores más bajos de MRE. En estas muestras se observa una alta precisión en la localización de las etiquetas cefalométricas, con una alineación adecuada entre los puntos predichos y los puntos de referencia. La correcta identificación de las estructuras anatómicas en estos casos evidencia la capacidad del modelo para generalizar de manera efectiva cuando las características anatómicas se encuentran claramente definidas y presentan un contraste adecuado.

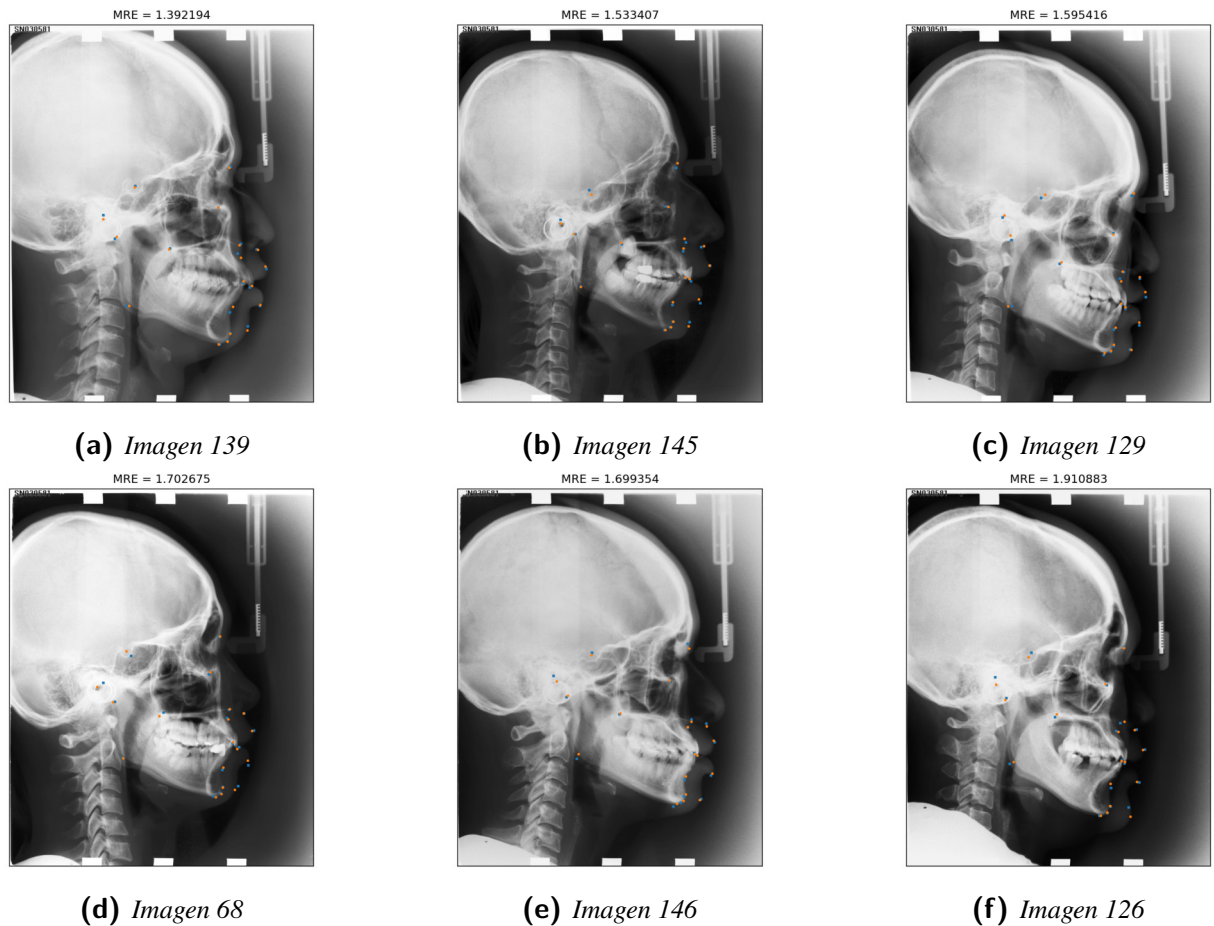


Figura 6.4: Imágenes del conjunto *Test1* con los seis valores más bajos de Error Radial Medio (MRE).

Por el contrario, la Figura 6.5 muestra las seis imágenes del conjunto *Test1* que presentan los valores más altos de MRE. En estos casos se identifican discrepancias notables entre las predicciones del modelo y las anotaciones de referencia, especialmente en regiones anatómicas complejas o con bajo

contraste. Dichas desviaciones corresponden a errores puntuales del sistema y coinciden con los valores atípicos identificados en el análisis estadístico del MRE por imagen.

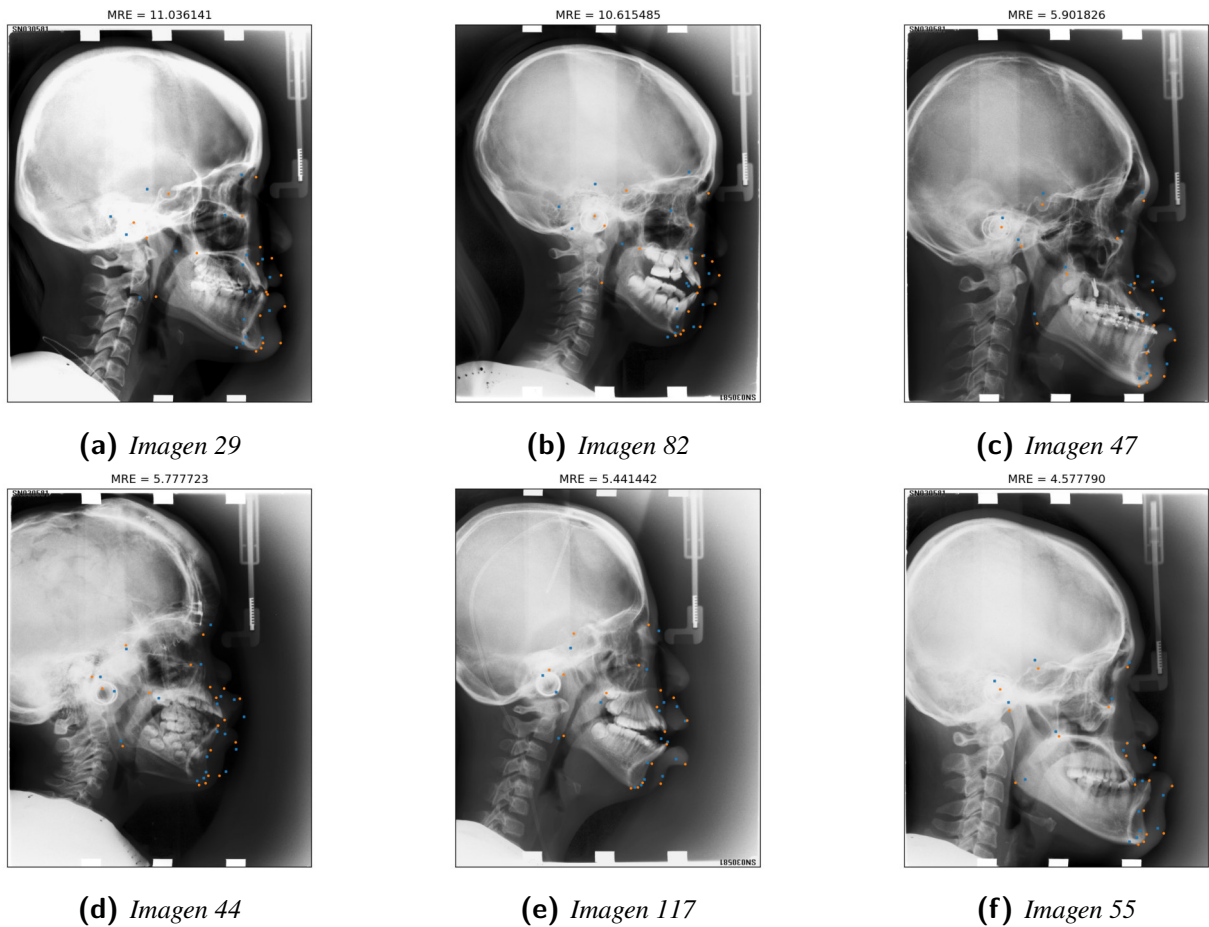


Figura 6.5: Imágenes del conjunto *Test1* con los seis valores más altos de Error Radial Medio (MRE).

La comparación entre ambos grupos de imágenes pone de manifiesto la variabilidad del desempeño del modelo a nivel individual y resalta la importancia de complementar las métricas cuantitativas con un análisis visual. Este enfoque permite identificar patrones asociados a errores elevados y proporciona información relevante para futuras mejoras en la arquitectura del modelo, el preprocesamiento de las imágenes o la calidad de las anotaciones cefalométricas.

6.3.2. Resultados por Imagen en el Conjunto Test2

La Figura 6.6 presenta la distribución del Error Radial Medio (MRE) por imagen para el conjunto de prueba *Test2*, conformado por 100 muestras. De forma análoga al conjunto anterior, cada punto

representa el MRE promedio calculado a partir de las 19 etiquetas cefalométricas de cada imagen, mientras que la línea horizontal discontinua indica el MRE promedio global, con un valor de 3.81 mm.

En este conjunto se observa una mayor dispersión de los errores y la presencia de varios picos pronunciados con valores superiores a 10 mm, los cuales corresponden a outliers severos. Estos casos indican situaciones en las que el modelo presenta desviaciones importantes respecto a la referencia, posiblemente asociadas a condiciones particulares de las imágenes o a una mayor complejidad anatómica en dichas muestras.

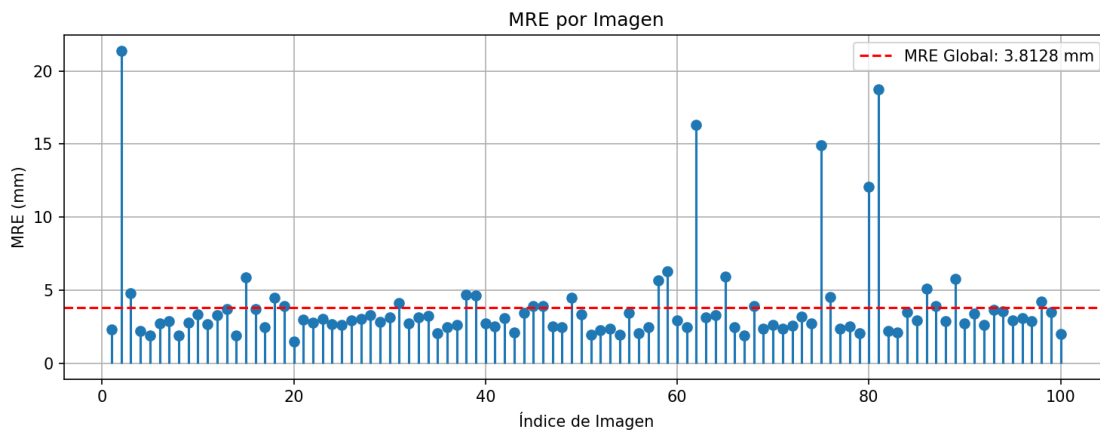


Figura 6.6: Distribución del Error Radial Medio (MRE) por imagen para el conjunto de prueba *Test2*. La línea discontinua representa el MRE promedio global del conjunto.

Análisis Cualitativo del Desempeño en el Conjunto *Test2*

Con el propósito de complementar el análisis cuantitativo del Error Radial Medio (MRE) obtenido para el conjunto de prueba *Test2*, se realizó un análisis cualitativo mediante la inspección visual de las imágenes que presentan el mejor y el peor desempeño del modelo. Este análisis permite evaluar de forma directa la correspondencia entre las predicciones generadas por la red neuronal y las anotaciones cefalométricas de referencia.

A partir de los valores reportados en la Tabla 6.4, se seleccionaron las seis imágenes con los valores más bajos de MRE, las cuales corresponden a los casos con mayor precisión en la localización de las etiquetas cefalométricas. Estas imágenes se muestran en la Figura 6.7. En dichos casos se observa una adecuada alineación entre los puntos predichos y los puntos de referencia, lo que indica un desempeño robusto del modelo cuando las estructuras anatómicas presentan una definición clara y un contraste adecuado.

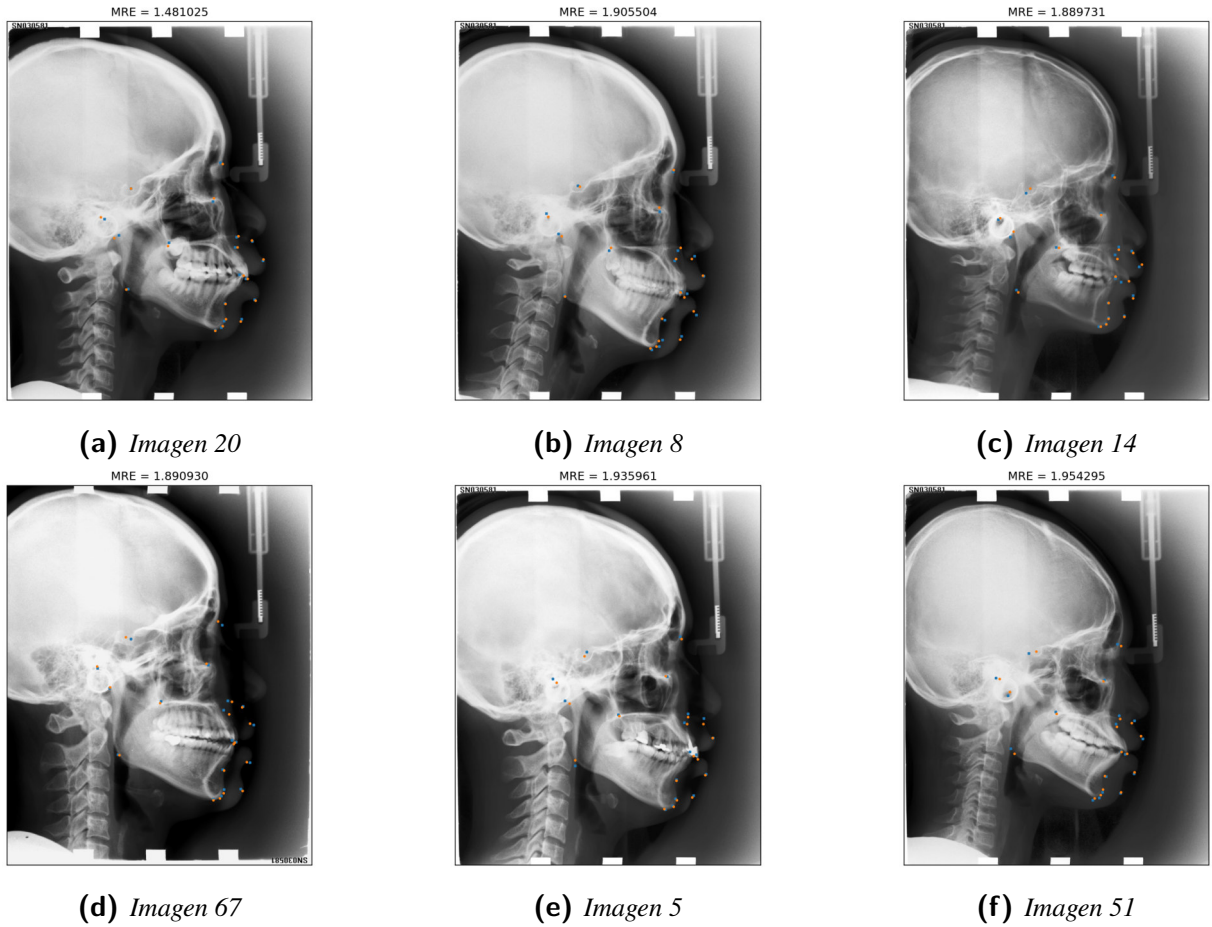


Figura 6.7: Imágenes del conjunto *Test2* con los seis valores más bajos de Error Radial Medio (MRE).

Por otro lado, se identificaron las seis imágenes con los valores más altos de MRE dentro del conjunto *Test2*, las cuales corresponden a casos atípicos en los que el modelo presenta desviaciones significativas respecto a la referencia. Estas imágenes se presentan en la Figura 6.8. En ellas se aprecian errores notorios en la localización de algunas etiquetas cefalométricas, particularmente en regiones anatómicas complejas o con bajo contraste, lo que explica los elevados valores de MRE registrados.

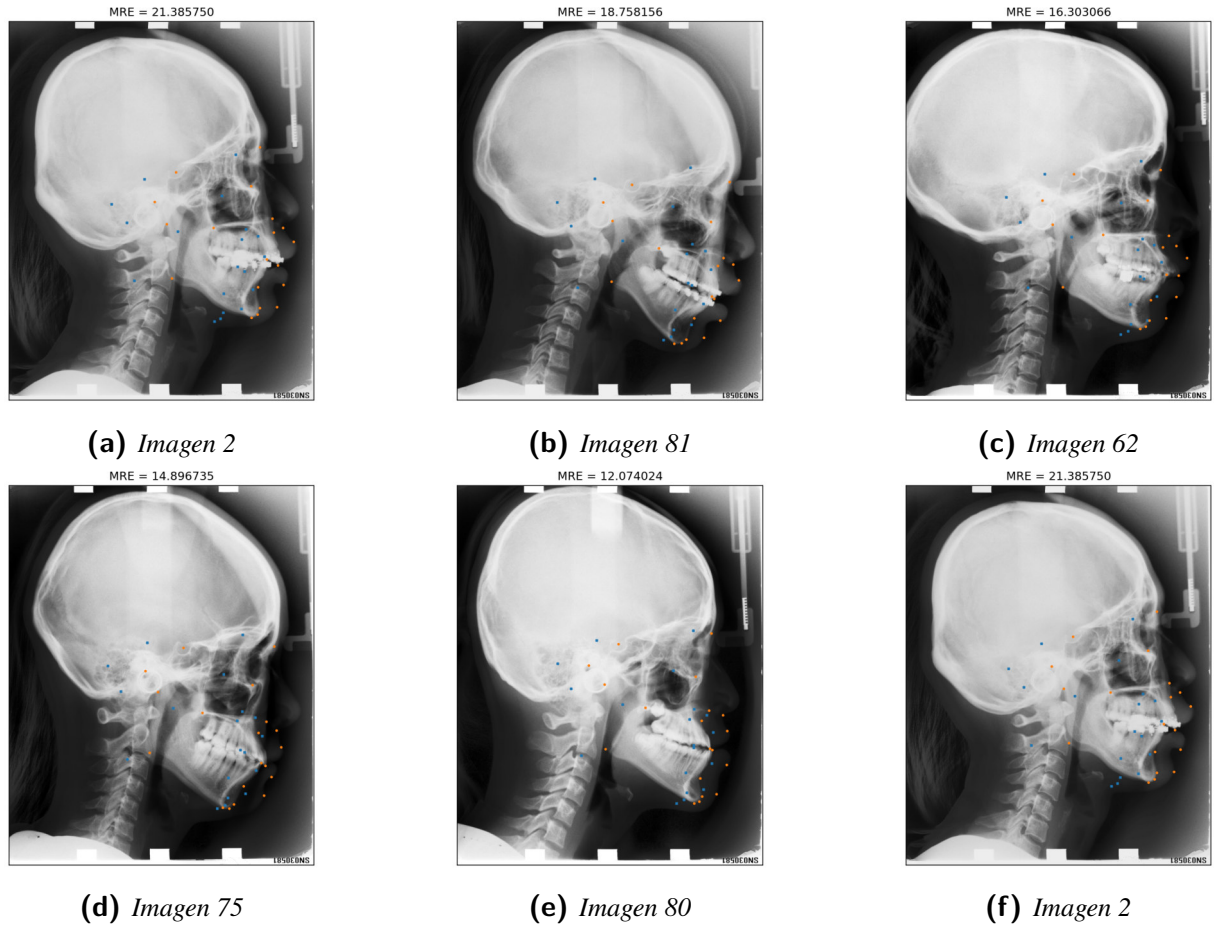


Figura 6.8: Imágenes del conjunto *Test2* con los seis valores más altos de Error Radial Medio (MRE).

La comparación entre los casos de bajo y alto error en el conjunto *Test2* evidencia una mayor dispersión en los valores de MRE en comparación con el conjunto *Test1*, así como la presencia de outliers severos. Este comportamiento sugiere que el modelo enfrenta mayores dificultades de generalización en este subconjunto, lo cual puede estar asociado a variaciones en la calidad de las imágenes, diferencias anatómicas más marcadas o inconsistencias en las anotaciones de referencia. No obstante, el buen desempeño observado en los casos de bajo error confirma la capacidad del modelo para realizar una localización precisa de puntos cefalométricos bajo condiciones favorables.

6.4. Análisis de Resultados de Localización por Etiqueta Cefalométrica

Con el objetivo de analizar el desempeño del modelo a nivel de cada punto cefalométrico, se calculó el Error Radial Medio (MRE) promedio para cada etiqueta en los conjuntos de prueba *Test1* y

Test2. Este análisis permite identificar aquellos puntos cuya localización resulta más precisa, así como aquellos que presentan mayor dificultad para el modelo.

La Tabla 6.5 presenta el MRE promedio obtenido para cada una de las etiquetas cefalométricas evaluadas. En términos generales, el conjunto *Test1* presenta un MRE global de 3.10 mm, mientras que en *Test2* dicho error aumenta a 3.81 mm, correspondiente a un incremento relativo del 23.0 % respecto al conjunto *Test1*, lo cual evidencia una mayor variabilidad anatómica inter-paciente y diferencias en las condiciones radiográficas de adquisición.

Etiqueta cefalométrica	MRE Promedio [mm]	
	Test1	Test2
1.- Sella turcica	3.75284819510018	4.42859391425897
2.- Nasion	3.24126776281455	4.02684865602129
3.- Orbitale	3.00897519115971	4.03996772474209
4.- Porion	3.87768231011651	5.35540035106731
5.- Subspinale	2.97395911011125	3.71998903227134
6.- Supramentale	2.65546788163166	2.85955027371976
7.- Pogonion	2.29928989069753	2.96038471835939
8.- Menton	2.41223445682595	2.80753342513088
9.- Gnathion	2.23741287992698	2.75425181427024
10.- Gonion	3.61053982673823	4.28566403405054
11.- Lower incisal incision	2.51754245177787	3.24765350165001
12.- Upper incisal incision	2.62239232611127	3.60812523573349
13.- Upper lip	2.99101837388186	4.27412821723367
14.- Lower lip	3.23121508021343	3.93825386312178
15.- Subnasale	2.83023146195681	3.70533081540525
16.- Soft tissue pogonion	5.01564273625773	3.66527028545858
17.- Posterior nasal spine	2.84695094135673	4.00957205595919
18.- Anterior nasal spine	3.07009999349115	3.82540333751207
19.- Articulare	3.66912274014353	4.93107236574266
Promedio	3.09809966	3.812789138

Tabla 6.5: MRE promedio obtenido para cada etiqueta cefalométrica.

Con el fin de complementar este análisis, la Figura 6.9 muestra la distribución del MRE promedio por etiqueta para el conjunto *Test1*. La mayoría de los landmarks presentan errores inferiores o cercanos al promedio global, destacando *Gnathion*, *Pogonion* y *Menton*. El landmark *Soft Tissue Pogonion* presenta el mayor error promedio y una alta variabilidad.

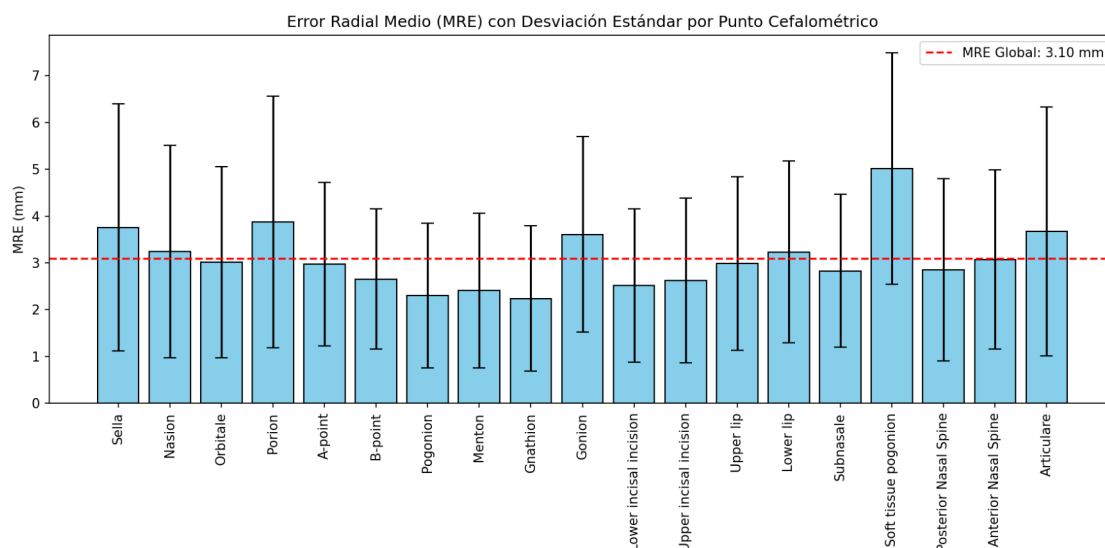


Figura 6.9: Error Radial Medio (MRE) promedio por punto cefalométrico con desviación estándar para el conjunto *Test1*.

De manera similar, la Figura 6.10 presenta los resultados correspondientes al conjunto *Test2*. Se observa un incremento generalizado del MRE y su desviación estándar. Landmarks como *Porion*, *Articulare* y *Upper Lip* muestran errores superiores al promedio global.

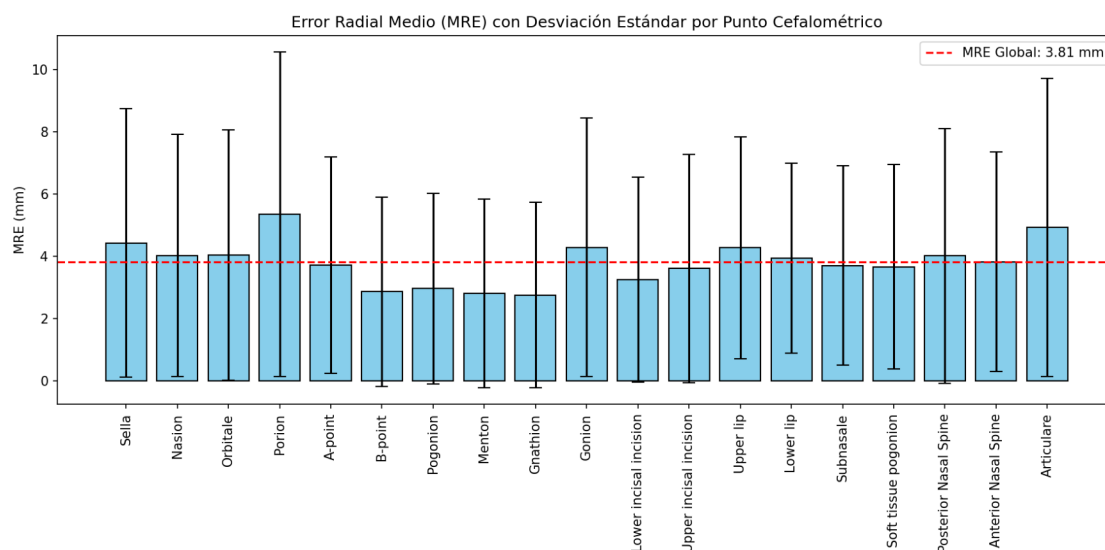


Figura 6.10: Error Radial Medio (MRE) promedio por punto cefalométrico con desviación estándar para el conjunto *Test2*.

6.4.1. Análisis del Error Radial Medio por Landmark Cefalométrico

La Figura 6.11 muestra la distribución del MRE por imagen para cada landmark en *Test1*. La mayoría de los puntos presentan errores inferiores a 3 mm, destacando Gnathion, Menton, Pogonion y Supramentale.

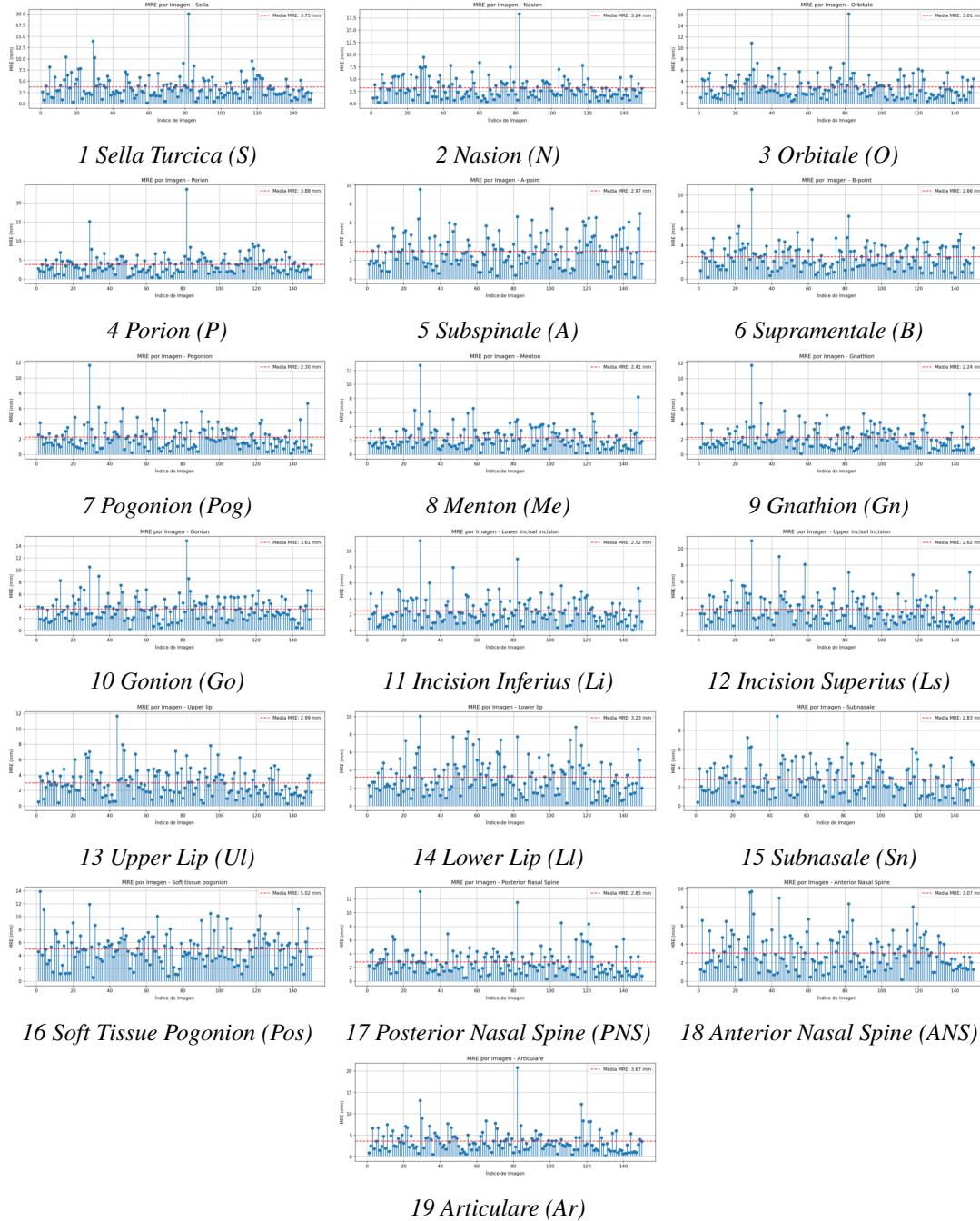


Figura 6.11: Distribución del Error Radial Medio (MRE) para cada punto cefalométrico en las 150 imágenes de prueba del conjunto de datos *Test1*.

En la Figura 6.12 se presentan los resultados del conjunto *Test2*, donde se observa un incremento del MRE conservando patrones de dispersión similares.

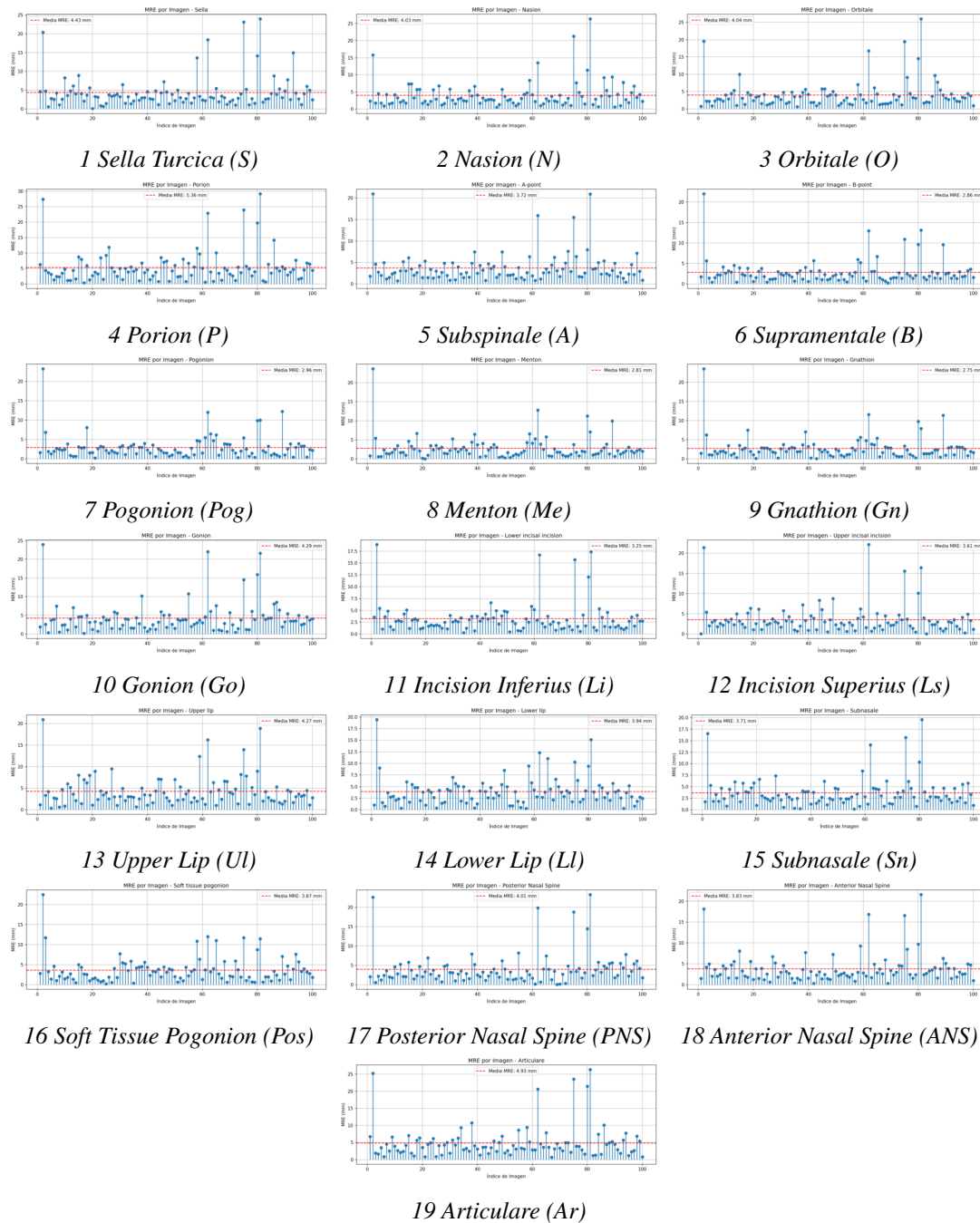


Figura 6.12: Distribución del Error Radial Medio (MRE) para cada punto cefalométrico en las 100 imágenes de prueba del conjunto de datos *Test2*.

En conjunto, los resultados confirman que los mayores errores se concentran en landmarks de alta complejidad anatómica y tejidos blandos, mientras que los landmarks mandibulares presentan el mejor

desempeño, confirmando la aplicabilidad clínica potencial del modelo.

6.5. Análisis de Resultados Generales considerando la Tasa de Detección Exitosa (SDR)

La Tasa de Detección Exitosa (Successful Detection Rate, SDR) es una métrica ampliamente utilizada en cefalometría computacional para evaluar la proporción de landmarks cuya localización automática se encuentra dentro de un umbral clínicamente aceptable respecto a la anotación de referencia. En este estudio se emplearon umbrales de 2, 2.5, 3 y 4 mm, los cuales permiten analizar el desempeño del sistema bajo distintos criterios de precisión clínica.

Las Tablas 6.6 y 6.7 presentan los resultados detallados del Error Radial Medio (MRE), la desviación estándar y los porcentajes de detecciones exitosas para cada landmark cefalométrico en los conjuntos *Test1* y *Test2*, respectivamente.

Tabla 6.6: Resultados del error radial medio (MRE), desviación estándar y porcentaje de detecciones dentro de distintos umbrales SDR (*Test1*).

Punto	MRE (mm)	Desviación Estándar (mm)	SDR 2 mm (%)	SDR 2.5 mm (%)	SDR 3 mm (%)	SDR 4 mm (%)
Sella	3.75	2.64	23.33	36.67	47.33	64.00
Nasion	3.24	2.27	33.33	41.33	54.67	70.00
Orbitale	3.01	2.04	37.33	47.33	58.67	74.00
Porion	3.88	2.69	20.67	31.33	43.33	59.33
A-point	2.97	1.74	35.33	44.00	59.33	75.33
B-point	2.66	1.50	38.00	50.67	62.00	85.33
Pogonion	2.30	1.55	50.00	64.00	74.00	86.00
Menton	2.41	1.65	48.67	61.33	69.33	88.00
Gnathion	2.24	1.55	54.00	63.33	75.33	88.67
Gonion	3.61	2.09	24.00	32.67	43.33	67.33
Lower incisal incision	2.52	1.64	41.33	62.00	71.33	84.67
Upper incisal incision	2.62	1.76	44.00	56.00	65.33	80.67
Upper lip	2.99	1.85	33.33	45.33	56.67	74.00
Lower lip	3.23	1.94	30.00	42.67	56.67	69.33
Subnasale	2.83	1.64	35.33	48.67	61.33	76.00
Soft tissue pogonion	5.02	2.47	10.00	16.67	19.33	37.33
Posterior nasal spine	2.85	1.95	40.67	52.67	65.33	79.33
Anterior nasal spine	3.07	1.91	33.33	46.67	58.67	71.33
Articulare	3.67	2.66	26.67	36.67	49.33	63.33
Promedios	3.10	1.98	34.70	46.32	57.44	73.37

Finalmente, la Tabla 6.8 resume los rangos promedio de SDR obtenidos para ambos conjuntos de prueba, evidenciando una disminución generalizada del desempeño en *Test2*, consistente con el incremento observado en el MRE global.

Tabla 6.7: Resultados del error radial medio (MRE), desviación estándar y porcentaje de detecciones dentro de distintos umbrales SDR (Test2).

Punto	MRE (mm)	Desviación Estándar (mm)	SDR 2 mm (%)	SDR 2.5 mm (%)	SDR 3 mm (%)	SDR 4 mm (%)
Sella	4.42	4.32	20	29	44	63
Nasion	4.02	3.89	29	41	52	66
Orbitale	4.03	4.03	29	39	47	65
Porion	5.35	5.21	21	25	32	48
A-point	3.71	3.48	29	42	55	70
B-point	2.85	3.04	48	63	74	86
Pogonion	2.96	3.05	42	56	67	85
Menton	2.80	3.02	47	62	70	83
Gnathion	2.75	2.97	50	59	73	88
Gonion	4.28	4.15	27	32	42	64
Lower incisal incision	3.24	3.29	39	46	64	81
Upper incisal incision	3.60	3.67	29	43	55	75
Upper lip	4.27	3.56	22	32	43	60
Lower lip	3.93	3.05	25	34	46	60
Subnasale	3.70	3.19	28	39	53	69
Soft tissue pogonion	3.66	3.28	36	43	50	69
Posterior Nasal Spine	4.00	4.08	26	40	48	67
Anterior Nasal Spine	3.82	3.53	27	41	54	72
Articulare	4.93	4.79	21	27	37	54
Promedios	3.81	3.66	31.32	41.74	52.95	69.74

Rango SDR (%)				
Conjunto	2 mm	2.5 mm	3 mm	4 mm
Test1 (150 imágenes)	34.70	46.32	57.44	73.37
Test2 (100 imágenes)	31.32	41.74	52.95	69.74

Tabla 6.8: Rango de SDR para los conjuntos de datos Test1 y Test2.

En conjunto, los resultados muestran que el modelo alcanza tasas de detección superiores al 70 % bajo el umbral clínico de 4 mm, mientras que los landmarks mandibulares bien definidos (Gn, Me, Pog, B) presentan consistentemente los mayores valores de SDR, confirmando la robustez del sistema en estructuras óseas de alta definición radiográfica.

6.6. Comparación de Resultados

Con el objetivo de evaluar el desempeño del modelo propuesto frente a métodos del estado del arte, se realizó una comparación directa utilizando el conjunto de datos IEEE ISBI 2015 en sus particiones Test1 y Test2. Las Tablas 6.10 y 6.11 resumen los valores reportados de la Tasa de Detección Exitosa (SDR) bajo distintos umbrales de error radial, así como el Error Radial Medio (MRE).

En la Tabla 6.10 se observa que los métodos más recientes alcanzan valores de SDR superiores al 85 % para el umbral $SDR < 2$ mm y cercanos o superiores al 95 % para $SDR < 4$ mm, con errores

MRE que oscilan típicamente entre 0.9 mm y 1.5 mm. Estas arquitecturas incorporan estrategias complejas como mecanismos de atención multi-cabeza, redes multiescala, grafos convolucionales (GCN) y decodificadores basados en mapas de calor, lo cual incrementa sustancialmente la profundidad de la red y el número de parámetros entrenables.

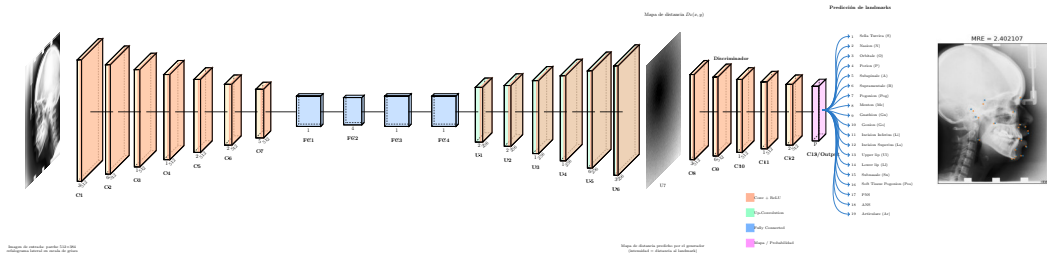


Figura 6.13: Arquitectura adversarial Encoder-Decoder de Dai et al. (2019), empleada como referencia representativa de redes profundas para localización cefalométrica.

Como ejemplo representativo, el trabajo de Dai et al. (2019) sobre redes adversariales encoder-decoder para localización cefalométrica (Fig. 6.13) no reporta explícitamente el número total de parámetros entrenables. No obstante, la Tabla 6.9 resume los detalles estructurales de su arquitectura, evidenciando un modelo de gran escala compuesto por: (i) siete capas convolucionales en el encoder (C1–C7) con hasta 512 filtros, (ii) cuatro capas totalmente conectadas de alta dimensionalidad (12288–400–12288–12288 neuronas), y (iii) siete capas de up-convolución en el decoder (U1–U7), además de un discriminador profundo con seis capas convolucionales (C8–C13). La presencia de capas totalmente conectadas de gran tamaño, junto con un encoder y decoder profundos y un discriminador adicional, sitúa esta arquitectura en el orden de decenas de millones de parámetros entrenables, comparable a otras redes modernas de tipo U-Net con atención o esquemas multiescala empleadas en el mismo benchmark ISBI 2015.

En contraste, el modelo propuesto alcanza un $\text{SDR} < 2$ mm de 34.70 % y un MRE de 3.09 mm. Aunque estos valores son inferiores a los obtenidos por los métodos de referencia, este comportamiento responde directamente al diseño deliberadamente compacto de la arquitectura, la cual prioriza la eficiencia computacional, la portabilidad y la viabilidad de implementación en hardware embebido de bajo consumo.

Una tendencia similar se observa en los resultados del conjunto Test2 (Tabla 6.11), donde los métodos del estado del arte mantienen $\text{SDR} < 2$ mm superiores al 70 % y MRE por debajo de 1.8 mm, mientras que el modelo propuesto presenta un $\text{SDR} < 2$ mm de 31.32 % y un MRE de 3.81 mm. La reducción en desempeño entre Test1 y Test2 refleja la mayor complejidad de este subconjunto y

confirma la sensibilidad del modelo a variaciones anatómicas y de adquisición, efecto que también es reportado por la literatura en arquitecturas más profundas.

El análisis por punto cefalométrico (mostrado en las Figuras de Test1 y Test2) evidencia que los errores más elevados se concentran en puntos con mayor variabilidad anatómica y menor contraste radiográfico, tales como Orbitale, Gonion, Pogonion y Porion. Estos puntos requieren modelar relaciones espaciales complejas, las cuales son abordadas con mayor eficacia por redes profundas con módulos de atención y multiresolución. Sin embargo, el modelo propuesto mantiene una tendencia de error consistente y estable, sin degradaciones abruptas, lo que indica un comportamiento robusto y predecible bajo restricciones de cómputo.

Desde el punto de vista computacional, la arquitectura compacta propuesta permite su ejecución directa sobre la tarjeta NVIDIA Jetson Nano, obteniéndose un tiempo promedio de inferencia de 1.23 segundos por imagen. Este tiempo es significativamente menor al requerido por arquitecturas profundas que dependen de GPU de alto rendimiento, y permite la operación del sistema en entornos autónomos, portátiles y de bajo consumo energético, abriendo la posibilidad de aplicaciones clínicas móviles, unidades de diagnóstico en campo y sistemas de apoyo en regiones con infraestructura limitada.

En consecuencia, el modelo propuesto no busca competir directamente en precisión absoluta con los enfoques de alto costo computacional. Por ejemplo, el modelo adversarial Encoder–Decoder de Dai *et al.* (2019), cuya estructura detallada se resume en la Tabla 6.9, contiene capas totalmente conectadas de gran dimensión (12288–400–12288–12288 neuronas) y bloques convolucionales profundos con hasta 512 filtros, lo que sitúa su red en el orden de decenas de millones de parámetros entrenables. De manera similar, arquitecturas modernas basadas en U-Net con atención o enfoques multiescala reportados en el mismo benchmark ISBI 2015 típicamente oscilan entre 30 y 100 millones de parámetros, mientras que modelos híbridos CNN–Transformer o variantes de HRNet pueden incluso superar los 100 millones de parámetros; sino ofrecer una alternativa equilibrada entre exactitud, portabilidad y eficiencia energética. Este enfoque resulta particularmente relevante en escenarios reales donde el hardware disponible, el consumo de energía y la necesidad de operación continua son factores críticos, posicionando al modelo como una solución práctica y funcional para sistemas de detección cefalométrica en plataformas embebidas.

6.6.1. Desarrollo de la Interfaz Gráfica de Usuario para el etiquetado cefalométrico

El diseño y desarrollo de una Interfaz Gráfica de Usuario (GUI, por sus siglas en inglés) especializada constituye un componente central en la implementación de herramientas computacionales orientadas al análisis cefalométrico asistido por computadora. En el contexto de aplicaciones médicas, una GUI debe garantizar no solo la correcta visualización de las imágenes radiográficas, sino también

Tabla 6.9: *Detalles de las capas de la arquitectura adversarial Encoder–Decoder propuesta.*

Capa	Filtros	Kernel	Stride	Padding
Generador – Encoder (Convoluciones C1–C7)				
C1	32	7x7	1	same
C2	64	5x5	1	same
C3	128	5x5	1	same
C4	128	3x3	1	same
C5	256	3x3	1	same
C6	256	3x3	1	same
C7	512	3x3	1	same
Capas totalmente conectadas (FC)				
FC1	12288	–	–	–
FC2	400	–	–	–
FC3	12288	–	–	–
FC4	12288	–	–	–
Generador – Decoder (Up-convoluciones U1–U7)				
U1	256	4x4	2	same
U2	256	4x4	2	same
U3	128	4x4	2	same
U4	128	4x4	2	same
U5	64	4x4	2	same
U6	32	4x4	2	same
U7	1	4x4	1	same
Discriminador (Convoluciones C8–C13)				
C8	32	7x7	1	same
C9	64	5x5	1	same
C10	128	5x5	1	same
C11	128	3x3	1	same
C12	256	3x3	1	same
C13	512	3x3	1	same

Tabla 6.10: Comparación del modelo propuesto con métodos del estado del arte en el conjunto IEEE ISBI 2015 (Test1)

Autor (Año)	Método	Rx & Puntos	SDR<2	SDR<2.5	SDR<3	SDR<4	MRE
Arik <i>et al.</i> [33] (2017)	Deep CNNs	400 & 19	75.37	80.91	84.32	88.25	–
Qian <i>et al.</i> [83] (2019)	Faster R-CNN mejorado	400 & 19	82.50	86.20	89.30	90.60	–
Song <i>et al.</i> [34] (2020)	ResNet50 + registro	400 & 19	86.40	91.70	94.80	97.80	–
Qian <i>et al.</i> [186] (2020)	Atención multi-cabeza	400 & 19	87.61	93.16	96.35	98.74	1.15
Lee <i>et al.</i> [80] (2020)	CNN bayesiana	400 & 19	82.11	88.63	92.28	95.96	1.53
Zeng <i>et al.</i> [90] (2021)	CNN en cascada	400 & 19	81.37	89.09	93.79	97.86	1.34
He <i>et al.</i> [91] (2021)	Invariancia traslacional 2 etapas	400 & 19	87.47	92.89	95.15	98.25	1.09
Lu <i>et al.</i> [99] (2022)	Multiescala + GCN	400 & 19	87.93	92.42	95.68	98.32	1.06
Lee <i>et al.</i> [102] (2022)	CNN single-pass	400 & 19	86.42	92.00	95.68	98.46	–
Zhao <i>et al.</i> [107] (2023)	YOLOv3 multiscala	400 & 19	80.84	93.75	98.14	–	–
Neejara <i>et al.</i> [115] (2023)	CephXNet	400 & 19	88.06	94.64	96.87	98.68	0.92
Chen <i>et al.</i> [119] (2023)	U-Net + MC Dropout	400 & 19	79.65	–	–	97.22	1.39
Khalid <i>et al.</i> [125] (2024)	Two-stage + MH refine	400 & 19	87.17	90.77	95.18	98.12	1.23
Wang <i>et al.</i> [128] (2024)	STL + Heatmap decoder	400 & 19	88.21	92.42	95.19	97.75	1.06
Modelo Propuesto (2025)	CNN propuesta	400 & 19	34.70	46.32	57.44	73.37	3.09

Tabla 6.11: Comparación del modelo propuesto con métodos del estado del arte en el conjunto IEEE ISBI 2015 (Test2)

Autor (Año)	Método	Rx & Puntos	SDR<2	SDR<2.5	SDR<3	SDR<4	MRE
Arik <i>et al.</i> [33] (2017)	Deep CNNs	400 & 19	67.68	74.16	79.11	84.63	–
Qian <i>et al.</i> [83] (2019)	Faster R-CNN mejorado	400 & 19	72.40	76.15	79.65	85.90	–
Song <i>et al.</i> [34] (2020)	ResNet50 + registro	400 & 19	74.00	81.30	87.50	94.30	–
Qian <i>et al.</i> [186] (2020)	Atención multi-cabeza	400 & 19	76.32	82.95	87.95	94.63	1.43
Lee <i>et al.</i> [80] (2020)	CNN bayesiana	400 & 19	–	–	–	–	1.43
Zeng <i>et al.</i> [90] (2021)	CNN en cascada	400 & 19	70.58	79.53	86.05	93.32	1.64
He <i>et al.</i> [91] (2021)	Invariancia traslacional 2 etapas	400 & 19	75.47	83.44	87.84	94.63	1.43
Lu <i>et al.</i> [99] (2022)	Multiescala + GCN	400 & 19	76.11	83.68	88.68	95.21	1.37
Lee <i>et al.</i> [102] (2022)	CNN single-pass	400 & 19	74.58	81.79	87.53	94.26	–
Zhao <i>et al.</i> [107] (2023)	YOLOv3 multiscala	400 & 19	85.75	92.87	96.66	–	–
Neejara <i>et al.</i> [115] (2023)	CephXNet	400 & 19	78.72	83.54	91.26	96.36	1.41
Chen <i>et al.</i> [119] (2023)	U-Net + MC Dropout	400 & 19	80.05	–	–	97.53	1.33
Khalid <i>et al.</i> [125] (2024)	Two-stage + MH refine	400 & 19	75.79	84.00	89.00	94.53	1.78
Wang <i>et al.</i> [128] (2024)	STL + Heatmap decoder	400 & 19	76.26	83.26	88.63	94.74	1.33
Dai <i>et al.</i> [187] (2019)	GAN Encoder-Decoder (mapas de distancia)	300 & 19	<40	<50	<60	80	–
Modelo Propuesto (2025)	CNN propuesta	400 & 19	31.32	41.74	52.95	69.74	3.81

proporcionar mecanismos de interacción precisos, intuitivos y robustos que permitan a los profesionales de la salud realizar tareas de anotación con un alto grado de confiabilidad y reproducibilidad. Bajo esta premisa, Calderón-Sastre *et al.* desarrollaron una GUI específica para apoyar a dentistas y ortodoncistas en el proceso de etiquetado semiautomático de Radiografías Laterales de Cráneo (LCR) aplicadas al análisis cefalométrico [188].

La arquitectura funcional de la GUI propuesta se basa en un flujo de trabajo claramente estructurado, el cual se ilustra en la Fig. 6.14. Este diagrama esquematiza las etapas fundamentales del proceso de etiquetado, que incluyen la adquisición y carga de la imagen radiográfica, la aplicación de técnicas

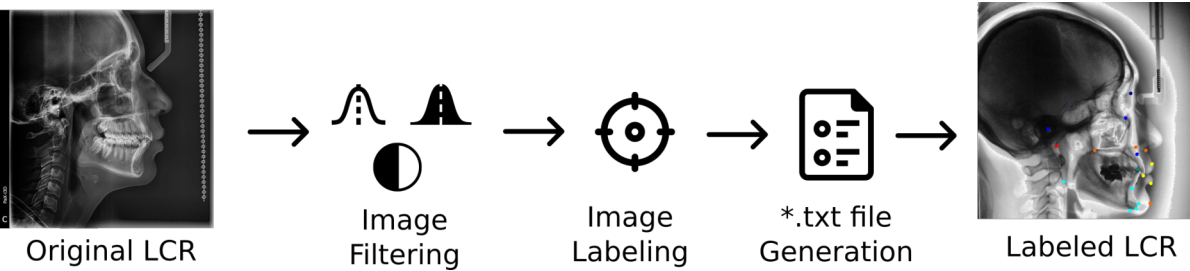


Figura 6.14: Tool Action flow.

de preprocesamiento digital, la selección manual de puntos cefalométricos y el almacenamiento estructurado de las coordenadas resultantes. Este enfoque modular garantiza que cada fase del procedimiento pueda ejecutarse de manera sistemática, reduciendo la variabilidad interobservador y mejorando la trazabilidad del proceso de anotación.

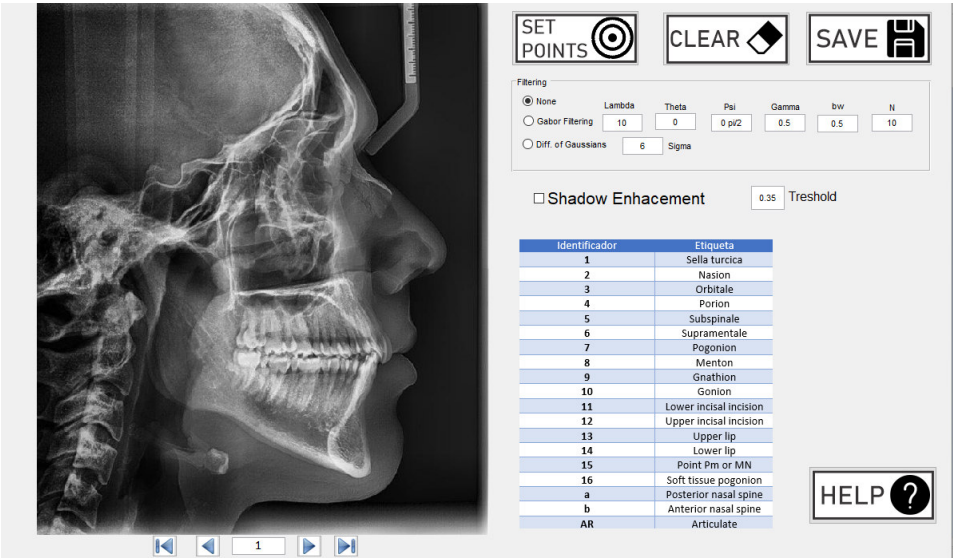


Figura 6.15: Medical Graphical User Interface for cephalometric labeling.

Desde el punto de vista de interacción humano-computadora, la GUI fue diseñada considerando principios de usabilidad y compatibilidad con el flujo de trabajo clínico estándar. Como se muestra en la Figura 6.15, la interfaz presenta un entorno visual organizado que integra de manera coherente la visualización de la radiografía, los controles de navegación entre imágenes, y los botones funcionales necesarios para iniciar el etiquetado (Set Points), corregir errores (Clear), guardar resultados (Save) y acceder a información de apoyo (Help). Esta disposición facilita que el especialista pueda concentrarse en la interpretación anatómica sin verse obstaculizado por complejidades técnicas del software.

Un aspecto distintivo de la GUI desarrollada es la integración directa de herramientas de preprocesamiento de imágenes dentro del mismo entorno de etiquetado. Estas herramientas incluyen

técnicas de mejora de contraste, filtrado por Diferencia de Gaussianas (DoG) y bancos de filtros de Gabor, las cuales permiten resaltar estructuras óseas y de tejido blando relevantes para la identificación de puntos cefalométricos. La posibilidad de aplicar estos métodos en tiempo real sobre la imagen de trabajo contribuye a mejorar la precisión del etiquetado y reduce la dependencia de la agudeza visual del experto, especialmente en regiones anatómicas con bordes poco definidos.

Para la localización de los puntos cefalométricos, la GUI emplea la función `ginput(z)` de Matlab™, que permite registrar coordenadas (x, y) mediante selección directa sobre la imagen con el cursor. Este procedimiento se repite hasta completar los 19 puntos de referencia definidos para el análisis cefalométrico. Una vez finalizado el proceso, las coordenadas son exportadas automáticamente a un archivo de texto estructurado, facilitando su posterior utilización en estudios de aprendizaje automático o análisis estadístico.

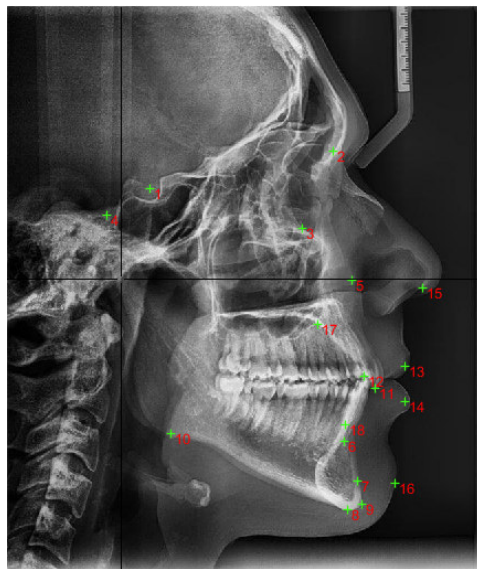


Figura 6.16: *CLR labeled using the proposed tool.*

La Figura 6.16 presenta un ejemplo representativo del resultado obtenido con la GUI, donde se observa una radiografía lateral de cráneo con los 19 puntos cefalométricos debidamente etiquetados. Este tipo de salida no solo valida la funcionalidad del sistema, sino que también demuestra su potencial para la generación sistemática de bases de datos anotadas de gran escala, esenciales para el desarrollo y evaluación de algoritmos automatizados de detección de landmarks cefalométricos.

En conjunto, el desarrollo de esta GUI representa una contribución significativa a la integración entre la práctica clínica y las herramientas computacionales avanzadas, proporcionando una plataforma metodológicamente sólida para el etiquetado preciso de imágenes cefalométricas y sentando las bases para la automatización progresiva de tareas de diagnóstico en odontología y ortodoncia [188].

6.7. Conclusiones Generales

El presente trabajo abordó el desarrollo, evaluación y análisis de un sistema automatizado para la detección de puntos cefalométricos basado en redes neuronales convolucionales profundas, orientado a aplicaciones de apoyo al diagnóstico en ortodoncia. A lo largo de la investigación se diseñó e implementó una arquitectura de aprendizaje profundo capaz de procesar radiografías cefalométricas laterales y predecir la localización espacial de 19 puntos anatómicos relevantes, evaluando su desempeño mediante métricas estandarizadas como el error radial medio (MRE) y la tasa de detección exitosa (SDR).

Los resultados experimentales obtenidos en los conjuntos de prueba Test1 y Test2 evidenciaron que el modelo propuesto es capaz de realizar una detección consistente de estructuras óseas mandibulares, alcanzando mayores tasas de detección en puntos como Gnathion, Menton y Pogonion, mientras que las estructuras de tejido blando y regiones de bajo contraste radiográfico continúan representando un reto significativo. Asimismo, se observó una disminución generalizada del desempeño en el conjunto Test2, lo cual pone de manifiesto la sensibilidad del sistema ante variaciones anatómicas, ruido y heterogeneidad en la adquisición de las imágenes.

Desde una perspectiva tecnológica y social, este trabajo demuestra la viabilidad de implementar sistemas de detección cefalométrica automatizada de bajo costo y bajo consumo energético, capaces de integrarse directamente a equipos de radiografía digital mediante una cámara de rayos X. Esta característica resulta particularmente relevante para clínicas odontológicas de recursos limitados, centros de salud comunitarios y unidades móviles, donde el acceso a software comercial de alto costo es restringido. La disponibilidad de sistemas computacionalmente eficientes permite no solo reducir el tiempo de análisis clínico, sino también disminuir la dependencia de personal altamente especializado para la obtención de mediciones cefalométricas preliminares.

No obstante, los resultados obtenidos también ponen en evidencia una limitación crítica: la necesidad de contar con bases de datos más extensas, balanceadas y representativas de la población mexicana. La mayoría de los conjuntos de datos utilizados en la literatura están conformados por radiografías de poblaciones asiáticas o europeas, cuyas características craneofaciales difieren significativamente de las de la población mexicana. Esta falta de representatividad limita la capacidad de generalización de los modelos y afecta directamente la confianza clínica en los diagnósticos generados de manera automática.

Actualmente, muchos odontólogos aún no pueden confiar plenamente en los sistemas automáticos de diagnóstico cefalométrico debido a variaciones morfológicas no modeladas adecuadamente, errores localizados en puntos críticos y a la ausencia de bases de datos locales con etiquetado experto suficiente.

Por lo tanto, se identifica como una línea prioritaria de trabajo futuro la creación de una base de datos nacional de radiografías cefalométricas mexicanas, con etiquetado clínico estandarizado, que permita entrenar y validar modelos específicamente adaptados a la morfología craneofacial de esta población.

En conclusión, este trabajo establece un primer paso hacia el desarrollo de herramientas inteligentes accesibles, eficientes y socialmente relevantes para el apoyo al diagnóstico ortodóntico en México. La consolidación de bases de datos representativas, junto con la incorporación de arquitecturas multiescala, mecanismos de atención y modelos bayesianos, permitirá avanzar hacia sistemas automatizados con mayor precisión, confiabilidad clínica y aceptación por parte de los profesionales de la salud.

6.8. Trabajo a Futuro

Una de las principales líneas de trabajo futuro derivadas de esta investigación es la creación de una base de datos pública de radiografías cefalométricas laterales con un número considerable de imágenes etiquetadas por especialistas en ortodoncia. Esta base de datos tendrá como objetivo constituirse como un referente nacional para el entrenamiento, evaluación y validación de sistemas computacionales e inteligencia artificial orientados al análisis cefalométrico automatizado.

Actualmente, los sistemas de detección cefalométrica basados en aprendizaje profundo aún no alcanzan el nivel de confiabilidad clínica requerido para sustituir o igualar los diagnósticos realizados por especialistas humanos, particularmente en poblaciones con características craneofaciales específicas, como la población mexicana. La carencia de conjuntos de datos amplios, balanceados y representativos limita la capacidad de generalización de los modelos y afecta directamente la confianza clínica en los diagnósticos generados de manera automática.

La disponibilidad de una base de datos pública con etiquetado clínico estandarizado permitirá no solo mejorar el desempeño de los modelos existentes, sino también establecer un marco de comparación objetivo entre diferentes arquitecturas y metodologías. Asimismo, esta iniciativa contribuirá a evaluar de manera cuantitativa el grado de confiabilidad de los sistemas automáticos frente al criterio de los especialistas, sentando las bases para el desarrollo de herramientas de apoyo diagnóstico seguras, confiables y adaptadas a la morfología craneofacial de la población mexicana.

Apéndice: Productos Académicos

6.9. Introducción

El presente apéndice documenta y clasifica los productos académicos generados durante el doctorado, organizándolos en líneas de investigación coherentes con la trayectoria científica desarrollada. Esta categorización permite visualizar la evolución temática, la consolidación de áreas de especialización y el impacto de las contribuciones realizadas en distintos dominios de la ingeniería, la inteligencia artificial y las tecnologías emergentes.

Los trabajos aquí incluidos abarcan investigación en: (i) inteligencia artificial aplicada a imágenes médicas, (ii) procesamiento de señales biomédicas y reconocimiento de patrones, (iii) sistemas AIoT, robótica y visión artificial, (iv) baterías y electromovilidad con técnicas de inteligencia artificial, y (v) modelado inteligente de sistemas energéticos y materiales semiconductores.

6.10. Criterios de Categorización

La clasificación se realizó con base en los siguientes criterios:

- **Dominio de aplicación:** medicina, energía, electromovilidad, robótica, interfaces humano-máquina, etc.
- **Metodología principal:** aprendizaje profundo, visión por computadora, procesamiento de señales, modelado físico, metaheurísticas, etc.
- **Tipo de contribución:** desarrollo de algoritmos, herramientas computacionales, modelos teóricos, o prototipos experimentales.
- **Impacto científico y tecnológico:** aporte al conocimiento, desarrollo de sistemas aplicados o innovación tecnológica.

Con base en estos criterios, los artículos se agrupan en cinco categorías principales.

6.11. Categoría 1: Inteligencia Artificial y Visión por Computadora en Medicina

6.11.1. Descripción de la línea

Esta línea de investigación se centra en el desarrollo de sistemas de Diagnóstico Asistido por Computadora (CAD) basados en técnicas de aprendizaje profundo y visión por computadora para el análisis automático de imágenes médicas. Incluye trabajos sobre localización de puntos cefalométricos, herramientas de etiquetado semiautomático y segmentación de tumores cerebrales mediante arquitecturas avanzadas de redes neuronales.

6.11.2. Productos académicos asociados

- **Automatic cephalometric landmark identification in lateral skull X-Rays using convolutional neural networks** (2021). Artículo científico – Revista de Aplicaciones de la Ingeniería.
- **Preprocessing and Labeling Tool for Lateral Skull X-Ray Images Applied to Cephalometric Analysis** (2020). Congreso IEEE ROPEC.
- **Cephalometric References' Localization in Lateral Skull Radiographs: A Systematic Review** (2020). Technical Report – Research in Computing Science.
- **Multi-Scale ConvNeXt for Robust Brain Tumor Segmentation in Multimodal MRI** (2026). Artículo científico – Technologies.

6.11.3. Aportaciones clave

- Desarrollo de modelos automáticos para reducir la carga de trabajo de especialistas clínicos.
- Implementación de arquitecturas modernas de deep learning para análisis de imágenes médicas.
- Creación de herramientas computacionales para apoyo en análisis cefalométrico.
- Segmentación precisa de tumores cerebrales con modelos multi-escala.

6.12. Categoría 2: Señales Biomédicas y Reconocimiento de Patrones

6.12.1. Descripción de la línea

Esta línea aborda el procesamiento de señales electromiográficas (sEMG) y datos inerciales para el reconocimiento de patrones mediante redes neuronales profundas. Se exploraron aplicaciones en escritura manuscrita asistida por tecnología y reconocimiento de trayectorias de movimiento humano.

6.12.2. Productos académicos asociados

- **Multi-Stroke handwriting character recognition based on sEMG using convolutional-recurrent neural networks** (2020). Artículo científico – Mathematical Biosciences and Engineering.
- **Reconocimiento de caracteres manuscritos a partir de señales EMG usando redes neuronales convolucionales y recurrentes de Elman** (2020). Technical Report – Research in Computing Science.
- **Reconocimiento de trayectorias del brazo usando sensores de aceleración y giroscopio** (2020). Technical Report – Research in Computing Science.

6.12.3. Aportaciones clave

- Integración de sensores wearables con modelos de aprendizaje profundo.
- Desarrollo de sistemas de reconocimiento de patrones en tiempo real.
- Aplicaciones potenciales en interfaces humano-máquina y rehabilitación.

6.13. Categoría 3: Inteligencia Artificial de las Cosas (AIoT), Robótica y Visión Artificial

6.13.1. Descripción de la línea

Esta línea integra inteligencia artificial con sensores, dispositivos embebidos y visión por computadora para desarrollar sistemas inteligentes con impacto social y tecnológico.

6.13.2. Productos académicos asociados

- **Detecting objects using color and depth segmentation with Kinect sensor** (2012). Procedia Technology.
- **Prototype of an AIoT System for Braille Literacy with Deep Learning (DevrAille)** (2024). IEEE ROPEC.
- **REAXION Ciencia y Tecnología Universitaria XXXIV: Prototipo de un sistema AIoT para lectura braille** (2024). Artículo en revista universitaria.
- **Standing up assist prototype controlled by voice commands** (2025). Artículo científico – Journal of Scientific and Technical Applications.

6.13.3. Aportaciones clave

- Desarrollo de prototipos funcionales con integración de IA y hardware.
- Aplicaciones enfocadas en inclusión y accesibilidad tecnológica.

- Implementación de sistemas en tiempo real con visión artificial y control por voz.

6.14. Categoría 4: Baterías, Electromovilidad y Sistemas Energéticos Inteligentes

6.14.1. Descripción de la línea

Esta línea aborda el diseño, modelado y optimización de sistemas de baterías de litio para aplicaciones en electromovilidad, incluyendo el desarrollo de sistemas de gestión de baterías (BMS) y modelos basados en aprendizaje profundo.

6.14.2. Productos académicos asociados

- **Design of a Battery Management System (BMS) Applied to Electromobility** (2025). Artículo científico – Journal of Scientific and Technical Applications.
- **Lithium Battery Enhancement Through Electrical Characterization and Optimization Using Deep Learning** (2025). Artículo científico – World Electric Vehicle Journal.
- **Optimization of Lithium-Ion Batteries Using Boltzmann Metaheuristics Systems** (2024). Capítulo de libro – Springer.

6.14.3. Aportaciones clave

- Desarrollo de sistemas de monitoreo en tiempo real para baterías.
- Aplicación de modelos híbridos basados en física e inteligencia artificial.
- Contribuciones a la seguridad y eficiencia en sistemas de electromovilidad.

6.15. Categoría 5: Energía Renovable, Semiconductores e Inteligencia Artificial

6.15.1. Descripción de la línea

Esta línea combina modelado físico de materiales semiconductores con aprendizaje profundo para el diseño y simulación de sistemas de generación termoeléctrica y almacenamiento energético.

6.15.2. Productos académicos asociados

- **Semiconductor-Deep Learning for the Design and Simulation of Solar Energy Reservoir-Powered Thermogenerators** (2025). Capítulo de libro – Springer.

- **Intelligent Thermoelectric Power Generation System based on Semiconductors and Deep Learning** (2025). Artículo científico – Journal of Technological Prototypes.

6.15.3. Aportaciones clave

- Integración de modelos físicos con redes neuronales para optimización energética.
- Simulación avanzada de sistemas termoeléctricos.
- Contribuciones al desarrollo de energías limpias con IA.

Bibliografía

- [1] United Nations Department of Economic and Social Affairs., *World population prospects 2019*. No. 141 in 1, United Nations, 2019.
- [2] P. J. Denning and T. G. Lewis, “Exponential laws of computing growth,” *Communications of the ACM*, vol. 60, no. 1, pp. 54–65, 2017.
- [3] M. Körner, C. H. Weber, S. Wirth, K.-J. Pfeifer, M. F. Reiser, and M. Treitl, “Advances in Digital Radiography: Physical Principles and System Overview,” *RadioGraphics*, vol. 27, pp. 675–686, 5 2007.
- [4] H. Fujita, “AI-based computer-aided diagnosis (AI-CAD): the latest review to read first,” *Radiological Physics and Technology*, vol. 13, no. 1, pp. 6–19, 2020.
- [5] M. L. Giger, H. P. Chan, and J. Boone, “Anniversary paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM,” *Medical Physics*, vol. 35, no. 12, pp. 5799–5820, 2008.
- [6] K. Doi, “Computer-aided diagnosis in medical imaging: Historical review, current status and future potential,” *Computerized Medical Imaging and Graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.
- [7] H. P. Chan, K. Doi, S. Galhotra, C. J. Vyborny, H. Macmahon, and P. M. Jokich, “Image feature analysis and computer aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography,” *Medical Physics*, vol. 14, no. 4, pp. 538–548, 1987.
- [8] M. L. Giger, K. Doi, H. MacMahon, S. J. Dwyer III, and R. H. Schneider, “Computerized Detection Of Lung Nodules In Digital Chest Radiographs,” in *SPIE 0767, Medical Imaging*, vol. 767, pp. 384–387, 1 1987.
- [9] M. L. Giger, K. Doi, and H. Macmahon, “Image feature analysis and computer aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields,” 1988.
- [10] K. R. Hoffmann, K. Doi, H.-P. Chan, L. Fencil, H. Fujita, and A. Muraki, “Automated Tracking Of The Vascular Tree In DSA Images Using A Double-Square-Box Region-Of-Search Algorithm,” in

- Proceedings Volume 0626, Application of Optical Instrumentation in Medicine XIV and Picture Archiving and Communication Systems* (S. J. Dwyer III and R. H. Schneider, eds.), vol. 626, p. 326, 6 1986.
- [11] H. Fujita, K. Doi, L. E. Fencil, and K. G. Chua, "Image feature analysis and computer aided diagnosis in digital radiography. 2. Computerized determination of vessel sizes in digital subtraction angiography," *Medical Physics*, vol. 14, no. 4, pp. 549–556, 1987.
- [12] T. W. Freer and M. J. Ulissey, "Screening mammography with computer-aided detection: Prospective study of 12,860 patients in a community breast center," *Radiology*, vol. 220, no. 3, pp. 781–786, 2001.
- [13] K. Doi, "Current status and future potential of computer-aided diagnosis in medical imaging," *British Journal of Radiology*, vol. 78, pp. 21–30, 2005.
- [14] A. Athanasiou, *Orthodontic Cephalometry*. Mosby-Wolfe, 1995.
- [15] K. H. Yu, A. L. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," *Nature Biomedical Engineering*, vol. 2, no. 10, pp. 719–731, 2018.
- [16] A. S. Ahuja, "The impact of artificial intelligence in medicine on the future role of the physician," *PeerJ*, vol. 2019, no. 10, 2019.
- [17] "Computer aided diagnosis in publications 2010-2019." [Online]. Available from: https://app.dimensions.ai/analytics/publication/overview/timeline?search_mode=content&search_text=Computer%20Aided%20Diagnosis&search_type=kws&search_field=full_search&local:indicator-y1=timeline-source-published, 1 2026.
- [18] V. Krunić, *Succeeding with AI: How to make AI work for your business*. NY, USA: Manning Publications, 1 ed., 2020.
- [19] D. S. Gill and F. B. Naini, *Orthodontics : Principles and Practice*. West Sussex, UK: John Wiley & Sons, 1 ed., 2011.
- [20] K. Kula and A. Ghoneima, *Cephalometry in Orthodontics: 2D and 3D*. Batavia, IL: Quintessence Publishing Co, Inc, 2018.
- [21] W. B. Downs, "Variations in facial relationships: Their significance in treatment and prognosis," *American Journal of Orthodontics*, vol. 34, 10 1948.
- [22] C. C. Steiner, "Cephalometrics for you and me," *American Journal of Orthodontics*, vol. 39, no. 10, pp. 729–755, 1953.
- [23] T. Rakosi, *Cephalometric Radiography*. Wolfe Medical Publications Ltd., 1 ed., 1982.

- [24] C.-W. Wang, C.-t. Huang, M.-c. Hsieh, C.-H. Li, S.-w. Chang, W.-c. Li, R. Vandaele, R. Maree, S. Jodogne, P. Geurts, C. Chen, G. Zheng, C. Chu, H. Mirzaalian, G. Hamarneh, T. Vrtovec, and B. Ibragimov, "Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X-Ray Images: A Grand Challenge," *IEEE Transactions on Medical Imaging*, vol. 34, pp. 1890–1900, Sept. 2015.
- [25] T. Foster, *A textbook of orthodontics*, vol. 92. Cambridge, Massachusetts, USA: Blackwell Scientific Publications, 3 ed., 1990.
- [26] T. Rakosi, *An Atlas and Manual of Cephalometric Radiography*. London: Wolfe Medical Publications Ltd., 1 ed., 1979.
- [27] J. Cardillo, "An Image Processing System for Locating Craniofacial Landmarks," *IEEE Transactions on Medical Imaging*, vol. 13, no. 2, pp. 275–289, 1994.
- [28] R. Leonardi, D. Giordano, and F. Maiorana, "An evaluation of cellular neural networks for the automatic identification of cephalometric landmarks on digital images," *BioMed Research International*, vol. 2009, 1 2009.
- [29] S. Baumrind and R. C. Frantz, "The reliability of head film measurements," *American Journal of Orthodontics*, vol. 60, pp. 111–127, 8 1971.
- [30] E. M. Ongkosuwito, "The reproducibility of cephalometric measurements: a comparison of analogue and digital methods," *The European Journal of Orthodontics*, vol. 24, pp. 655–665, 12 2002.
- [31] S. AlBarakati, K. Kula, and A. Ghoneima, "The reliability and reproducibility of cephalometric measurements: a comparison of conventional and digital methods," *Dentomaxillofacial Radiology*, vol. 41, pp. 11–17, 1 2012.
- [32] K. Sayinsu, F. Isik, G. Trakyalı, and T. Arun, "An evaluation of the errors in cephalometric measurements on scanned cephalometric images and conventional tracings," *The European Journal of Orthodontics*, vol. 29, pp. 105–108, 2 2007.
- [33] S. Ö. Arik, B. Ibragimov, and L. Xing, "Fully automated quantitative cephalometry using convolutional neural networks," *Journal of Medical Imaging*, vol. 4, p. 014501, 1 2017.
- [34] Y. Song, X. Qiao, Y. Iwamoto, and Y. wei Chen, "Automatic cephalometric landmark detection on x-ray images using a deep-learning method," *Applied Sciences*, vol. 10, p. 2547, 4 2020.
- [35] M. OSTERGAARD, B. EBJERG, and M. SZKUDLAREK, "Imaging in early rheumatoid arthritis: roles of magnetic resonance imaging, ultrasonography, conventional radiography and

- computed tomography,” *Best Practice & Research Clinical Rheumatology*, vol. 19, pp. 91–116, 2 2005.
- [36] D. A. Tyndall, J. B. Price, S. Tetradis, S. D. Ganz, C. Hildebolt, and W. C. Scarfe, “Position statement of the american academy of oral and maxillofacial radiology on selection criteria for the use of radiology in dental implantology with emphasis on cone beam computed tomography,” *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 113, pp. 817–826, 6 2012.
- [37] J. W. Casselman, K. Gieraerts, D. Volders, J. Delanote, K. Mermuys, B. D. Foer, and G. Swennen, “Cone beam ct: non-dental applications,” *Journal of the Belgian Society of Radiology*, vol. 96, p. 333, 11 2013.
- [38] N. Shah, N. Bansal, and A. Logani, “Recent advances in imaging technologies in dentistry,” *World journal of radiology*, vol. 6, pp. 794–807, 10 2014.
- [39] S. Laughlin and W. Montanera, “Central nervous system imaging,” *Postgraduate Medicine*, vol. 104, pp. 73–88, 11 1998.
- [40] S. Mastrogiacomo, W. Dou, J. A. Jansen, and X. F. Walboomers, “Magnetic resonance imaging of hard tissues and hard tissue engineered bio-substitutes,” *Molecular Imaging and Biology*, vol. 21, pp. 1003–1019, 12 2019.
- [41] C. Maspero, A. Abate, F. Bellincioni, D. Cavagnetto, V. Lanteri, A. Costa, and M. Farronato, “Comparison of a tridimensional cephalometric analysis performed on 3t-mri compared with cbct: a pilot study in adults,” *Progress in Orthodontics*, vol. 20, p. 40, 12 2019.
- [42] S. Shukla, A. Chug, and K. I. Afrashtehfar, “Role of cone beam computed tomography in diagnosis and treatment planning in dentistry: An update.,” *Journal of International Society of Preventive & Community Dentistry*, vol. 7, pp. S125–S136, 11 2017.
- [43] R. Olszewski, G. Cosnard, B. Macq, P. Mahy, and H. Reyhler, “3d ct-based cephalometric analysis: 3d cephalometric theoretical concept and software,” *Neuroradiology*, vol. 48, pp. 853–862, 11 2006.
- [44] N. Ogawa, Y. Miyazaki, M. Kubota, J. C. Huang, A. J. Miller, and K. Maki, “Application of cone beam ct 3d images to cephalometric analysis,” *Orthodontic Waves*, vol. 69, pp. 138–150, 12 2010.
- [45] A. Juerchott, C. Freudlsperger, S. Zingler, M. A. Saleem, J. M. E. Jende, C. J. Lux, M. Bendszus, S. Heiland, and T. Hilgenfeld, “In vivo reliability of 3d cephalometric landmark determination on magnetic resonance imaging: a feasibility study,” *Clinical Oral Investigations*, vol. 24, pp. 1339–1349, 3 2020.

- [46] R. Neeraja and L. J. Anbarasi, "A critical review of artificial intelligence based techniques for automatic prediction of cephalometric landmarks," *Artificial Intelligence Review*, vol. 58, 5 2025.
- [47] J. K. Liu, Y. T. Chen, and K. S. Cheng, "Accuracy of computerized automatic identification of cephalometric landmarks," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 118, no. 5, pp. 535–540, 2000.
- [48] V. Grau, M. Alcañiz, M. C. Juan, C. Monserrat, and C. Knoll, "Automatic localization of cephalometric landmarks," *Journal of Biomedical Informatics*, vol. 34, no. 3, pp. 146–156, 2001.
- [49] V. Ciesielski and A. Innes, "Genetic programming for landmark detection in cephalometric radiology images," *International Journal of Knowledge-based and Intelligent Engineering Systems*, vol. 7, pp. 164–171, 2003.
- [50] S. Rueda and M. Alcañiz, "An approach for the automatic cephalometric landmark detection using mathematical morphology and active appearance models," *Med Image Comput Comput Assist Interv.*, pp. 159–166, 2006.
- [51] H. Mohseni and S. Kasaei, "Automatic localization of cephalometric landmarks," in *2007 IEEE International Symposium on Signal Processing and Information Technology*, pp. 396–401, IEEE, 12 2007.
- [52] R. Kafieh, A. Mehri, and S. Sadri, "Automatic landmark detection in cephalometry using a modified active shape model with sub image matching," in *2007 International Conference on Machine Vision*, pp. 73–78, IEEE, 12 2007.
- [53] M. A. Mosleh, M. S. Baba, N. Himazian, and B. M. AL-Makramani, "An image processing system for cephalometric analysis and measurements," in *2008 International Symposium on Information Technology*, pp. 1–8, IEEE, 8 2008.
- [54] A. A. Pouyan and M. Farshbaf, "Cephalometric landmarks localization based on histograms of oriented gradients," in *2010 International Conference on Signal and Image Processing*, pp. 1–6, IEEE, 12 2010.
- [55] P. Vuinić, Z. Trpovski, and I. Čepan, "Automatic landmarking of cephalograms using active appearance models," *European Journal of Orthodontics*, vol. 32, no. 3, pp. 233–241, 2010.
- [56] S. Shahidi, M. Oshagh, F. Gozin, P. Salehi, and S. M. Danaei, "Accuracy of computerized automatic identification of cephalometric landmarks by a designed software," *Dentomaxillofacial Radiology*, vol. 42, 1 2013.

- [57] M. Majd and F. Shoeleh, "A novel hybrid approach for cephalometric landmark detection," tech. rep., Shiraz University, 6 2015.
- [58] J. Montúfar, M. Romero, and R. J. Scougall-Vilchis, "Automatic 3-dimensional cephalometric landmarking based on active shape models in related projections," *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 153, pp. 449–458, 3 2018.
- [59] A. Kaur and C. Singh, "Automatic cephalometric landmark detection using Zernike moments and template matching," *Signal, Image and Video Processing*, vol. 9, no. 1, pp. 117–132, 2015.
- [60] I. P. Sari, R. Widayati, M. Priaminiarti, D. Danudirdjo, and T. L. Mengko, "Initial estimation of landmark location for automated cephalometric analysis using template matching method," in *2015 4th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*, pp. 159–162, IEEE, 11 2015.
- [61] L. Favaedi and M. Petrou, "Cephalometric landmarks identification using probabilistic relaxation," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 4391–4394, IEEE, 8 2010.
- [62] S. Vasamsetti, V. Sardana, P. Kumar, and H. Sardana, "Feature extraction through wavelet decomposition for automatic detection of landmarks on cephalogram images," 4 2017.
- [63] B. C. Neelapu, O. P. Kharbanda, V. Sardana, A. Gupta, S. Vasamsetti, R. Balachandran, and H. K. Sardana, "Automatic localization of three-dimensional cephalometric landmarks on cbct images by extracting symmetry features of the skull," *Dento maxillo facial radiology*, vol. 47, p. 20170054, 2 2018.
- [64] P. Vucinic, Z. Trpovski, and I. Scepan, "Automatic landmarking of cephalograms using active appearance models," *The European Journal of Orthodontics*, vol. 32, pp. 233–241, 6 2010.
- [65] D. P. Jetwani, S. Kumar, and H. Sardana, "Cephalometric landmark identification using fuzzy wavelet edge detector," in *2011 IEEE International Symposium on Medical Measurements and Applications*, pp. 349–353, 2011.
- [66] S. Sanei, P. Sanaei, and M. Zahabsaniei, "Cephalogram analysis applying template matching and fuzzy logic," *Image and Vision Computing*, vol. 18, no. 1, pp. 39–48, 1999.
- [67] V. Rakhshan, H. Rakhshan, and A. Sheibaninia, "Developing an automatic lateral cephalometric landmark identification program and evaluating its performance.," *International journal of computerized dentistry*, vol. 12, pp. 327–43, 2009.

- [68] B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, “Automatic cephalometric X-ray landmark detection by applying game theory and random forests,” *ISBI Int. Symp. Biomed. Imag. 2014, Automat. Cephalometric X-Ray Landmark Detection Challenge*, pp. 1–8, 2014.
- [69] C. Chu, C. Chen, C. Wang, C. Huang, C. Li, L. Nolte, and G. Zheng, “Fully Automatic Cephalometric X-Ray Landmark Detection Using Random Forest Regression and Sparse Shape Composition,” *Proc. ISBI Int. Symp. Biomed. Imag. 2014, Automat. Cephalometric X-Ray Landmark Detection Challenge*, pp. 9–16, 2014.
- [70] R. Vandaele, R. Marée, S. Jodogne, and P. Geurts, “Automatic cephalometric x-ray landmark detection challenge 2014: A tree-based approach,” *ISBI Int. Symp. Biomed. Imag. 2014, Automat. Cephalometric X-Ray Landmark Detection Challenge*, pp. 37–44, 2014.
- [71] H. Mirzaalian and G. Hamarneh, “Automatic Globally-Optimal Pictorial Structures with Random Decision Forest Based Likelihoods For Cephalometric X-Ray Landmark Detection,” *Automatic Cephalometric X-Ray Landmark Detection Challenge 2014*, 2014.
- [72] C. Lindner and T. F. Cootes, “Fully automatic cephalometric evaluation using random forest regression-voting,” *Isbi*, pp. 1–8, 2015.
- [73] C. Lindner, C.-w. Wang, C.-t. Huang, C.-h. Li, S.-W. Chang, and T. F. Cootes, “Fully Automatic System for Accurate Localisation and Analysis of Cephalometric Landmarks in Lateral Cephalograms,” *Scientific Reports*, vol. 6, p. 33581, 9 2016.
- [74] H.-W. Hwang, J.-H. Park, J.-H. Moon, Y. Yu, H. Kim, S.-B. Her, G. Srinivasan, M. N. A. Aljanabi, R. E. Donatelli, and S.-J. Lee, “Automated identification of cephalometric landmarks: Part 2- might it be better than human?,” *The Angle Orthodontist*, vol. 90, pp. 69–76, 1 2020.
- [75] S. Chakrabartty, M. Yagi, T. Shibata, and G. Cauwenberghs, “Robust cephalometric landmark identification using support vector machines,” in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, vol. 2, pp. II–825, 2003.
- [76] B. Romaniuk, M. Desvignes, M. Revenu, M.-J. D. Linear, B. Romaniuk, M. Desvignes, M. Revenu, and M. Deshayes, “Linear and non-linear model for statistical localization of landmarks,” *Hal Open Science*, vol. 1, pp. 393–396, 1 2002.
- [77] C. Chu, C. Chen, C.-W. Wang, C.-T. Huang, C.-H. Li, L.-P. Nolte, and G. Zheng, “Fully automatic cephalometric x-ray landmark detection using random forest regression and sparse shape composition,” tech. rep., Institute for Surgical Technology and Biomechanics, University of Bern., 5 2015.

- [78] C. Chen, C. Wang, C. Huang, C. Li, and G. Zheng, “Fully-Automatic Landmark detection in Skull X-Ray images,” *Proc. ISBI Int. Symp. Biomed. Imag. 2014, Automat. Cephalometric X-Ray Landmark Detection Challenge*, pp. 1–8, 2014.
- [79] B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, “Computerized Cephalometry by Game Theory with Shape- and Appearance-Based Landmark Refinement,” *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI) 2015 – Grand Challenges in Dental X-ray Image*, pp. 1–8, 2015.
- [80] J.-H. Lee, H.-J. Yu, M. ji Kim, J.-W. Kim, and J. Choi, “Automated cephalometric landmark detection with confidence regions using bayesian convolutional neural networks,” *BMC Oral Health*, vol. 20, p. 270, 12 2020.
- [81] S. H. Kang, K. Jeon, H.-J. Kim, J. K. Seo, and S.-H. Lee, “Automatic three-dimensional cephalometric annotation system using three-dimensional convolutional neural networks: a developmental trial,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 8, no. 2, pp. 210–218, 2020.
- [82] A. Gupta, O. P. Kharbanda, V. Sardana, R. Balachandran, and H. K. Sardana, “A knowledge-based algorithm for automatic detection of cephalometric landmarks on cbct images,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, pp. 1737–1752, 11 2015.
- [83] J. Qian, M. Cheng, Y. Tao, J. Lin, and H. Lin, “Cephanet: An improved faster r-cnn for cephalometric landmark detection,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 868–871, IEEE, 4 2019.
- [84] S. Nishimoto, Y. Sotsuka, K. Kawai, H. Ishise, and M. Kakibuchi, “Personal computer-based cephalometric landmark detection with deep learning, using cephalograms on the internet,” *Journal of Craniofacial Surgery*, vol. 30, pp. 91–95, 1 2019.
- [85] F. Kunz, A. Stellzig-Eisenhauer, F. Zeman, and J. Boldt, “Artificial intelligence in orthodontics,” *Journal of Orofacial Orthopedics / Fortschritte der Kieferorthopädie*, vol. 81, pp. 52–68, 1 2020.
- [86] C. Lee, C. Tanikawa, J.-Y. Lim, and T. Yamashiro, “Deep learning based cephalometric landmark identification using landmark-dependent multi-scale patches,” *arXiv*, vol. 1, pp. 1–5, 6 2019.
- [87] Y. Song, X. Qiao, Y. Iwamoto, and Y.-w. Chen, “Automatic cephalometric landmark detection on x-ray images using a deep-learning method,” *Applied Sciences*, vol. 10, no. 7, 2020.
- [88] J. Qian, M. Cheng, Y. Tao, J. Lin, and H. Lin, “Cephanet: An improved faster r-cnn for cephalometric landmark detection,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 868–871, 2019.

- [89] M.-S. Kim, Y. Kim, H. Choi, W. Kim, S. Park, D. Lee, D. K. Kim, H. J. Kim, H. Choi, D.-W. Hyun, J.-Y. Lee, E. Y. Choi, D.-S. Lee, J.-W. Bae, and I. Mook-Jung, "Transfer of a healthy microbiota reduces amyloid and tau pathology in an alzheimer's disease animal model," *Gut*, vol. 69, pp. 283–294, 2 2020.
- [90] M. Zeng, Z. Yan, S. Liu, Y. Zhou, and L. Qiu, "Cascaded convolutional networks for automatic cephalometric landmark detection," *Medical Image Analysis*, vol. 68, p. 101904, 2 2021.
- [91] T. He, J. Guo, W. Tang, W. Zeng, P. He, F. Zeng, and Z. Yi, "Cascade-refine model for cephalometric landmark detection in high-resolution orthodontic images," *Knowledge-Based Systems*, vol. 265, p. 110332, 4 2023.
- [92] A. Yueyuan and W. Hong, "Swin transformer combined with convolutional encoder for cephalometric landmarks detection," in *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pp. 184–187, 2021.
- [93] N. R and L. J. Anbarasi, "A review on automatic cephalometric landmark identification using artificial intelligence techniques," in *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 572–577, 2021.
- [94] K. Oh, I.-S. Oh, V. N. T. Le, and D.-W. Lee, "Deep anatomical context feature learning for cephalometric landmark detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 806–817, 2021.
- [95] G. Bulatova, B. Kusnoto, V. Grace, T. P. Tsay, D. M. Avenetti, and F. J. C. Sanchez, "Assessment of automatic cephalometric landmark identification using artificial intelligence," *Orthodontics & Craniofacial Research*, vol. 24, pp. 37–42, 12 2021.
- [96] L. Wang, L. Ma, Y. Li, K. Niu, and Z. He, "A dcnn system based on an iterative method for automatic landmark detection in cephalometric x-ray images," *Biomedical Signal Processing and Control*, vol. 68, p. 102757, 7 2021.
- [97] P. K. Reddy, A. Kanakatte, J. Gubbi, M. Poduval, A. Ghose, and B. Purushothaman, "Anatomical landmark detection using deep appearance-context network," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3569–3572, IEEE, 11 2021.
- [98] F. Schwendicke, A. Chaurasia, L. Arsiwala, J.-H. Lee, K. Elhennawy, P.-G. Jost-Brinkmann, F. Demarco, and J. Krois, "Deep learning for cephalometric landmark detection: systematic review and meta-analysis," *Clinical Oral Investigations*, vol. 25, pp. 4299–4309, 7 2021.

- [99] G. Lu, Y. Zhang, Y. Kong, C. Zhang, J.-L. Coatrieux, and H. Shu, “Landmark localization for cephalometric analysis using multiscale image patch-based graph convolutional networks,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, pp. 3015–3024, 7 2022.
- [100] M. Ugurlu, “Performance of a convolutional neural network- based artificial intelligence algorithm for automatic cephalometric landmark detection,” *Turkish Journal of Orthodontics*, pp. 94–100, 6 2022.
- [101] M. Moreno and S. Gebeile-Chauty, “Étude comparative de deux logiciels de détection de points de repère céphalométriques par intelligence artificielle,” *L’Orthodontie Française*, vol. 93, pp. 41–61, 3 2022.
- [102] M. Lee, M. Chung, and Y.-G. Shin, “Cephalometric landmark detection via global and local encoders and patch-wise attentions,” *Neurocomputing*, vol. 470, pp. 182–189, 1 2022.
- [103] I. El-Feghi, M. Sid-Ahmed, and M. Ahmadi, “Automatic localization of craniofacial landmarks for assisted cephalometry,” *Pattern Recognition*, vol. 37, pp. 609–621, 3 2004.
- [104] J. Yao, W. Zeng, T. He, S. Zhou, Y. Zhang, J. Guo, and W. Tang, “Automatic localization of cephalometric landmarks based on convolutional neural network,” *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 161, pp. e250–e259, 3 2022.
- [105] A. Innes, V. Ciesielski, J. Mamutil, and S. John, “Innes,” in *Landmark Detection for Cephalometric Radiology Images using Pulse Coupled Neural Networks* (I. H. Arabnia and Y. Mun, eds.), pp. 511–517, 6 2002.
- [106] B. Ristau, M. Coreil, A. Chapple, P. Armbruster, and R. Ballard, “Comparison of audaxceph®’s fully automated cephalometric tracing technology to a semi-automated approach by human examiners,” *International Orthodontics*, vol. 20, p. 100691, 12 2022.
- [107] C. Zhao, Z. Yuan, S. Luo, W. Wang, Z. Ren, X. Yao, and T. Wu, “Automatic recognition of cephalometric landmarks via multi-scale sampling strategy,” *Heliyon*, vol. 9, p. e17459, 6 2023.
- [108] X. Dai, H. Zhao, T. Liu, D. Cao, and L. Xie, “Locating anatomical landmarks on 2d lateral cephalograms through adversarial encoder-decoder networks,” *IEEE Access*, vol. 7, pp. 132738–132747, 2019.
- [109] W. Hong, S.-M. Kim, J. Choi, J. Ahn, J.-Y. Paeng, and H. Kim, “Automated cephalometric landmark detection using deep reinforcement learning,” *Journal of Craniofacial Surgery*, vol. 34, pp. 2336–2342, 11 2023.

- [110] J. V. Weingart, S. Schlager, M. C. Metzger, L. S. Brandenburg, A. Hein, R. Schmelzeisen, F. Bamberg, S. Kim, E. Kellner, M. Reisert, and M. F. Russe, “Automated detection of cephalometric landmarks using deep neural patchworks,” *Dentomaxillofacial Radiology*, vol. 52, 9 2023.
- [111] H. Ye, Z. Cheng, N. Ungvijanpunya, W. Chen, L. Cao, and Y. Gou, “Is automatic cephalometric software using artificial intelligence better than orthodontist experts in landmark identification?,” *BMC Oral Health*, vol. 23, p. 467, 7 2023.
- [112] Q. Chang, Z. Wang, F. Wang, J. Dou, Y. Zhang, and Y. Bai, “Automatic analysis of lateral cephalograms based on high-resolution net,” *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 163, pp. 501–508.e4, 4 2023.
- [113] G. S. Duran, Şule Gökmen, K. G. Topsakal, and S. Görgülü, “Evaluation of the accuracy of fully automatic cephalometric analysis software with artificial intelligence algorithm,” *Orthodontics & Craniofacial Research*, vol. 26, pp. 481–490, 8 2023.
- [114] H. Bao, K. Zhang, C. Yu, H. Li, D. Cao, H. Shu, L. Liu, and B. Yan, “Evaluating the accuracy of automated cephalometric analysis based on artificial intelligence,” *BMC Oral Health*, vol. 23, p. 191, 4 2023.
- [115] R. Neeraja and L. J. Anbarasi, “Cephxnet: A deep convolutional squeeze-and-excitation model for landmark prediction on lateral cephalograms,” *IEEE Access*, vol. 11, pp. 90780–90800, 2023.
- [116] H. Lee, J. M. Cho, S. Ryu, S. Ryu, E. Chang, Y.-S. Jung, and J.-Y. Kim, “Automatic identification of posteroanterior cephalometric landmarks using a novel deep learning algorithm: a comparative study with human experts,” *Scientific Reports*, vol. 13, p. 15506, 9 2023.
- [117] Q. Zhu, Y. Bi, D. Wang, X. Chu, J. Chen, and Y. Wang, “Coordinated transformer with position and sample-aware central loss for anatomical landmark detection,” *arXiv*, pp. 1–10, 5 2023.
- [118] Q. Feng, S. Liu, J. xiang Peng, T. Yan, H. Zhu, Z. jun Zheng, and H. chao Feng, “Deep learning-based automatic sella turcica segmentation and morphology measurement in x-ray images,” *BMC Medical Imaging*, vol. 23, p. 41, 3 2023.
- [119] J. Chen, H. Che, J. Sun, Y. Rao, and J. Wu, “An automatic cephalometric landmark detection method based on heatmap regression and monte carlo dropout,” in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1–4, IEEE, 7 2023.
- [120] F. Wu, L. Chen, B. Feng, and P. Hao, “Multi-scale hybrid transformer network with grouped convolutional embedding for automatic cephalometric landmark detection,” in *Computer-Aided*

- Design and Computer Graphics: 18th International Conference, CAD/Graphics 2023, Shanghai, China, August 19–21, 2023, Proceedings*, (Berlin, Heidelberg), p. 250–265, Springer-Verlag, 2023.
- [121] M. Juneja, P. Garg, R. Kaur, P. Manocha, Prateek, S. Batra, P. Singh, S. Singh, and P. Jindal, “A review on cephalometric landmark detection techniques,” *Biomedical Signal Processing and Control*, vol. 66, p. 102486, 4 2021.
- [122] J. Londono, S. Ghasemi, A. H. Shah, A. Fahimipour, N. Ghadimi, S. Hashemi, Z. Khurshid, and M. Dashti, “Evaluation of deep learning and convolutional neural network algorithms accuracy for detecting and predicting anatomical landmarks on 2d lateral cephalometric images: A systematic review and meta-analysis,” *The Saudi Dental Journal*, vol. 35, pp. 487–497, 7 2023.
- [123] L. CAO, H. HE, and F. HUA, “Deep learning algorithms have high accuracy for automated landmark detection on 2d lateral cephalograms,” *Journal of Evidence-Based Dental Practice*, vol. 22, p. 101798, 12 2022.
- [124] L. Lin, B. Tang, L. Cao, J. Yan, T. Zhao, F. Hua, and H. He, “The knowledge, experience, and attitude on artificial intelligence-assisted cephalometric analysis: Survey of orthodontists and orthodontic students,” *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 164, pp. e97–e105, 10 2023.
- [125] M. A. Khalid, A. Khurshid, K. Zulfiqar, U. Bashir, and M. M. Fraz, “A two-stage regression framework for automated cephalometric landmark detection incorporating semantically fused anatomical features and multi-head refinement loss,” *Expert Systems with Applications*, vol. 255, p. 124840, 12 2024.
- [126] M. Strunga, D. S. Ballová, J. Tomášik, Ľubica Oravcová, Ľuboš Danišovič, and A. Thurzo, “Ai-automated cephalometric tracing: A new normal in orthodontics?,” in *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)*, pp. 1–4, IEEE, 2 2024.
- [127] R. Khan, M. A. Khalid, K. Zulfiqar, U. Bashir, and M. M. Fraz, “Enhancing cephalometric landmark detection with a two-stage cascaded cnn on multi-resolution multi-modal data,” in *Medical Image Understanding and Analysis: 28th Annual Conference, MIUA 2024, Manchester, UK, July 24–26, 2024, Proceedings, Part II*, (Berlin, Heidelberg), p. 3–18, Springer-Verlag, 2024.
- [128] X. Wang, E. Rigall, X. An, Z. Li, Q. Cai, S. Zhang, and J. Dong, “A new benchmark and low computational cost localization method for cephalometric analysis,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, pp. 8939–8952, 9 2024.

- [129] R. Rojas, *Neural Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1996.
- [130] J. Keller M., D. Liu, and D. B. Fogel, *Fundamentals of Computational Intelligence*. John Wiley & Sons, 1 ed., 2016.
- [131] F. Chollet, *Deep learning with Python*. New York, New York, USA: Manning Publications, 1 ed., 2017.
- [132] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2016.
- [133] A. Burkov, *The Hundred-Page Machine Learning Book*. Andriy Burkov, 2019.
- [134] S. J. Russell and P. Norvig, *Artificial Intelligence a Modern Approach*. England: Pearson, 3 ed., 2016.
- [135] A. Wnuk, A. Davis, C. Parks, D. Halber, D. Kelly, G. Zyla, K. Hopkin, K. Weintraub, J. M. Beverly, K. Safia Sheikh, L. Wesse, L. Chiu, M. Fessenden, M. Galinato, M. Richardson, S. Blumenrath, and S. Rojahn, *BrainFacts a primer on the brain and nervous system*, vol. 8. Washington, DC: Society for Neuroscience, 8 ed., 2018.
- [136] N. Yadav, A. Yadav, and M. Kumar, *An Introduction to Neural Network Methods for Differential Equations*. SpringerBriefs in Applied Sciences and Technology, Dordrecht: Springer Netherlands, 2015.
- [137] B. Müller, J. Reinhardt, and M. T. Strickland, *Neural Networks*. Physics of Neural Networks, Berlin, Heidelberg: Springer Berlin Heidelberg, 1995.
- [138] Stanford University, “Harnessing the Power of Data in Health,” *Stanford Medicine Health Trends Report*, no. June, p. 20, 2017.
- [139] P. Cull, “The mathematical biophysics of Nicolas Rashevsky,” *BioSystems*, vol. 88, no. 3, pp. 178–184, 2007.
- [140] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The Bulletin of Mathematical Biophysics*, vol. 5, 12 1943.
- [141] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychological Review*, vol. 65, no. 6, 1958.
- [142] M. Minsky and S. A. Papert, *Perceptrons: An Introduction to Computational Geometry*. The MIT Press, 2017.
- [143] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, pp. 533–536, 10 1986.

- [144] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to digit recognition," 1989.
- [145] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [146] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.
- [147] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems* (Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, eds.), vol. 27, pp. 2672–2680, Curran Associates, Inc., 2014.
- [148] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jess, *Neural Network Design*. Stillwater, OK, USA: Martin Hagan, 2nd ed., 2014.
- [149] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019.
- [150] S. Ruder, "An overview of gradient descent optimization algorithms," *CoRR*, vol. abs/1609.04747, 2016.
- [151] "University of toronto's introduction to neural networks and machine learning lecture 6a: Overview of mini-batch gradient descent." [Online]. Available from: http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf, 2014.
- [152] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, 2015.
- [153] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, p. 2121–2159, July 2011.
- [154] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [155] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 5 2015.
- [156] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess,

- J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” *arXiv*, 2020.
- [157] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv*, 2020.
- [158] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis, “Improved protein structure prediction using potentials from deep learning,” *Nature*, vol. 577, no. 7792, pp. 706–710, 2020.
- [159] K. A. Lamkin-Kennard and M. B. Popovic, “Sensors: Natural and Synthetic Sensors,” in *Biomechatronics*, pp. 81–107, Elsevier, 2019.
- [160] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *The Journal of Physiology*, vol. 160, pp. 106–154, 1 1962.
- [161] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex,” *The Journal of Physiology*, vol. 195, no. 1, pp. 215–243, 1968.
- [162] K. Fukushima, “Cognitron: A self-organizing multilayered neural network,” *Biological Cybernetics*, vol. 20, no. 3-4, pp. 121–136, 1975.
- [163] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biological Cybernetics*, vol. 36, pp. 193–202, 4 1980.
- [164] M. N. O. Sadiku and W. H. Ali, *Signals and Systems A primer with Matlab*, vol. 1. CRC Press, 1 ed., 2016.
- [165] F. Chollet *et al.*, “Keras.” <https://keras.io>, 2015.
- [166] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into Deep Learning*. Cambridge University Press, 2020. <https://d2l.ai>.
- [167] M. D. Zeiler and R. Fergus, “Stochastic pooling for regularization of deep convolutional neural networks,” *1st International Conference on Learning Representations, ICLR 2013 - Conference Track Proceedings*, pp. 1–9, 2013.
- [168] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *International Conference on Learning Representations*, 2015.
- [169] T. Dozat, “Incorporating Nesterov Momentum into Adam,” *ICLR Workshop*, no. 1, pp. 2013–2016, 2016.

- [170] S. Albelwi and A. Mahmood, “A framework for designing the architectures of deep Convolutional Neural Networks,” *Entropy*, vol. 19, no. 6, 2017.
- [171] J. Bergstra and Y. Bengio, “Random Search for Hyper-Parameter Optimization,” *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.
- [172] L. N. Smith and N. Topin, “Deep Convolutional Neural Network Design Patterns,” *CoRR*, vol. abs/1611.00847, 2016.
- [173] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A Large-Scale Hierarchical Image Database,” in *CVPR09*, 2009.
- [174] A. Krizhevsky, “Learning Multiple Layers of Features from Tiny Images,” tech. rep., University of Toronto, 2009.
- [175] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [176] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, “Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.
- [177] A. Farooq, S. Anwar, M. Awais, and S. Rehman, “A deep CNN based multi-class classification of Alzheimer’s disease using MRI,” in *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 1–6, IEEE, 10 2017.
- [178] W. Sun, T. L. B. Tseng, J. Zhang, and W. Qian, “Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data,” *Computerized Medical Imaging and Graphics*, vol. 57, pp. 4–9, 2017.
- [179] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, “Convolutional Neural Networks for Diabetic Retinopathy,” *Procedia Computer Science*, vol. 90, no. July, pp. 200–205, 2016.
- [180] C. Chen, C. Qin, H. Qiu, G. Tarroni, J. Duan, W. Bai, and D. Rueckert, “Deep Learning for Cardiac Image Segmentation: A Review,” *Frontiers in Cardiovascular Medicine*, vol. 7, 3 2020.
- [181] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images,” *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1240–1251, 5 2016.

- [182] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, “Regression convolutional neural network for improved simultaneous EMG control,” *J Neural Eng.*, vol. 16, p. 036015, 6 2019.
- [183] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.
- [184] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015. Software available from tensorflow.org.
- [185] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [186] J. Qian, W. Luo, M. Cheng, Y. Tao, J. Lin, and H. Lin, “Cephann: A multi-head attention network for cephalometric landmark detection,” *IEEE Access*, vol. 8, pp. 112633–112641, 2020.
- [187] X. Dai, H. Zhao, T. Liu, D. Cao, and L. Xie, “Locating anatomical landmarks on 2d lateral cephalograms through adversarial encoder-decoder networks,” *IEEE Access*, vol. 7, pp. 132738–132747, 2019.
- [188] E. Calderon-Sastre, J. L. Lopez-Ramirez, J. Ruiz-Pinales, J. G. Avina-Cervantes, M. A. Ibarra-Manzano, and J. C. Garnica-Palazuelos, “Preprocessing and labeling tool for lateral skull x-ray images applied to cephalometric analysis,” in *2020 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, vol. 4, pp. 1–6, 2020.