

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL

Unidad Tamaulipas

**Un método para la segmentación
semántica de ultrasonografías de
mama basado en cómputo
neuronal**

Tesis que presenta:

Luis Eduardo Aguilar Camacho

Para obtener el grado de:

**Maestro en Ciencias en Ingeniería y
Tecnologías Computacionales**

Director de la Tesis:
Dr. Wilfrido Gómez Flores

© Derechos reservados por
Luis Eduardo Aguilar Camacho
2020

This research was partially funded by project number FidSC2018/145 from "Fondo SEP-Cinvestav 2018"

La tesis presentada por Luis Eduardo Aguilar Camacho fue aprobada por:

Dr. José Juan García Hernández

Dr. Gregorio Toscano Pulido

Dr. Wilfrido Gómez Flores, Director

Cd. Victoria, Tamaulipas, México., 23 de Octubre de 2020

A mis padres

Agradecimientos

- Agradezco infinitamente a mis padres José Luis Aguilar Bustos y Juana Idalia Camacho López por el apoyo y la confianza que me brindaron para vivir una experiencia más en mi vida.
- Agradezco al Dr. Wilfrido Gómez Flores por su amistad, apoyo, paciencia, enseñanza y esfuerzo.
- Agradezco al Dr. Juan Humberto Sossa Azuela y al Instituto Politécnico Nacional por permitirme realizar una estancia para realizar esta investigación.
- Agradezco a mis sinodales, Dr. Gregorio Toscano Pulido y Dr. José Juan Hernández García, por sus observaciones y recomendaciones que contribuyeron a mejorar esta investigación.
- Agradezco a CONACyT por el apoyo económico brindado que me permitió concentrarme en mis estudios y al CINVESTAV-Tamaulipas por la oportunidad de realizar estudios de posgrado.
- Agradezco al proyecto FidSC2018/145 del Fondo Sep-Cinvestav por el apoyo económico para realizar esta investigación.
- Agradezco a mis compañeros y amigos Yeyo, Fernando, Diana, Armando, Iván, Melissa, Ricardo, Shanel, Mogui, Elías y Tipa, por su amistad, apoyo y consejos.

Índice General

Índice General	I
Índice de Figuras	v
Índice de Tablas	vii
Índice de Algoritmos	ix
Publicaciones	ix
Resumen	xi
Abstract	xiii
1. Introducción	1
1.1. Antecedentes	1
1.1.1. Cáncer de mama	1
1.1.2. Sistemas CAD	3
1.2. Planteamiento del problema	6
1.3. Preguntas de investigación	8
1.4. Hipótesis	8
1.5. Objetivos general y particulares	8
1.5.1. General	8
1.5.2. Particulares	9
1.5.3. Metodología	9
1.5.4. Organización de la tesis	10
2. Marco Teórico	11
2.1. Definición de segmentación semántica	11
2.2. Pre-procesamiento y extracción de características	14
2.2.1. Pre-procesamiento	14
2.2.2. Extracción de características	16
2.2.2.1. Características basada en contraste	16
2.2.2.2. Filtros de Log-Gabor	20
2.2.2.3. Momentos de Hu	21
2.2.2.4. Momentos de Zernike	25
2.2.2.5. Transformada de Hermite	26
2.2.2.6. Ranklets	28
2.2.2.7. Auto-Información mutua	30
2.2.2.8. Matriz de co-ocurrencia de niveles de gris	33

2.3.	Reducción de dimensionalidad	34
2.3.1.	Análisis de Componentes Principales Asimétrico	35
2.3.2.	Mínima Redundancia Máxima Relevancia	37
2.4.	Redes Neuronales Artificiales	38
2.4.1.	Modelo matemático	38
2.4.2.	Descenso de gradiente estocástico	39
2.4.3.	Tasa de aprendizaje y momentum	41
2.4.4.	Funciones de activación	42
2.4.4.1.	Leaky ReLU	43
2.4.4.2.	Softmax	43
2.4.5.	Perceptrón multicapa	44
2.4.6.	Redes Neuronales Convolucionales	46
2.5.	Transferencia de aprendizaje	49
2.6.	Aumento artificial de datos	50
2.7.	Índices de validación	50
2.7.1.	Exactitud	51
2.7.2.	Coefficiente de correlación de Matthews	51
2.7.3.	Sensibilidad	52
2.7.4.	Especificidad	52
2.7.5.	Exactitud balanceada	52
2.7.6.	Medida F1	53
2.7.7.	Índice Jaccard	53
2.8.	Validación cruzada de k dobles	53
2.9.	Pruebas de significación estadística	53
2.9.1.	Prueba de Shapiro-Wilk	54
2.9.2.	Prueba de Wilcoxon	55
2.9.3.	Prueba de Kruskal-Wallis con corrección Bonferroni	55
2.10.	Sumario	55
3.	Estado del Arte	57
3.1.	Enfoque Convencional	57
3.2.	Enfoque Convolutivo	60
3.3.	Sumario	63
4.	Metodología	65
4.1.	Conjunto de imágenes de BUS	65
4.2.	Marco experimental	66
4.3.	Experimentación	67
4.3.1.	Enfoque convencional	68
4.3.2.	Enfoque convolutivo	72
4.4.	Validación	75
4.5.	Infraestructura computacional	76
4.6.	Sumario	77

- 5. Resultados** **79**
- 5.1. Enfoque Convencional 79
- 5.2. Enfoque Convolucional 85
- 5.3. Sumario 88

- 6. Conclusiones y trabajo futuro** **89**
- 6.1. Conclusiones 89
- 6.2. Contribución 91
- 6.3. Trabajo futuro 91

Índice de Figuras

1.3. Etapas básicas de un sistema CADe para ultrasonido de mama.	4
2.2. Pre-procesamiento de una imagen BUS y el efecto en su histograma de intensidad. Superior: Imagen BUS original. Medio: Imagen BUS filtrada. Inferior: Imagen BUS normalizada.	15
2.7. Banco de filtros de log-Gabor en el dominio de la frecuencia con diferentes valores de ρ y θ	21
2.9. Características de momentos de Hu.	25
2.10. Polinomios de Zernike para $n \in \{3, 4, 5\}$ y $l \in \{0, \dots, 5\}$	27
2.11. Descomposición de una imagen BUS con la transformada de Hermite con distintos valores de orden n	29
2.12. (a) Para una imagen de tamaño 11×11 , la máscara <i>ranklet</i> con resolución $r = 4$ genera 49 ventanas traslapadas centradas en los puntos grises. (b) Orientaciones de la máscara <i>ranklet</i> en: horizontal (h), vertical (v) y diagonal (d), con sus correspondiente subconjuntos de tratamiento \mathcal{T} y control \mathcal{C}	30
2.13. Características de textura basadas en <i>ranklet</i> con cuatro resoluciones.	31
2.14. Características de textura basadas en auto-información mutua.	32
2.15. Características de textura basadas en GLCM.	35
2.16. Morfología de una neurona biológica.	39
2.17. Analogía de una neurona biológica y una artificial.	40
2.18. Decaimiento de la tasa de aprendizaje de acuerdo con la Ecuación 2.77.	42
2.19. Diferencias entre las funciones de activación.	44
2.20. Ejemplo de una red perceptrón multicapa (MLP).	45
2.21. Diagrama básico de una red neuronal convolucional codificador-decodificador para la segmentación semántica de imágenes.	48
2.22. Modelo general de las arquitecturas CNN utilizadas en esta investigación.	48
2.23. Ejemplo de operaciones de aumento artificial de datos.	50
2.24. Matriz de confusión para un problema de dos clases.	51
2.25. Distribución del conjunto de datos para los conjuntos de entrenamiento y validación en el método validación cruzada de k dobles.	54
4.1. Ejemplo de imágenes BUS con la segmentación manual en rojo, donde a) y b) son tumores benignos, c) y d) son tumores malignos.	66
4.2. Flujo de procesos para el entrenamiento y validación de un CADe con enfoque convencional.	67
4.3. Flujo de procesos para el entrenamiento y validación de un CADe con enfoque convolucional.	67
4.4. Proceso de extracción de características.	69

4.5.	Cálculo de la media local $I_{\mu}(x, y)$ a partir de una imagen de característica local y agrupación tipo <i>max-pooling</i> con tamaño $w_p = 2$. En las regiones coloreadas se observa que para un <i>stride</i> de 2 se analiza una ventana de 2×2	70
4.6.	Proceso de entrenamiento para la segmentación semántica de imágenes BUS con una MLP.	71
4.7.	Etapas de post-procesamiento.	73
4.8.	Proceso de entrenamiento para la segmentación semántica de imágenes BUS con una CNN.	75
5.1.	Ejemplos de segmentación del tumor con enfoque convencional.	84
5.2.	Ejemplos de segmentación del tumor con enfoque convolucional.	86

Índice de Tablas

2.1. Características de textura basadas en GLCM.	34
3.1. Desempeño de segmentación de enfoques convencionales.	60
3.2. Desempeño de segmentación de métodos basados en el enfoque convolucional.	63
4.1. Características generales de las arquitecturas CNN utilizadas en esta investigación.	68
4.2. Características locales calculadas a partir de métodos descritos en el Capítulo 2.	68
4.3. Parámetros de entrenamiento del enfoque convolucional.	75
5.1. Resultados de segmentación de la MLP con agrupamiento <i>mean-pooling</i>	80
5.2. Resultados de segmentación de la MLP con agrupamiento <i>max-pooling</i>	80
5.3. Prueba de Wilcoxon entre APCA y MRMR por tamaño de ventana con agrupación <i>mean-pooling</i>	81
5.4. Prueba de Wilcoxon entre APCA y MRMR por tamaño de ventana con agrupación <i>max-pooling</i>	81
5.5. Prueba de Kruskal-Wallis entre métodos para agrupación <i>mean-pooling</i>	82
5.6. Prueba de Kruskal-Wallis entre métodos para agrupación <i>max-pooling</i>	82
5.7. Porcentaje de reducción de características.	83
5.8. Resultados de la MLP con <i>max-pooling</i> y post-procesamiento.	83
5.9. Resultados de segmentación por tipo de lesión.	84
5.10. Resultados de segmentación con cuatro arquitecturas CNN.	85
5.11. Prueba de Kruskal-Wallis entre arquitecturas CNN.	85
5.12. Resultados de segmentación por tipo de tumor.	87
5.13. Prueba de Kruskal-Wallis entre enfoques convencional y convolucional.	87

Publicaciones

Luis Eduardo Aguilar Camacho, Wilfrido Gómez Flores, Juan Humberto Sossa Azuela, *A comparative study of neural computing approaches for semantic segmentation of breast ultrasound images*, in The XXVII edition of the Brazilian Congress of Biomedical Engineering (CBEB 2020), Vitória, Brazil, October 2020.

Un método para la segmentación semántica de ultrasonografías de mama basado en cómputo neuronal

por

Luis Eduardo Aguilar Camacho

Unidad Tamaulipas

Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, 2020

Dr. Wilfrido Gómez Flores, Director

Los sistemas de diagnóstico asistido por computadora (CADx, por sus siglas en inglés) tienen el potencial de coadyuvar al especialista en el análisis de imágenes mamográficas para la recomendación de la conducta clínica de los pacientes. Una etapa importante en un sistema CADx es la detección y segmentación del tumor para su posterior análisis, lo cual se realiza mediante sistemas de detección asistido por computadora (CAdE, por sus siglas en inglés). Esta tarea se puede realizar desde el enfoque de aprendizaje supervisado, de manera que cada píxel de la imagen es clasificado, lo cual se denomina segmentación semántica. Comúnmente se utilizan redes neuronales artificiales (ANN, por sus siglas en inglés) para crear modelos de segmentación semántica para distinguir entre dos clases: tumor y tejido adyacente. En la literatura, se pueden encontrar dos tipos de enfoques para los sistemas CAdE basados en ANN: convencional y convolucional. Los enfoques convencionales requieren de un conjunto de características definido por un experto para el entrenamiento de una ANN, mientras que el enfoque convolucional no requiere al experto, ya que la extracción de características es automática. En este proyecto de investigación se propone un método de segmentación semántica de ultrasonografías de mama basado en cómputo neuronal, el cual se define a partir de un estudio comparativo entre los enfoques convencional y convolucional.

A method for semantic segmentation of breast ultrasonography based on neuronal computation

by

Luis Eduardo Aguilar Camacho

Cinvestav Tamaulipas

Research Center for Advanced Study from the National Polytechnic Institute, 2020

Dr. Wilfrido Gómez Flores, Advisor

Computer-aided diagnostic (CADx) systems have the potential to assist the specialist in the analysis of mammographic images in recommending the clinical conduct of patients. An important stage in a CADx system is the detection and segmentation of the tumor for subsequent analysis, which is carried out using computer-aided detection systems (CADE). This task can be carried out from the supervised learning approach, so that each pixel in the image is classified, which is called semantic segmentation. Artificial neural networks (ANNs) are commonly used to create semantic segmentation models to distinguish between two classes: tumor and adjacent tissue. In the literature, two types of approaches can be found for ANN-based CADE systems: conventional and convolutional. Conventional approaches require a set of features selected by an expert for ANN training, while the convolutional approach does not require an expert, since feature extraction is automatic. In this research project, a semantic segmentation method for breast ultrasonography based on neuronal computation is proposed, which is defined from a comparative study between conventional and convolutional approaches.

1

Introducción

En este capítulo se presenta una breve introducción y antecedentes del cáncer de mama, el problema de la segmentación semántica de tumores de mama, y los sistemas de detección asistidos por computadora. Además, se presenta el planteamiento del problema, la hipótesis y los objetivos a cumplir en esta tesis.

1.1 Antecedentes

1.1.1 Cáncer de mama

El cáncer de mama es un proceso oncológico donde las células de la glándula mamaria se multiplican con rapidez y sin control para formar tumores cancerosos (o malignos), las cuales pueden diseminarse a otras partes del cuerpo [5]. Actualmente, esta patología se considera como la primera causa de muerte por tumor maligno en mujeres a nivel mundial [16]. Para realizar el diagnóstico de este tipo de cáncer se requieren sistemas de imagen y profesionales expertos en patología mamaria.

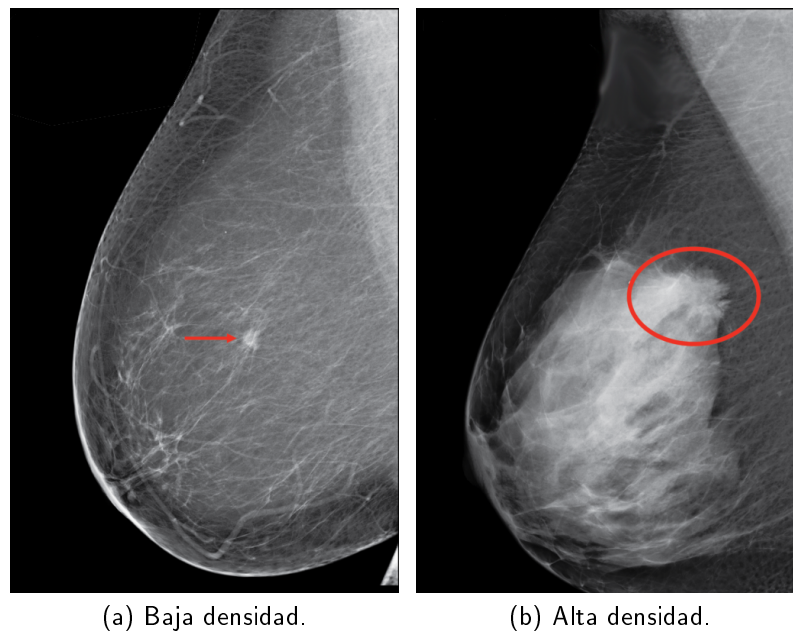
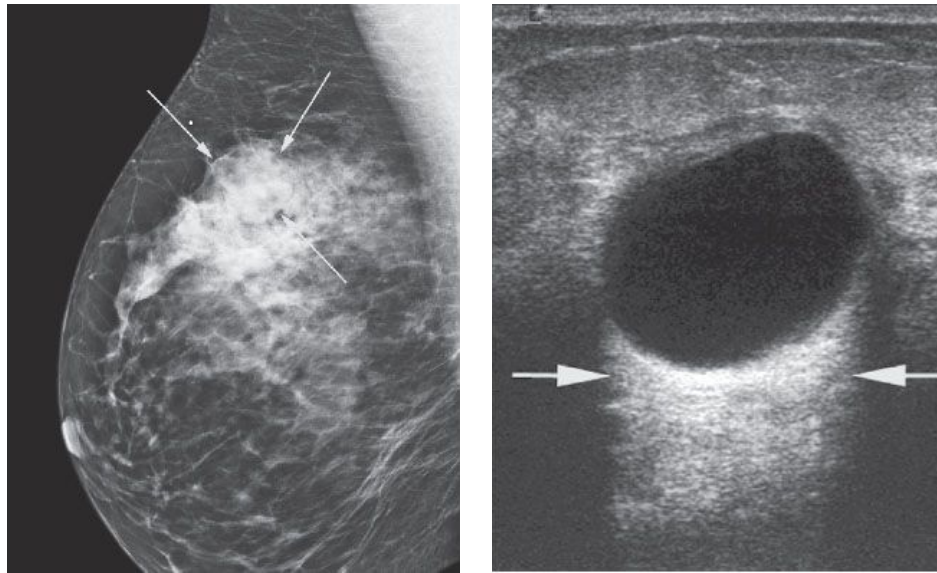


Figura 1.1: Mamografía con diferente densidad de tejido mamario. En rojo se señala la localización de tumores.

Hoy en día se cuenta con técnicas como la mamografía, ultrasonido, resonancia magnética nuclear y biopsia, para la detección y prevención del cáncer de mama [42].

La mamografía (rayos-X) es el principal método imagenológico para la detección temprana del cáncer de mama, ya que es capaz de revelar estructuras en el orden de los micrómetros como las microcalcificaciones [42]. Sin embargo, este método emite radiación ionizante y comprime la mama para obtener imágenes de calidad. Además, la sensibilidad de la mamografía es aproximadamente 63 % en mamas densas y 87 % en mamas con mayor contenido graso, mientras que su especificidad está entre 87 % y 99 %, respectivamente [2]. La Figura 1.1 muestra dos mamografías que presentan tejidos mamarios con baja y alta densidad. Como se observa en la Figura 1.1b, la densidad de la mama dificulta la detección de un tumor debido a la alta densidad del tejido mamario. Por otro lado, el ultrasonido permite obtener imágenes útiles para caracterizar lesiones, destacando su bajo costo, accesibilidad y ausencia de radiaciones ionizantes. De este modo, ayuda a diferenciar entre lesiones benignas y malignas debido a su capacidad para caracterizar nódulos mamarios y permite su identificación en mamas densas. La Figura 1.2 muestra que el ultrasonido puede detectar

anormalidades ocultas en la mamografía en una paciente con tejido mamario denso. Ambos métodos en conjunto pueden mejorar la precisión del diagnóstico para la detección temprana del cáncer de mama.



(a) Mamografía con una anomalía oculta debido al tejido mamario denso.

(b) Detección de la misma anomalía en (a) por ultrasonido.

Figura 1.2: Detección de anomalías por mamografía y ultrasonido de mama.

1.1.2 Sistemas CAD

En la última década, se han desarrollado sistemas de diagnóstico asistido por computadora (CAD, por sus siglas en inglés) con el objetivo de reducir la variabilidad inter- e intra-observador en el diagnóstico de tumores de mama, lo cual se puede deber a fatiga, falta de experiencia o negligencia [63]:

- **Variabilidad inter-observador:** diferencia que existe entre las interpretaciones de distintos expertos sobre una misma imagen.
- **Variabilidad intra-observador:** diferencia que existe en la interpretación de un experto cuando revisa la misma imagen en distintos instantes de tiempo.

Cabe señalar que los sistemas CAD no son diseñados para reemplazar el diagnóstico de los radiólogos, sino para mejorar la precisión en la detección y clasificación del tumor. Los sistemas CAD pueden identificar regiones con sospecha de malignidad debido a su capacidad de caracterizar nódulos mamarios mediante técnicas de análisis de imágenes y reconocimiento de patrones, y pueden clasificarse en dos tipos de sistemas de acuerdo con la tarea que realizan [17]:

- **CADe**: Sistemas de detección asistido por computadora orientados a la localización de lesiones en imágenes médicas.
- **CADx**: Sistemas de diagnóstico asistido por computadora, los cuales realizan la caracterización de las lesiones, por ejemplo, la distinción entre tumores benignos y malignos.

Los sistemas CADe que se emplean en el análisis de imágenes de mama detectan anomalías en el patrón del tejido mamario independientemente si se trata de un cáncer o un tumor benigno, es decir, distingue entre las regiones de tumor (primer plano) y tejido adyacente (fondo). Si se utiliza aprendizaje supervisado para realizar esta tarea, entonces se deriva el concepto denominado segmentación semántica, el cual tiene como objetivo dar un significado único a cada píxel de la imagen. Generalmente, los sistemas CADe de ultrasonido para la segmentación semántica de tumores de mama involucran cuatro etapas [12], como se muestra en la Figura 1.3:

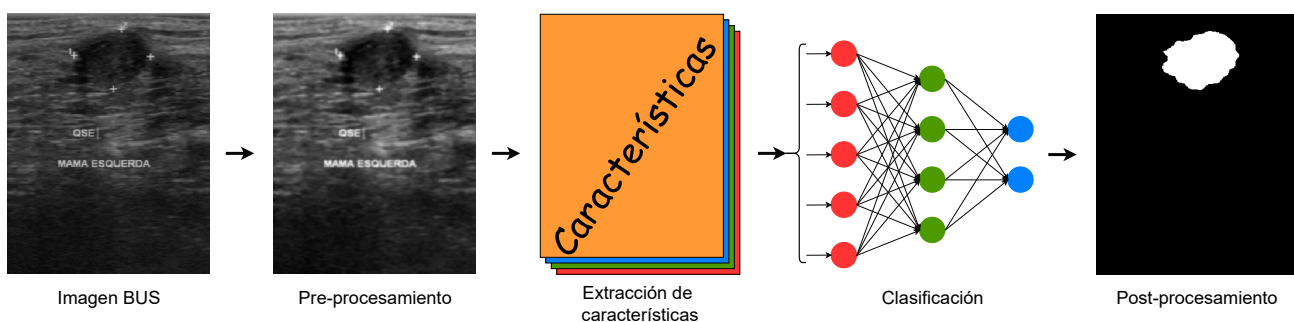


Figura 1.3: Etapas básicas de un sistema CADe para ultrasonido de mama.

1. **Pre-procesamiento**: las principales limitaciones de las imágenes de ultrasonido de mama (BUS, por sus siglas en inglés) son el bajo contraste y la presencia de ruido. La tarea del

pre-procesamiento es mejorar la calidad de la imagen, reduciendo el ruido y preservando las características importantes de las estructuras de interés.

2. **Extracción de características:** se calcula un conjunto de atributos o rasgos de cada píxel de la imagen para distinguir entre clases tumor y fondo. Esto puede involucrar atributos de contraste, textura, espaciales, entre otros.
3. **Clasificación:** a partir de las características extraídas, los píxeles se clasifican en tumor y fondo con algún método de clasificación supervisada para obtener la imagen segmentada.
4. **Pos-procesamiento:** se reducen los falsos positivos y se refina la solución de la segmentación del tumor.

Actualmente, la segmentación semántica de tumores de mama se puede abordar mediante el uso de redes neuronales artificiales (ANN, por sus siglas en inglés), ya que aprenden a partir de un conjunto de datos de entrenamiento. En la literatura, los sistemas CAdE basados en ANN se pueden categorizar en dos enfoques principales [58]:

- **Enfoque convencional:** sigue la metodología tradicional del reconocimiento de objetos, en el que un clasificador recibe las características extraídas de una imagen BUS, las cuales fueron definidas por un experto bajo el supuesto que son descriptivas.
- **Enfoque convolucional:** está basado en el uso de redes neuronales convolucionales (CNN, por sus siglas en inglés) [51, 67], que a diferencia del enfoque convencional, las características descriptivas de la imagen BUS se aprenden automáticamente sin la necesidad de un experto que las defina.

Es importante destacar que el uso de las arquitecturas CNN requiere de recursos de cómputo especializado como unidades de procesamiento gráfico (GPUs, por sus siglas en inglés) o arreglos de compuertas lógicas programables en campo (FPGAs, por sus siglas en inglés) para acelerar

el entrenamiento y ajuste de los pesos de la CNN en tiempos de cómputo razonables. El uso de FPGAs comparado con el de las GPUs tiene sus ventajas como: alto rendimiento con baja energía, alto paralelismo, flexibilidad para ser reconfigurable y seguridad. Sin embargo, también muestra desventajas debido al costo de tiempo de reconfiguración en la configuración de FPGA para diferentes diseños, además de la dificultad de programación utilizando lenguajes de programación de hardware, el cual costaría tiempo dominar [59]. Por esta razón, la mayoría de los proyectos de investigación seleccionan GPUs para realizar el entrenamiento, y pocos proyectos eligen FPGAs para entrenar el modelo [34]. Además, una CNN demanda una cantidad excesiva de imágenes para realizar el entrenamiento de la misma, por lo que se suele utilizar transferencia de aprendizaje para aprovechar arquitecturas CNN pre-entrenadas. La transferencia de aprendizaje se utiliza como punto inicial cuando se quiere obtener un modelo capaz de realizar una tarea diferente a la que realiza originalmente [67], y esto ocurre cuando es limitada la fuente de datos de entrenamiento [44]. En esta investigación se propondrá un método de segmentación semántica de ultrasonografías de mama basado en cómputo neuronal, el cual se definirá a partir de un estudio comparativo entre los enfoques convencional y convolucional.

1.2 Planteamiento del problema

Actualmente, los sistemas CADe han sido de gran utilidad para la detección de tumores de mama mediante la segmentación semántica de imágenes BUS. Estos sistemas funcionan como segunda opinión para los expertos, de tal forma que coadyuvan en la detección automática de tumores. Realizar el proceso de segmentación puede abordarse desde dos enfoques distintos, denominados convencional y convolucional [58].

Para llevar a cabo la segmentación mediante el enfoque convolucional, se requiere el uso de GPUs para acelerar el proceso de entrenamiento de la red. Además, entrenar una CNN desde cero requiere grandes conjuntos de imágenes BUS para capturar diferentes variaciones de las mismas durante el

entrenamiento. Sin embargo, existen arquitecturas CNN pre-entrenadas capaces de adecuarse para realizar una nueva tarea de segmentación, de tal forma que se evita entrenar una CNN desde cero. Los enfoques basados en CNNs han demostrado un buen desempeño para la segmentación y clasificación de tumores, gracias a su capacidad de aprender automáticamente características con alto nivel de abstracción. Una CNN puede tener decenas o cientos de niveles de profundidad compuestas por capas de convolución que extraen mapas de características. En los primeros niveles de profundidad de la CNN se aprenden características generales, tales como brillos, colores, bordes, entre otros, mientras que en niveles más profundos se aprenden características más complejas que describen mejor a los objetos. En cada capa de convolución se aplican filtros con distintas resoluciones, y la salida de cada imagen convolucionada representa un mapa de características que se emplea como entrada para la siguiente capa [24, 25, 63, 67].

Por otro lado, los métodos convencionales [27, 35, 52, 60] suelen realizar la segmentación de tumores con menor uso de recursos computacionales, ya que se basan en la experiencia de un experto para realizar la extracción de características bajo el supuesto que son discriminantes entre diferentes clases e invariantes para la misma clase [20]. Existen diversos métodos en la literatura que realizan la extracción de características, las cuales pueden ser detectores de bordes, filtros de texturas, técnicas espaciales como momentos geométricos, diferencias de brillo y contraste, entre otros. El desafío de los métodos convencionales consiste en seleccionar características relevantes para discriminar entre tejido normal y tumoral.

De acuerdo a una revisión del estado del arte, actualmente no existe un estudio que compare el desempeño de sistemas CADe basado en los enfoques convencional y convolucional, en donde se determine si ambos enfoques son competitivos para la detección y segmentación de tumores de mama en imágenes BUS. Aunque en la literatura de visión por computadora se ha demostrado la superioridad de las CNN sobre métodos clásicos en distintas aplicaciones, en el caso de segmentación de imágenes BUS no se tiene evidencia de las ventajas y limitantes que tienen ambos enfoques evaluados sobre una misma base de imágenes. Con lo anterior, surge la duda de si es justificable

utilizar una CNN, que emplea recursos computacionales de alto desempeño, o si es suficiente utilizar un enfoque convencional, el cual requiere una infraestructura computacional moderada, para obtener resultados de segmentación equiparables.

1.3 Preguntas de investigación

De lo antes expuesto se derivan las siguientes preguntas al término de la investigación:

- ¿Se obtienen resultados de segmentación equiparables entre un enfoque convolucional y uno convencional para la segmentación semántica de tumores en imágenes BUS?
- ¿Es justificable el uso de recursos computacionales de alto desempeño para el entrenamiento de una CNN para la segmentación de tumores en imágenes BUS?
- ¿Cuáles son las características que proporcionan información relevante que permite distinguir a cada píxel en las clases tumor y tejido adyacente en un enfoque convencional?

1.4 Hipótesis

Es posible obtener desempeños equiparables para la segmentación semántica de imágenes BUS entre los enfoques de cómputo neuronal convencional y convolucional.

1.5 Objetivos general y particulares

1.5.1 General

Obtener un método de segmentación semántica de ultrasonografías de mama basado en cómputo neuronal a partir de un estudio comparativo entre los enfoques convencional y convolucional.

1.5.2 Particulares

- Definir las arquitecturas de redes neuronales convolucionales pre-entrenadas para el aprendizaje automático de características.
- Definir las características locales que describan a cada píxel de la imagen y la configuración de red neuronal artificial para el enfoque convencional.
- Comparar los desempeños de segmentación de los enfoques convencional y convolucional desarrollados mediante índices de similitud.

1.5.3 Metodología

Para cumplir con los objetivos planificados de esta tesis, se siguió una metodología de investigación dividida en tres etapas:

1. **Estudio e implementación de técnicas para extracción de características y reducción de dimensionalidad.** Se estudiaron técnicas que permiten extraer rasgos de los píxeles para crear un modelo que permita clasificarlos.
 - Se revisó el estado del arte relacionado con técnicas de extracción de características.
 - Se revisaron técnicas de reducción de dimensionalidad de características.
 - Se implementaron técnicas para la extracción de características y reducción de dimensionalidad de las características de los píxeles.
2. **Estudio e implementación de técnicas basadas en cómputo neuronal para segmentación de tumores de mama en imágenes de BUS.** Se implementó una ANN tipo perceptrón multicapa para el enfoque convencional, así como distintas arquitecturas para el enfoque convolucional.

- Se revisó el estado del arte relacionado con sistemas CADe para la segmentación de tumores de mama en imágenes BUS para ambos enfoques.
- Se implementaron los enfoques convencional y convolucional para la segmentación de tumores en imágenes de BUS.

3. **Evaluación de los métodos propuestos.** Fue evaluado el rendimiento de los métodos realizados para ambos enfoques.

- Se comparó el método del enfoque convencional con los métodos del enfoque convolucional con base en índices de similitud con la segmentación de referencia.
- Se realizaron pruebas de significación estadística para definir el método definitivo propuesto en esta tesis.

1.5.4 Organización de la tesis

Esta tesis consta de seis capítulos iniciando con un capítulo introductorio para poner contextualizar al lector. El Capítulo 2 presenta una introducción de las técnicas utilizadas en este estudio para cada una de las etapas, pre-procesamiento, extracción de características, reducción de características, clasificación y post-procesamiento. También se describen los métodos e índices de validación utilizados para medir el rendimiento de segmentación. El Capítulo 3 muestra un resumen de los trabajos relevantes realizados para la segmentación de tumores de mama en ultrasonografías basados en los enfoques convencional y convolucional. El Capítulo 4 presenta el desarrollo de los métodos para cada enfoque. El Capítulo 5 muestra los resultados experimentales que se obtuvieron con ambos enfoques. Finalmente, el Capítulo 6 presenta las conclusiones obtenidas de este trabajo de investigación y define posibles trabajos futuros sobre este tema.

2

Marco Teórico

En este capítulo se presentan los métodos relacionados con la segmentación semántica de ultrasonidos de mama. Además, se describen las técnicas de pre-procesamiento, extracción de características y pos-procesamiento utilizadas en el estudio. También se proveen los detalles del cómputo neuronal y sus enfoques. Finalmente, se explican los métodos e índices de similitud para medir el desempeño de segmentación.

2.1 Definición de segmentación semántica

La segmentación de imágenes es el proceso de dividir una imagen $I(x, y)$ en sus regiones constituyentes hasta un nivel de subdivisión en el que se aislen los objetos de interés [22]. Sea R la representación completa de una imagen, entonces la segmentación es un proceso que divide R en n subregiones, R_1, R_2, \dots, R_n , tal que:

1. $\bigcup_{i=1}^n R_i = R$.

2. R_i es una región conectada, $i = 1, 2, \dots, n$.
3. $R_i \cap R_j = \emptyset, \forall i, j = 1, \dots, n, \text{ y } i \neq j$.
4. $P(R_i) = \text{verdadero}, \forall i = 1, \dots, n$.
5. $P(R_i \cup R_j) = \text{falso}$, para cualquier región adyacente R_i y R_j .

La primera condición indica que la unión de todas las subregiones reconstruyen la imagen completa. La segunda condición requiere que todos los puntos en una región están conexos en 4-conectividad¹ ú 8-conectividad. La tercera condición indica que dos regiones adyacentes deben ser disjuntas, es decir, su unión debe ser el conjunto vacío. La cuarta condición establece que todos los píxeles en una región conexa deben poseer la misma propiedad de similitud, donde P es un predicado lógico definido sobre los puntos en el conjunto R_i . La quinta condición indica que dos regiones adyacentes deben tener diferentes propiedades de similitud de acuerdo con el predicado lógico P .

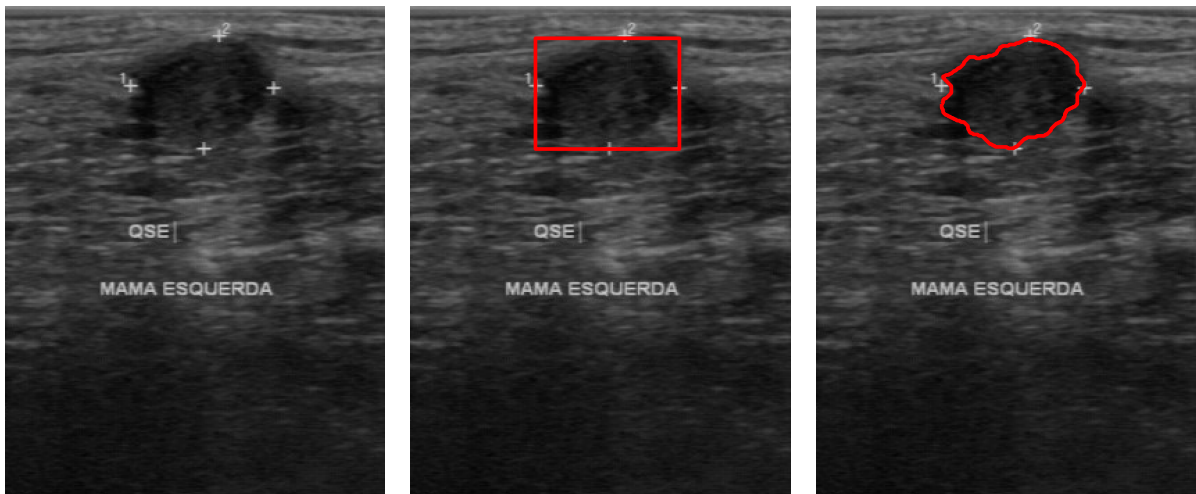
En el campo de visión por computadora, la segmentación semántica es el proceso de dividir una imagen digital $I(x, y)$ en grupos de píxeles con el objetivo de cambiar su representación en otra más significativa. En otras palabras, se asocia una etiqueta de clase a cada uno de los píxeles en la imagen para identificar grupos de píxeles que compartan la misma etiqueta de clase.

Actualmente, existen algoritmos de segmentación no supervisada basado en bordes, similitud de intensidad, texturas, entre otros. Sin embargo, debido a que la segmentación semántica utiliza aprendizaje supervisado, suele ser más precisa que los algoritmos no supervisados de segmentación de objetos. Esta característica favorece a la segmentación semántica para que sea muy utilizada en aplicaciones como conducción autónoma, inspección industrial, análisis de imágenes médicas, visión en robots, entre otros.

Por otro lado, la detección de objetos es una técnica de visión por computadora para localizar instancias de objetos en imágenes, dando como resultado una ventana que localiza el objeto

¹Se presenta 4-conectividad cuando dos píxeles adyacentes forman parte del mismo objeto si están conectados a lo largo de la dirección horizontal o vertical, mientras que 8-conectividad es cuando se conectan a lo largo de la dirección horizontal, vertical o diagonal.

detectado, mientras que la segmentación delimita el contorno de los objetos de distintas categorías a nivel de píxel, dando como resultado cualquier forma arbitraria [19]. La Figura 2.1 muestra la diferencia entre detección y segmentación del tumor en una imagen BUS.



(a) Imagen BUS.

(b) Detección del tumor.

(c) Segmentación del tumor.

Figura 2.1: Diferencia entre detección y segmentación de tumores en imágenes BUS.

Para realizar la segmentación semántica de una imagen digital $I(x, y)$, se requiere de un algoritmo capaz de clasificar cada píxel que contiene la imagen. Las ANN son ampliamente utilizadas para clasificación de patrones debido a que cuenta con una etapa de aprendizaje para generar un modelo de clasificación, lo cual permite que una ANN aprenda a distinguir entre clases de píxeles. Entonces, desde el punto de vista de cómputo neuronal, la segmentación semántica se puede abordar desde dos enfoques: convencional y convolucional.

El enfoque convencional es aquel que sigue la metodología clásica del reconocimiento de objetos, en el que un clasificador recibe como entrada las características de la imagen, las cuales fueron definidas por un experto. Cabe mencionar que este enfoque puede utilizar cualquier clasificador siempre y cuando sea un método de aprendizaje supervisado [27, 35, 52, 60]. En cambio, el enfoque convolucional realiza operaciones de convolución para extraer y aprender características de la imagen de manera automática, donde en las primeras capas extrae características generales como contornos,

brillo, forma, entre otros. En capas más profundas se extraen características más sutiles que permiten diferenciar de manera eficiente a los objetos de interés [24, 25, 63, 67].

2.2 Pre-procesamiento y extracción de características

Una de las tareas más importantes para mejorar los resultados de segmentación semántica, es el pre-procesamiento y extracción de características de la imagen, para posteriormente entrenar una ANN para la clasificación de los píxeles [4].

2.2.1 Pre-procesamiento

La etapa de pre-procesamiento aplica un conjunto de técnicas a las imágenes de entrada con el objetivo de mejorar la calidad de las mismas, alterando sus valores y preservando la información relevante de la estructura de los objetos. Esta etapa es indispensable para la extracción de características de la imagen BUS, ya que presentan dos características distintivas que impactan en la exactitud de la segmentación del tumor: artefacto *speckle* y bajo contraste entre el tumor y el tejido adyacente [49].

El filtro Gaussiano es un filtro paso bajo ampliamente utilizado para reducir ruido y suavizar detalles irrelevantes en la imagen en escalas menores a las escalas de los objetos de interés. El filtrado de la imagen de entrada se expresa como:

$$I_f(x, y) = I(x, y) * G_\sigma(x, y), \quad (2.1)$$

donde “*” denota el operador de convolución, y $G_\sigma(x, y)$ es un *kernel* Gaussiano de tamaño $w \times w$ y un ancho de banda σ .

Una vez filtrada la imagen de entrada, se expande el rango dinámico de la imagen para obtener valores de intensidad en el rango $[0, 1]$, lo cual es un proceso de normalización con el objetivo de

tener una imagen con mayor contraste [22], de forma que:

$$I_c(x, y) = \frac{I_f(x, y) - I_{min}}{I_{max} - I_{min}}, \quad (2.2)$$

donde I_{min} y I_{max} son los valores de intensidad mínimo y máximo de la imagen filtrada $I_f(x, y)$. La Figura 2.2 muestra la transformación que sufre la imagen y su histograma al aplicar cada etapa del pre-procesamiento.

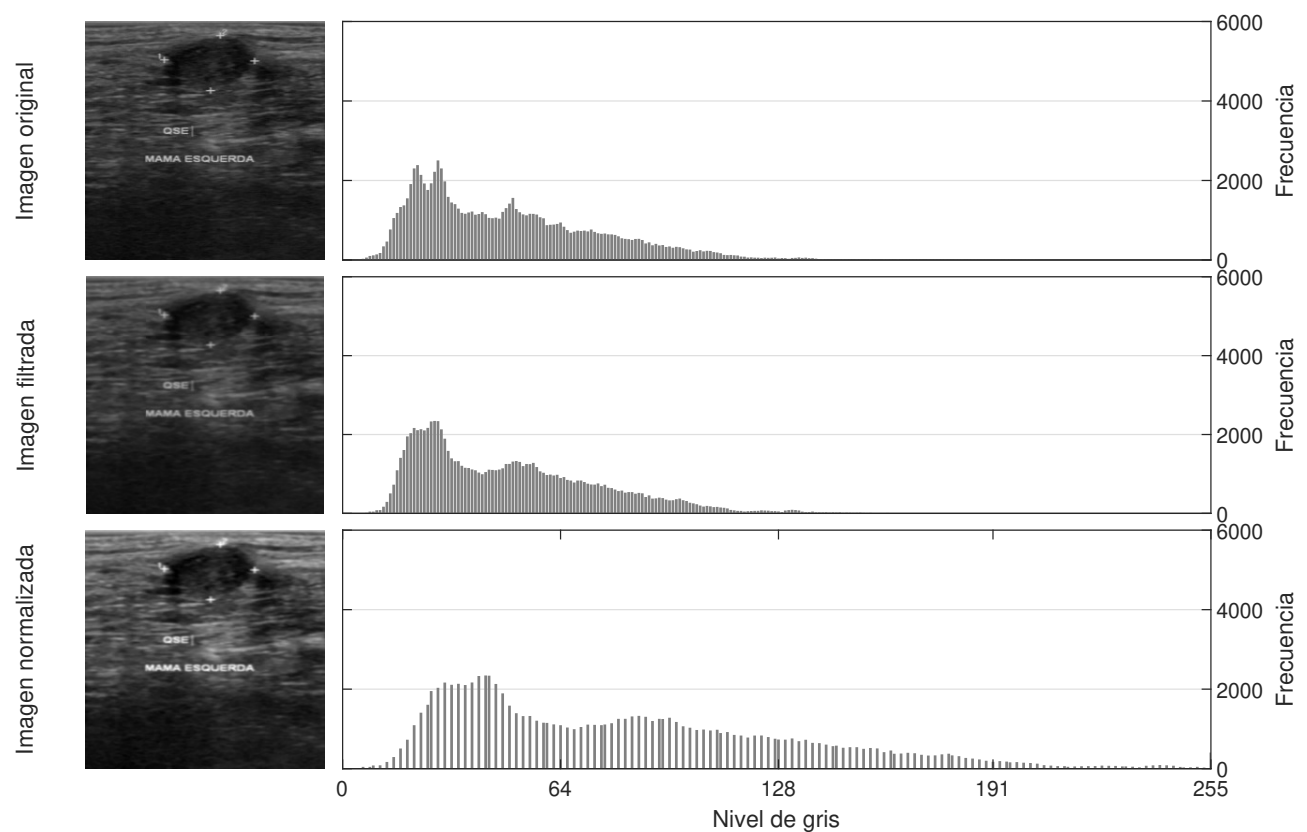


Figura 2.2: Pre-procesamiento de una imagen BUS y el efecto en su histograma de intensidad. Superior: Imagen BUS original. Medio: Imagen BUS filtrada. Inferior: Imagen BUS normalizada.

2.2.2 Extracción de características

Para realizar la clasificación de cada uno de los píxeles, es necesario extraer rasgos únicos que permitan diferenciar entre clases. De este modo, técnicas como filtros de Log-Gabor, momentos de Hu, polinomios de Hermite, entre otros, son apropiados para realizar dicha tarea.

2.2.2.1. Características basada en contraste

El contraste se puede definir como la diferencia de intensidad relativa que existe entre un punto de una imagen y sus vecinos. La variación de contraste es directamente proporcional a la diferencia de intensidad entre un objeto y su fondo. A mayor diferencia en intensidad, mayor contraste, y viceversa, a menor diferencia de intensidad, menor contraste. De este modo, el contraste local se define como la variación de intensidad en un vecindario de tamaño $w \times w$, el cual se puede medir desde dos perspectivas: diferencias entre valores de intensidad y estadísticos de dispersión.

Las medidas basadas en diferencias utilizan el máximo y mínimo locales definidos por la dilatación y erosión, respectivamente. Ambos son operadores morfológicos primarios y se pueden derivar operaciones más complejas. Ambas operaciones relacionan los píxeles de la imagen con una ventana llamada elemento estructurante, por lo que una operación morfológica sobre una imagen implica una operación de enmascaramiento con el elemento estructurante. La Figura 2.3 muestra máscaras de distintos elementos estructurantes.

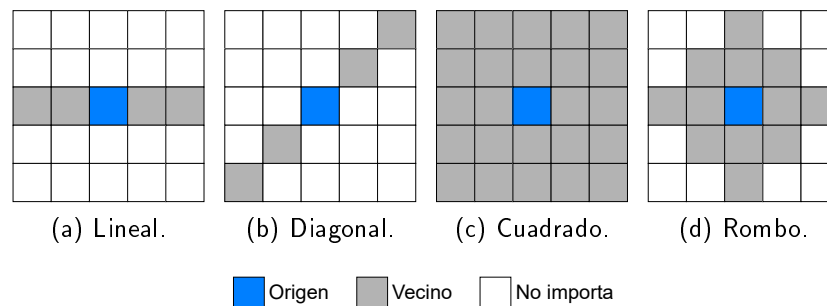


Figura 2.3: Elementos estructurantes de vecindario 5×5 .

La erosión consiste en tomar el valor mínimo de la región que enmascara el elemento estructurante

b cuando su origen está sobre el píxel $I(x, y)$:

$$[I \ominus b](x, y) = \min_{(s,t) \in b} \{I(x + s, y + t)\}. \quad (2.3)$$

Caso contrario a la erosión, la dilatación consiste en tomar el valor máximo de la región que enmascara el elemento estructurante b cuando su origen está sobre el píxel $I(x, y)$:

$$[I \oplus b](x, y) = \max_{(s,t) \in b} \{I(x - s, y - t)\}. \quad (2.4)$$

La Figura 2.4 muestra el resultado de aplicar operaciones morfológicas a la imagen BUS usando un elemento estructurante cuadrado, dando como resultado las imágenes $I_d(x, y)$ y $I_e(x, y)$ para la dilatación y erosión, respectivamente.



(a) Imagen BUS original.

(b) Dilatación.

(c) Erosión.

Figura 2.4: Operaciones morfológicas sobre una imagen BUS para el cálculo de mínimos y máximos locales definidos por la erosión y dilatación, respectivamente.

Con base en estos operadores morfológicos se calculan los gradientes morfológicos, los cuales detectan los cambios locales de intensidad que describen el contraste local [55]. Se pueden extraer tres tipos de gradientes morfológicos:

- Gradiente interno:

$$I_{\nabla_i}(x, y) = I(x, y) - (I(x, y) \ominus b). \quad (2.5)$$

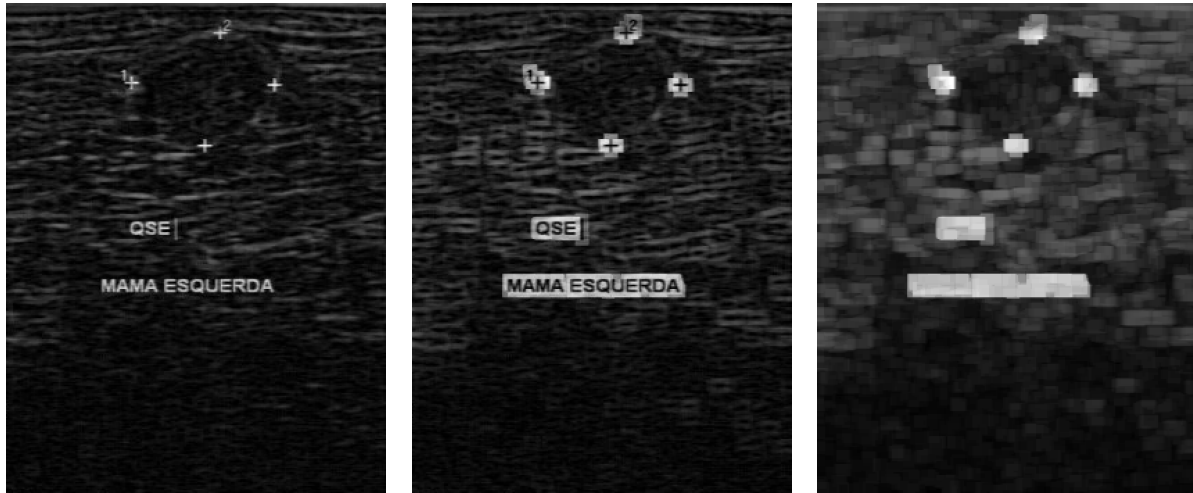
- Gradiente externo:

$$I_{\nabla_e}(x, y) = (I(x, y) \oplus b) - I(x, y). \quad (2.6)$$

- Gradiente de Beucher:

$$I_{\nabla_b}(x, y) = (I(x, y) \oplus b) - (I(x, y) \ominus b). \quad (2.7)$$

La Figura 2.5 muestra el resultado de aplicar los gradientes morfológicos a una imagen BUS.



(a) Gradiente interno.

(b) Gradiente externo.

(c) Gradiente de Beucher.

Figura 2.5: Operaciones de gradientes morfológicos sobre una imagen BUS para calcular el contraste local usando un elemento estructurante cuadrado de tamaño 3×3 .

Dada una imagen $I(x, y)$, la media local obtiene una versión suavizada $I_\mu(x, y)$, donde en el punto (x, y) se calcula la media de intensidad de sus vecinos en una ventana de tamaño $w \times w$, lo cual se calcula como:

$$I_\mu(x, y) = I(x, y) * \mu(x, y), \quad (2.8)$$

donde $\mu(x, y)$ es el *kernel* de convolución:

$$\mu(x, y) = \frac{1}{w^2} \mathbb{1}_w, \quad (2.9)$$

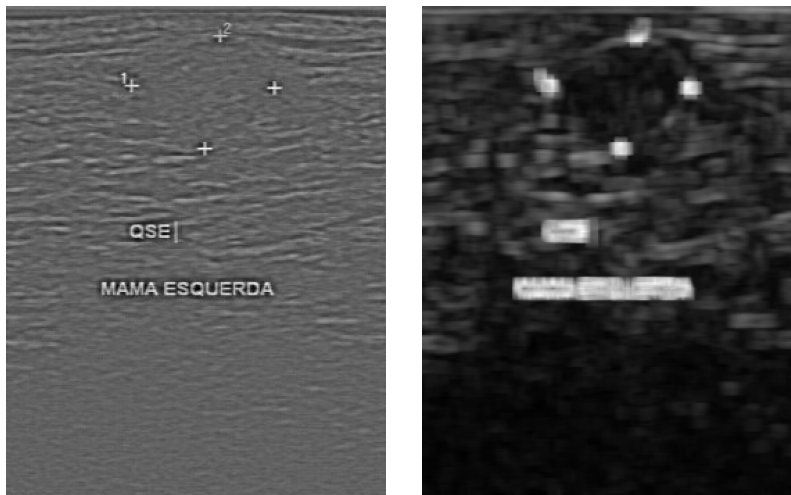
donde $\mathbb{1}_w$ es una matriz cuadrada de unos de tamaño $w \times w$. A partir de la imagen de medias locales, se calcula el contraste local como:

$$I_{\nabla\mu} = I(x, y) - I_{\mu}(x, y). \quad (2.10)$$

También, la desviación estándar local mide el contraste en el punto (x, y) , la cual se calcula como:

$$I_{\sigma}(x, y) = [I_{\nabla\mu}(x, y) * \sigma(x, y)]^{1/2}, \quad (2.11)$$

donde $\sigma(x, y) = \frac{1}{w^2-1} \mathbb{1}_w$ es un *kernel* de convolución de tamaño $w \times w$. La Figura 2.6 muestra las imágenes de contraste local con base en la media local y desviación estándar local, considerando un tamaño $w = 3$.



(a) Diferencia con la media local. (b) Desviación estándar local.

Figura 2.6: Características de contraste basadas en estadísticos.

Con base en las medidas de contraste anteriores, el vector de características basadas en contraste

queda como:

$$\mathbf{x}_{\text{con}} = [I_{\nabla_e}(x, y), I_{\nabla_i}(x, y), I_{\nabla_b}(x, y), I_{\nabla_\mu}(x, y), I_\sigma(x, y)]^T. \quad (2.12)$$

2.2.2.2. Filtros de Log-Gabor

Los filtros de Gabor han sido utilizados ampliamente para extraer información de textura de las imágenes debido a que emulan las características psicofísicas de los campos receptivos simples del sistema visual humano [43].

Los filtros log-Gabor, propuestos por Field en 1987, son una extensión mejorada de los filtros de Gabor [43]. Estos filtros son definidos en el dominio de la frecuencia como funciones Gaussianas desplazadas desde el origen. Presentan una componente de corriente continua nula y pueden producir filtros con una extensión espacial mínima en un esquema multiresolución con escalas en octavas. Los filtros log-Gabor se generan mediante el producto de dos componentes: filtros radiales (\hat{G}_ρ) y filtros angulares (\hat{G}_θ), de la siguiente forma:

$$\hat{G}(\rho, \theta) = \exp\left(-\frac{1}{2}\left(\frac{\log\left(\frac{\rho}{u_0}\right)}{\log\left(\frac{\alpha\rho}{u_0}\right)}\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{\theta - \theta_0}{\alpha_\theta}\right)^2\right) \quad (2.13)$$

donde (ρ, θ) son coordenadas polares, u_0 es la frecuencia central, θ_0 es el ángulo de orientación, α_ρ y α_θ determinan la escala y el ancho de banda angular, respectivamente.

El filtrado de la imagen $I(x, y)$ se realiza en el dominio de la frecuencia como:

$$I_{\rho\theta}(x, y) = \left| \mathcal{F}^{-1} \left\{ \mathcal{F}(I_c(x, y)) \cdot \hat{G}(\rho, \theta) \right\} \right| \quad (2.14)$$

donde \mathcal{F} y \mathcal{F}^{-1} es la transformada de Fourier y su inversa, respectivamente, y $|\cdot|$ denota la magnitud.

La Figura 2.7 muestra un ejemplo de los filtros que se generan a partir de distintos valores para ρ y θ , y la Figura 2.8 muestra el resultado de aplicar el banco de filtros de log-Gabor sobre una imagen BUS considerando cuatro orientaciones θ y cuatro escalas ρ para generar un vector de características

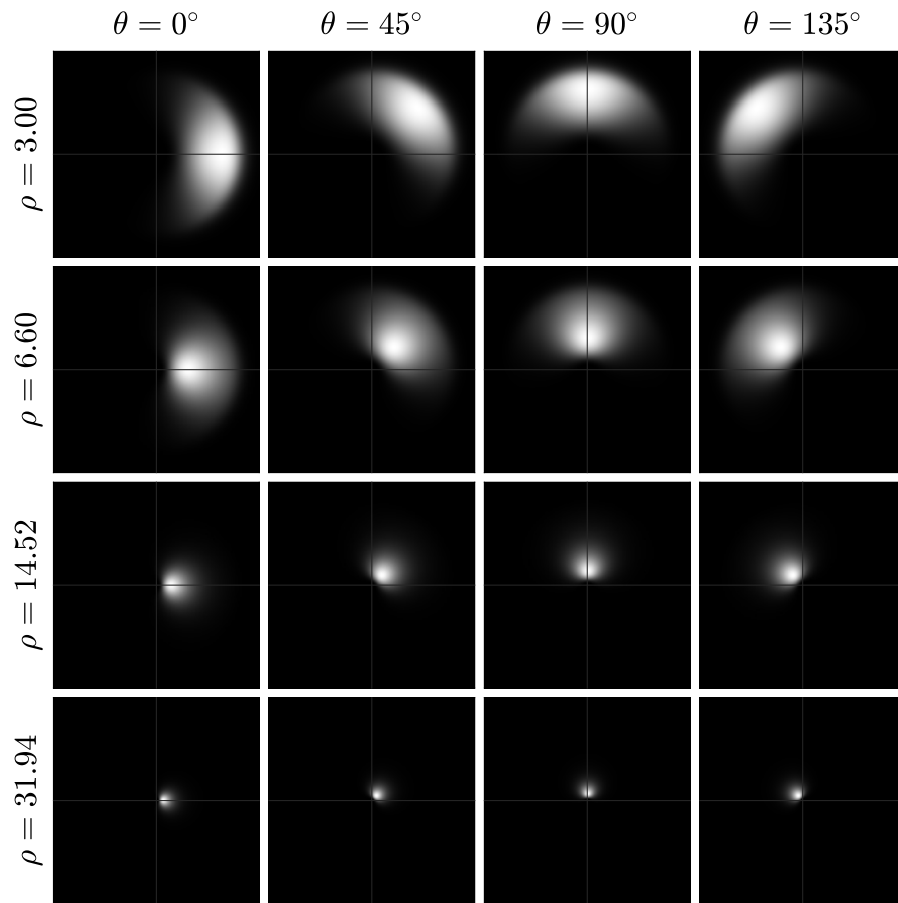


Figura 2.7: Banco de filtros de log-Gabor en el dominio de la frecuencia con diferentes valores de ρ y θ .

con dieciséis elementos. De este modo, las imágenes obtenidas forman el conjunto de características de log-Gabor, considerando ρ_n escalas y θ_n orientaciones como:

$$\mathbf{x}_{\text{gab}} = [I_{\rho_1, \theta_1}(x, y), I_{\rho_1, \theta_2}(x, y), \dots, I_{\rho_n, \theta_n}(x, y)]^T. \quad (2.15)$$

2.2.2.3. Momentos de Hu

Los momentos invariantes de Hu son comúnmente utilizados en reconocimiento de objetos, los cuales realizan una descripción invariante a las traslaciones, escalamientos y rotaciones del objeto.

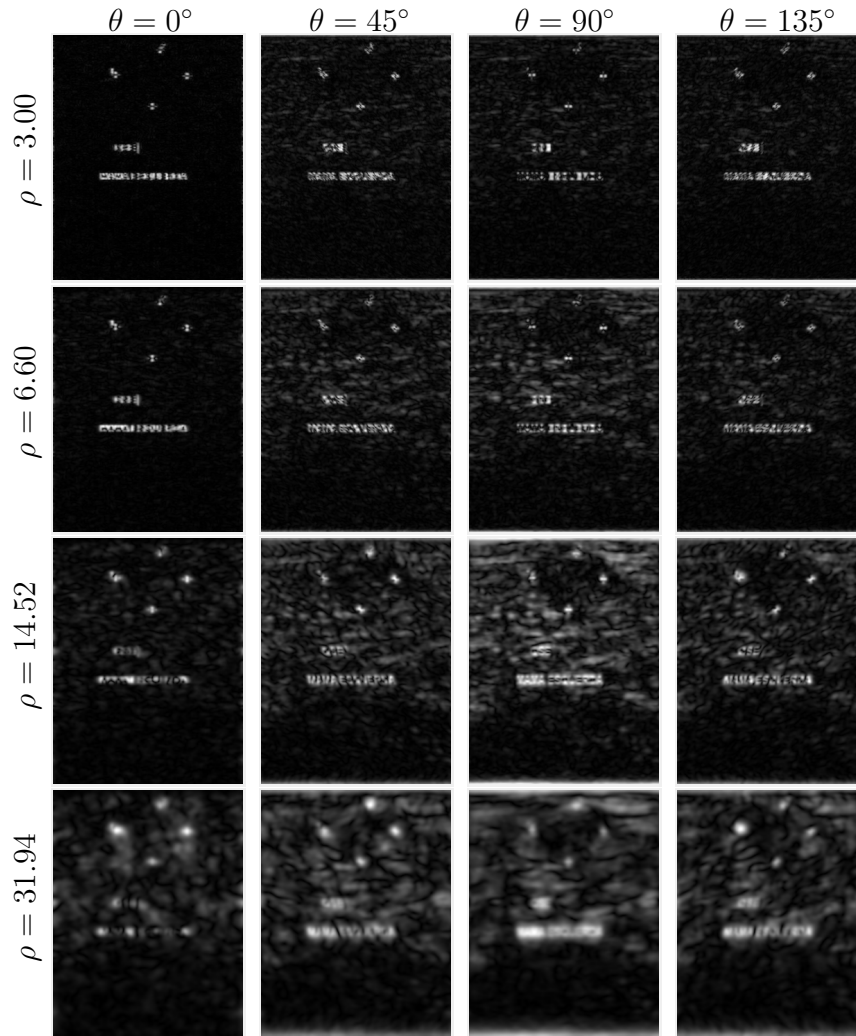


Figura 2.8: Características de log-Gabor a partir del banco de filtros en la Figura 2.7.

Estos momentos consisten de siete invariantes denotados como ϕ_1, \dots, ϕ_7 . Los momentos de Hu de una imagen $I(x, y)$ son obtenidos a partir de los momentos geométricos de orden $p + q$, los cuales, para una imagen de tamaño $M \times N$ se definen como [3]:

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p y^q f(x, y), \quad (2.16)$$

donde $p = 0, 1, 2, \dots$ y $q = 0, 1, 2, \dots$ son enteros positivos. El centro de masa (\bar{x}, \bar{y}) del objeto se calcula como:

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \text{y} \quad \bar{y} = \frac{m_{01}}{m_{00}}. \quad (2.17)$$

A partir de los momentos geométricos, los momentos centrales hasta de orden tres se calculan como:

$$\mu_{00} = m_{00}, \quad (2.18)$$

$$\mu_{10} = \mu_{01} = 0, \quad (2.19)$$

$$\mu_{20} = m_{20} - \bar{x}m_{10}, \quad (2.20)$$

$$\mu_{02} = m_{02} - \bar{y}m_{01}, \quad (2.21)$$

$$\mu_{11} = m_{11} - \bar{x}m_{01} = m_{11} - \bar{y}m_{10}, \quad (2.22)$$

$$\mu_{21} = m_{21} - 2\bar{x}m_{11} - \bar{y}m_{20} + 2m_{01}\bar{x}^2, \quad (2.23)$$

$$\mu_{12} = m_{12} - 2\bar{y}m_{11} - \bar{x}m_{02} + 2m_{10}\bar{y}^2, \quad (2.24)$$

$$\mu_{30} = m_{30} - 3\bar{x}m_{20} + 2m_{10}\bar{x}^2, \quad (2.25)$$

$$\mu_{03} = m_{03} - 3\bar{y}m_{02} + 2m_{01}\bar{y}^2. \quad (2.26)$$

Los momentos centrales normalizados de orden $p+q$ son invariantes a traslaciones y escalamientos de la imagen, y se calculan a partir de los momentos centrales como:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad (2.27)$$

donde

$$\gamma = \frac{p+q}{2}, \quad \text{donde } (p+q) \geq 2. \quad (2.28)$$

Finalmente, a partir de los momentos centrales normalizados se calculan los siete momentos de Hu, los cuales son invariantes a rotaciones:

$$\phi_1 = \eta_{20} + \eta_{02}, \quad (2.29)$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \quad (2.30)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2, \quad (2.31)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2, \quad (2.32)$$

$$\begin{aligned} \phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2], \end{aligned} \quad (2.33)$$

$$\phi_6 = (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}), \quad (2.34)$$

$$\begin{aligned} \phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + \\ (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]. \end{aligned} \quad (2.35)$$

La Figura 2.9 muestra las características extraídas sobre la imagen BUS al aplicar los momentos de Hu. Cabe mencionar que los momentos de Hu se calculan en un vecindario de tamaño $w \times w$ para extraer información geométrica local. De tal forma que el conjunto de características de Hu queda como:

$$\mathbf{x}_{\text{hu}} = [I_{\phi_1}(x, y), I_{\phi_2}(x, y), I_{\phi_3}(x, y), I_{\phi_4}(x, y), I_{\phi_5}(x, y), I_{\phi_6}(x, y), I_{\phi_7}(x, y)]^T. \quad (2.36)$$

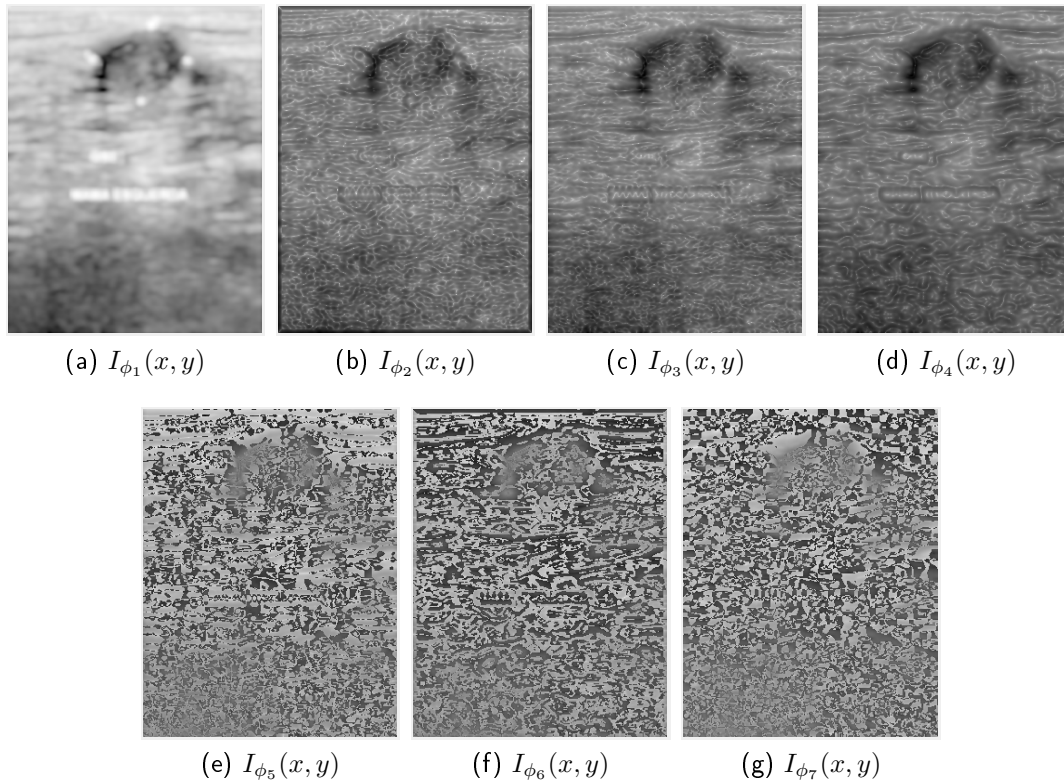


Figura 2.9: Características de momentos de Hu.

2.2.2.4. Momentos de Zernike

Los momentos de Zernike han sido utilizados como una herramienta del proceso de reconocimiento de patrones para la extracción de características, los cuales han demostrado un buen desempeño en imágenes digitales como satelitales, ultrasonido, microscópicas, entre otros. Estos momentos utilizan como base los polinomios del mismo nombre, desarrollados por Frits Zernike en 1934. Las principales características de los momentos de Zernike son: robustez al ruido, invariancia a la rotación y fácil reconstrucción [15].

Los momentos de Zernike son un conjunto de polinomios ortogonales definidos en un círculo unitario, expresados como:

$$V_n^l(x, y) = V_n^l(\rho \sin \theta, \rho \cos \theta) = R_n^l(\rho) e^{il\theta}, \quad (2.37)$$

donde $R_n^l(\rho)$ es la parte radial, $e^{il\theta}$ es la parte angular, n es el grado del polinomio y l es el parámetro de dependencia angular. La coordenada ρ es la distancia radial normalizada y θ es el ángulo desde el eje y , en donde l y n son enteros, $l \geq 0$, $l < 0$ y $n \geq 0$. Los números n y l están sujetos a las siguientes restricciones: $n \geq |l|$ y $n - |l|$ siempre es un número par. Los polinomios radiales están expresados por:

$$R_n^l(\rho) = \sum_{s=0}^{\frac{n-|l|}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|l|}{2} - s\right)! \left(\frac{n-|l|}{2} - s\right)!} \rho^{n-2s}. \quad (2.38)$$

La Figura 2.10 muestra un ejemplo de las imágenes extraídas a partir de una imagen BUS utilizando los momentos de Zernike. Cabe mencionar que los momentos de Zernike se calculan en vecindarios de $w \times w$ para extraer la textura local. Considerando $n \in \{3, 4, 5\}$ y $l \in \{0, \dots, 5\}$, se genera un conjunto de características de ocho elementos. De este modo el vector queda como:

$$\mathbf{x}_{\text{zrk}} = \left[I_{V_3^1}(x, y), I_{V_3^3}(x, y), \dots, I_{V_5^5}(x, y) \right]^T. \quad (2.39)$$

2.2.2.5. Transformada de Hermite

La transformada de Hermite es un modelo de representación que incorpora algunas propiedades de percepción visual como modelar los campos receptivos presentes en el ojo humano [40]. Esta transformada es una herramienta eficiente para la detección y extracción de características en la imagen como bordes y líneas, utilizando ventanas Gaussianas, las cuales, al ser isotrópicas, se puede escribir que $\sigma = \sigma_x = \sigma_y$:

$$W(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{\sigma^2}}. \quad (2.40)$$

Los polinomios de Hermite se calculan como:

$$G_{n-m,m}(x, y) = \frac{1}{\sqrt{2^n(n-m)!m!}} H_{n-m}\left(\frac{x}{\sigma}\right) H_m\left(\frac{y}{\sigma}\right), \quad (2.41)$$

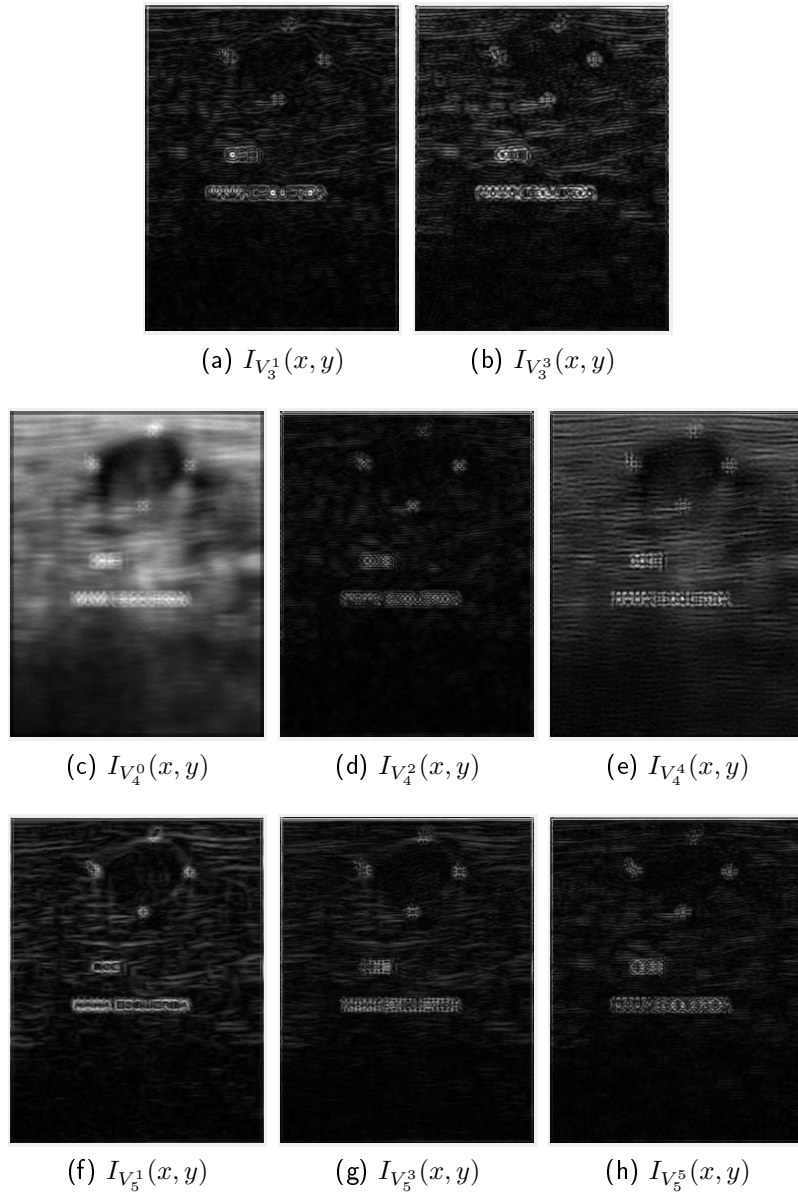


Figura 2.10: Polinomios de Zernike para $n \in \{3, 4, 5\}$ y $l \in \{0, \dots, 5\}$.

donde H_{n-m} y H_m son los polinomios de Hermite de orden $n - m$ y m , respectivamente. La forma matemática de los polinomios de Hermite está dada por [1]:

$$H_n(x) = (-1)^n \exp(x^2) \frac{d^n \exp(-x^2)}{dx^n}, \quad \text{para } n = 0, 1, 2, \dots, \quad (2.42)$$

donde el operador $\frac{d^n}{dx^n}$ denota la n -ésima derivada de la función. A partir de las Ecuaciones 2.39 y 2.40, el banco de filtros de Hermite se define como:

$$f_{H_{n-m,m}} = \frac{G_{n-m,m}(x, y) \cdot W(x, y)}{\sum_{i,j} W(x, y)}. \quad (2.43)$$

Finalmente, las características de Hermite se obtienen mediante la convolución de la imagen $I(x, y)$ con el banco de filtros $f_{H_{n-m,m}}$:

$$I_{n-m,m}(x, y) = I(x, y) * f_{H_{n-m,m}}(x, y), \quad (2.44)$$

donde $n \in \{0, 1, 2, 3\}$ y $m \in \{0, \dots, n\}$. La Figura 2.11 muestra un ejemplo de las imágenes generadas a partir de la transformada de Hermite. Además, el vector de características queda como:

$$\mathbf{x}_{\text{her}} = [I_{0,1}(x, y), I_{1,0}(x, y), \dots, I_{n-m,m}(x, y)]^T. \quad (2.45)$$

2.2.2.6. Ranklets

La transformada *Ranklet* es un análisis no paramétrico, con orientación selectiva y multiresolución, para caracterizar texturas de una imagen de entrada $I(x, y)$. Esta transformada produce un conjunto de imágenes definidas como *ranklets*, donde una máscara *ranklet* está definida por los parámetros de resolución y orientación [41]. De esta forma, el número total de imágenes *ranklet* está dado por la multiplicación del número de resoluciones y el número de orientaciones. Para el análisis multiresolución, se debe definir un conjunto de valores r , comúnmente en potencias de 2.

Por otro lado, el análisis de orientación selectiva se realiza dividiendo los píxeles dentro de una máscara *ranklet* en dos subconjuntos: tratamiento (\mathcal{T}) y control (\mathcal{C}). De acuerdo con la orientación se nombran los subconjuntos: \mathcal{T}_v y \mathcal{C}_v para la orientación vertical, \mathcal{T}_h y \mathcal{C}_h para la orientación horizontal,

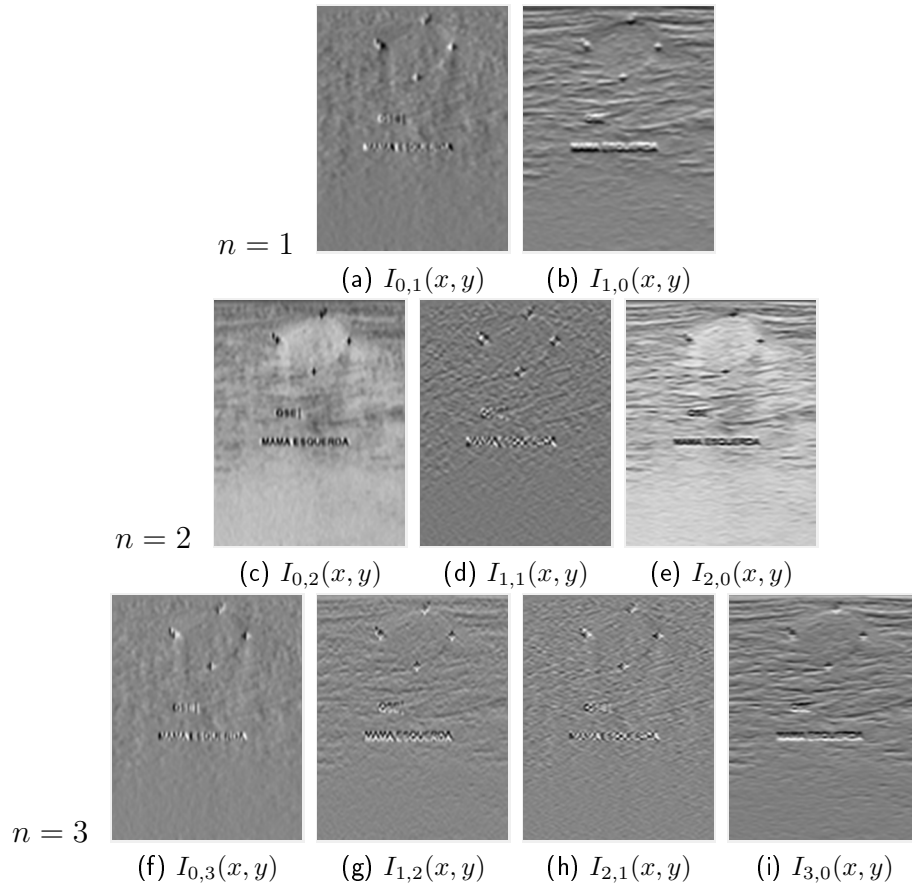


Figura 2.11: Descomposición de una imagen BUS con la transformada de Hermite con distintos valores de orden n .

y \mathcal{T}_d y \mathcal{C}_d para la orientación diagonal. La Figura 2.12 muestra un ejemplo del concepto de resolución y orientación para una máscara ranklet con $r = 4$. El análisis no paramétrico se realiza en el vecindario cubierto por la máscara *ranklet*, de modo que los coeficientes de *ranklet* se calculan como:

$$R_j = \frac{W^j}{\frac{n}{2}} - 1, \quad (2.46)$$

donde $j \in \{h, v, d\}$ es la orientación y

$$W = \sum_{p \in \mathcal{T}} \pi(p) - \frac{n(n+1)}{2}, \quad (2.47)$$

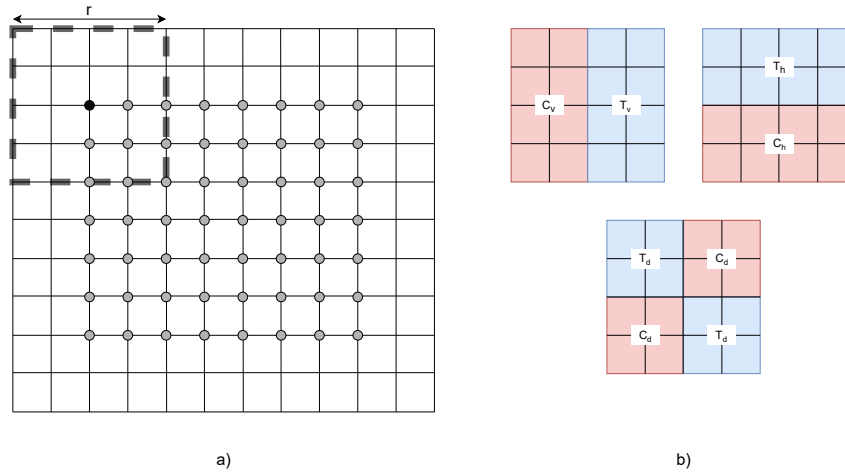


Figura 2.12: (a) Para una imagen de tamaño 11×11 , la máscara *ranklet* con resolución $r = 4$ genera 49 ventanas traslapadas centradas en los puntos grises. (b) Orientaciones de la máscara *ranklet* en: horizontal (*h*), vertical (*v*) y diagonal (*d*), con sus correspondiente subconjuntos de tratamiento \mathcal{T} y control \mathcal{C} .

donde $\pi(p)$ es el valor de rango del píxel p que pertenece al subconjunto \mathcal{T} , y $n = r^2$ es el número de píxeles en el vecindario. El valor del coeficiente de *ranklet* se encuentra en el rango $[-1, 1]$. Posteriormente, se promedian las tres orientaciones *ranklet* a una misma resolución, de este modo se obtiene:

$$I_r(x, y) = \frac{I_{hr}(x, y) + I_{vr}(x, y) + I_{dr}(x, y)}{3}. \quad (2.48)$$

El vector de características se define para las resoluciones $r = [2, 4, 8, 16]$ como:

$$\mathbf{x}_{\text{rnk}} = [I_{r_2}(x, y), I_{r_4}(x, y), I_{r_8}(x, y), I_{r_{16}}(x, y)]^T. \quad (2.49)$$

2.2.2.7. Auto-Información mutua

La auto-información mutua mide el grado de similitud entre desplazamientos sucesivos de las imágenes *ranklet* en las direcciones horizontal y vertical [21]. Debido a que la información mutua se calcula sobre variables discretas, es necesario cuantizar cada coeficiente *ranklet* en el rango $[1, Q]$

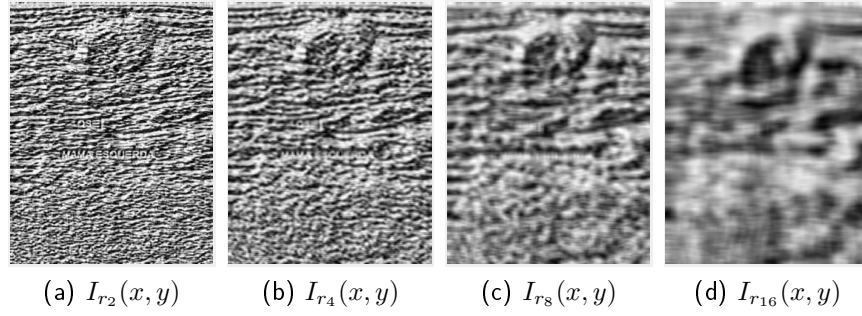


Figura 2.13: Características de textura basadas en *ranklet* con cuatro resoluciones.

mediante:

$$R'_j = \lfloor R_j K + K + 1 \rfloor, \quad (2.50)$$

donde $\lfloor \cdot \rfloor$ representa la función piso, $K = (Q - 1)/2$, y Q es el número de niveles de cuantización.

Dada una imagen *ranklet* cuantizada $I_{\text{rnk}}(x, y)$ de tamaño $M \times N$ píxeles (con $Q = 256$), la auto-información mutua en la dirección horizontal se define como:

$$A_x = \sum_{i=1}^{M-1} \tilde{I}(I_{\text{rnk}}(k_l, y); I_{\text{rnk}}(k_r, y)), \forall y : 1 \leq y \leq N, \quad (2.51)$$

donde $k_l = 1, 2, \dots, M - i$ y $k_r = i + 1, i + 2, \dots, M$ indica los desplazamientos izquierda y derecha, respectivamente, mientras que para la dirección vertical se define como:

$$A_y = \sum_{i=1}^{N-1} \tilde{I}(I_{\text{rnk}}(x, k_u); I_{\text{rnk}}(x, k_d)), \forall x : 1 \leq x \leq M, \quad (2.52)$$

donde $k_u = 1, 2, \dots, N - i$ y $k_d = i + 1, i + 2, \dots, N$ indica los desplazamientos arriba y abajo, respectivamente. Para calcular A_x y A_y , la información mutua normalizada se calcula como:

$$\tilde{I}(X; Y) = \frac{I(X; Y)}{\sqrt{H(X)H(Y)}}, \quad (2.53)$$

donde $H(X)$ y $H(Y)$ representan la entropía para las variables discretas X y Y , e $I(X; Y)$ es la

información mutua entre dos variables X y Y definida como:

$$I(X;Y) = H(X) - H(X | Y), \quad (2.54)$$

donde $H(X | Y)$ es la entropía conjunta. Las características extraídas se calculan de manera local en vecindarios de $w \times w$, generando un vector de ocho elementos a partir de las cuatro resoluciones de ranklets:

$$\mathbf{x}_{\text{ami}} = [I_{Ax_2}(x, y), I_{Ay_2}(x, y), I_{Ax_4}(x, y), I_{Ay_4}(x, y), I_{Ax_8}(x, y), I_{Ay_8}(x, y), I_{Ax_{16}}(x, y), I_{Ay_{16}}(x, y)]. \quad (2.55)$$

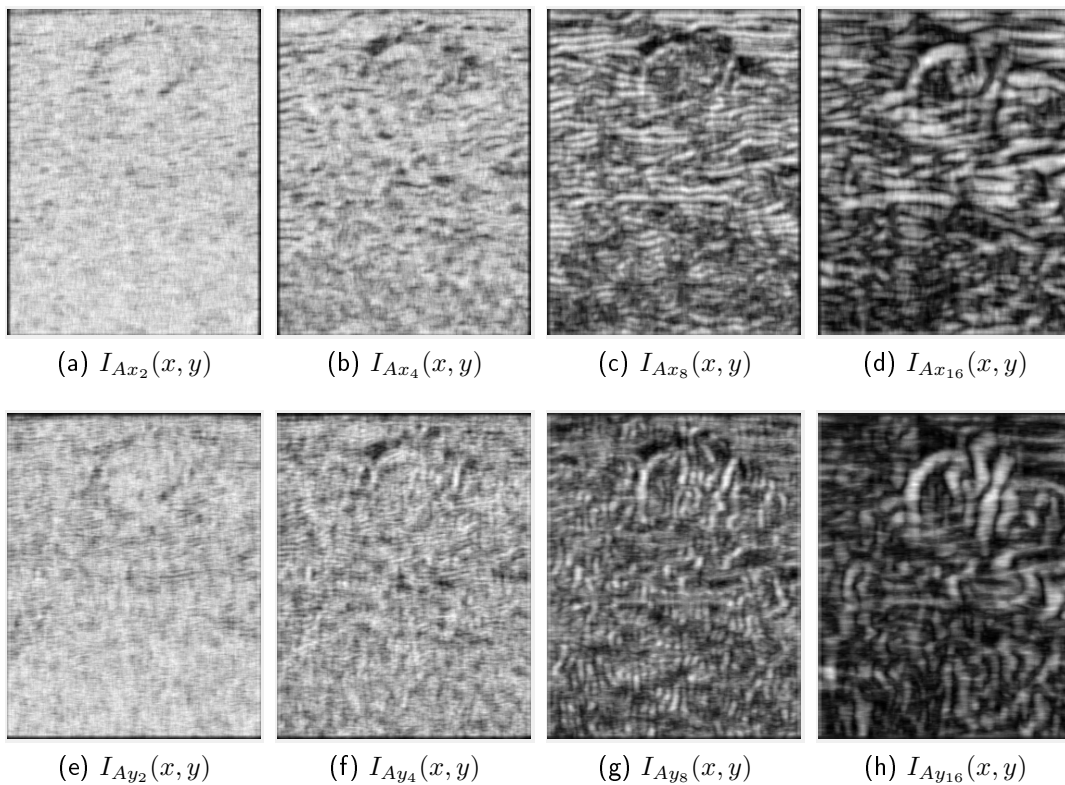


Figura 2.14: Características de textura basadas en auto-información mutua.

2.2.2.8. Matriz de co-ocurrencia de niveles de gris

A partir de una imagen $I(x, y)$ con L niveles de intensidad, se obtienen características basadas en la matriz de co-ocurrencia de niveles de gris (GLCM, por sus siglas en inglés) [65], la cual registra la frecuencia relativa entre dos niveles de gris p y q a una distancia d y una orientación θ :

$$c(i, j) = || [(x_1, y_1), (x_2, y_2)] \mid x_2 - x_1 = d \cos \theta, y_2 - y_1 = d \sin \theta, I(x_1, y_1) = i, I(x_2, y_2) = j ||, \quad (2.56)$$

donde $I(x_1, y_1)$ y $I(x_2, y_2)$ son dos píxeles, y $|| \cdot ||$ denota el número de pares de píxeles que satisfacen la condición. Además, la imagen $I(x, y)$ se cuantifica a Q niveles de intensidad, con $Q < L$:

$$g(x, y) = \lfloor I(x, y) \cdot K_1 + K_2 \rfloor, \quad (2.57)$$

donde $\lfloor \cdot \rfloor$ representa la función piso y $K_1 = Q / (I_{\text{máx}} - I_{\text{mín}})$ y $K_2 = 1 - K_1 \cdot I_{\text{mín}}$. Las frecuencias relativas en la GLCM se normalizan dividiendo todas las entradas de la matriz por la suma de todos sus elementos para obtener la probabilidad relativa:

$$p(i, j) = \frac{c(i, j)}{\sum_{s=1}^Q \sum_{t=1}^Q c(s, t)}, \quad (2.58)$$

tal que

$$\sum_{i=1}^Q \sum_{j=1}^Q p(i, j) = 1. \quad (2.59)$$

A partir de la GLCM se pueden extraer rasgos de textura, los cuales requieren calcular los estadísticos $\mu_i = \sum_{i,j} i \cdot p(i, j)$, $\mu_j = \sum_{i,j} j \cdot p(i, j)$, $\sigma_i^2 = \sum_{i,j} (i - \mu_i)^2 p(i, j)$ y $\sigma_j^2 = \sum_{i,j} (j - \mu_j)^2 p(i, j)$. En la Tabla 2.1 se resumen las siete características de textura basadas en GLCM usadas en esta investigación, las cuales se calculan en vecindarios de tamaño $w \times w$ para obtener texturas locales y la Figura 2.15 muestra dichas características. De este modo, el conjunto de características se define

Característica	Ecuación
Energía	$\sum_{i,j} p(i,j)^2$
Entropía	$-\sum_{i,j} p(i,j) \log_2(p(i,j))$
Correlación	$\sum_{i,j} \frac{(i \cdot j)p(i,j) - \mu_i \mu_j}{\sigma_i \sigma_j}$
Contraste	$\sum_{i,j} i - j ^2 p(i,j)$
Varianza	$\sum_{i,j} (i - \mu_{i,j})^2 p(i,j)$
Homogeneidad	$\sum_{i,j} \frac{p(i,j)}{1+ i-j ^2}$
Máxima Probabilidad	$\max_{i,j} p(i,j)$

Tabla 2.1: Características de textura basadas en GLCM.

como:

$$\mathbf{x}_{\text{glcm}} = [I_{\text{egi}}(x, y), I_{\text{ent}}(x, y), I_{\text{cor}}(x, y), I_{\text{con}}(x, y), I_{\text{var}}(x, y), I_{\text{hom}}(x, y), I_{\text{max}}(x, y)]^T. \quad (2.60)$$

2.3 Reducción de dimensionalidad

En el campo de reconocimiento de patrones, generalmente debe enfrentarse el problema de alta dimensionalidad del espacio de características, lo cual aumenta el costo computacional del entrenamiento del clasificador y además se tiene el riesgo de sobre-entrenamiento debido a la generación de espacios dispersos. La reducción de dimensionalidad involucra identificar y eliminar variables irrelevantes, de tal forma que se mejora el rendimiento de clasificación y acelera el tiempo de entrenamiento.

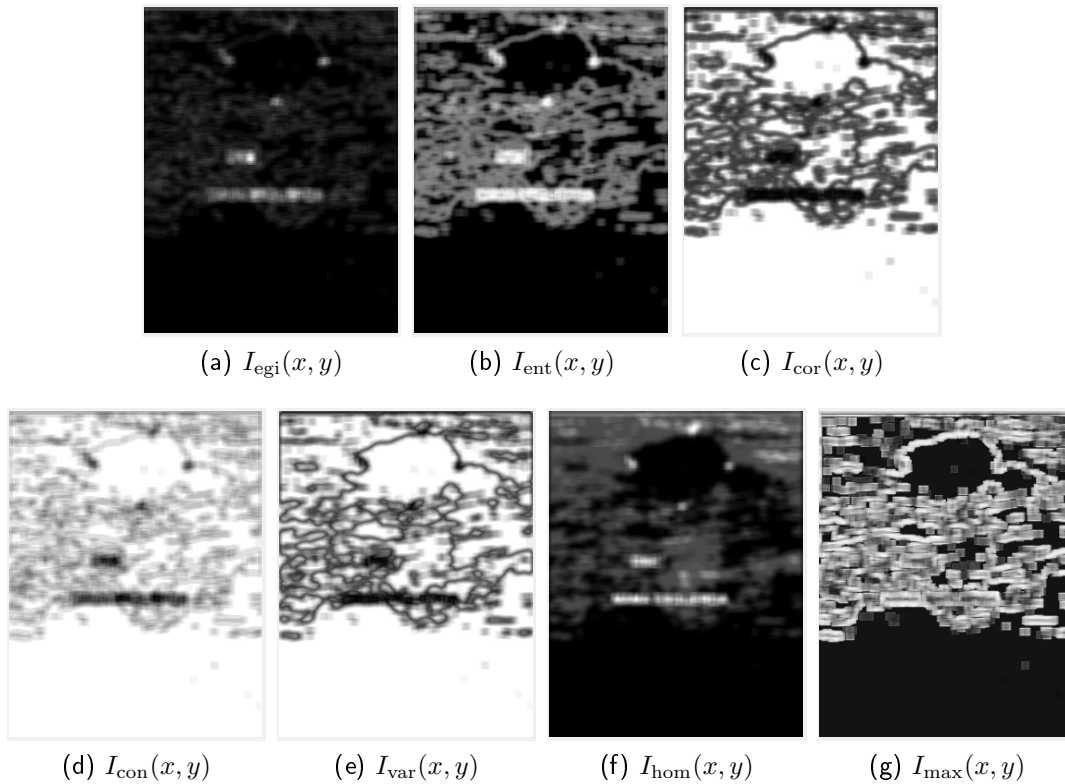


Figura 2.15: Características de textura basadas en GLCM.

2.3.1 Análisis de Componentes Principales Asimétrico

El análisis de componentes principales asimétrico (APCA, por sus siglas en inglés) es una técnica de reducción de dimensionalidad basada en el mapeo de un espacio de características de alta dimensionalidad a uno de menor dimensionalidad. APCA es apropiado para problemas de clasificación con desbalance de clases como el que se aborda en esta investigación, en donde se tiene menor cantidad de píxeles de clase tumor que de clase fondo [29].

APCA combina el análisis de componentes principales (PCA, por sus siglas en inglés) seguido de el análisis discriminante lineal (LDA, por sus siglas en inglés). PCA maximiza la capacidad de reconstrucción de la estructura intrínseca de las características, mientras que LDA maximiza el poder discriminante de las características.

Dado un espacio de características d -dimensional con n muestras, donde la clase positiva ω_+

tiene n_+ muestras y la clase ω_- tiene n_- , tal que $n = n_+ + n_-$. Se calculan los vectores de medias clase-condicionales \mathbf{m}_+ y \mathbf{m}_- , así como las matrices de covarianza Σ_+ y Σ_- . Entonces, la matriz de covarianza promedio se calcula como:

$$\Sigma_m = \frac{1}{n} [n_+(\mathbf{m}_+ - \mathbf{m})(\mathbf{m}_+ - \mathbf{m})^T + n_-(\mathbf{m}_- - \mathbf{m})(\mathbf{m}_- - \mathbf{m})^T], \quad (2.61)$$

donde \mathbf{m} es el vector de medias global de todas las muestras. Posteriormente, la matriz de covarianza combinada se calcula como:

$$\Sigma_c = \frac{n_-}{n}\Sigma_+ + \frac{n_+}{n}\Sigma_- + \Sigma_m. \quad (2.62)$$

La matriz de pesos PCA se calcula a partir del problema de descomposición de valores y vectores propios $\Sigma_c = \Phi\Lambda\Phi^T$, extrayendo los m_{pca} vectores propios \mathbf{w}_{pca} a partir de Φ correspondientes con los m_{pca} valores propios más grandes en Λ , tal que el valor de m reconstruye el 99 % de la estructura intrínseca del espacio de características.

Para calcular la matriz de pesos LDA, primeramente las matrices de covarianza Σ_+ , Σ_- y Σ_m son mapeadas a un espacio m_{pca} -dimensional como:

$$\hat{\Sigma}_+ = \mathbf{w}_{pca}^T \Sigma_+ \mathbf{w}_{pca}, \quad (2.63)$$

$$\hat{\Sigma}_- = \mathbf{w}_{pca}^T \Sigma_- \mathbf{w}_{pca}, \quad (2.64)$$

$$\hat{\Sigma}_m = \mathbf{w}_{pca}^T \Sigma_m \mathbf{w}_{pca}. \quad (2.65)$$

Después se resuelve el problema generalizado de descomposición de valores y vectores propios $(\hat{\Sigma}_+ + \gamma\hat{\Sigma}_m) = (\hat{\Sigma}_+ + \beta\hat{\Sigma}_-)\Phi\Lambda$, para obtener los m_{lda} vectores propios a partir de Φ correspondientes con los m_{lda} valores propios más grandes en Λ , tal que el valor de m_{lda} mantiene el 99 % de la información discriminante. El parámetro $\beta = 0.95$ es un peso de asimetría entre las dos

clases, y el parámetro $\gamma = 10$ pondera la información discriminante. Finalmente la matriz de pesos APCA queda definida como:

$$\mathbf{W}_{\text{apca}} = \mathbf{W}_{\text{pca}} \mathbf{W}_{\text{lda}}. \quad (2.66)$$

Entonces un patrón de entrada $\mathbf{x} \in \mathbb{R}^d$ es mapeado a un espacio $\mathbb{R}^{m_{\text{lda}}}$ (con $m_{\text{lda}} < d$) como

$$\hat{\mathbf{x}} = \mathbf{w}_{\text{apca}}^T \mathbf{x}. \quad (2.67)$$

2.3.2 Mínima Redundancia Máxima Relevancia

El método de mínima redundancia máxima relevancia (MRMR) ordena las características de manera descendente desde la más relevante con respecto a la distribución de clases, hasta la característica menos relevante [46]. La relevancia y redundancia se cuantifican en términos de información mutua entre dos variables, definida en la Ecuación 2.53.

Dado un conjunto de características d -dimensional, denotado como X , donde cada muestra está asociada a una etiqueta de clase en el conjunto Y . La relevancia mide el nivel de dependencia entre la i -ésima característica y la variable de clase, de modo que se maximiza la condición:

$$\text{máx } D(S_m, Y), \quad D = \frac{1}{m} \sum_{X_i \in S_m} I(X_i; Y), \quad (2.68)$$

mientras que la redundancia identifica las características que son mutuamente exclusivas, de modo que se minimiza la condición:

$$\text{mín } R(S_m), \quad R = \frac{1}{m^2} \sum_{X_i, X_j \in S_m} I(X_i; X_j). \quad (2.69)$$

Entonces el criterio MRMR se obtiene combinando D y R en una sola función de costo que se maximiza con el siguiente algoritmo incremental:

$$\max_{X_i \in S_m} \left[I(X_i; Y) - \frac{1}{m} \sum_{X_j \in S} I(X_i; X_j) \right], \quad (2.70)$$

donde S_m es el subconjunto con m características que no han sido evaluadas, y S es el subconjunto con $d-m$ características ordenadas. Este proceso se repite hasta que $S_m = \emptyset$. En la primera iteración, $S = \emptyset$; por tanto, la primera característica seleccionada es aquella que maximiza la relevancia definida en la Ecuación 2.68.

Después del ordenamiento de características con la Ecuación 2.70, se realiza una búsqueda hacia adelante para seleccionar las primeras m características que minimizan el error de clasificación, tal que el número de características seleccionadas es $m < d$.

2.4 Redes Neuronales Artificiales

Las redes neuronales artificiales (ANN) son modelos matemáticos inspirados en el funcionamiento de las neuronas biológicas, cuyo propósito es mapear datos de entrada a una salida. En la Figura 2.16 se muestra la morfología de una neurona biológica. La neurona está compuesta por un cuerpo celular llamado soma, las dendritas que transmiten impulsos eléctricos provenientes de otras neuronas hacia el soma y un axón que transmite impulsos eléctricos hacia otras neuronas. El enlace electro-químico para la transmisión de estos impulsos se denomina *sinapsis*.

A partir de este modelo biológico se definieron las bases de las ANN actuales, definidas como modelos matemáticos.

2.4.1 Modelo matemático

Si se realiza la analogía entre una neurona biológica y una artificial, las dendritas introducen impulsos hacia el cuerpo de la neurona, si sobrepasa un umbral la neurona se excita y transmite un impulso eléctrico a través del axón hacia otras neuronas. Este modelo de la neurona fue nombrado

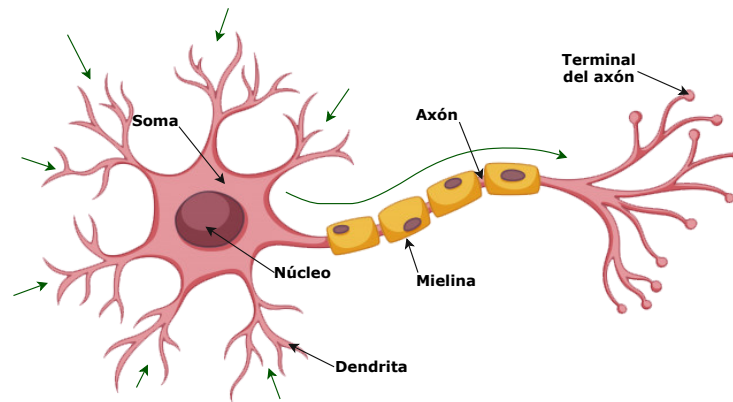


Figura 2.16: Morfología de una neurona biológica.

como modelo *perceptrón*. La salida de un perceptrón para un patrón de entrada $\mathbf{x} = [x_1, \dots, x_d]$ está definido como:

$$g(\mathbf{x}) = \varphi \left(w_0 + \sum_{i=0}^d w_i x_i \right), \quad (2.71)$$

donde w_0 es el bias, w_i es el i -ésimo peso sináptico para la i -ésima neurona, y

$$\varphi(u) = \begin{cases} +1 & \text{si } u > 0 \\ -1 & \text{otro caso} \end{cases} \quad (2.72)$$

es una función de activación de tipo escalón, donde u representa la sumatoria ponderada en 2.71.

En la Figura 2.17 se observa la analogía entre los modelos biológico y artificial.

2.4.2 Descenso de gradiente estocástico

El descenso de gradiente es uno de los algoritmos de optimización más populares en disciplinas como aprendizaje para máquinas, procesamiento adaptativo de señales, optimización de parámetros de modelos, entre otras, particularmente por su uso extensivo en el campo de las redes neuronales con el propósito de minimizar una función de costo $J(\mathbf{w})$ dado un vector de parámetros $\mathbf{w} = [w_1, \dots, w_d]$.

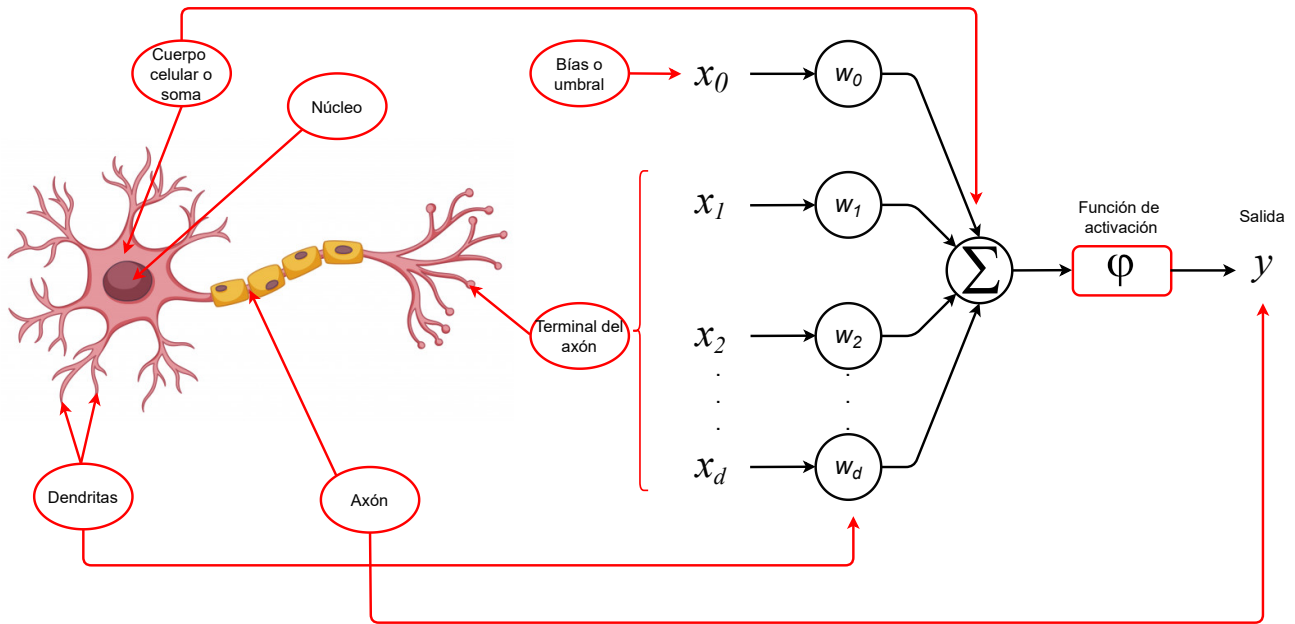


Figura 2.17: Analogía de una neurona biológica y una artificial.

El gradiente es un vector compuesto con las d derivadas parciales de la función J , definido como:

$$\nabla J(\mathbf{x}) = \left[\frac{\partial J(\mathbf{x})}{\partial w_1}, \dots, \frac{\partial J(\mathbf{x})}{\partial w_d} \right]^T. \quad (2.73)$$

El algoritmo de descenso de gradiente busca generar una secuencia de vectores que converja a la solución que minimiza $J(\mathbf{w})$, definidos como:

$$\mathbf{w}(1), \quad \text{arbitrario}, \quad (2.74)$$

$$\mathbf{w}(k+1) = \mathbf{w}(k) - \eta(k) \nabla J(\mathbf{w}(k)), \quad k \geq 1 \quad (2.75)$$

donde η es un factor de escala positivo llamado tasa de aprendizaje. En el campo de ANNs, al vector de parámetros \mathbf{w} se le conoce como vector de pesos sinápticos. Existe una versión llamada descenso de gradiente estocástico, donde el gradiente se actualiza a partir de un subconjunto de muestras tomadas aleatoriamente (denominado *mini-batch*) con el objetivo de introducir ruido al sistema para sacar a la solución de óptimos locales y agilizar el tiempo de entrenamiento.

2.4.3 Tasa de aprendizaje y momentum

El valor de la tasa de aprendizaje (η) controla el tamaño del cambio de los pesos en cada iteración. Se deben evitar dos extremos: un ritmo de aprendizaje demasiado pequeño puede ocasionar una disminución en la velocidad de convergencia y la posibilidad de acabar atrapado en un mínimo local; en cambio, un ritmo de aprendizaje demasiado grande puede conducir a inestabilidades en la función de error, lo cual evitará que se produzca la convergencia debido a que se darán saltos en torno al mínimo global sin alcanzarlo.

El factor de *momentum* le da un empuje constante para que la solución siempre avance hacia el óptimo global. Además, incluye una fracción α de la actualización del peso en una época anterior:

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \alpha \mathbf{w}(t-1) \quad (2.76)$$

Este valor α suele tomar un valor próximo a la unidad (por ejemplo, 0.9). *Nesterov momentum* es una versión del *momentum* en donde los incrementos disminuyen cuando el algoritmo se acerca al óptimo global y se define como:

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \alpha \cdot (\alpha \mathbf{w}(t-1) + \Delta \mathbf{w}) + \Delta \mathbf{w} \quad (2.77)$$

Para mejorar la convergencia, se pueden usar técnicas de reinicio y decaimiento en la tasa de aprendizaje para el descenso de gradiente estocástico [37]:

$$\eta_t = \eta_{min} + \frac{1}{2} (\eta_{max} - \eta_{min}) \left(1 + \cos \left(\frac{T_{cur} \pi}{T_i} \right) \right), \quad (2.78)$$

donde η_{min} y η_{max} son los rangos de la tasa de aprendizaje, T_{cur} indica cuántas épocas se han realizado desde el último reinicio y T_i es el número de épocas que deben pasar para actualizar η_{min}

y η_{max} :

$$\eta_{min} = \lambda \eta_{min} \quad (2.79)$$

$$\eta_{max} = \lambda \eta_{max} \quad (2.80)$$

donde λ es el decaimiento de la tasa de aprendizaje. La Figura 2.18 muestra cómo en cada reinicio de la tasa de aprendizaje se tiene un decaimiento.

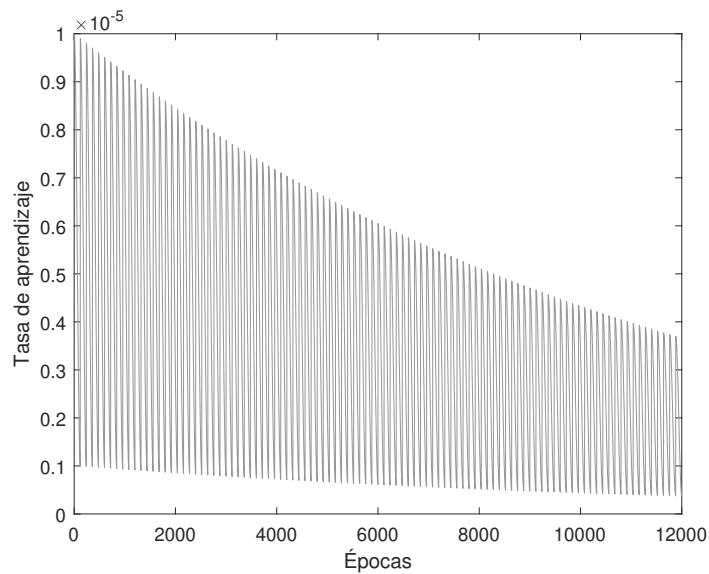


Figura 2.18: Decaimiento de la tasa de aprendizaje de acuerdo con la Ecuación 2.77.

2.4.4 Funciones de activación

Las funciones de activación acotan la salida de las neuronas, determinando el grado de activación dependiendo de si el patrón de entrada de cada neurona es relevante para la predicción del modelo. Las funciones de activación también ayudan a normalizar la salida de cada neurona.

2.4.4.1. Leaky ReLU

La función Leaky ReLU es una variante de la función ReLU (*rectified linear unit*), la cual transforma los valores introducidos multiplicando los negativos por un coeficiente rectificativo α y dejando los positivos según entran:

$$f(x) = \begin{cases} \alpha x, & \text{para } x < 0 \\ x, & \text{para } x \geq 0 \end{cases} \quad (2.81)$$

y su derivada es:

$$f'(x) = \begin{cases} \alpha, & \text{para } x < 0 \\ 1, & \text{para } x \geq 0 \end{cases} \quad (2.82)$$

Normalmente en las ANN se consideran la función sigmoïdal como función de activación y que se define como:

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \quad (2.83)$$

A diferencia de la función Leaky ReLU, la función Sigmoïdal tiene el inconveniente de desvanecimiento del gradiente cuando se aproxima a los extremos. La función Leaky ReLU permite que el gradiente no se estanque para valores positivos y tendrá un avance pequeño para valores negativos [45]. La Figura 2.19 muestra que la función Leaky ReLU no se estanca en los extremos y conserva valores negativos.

2.4.4.2. Softmax

La función Softmax es una de las funciones más conocidas en ciencia e ingeniería y ha sido utilizada en áreas como la teoría de juegos, el aprendizaje por refuerzo y el aprendizaje automático [18], la cual es una extensión de la función sigmoïdal [7]. Esta función transforma las

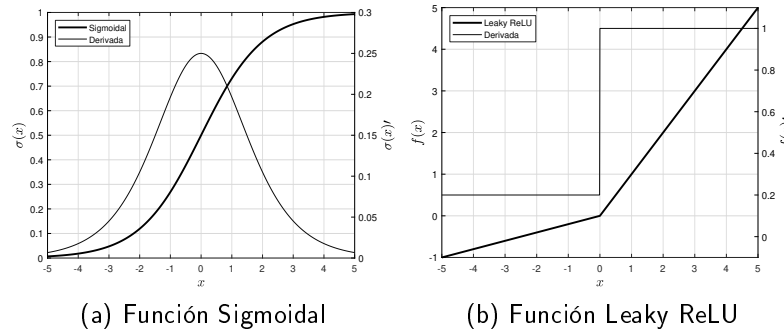


Figura 2.19: Diferencias entre las funciones de activación.

salidas en estimaciones de probabilidad posteriores para las clases mediante:

$$\sigma(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^c \exp(x_j)}, \quad i = 1, \dots, c, \quad (2.84)$$

donde c es el número total de nodos en la capa de salida de la red.

2.4.5 Perceptrón multicapa

Actualmente, el perceptrón multicapa es una de las arquitecturas más utilizadas en aprendizaje para máquinas debido a la capacidad de resolver problemas de clasificación no lineales. Generalmente, esta arquitectura consta de una capa de entrada, una capa oculta y una capa de salida, las cuales están formadas por un conjunto de neuronas. Las neuronas de la capa de entrada se encargan de recibir un patrón de entrada $\mathbf{x} = [x_1, x_2, \dots, x_d]^T$, y propagar la señal a las neuronas de la capa oculta. La última capa actúa como salida de la red, proporcionando al exterior la respuesta al patrón de entrada en términos de probabilidades posteriores para cada clase.

La Figura 2.20 muestra un ejemplo de una perceptrón multicapa, la cual tiene d neuronas de entrada que se encargan de recibir el patrón con d características conectadas entre sí a la capa oculta y finalmente conectadas a las neuronas en la capa de salida. De modo que la salida en el k -ésimo

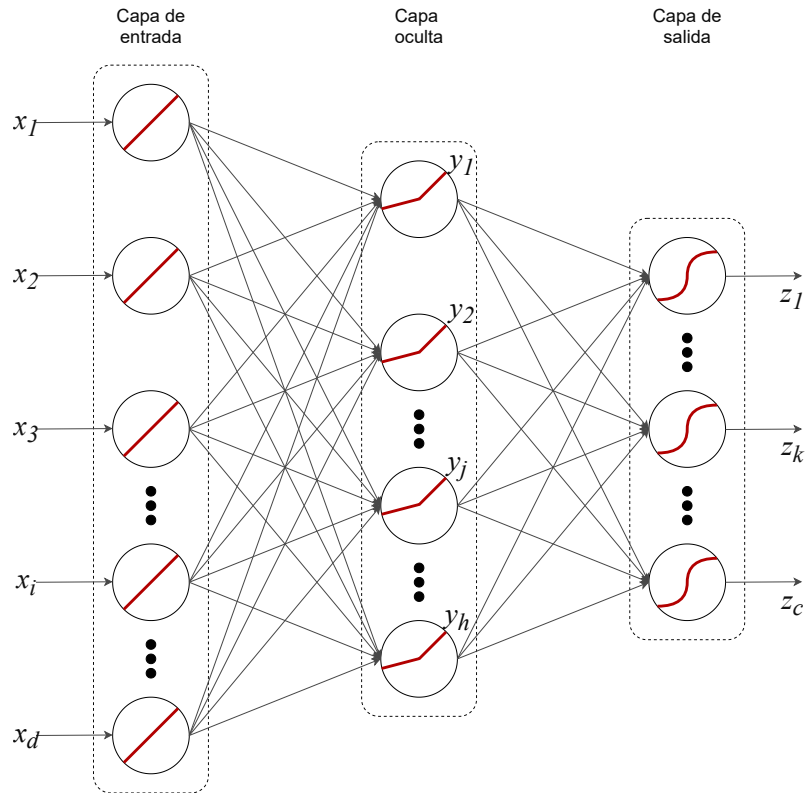


Figura 2.20: Ejemplo de una red perceptrón multicapa (MLP).

nodo de la red se define como:

$$z_k = \sigma \left[w_{k0} + \sum_{j=1}^h w_{kj} \phi \left(w_{j0} + \sum_{i=1}^d w_{ji} x_i \right) \right], \quad (2.85)$$

donde $\sigma[\cdot]$ y $\phi(\cdot)$, son las funciones de activación Softmax y Leaky ReLU, respectivamente. El valor que se obtiene en la salida de la red es la probabilidad posterior para cada clase, por lo que la suma de todas las salidas es la unidad.

Para evaluar el resultado de la clasificación del perceptrón multicapa se utiliza la función de costo de entropía cruzada, definida como:

$$J(\mathbf{w}) = - \sum_{k=1}^c t_k \log(z_k), \quad (2.86)$$

donde t y z son los valores de referencia y salida de la red, respectivamente, y \mathbf{w} representa a todos los pesos sinápticos de la red. El algoritmo de entrenamiento de *backpropagation* está basado en la técnica de descenso de gradiente, el cual cambia en la dirección que reduce el error. Este diferencial se obtiene mediante:

$$\Delta \mathbf{w} = -\eta \frac{\partial J}{\partial \mathbf{w}}, \quad (2.87)$$

donde η representa la tasa de aprendizaje. Una vez que se tiene el diferencial de los pesos, se procede a la actualización en la t -ésima iteración con la Ecuación 2.77.

2.4.6 Redes Neuronales Convolucionales

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) son un tipo de redes neuronales profundas muy populares en el campo del aprendizaje profundo, el cual es un tipo de aprendizaje automático en el que un modelo aprende a realizar tareas de clasificación a partir de imágenes, vídeos, textos o sonidos. Las CNNs son similares a las redes neuronales multicapa; sin embargo, debido a que su aplicación es realizada en matrices bidimensionales, son muy efectivas para localizar patrones en imágenes con el objetivo de reconocer objetos.

Actualmente, el uso de las CNNs se ha vuelto cada vez más popular debido a que eliminan la necesidad de la extracción manual de características para la etapa del entrenamiento. Además, generan buenos resultados de reconocimiento y pueden ser reutilizadas para nuevas tareas de clasificación. Las CNNs pueden tener decenas o cientos de capas que aprenden a detectar diferentes características de una imagen.

Es importante mencionar que las técnicas más avanzadas de aprendizaje profundo para segmentación provienen de un precursor común denominado red completamente convolucional (FCN, por sus siglas en inglés), donde la salida de la red es un puntaje de clasificación para toda la imagen. La idea es aprovechar las FCNs como extractores de características y reemplazar las capas completamente conectadas por capas convolucionales para generar mapas especiales en lugar

de puntajes de clasificación de modo que estos mapas se muestrean hacia arriba (*upsampling*) usando deconvoluciones para generar salidas etiquetadas por píxel, dando origen a las CNNs para segmentación.

Una CNN para segmentación consiste en dos partes: codificación y decodificación, seguido por un clasificador. La parte de codificación, se encarga de aplicar capas de convolución seguido de una función de activación y una capa de agrupamiento para extraer rasgos de la imagen y reducir el tamaño de la imagen, respectivamente. A este proceso se le denomina *down sampling*. Este proceso se repite sucesivamente hasta obtener el vector de características de la imagen. Por otro lado, la parte de decodificación, realiza el proceso inverso, es decir, reescala la imagen a un tamaño mayor seguida de una convolución y una función de activación, así sucesivamente hasta obtener el tamaño original de la imagen, donde la salida de la capa de agrupamiento es la entrada a una capa de convolución. A este proceso se le denomina *up sampling*. De acuerdo al número de capas de *down sampling* se tendrá el mismo número de capas de *up sampling*. La profundidad de la red depende del número de capas de *down sampling*. Además, por cada nivel de profundidad se obtiene información diferente de la imagen, por ejemplo, las primeras capas de una CNN resalta las características de curvas y bordes. En la Figura 2.21 se muestra el proceso general para segmentación utilizando una red CNN.

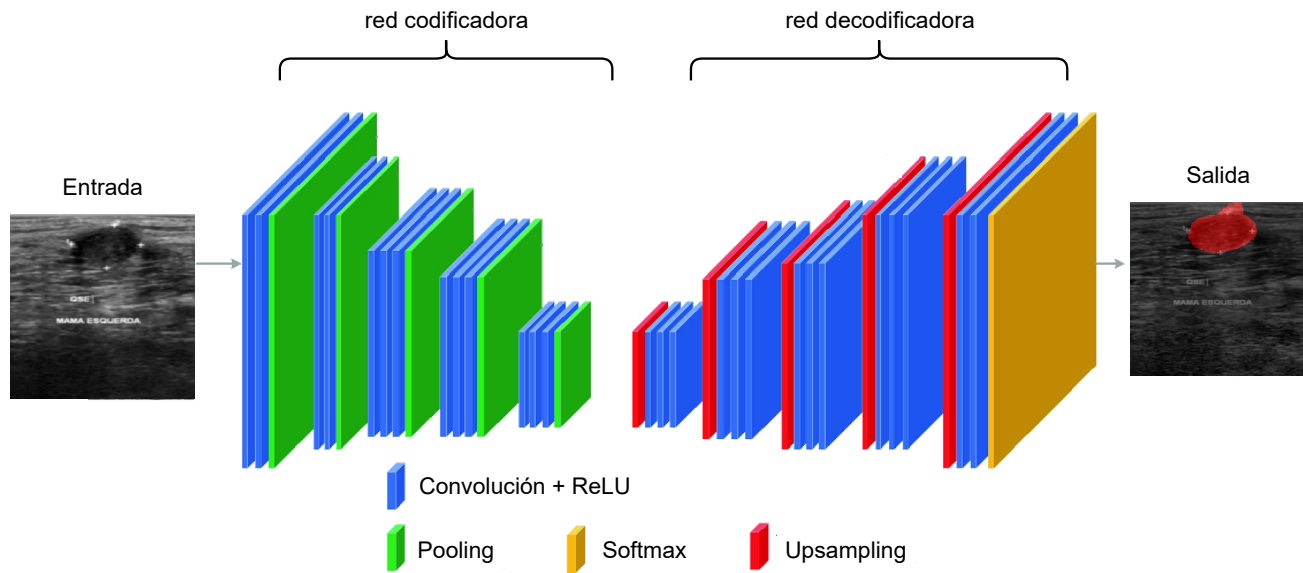


Figura 2.21: Diagrama básico de una red neuronal convolucional codificador-decodificador para la segmentación semántica de imágenes.

Cabe mencionar que existen distintos tipos de arquitecturas pre-entrenadas, las cuales están diseñadas para reconocer y segmentar una serie de objetos definidos. Sin embargo, es posible utilizar estas arquitecturas como extractores de características para poder segmentar un nuevo objeto. La Figura 2.22 ilustra de manera general cómo están diseñadas algunas de las arquitecturas más usadas en visión por computadora.

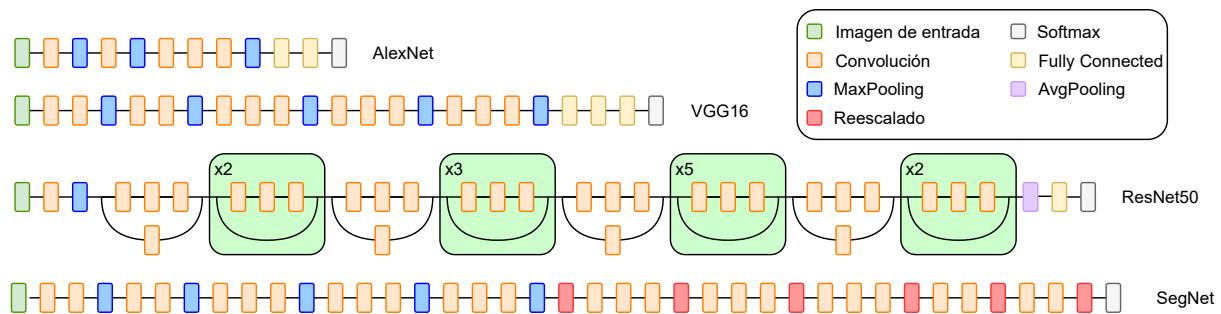


Figura 2.22: Modelo general de las arquitecturas CNN utilizadas en esta investigación.

Cada una de las arquitecturas CNN persigue el mismo objetivo, mediante el uso de filtros convolucionales se extraen rasgos característicos de cada imagen que le permita aprender a la CNN.

Sin embargo, se diferencian por el número de capas que tienen para extraer características, lo cual se denomina profundidad de la red.

2.5 Transferencia de aprendizaje

Una suposición de las metodologías tradicionales de aprendizaje automático es que los datos de entrenamiento y los datos de prueba se toman del mismo dominio, de modo que el espacio de características de entrada y las características de distribución de datos son las mismas [44]. Sin embargo, en algunos escenarios de aprendizaje automático del mundo real, esta suposición no se cumple. Por lo tanto, existe la necesidad de crear modelos entrenados con datos obtenidos más fácilmente de diferentes dominios. Esta metodología se conoce como transferencia de aprendizaje. La transferencia de aprendizaje (*transfer learning*) es un método de aprendizaje para máquinas donde un modelo desarrollado para una tarea específica es utilizado como punto inicial para un modelo que realice una segunda tarea completamente diferente [67]. Este método es un enfoque popular en el aprendizaje profundo en el que se utilizan modelos pre-entrenados en tareas de visión por computadora para acelerar el entrenamiento y mejorar el rendimiento de su modelo, reemplazando las capas finales para mapear el problema de clasificación nuevo y realizar un ajuste fino mediante *backpropagation* [36]. Generalmente, la transferencia de aprendizaje ocurre cuando los datos de entrenamiento son limitados o difíciles de recopilar. Actualmente, hay muchas aplicaciones de aprendizaje automático en las que la transferencia de aprendizaje se ha aplicado con éxito como en clasificación de imágenes, clasificación de actividades humanas, clasificación de texto en varios idiomas, entre otras más [44].

2.6 Aumento artificial de datos

El aumento de datos es el proceso de incrementar artificialmente la cantidad y diversidad del conjunto de datos de entrenamiento aplicando transformaciones a las imágenes de manera aleatoria, con el objetivo de crear nuevos datos a partir de los originales sin la necesidad de recopilar datos nuevos. Por lo tanto, esta técnica permite entrenar mejor el modelo evitando el sobreajuste del mismo, de modo que pueda tener una mayor variabilidad de muestras en el conjunto de entrenamiento utilizando repetidamente el conjunto de entrenamiento [53]. Las operaciones usadas comúnmente son transformaciones geométricas como: rotación, traslación, recorte, escalamiento, espejo (*flip*) y cambiar el nivel de intensidad. Este método conduce a mejoras sustanciales en la segmentación de imágenes médicas cuando el conjunto de datos de entrenamiento disponible es limitado [10]. La Figura 2.23 muestra algunos ejemplos del aumento artificial de datos.

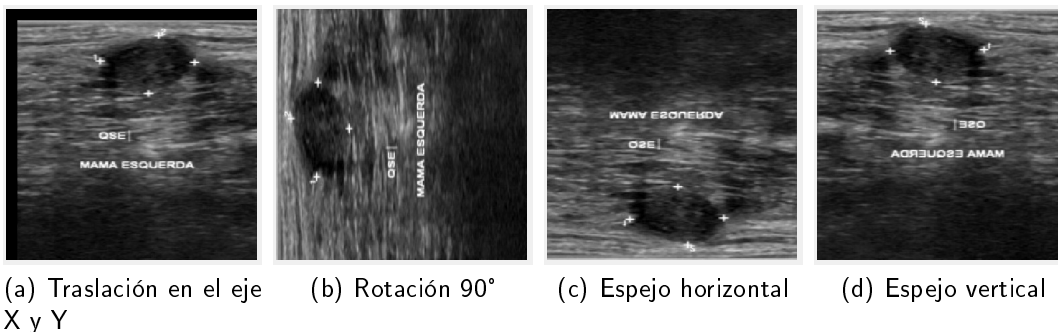


Figura 2.23: Ejemplo de operaciones de aumento artificial de datos.

2.7 Índices de validación

Los índices de validación se encargan de medir el desempeño de un modelo de clasificación, es decir, cuantificar la capacidad para distinguir entre píxeles correspondientes a la clase tumor y tejido adyacente. Cada uno de los píxeles cuenta con su etiqueta de referencia (*ground-truth*), la cual en conjunto de la clase predicha se puede construir la matriz de confusión como se muestra en la

Figura 2.24.

		Etiqueta de referencia	
		Tumor	Tejido adyacente
Etiqueta predicha por el clasificador	Tumor	Verdadero positivo (<i>TP</i>)	Falso positivo (<i>FP</i>)
	Tejido adyacente	Falso negativo (<i>FN</i>)	Verdadero negativo (<i>TN</i>)

Figura 2.24: Matriz de confusión para un problema de dos clases.

La matriz de confusión resume el conteo de aciertos y errores para cada clase y a partir de ellos se calculan los índices de desempeño de segmentación. A continuación se presentan índices de desempeño de clasificación usados para evaluar los modelos de segmentación semántica. Todos los índices deben tender a la unidad para indicar un desempeño de segmentación adecuado.

2.7.1 Exactitud

La exactitud (*ACC*, por sus siglas en inglés) representa la proporción entre el número de predicciones correctas (tanto positivas como negativas) y el total de predicciones, y se calcula como:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}. \quad (2.88)$$

2.7.2 Coeficiente de correlación de Matthews

El coeficiente de correlación de Matthews (*MCC*, por sus siglas en inglés) se utiliza como una medida de la calidad de las clasificaciones basadas en dos clases. Tiene en cuenta los verdaderos y falsos positivos y negativos por separado, y se considera generalmente como una medida equilibrada

utilizable en casos de desbalance entre clases. Como su nombre indica es una medida de correlación que puede tomar valores entre -1 y +1, donde un coeficiente de correlación de +1 representa una predicción perfecta, 0 una predicción aleatoria y -1 una predicción inversa. El MCC se calcula como:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (2.89)$$

2.7.3 Sensibilidad

Indica la capacidad del clasificador para identificar correctamente píxeles correspondientes a la clase tumor. La sensibilidad se define como:

$$SEN = \frac{TP}{TP + FN}. \quad (2.90)$$

2.7.4 Especificidad

La especificidad mide la capacidad del clasificador para identificar correctamente píxeles del fondo o tejido adyacente al tumor. La especificidad se define como:

$$SPE = \frac{TN}{TN + FP}. \quad (2.91)$$

2.7.5 Exactitud balanceada

Al igual que la exactitud, se puede usar al evaluar qué tan bueno es un clasificador, relacionando la sensibilidad y la especificidad. La exactitud balanceada se usa especialmente cuando las clases se encuentran desbalanceadas y se define como:

$$BAC = \frac{SEN + SPE}{2}. \quad (2.92)$$

2.7.6 Medida F1

La medida F1 combina las medidas de precisión y exhaustividad para devolver una medida de calidad más general del modelo. El valor F1 se define como:

$$F_1 = \frac{2TP}{2TP + FP + FN}. \quad (2.93)$$

2.7.7 Índice Jaccard

El índice Jaccard es una medida de similitud entre dos conjuntos, cuanto mayor sea la similitud de estos dos conjuntos, mayor será el valor del índice. El índice Jaccard (*Intersection over Union*) se define como:

$$IoU = \frac{TP}{TP + FP + FN}. \quad (2.94)$$

2.8 Validación cruzada de k dobleces

El método de validación cruzada de k dobleces es un proceso iterativo, el cual consiste en dividir el conjunto de imágenes de forma aleatoria en k grupos de aproximadamente el mismo tamaño. El primer grupo se emplea como el conjunto de prueba, mientras que los otros $k - 1$ grupos se emplean para entrenar el modelo, este proceso se repite k veces utilizando un grupo distinto como prueba en cada iteración. Este proceso genera k resultados cuyo promedio se emplea como resultado final.

2.9 Pruebas de significación estadística

Las pruebas estadísticas establecen una hipótesis nula y una hipótesis alternativa que son mutuamente excluyentes. La hipótesis nula es una suposición sobre un suceso en relación a algún parámetro (o parámetros) de una población o muestra.

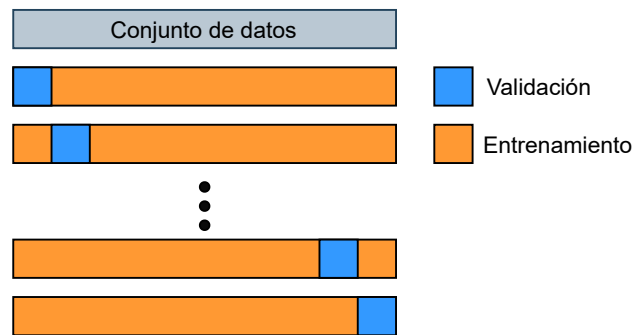


Figura 2.25: Distribución del conjunto de datos para los conjuntos de entrenamiento y validación en el método validación cruzada de k dobleces.

La salida de una prueba estadística es un valor- p que indica la probabilidad de que un valor estadístico calculado sea posible dada una hipótesis nula. Si el valor- p cumple con la condición de ser menor que un nivel de significancia impuesto (comúnmente 0.05), este se considera como un resultado estadísticamente significativo y, por tanto, permite rechazar la hipótesis nula y favorecerá la hipótesis alternativa. En caso contrario, la hipótesis nula será aceptada. A continuación se mencionan las pruebas de significación estadística para el análisis estadístico de los resultados.

2.9.1 Prueba de Shapiro-Wilk

La prueba de Shapiro-Wilk es una prueba de normalidad de los datos que permite decidir si se debe aplicar un método paramétrico o no paramétrico para el análisis de los valores obtenidos en la experimentación [56]. Esta prueba establece las siguientes hipótesis:

- Hipótesis nula: Los datos son normalmente distribuidos.
- Hipótesis alternativa: Los datos no son normalmente distribuidos.

Si se rechaza la hipótesis nula, entonces se opta por realizar pruebas no paramétricas para conocer si dos o más muestras son significativamente diferentes.

2.9.2 Prueba de Wilcoxon

La prueba de Wilcoxon es un método no paramétrico que contrasta dos muestras o grupos relacionados, dando a conocer si existen o no diferencias significativas entre ambas [56]. Asumiendo que las dos muestras vienen de la misma distribución, esta prueba establece las siguientes hipótesis:

- Hipótesis nula: Las medianas de dos grupos son iguales.
- Hipótesis alternativa: Las medianas de dos grupos son diferentes.

2.9.3 Prueba de Kruskal-Wallis con corrección Bonferroni

La prueba de Kruskal-Wallis es un método no paramétrico que compara las medianas de los grupos para probar si existe diferencias significativas, asumiendo que los grupos provienen de la misma distribución [56]. Esta prueba establece las siguientes hipótesis:

- Hipótesis nula: La mediana de los grupos es la misma.
- Hipótesis alternativa: Al menos uno de los grupos tiene mediana diferente.

Además, se utiliza la corrección de Bonferroni para saber entre qué pares de grupos se presenta dicha diferencia

2.10 Sumario

En este capítulo se presentó una introducción a la segmentación semántica y cómo se diferencia de la detección, seguido de los métodos de mejoramiento de contraste y la colección de técnicas para la extracción de características a partir de las imágenes BUS. Posteriormente se mostraron los métodos de reducción de dimensionalidad de características. También, se muestra la descripción de cada uno de los enfoques, convencional y convolucional, y los parámetros involucrados para cada

enfoque. Finalmente, se introdujeron los índices de validación y las pruebas de significancia estadística que se utilizan comúnmente para evaluar el rendimiento de segmentación de los sistemas CAdE.

3

Estado del Arte

En este capítulo se describen los tipos de sistemas CADe relacionados con la detección de tumores en imágenes de ultrasonido de mama (BUS, por sus siglas en inglés), los cuales se presentan con los enfoques de redes neuronales convencional y convolucional.

3.1 Enfoque Convencional

Como se mencionó en la Sección 1.1.2, el enfoque convencional es aquel que recibe características extraídas de una imagen BUS a partir del supuesto de un experto. Debido a que una imagen BUS está degradada con artefacto *speckle* y bajo contraste, la segmentación del tumor es una tarea desafiante [27, 33, 35, 50]. Por esta razón, las características basadas en contraste no son suficientes, por lo que se complementan con técnicas que tienen la capacidad de extraer características como textura, momentos geométricos, etc. Este enfoque parte desde la metodología tradicional del reconocimiento de objetos, en el cual se suele aplicar una etapa de preprocesamiento para mejorar la calidad de la imagen. El artefacto *speckle* es un tipo de ruido multiplicativo que aparece en imágenes de

ultrasonido, el cual degrada la calidad de la imagen, aumentando la dificultad de detectar detalles finos en las imágenes de BUS.

Trabajos como el de Gómez *et al.* [60], Marcomini *et al.* [39], Rodrigues *et al.* [50] y Huang *et al.* [27] utilizaron filtros de mediana, de media y Wiener, inclusive técnicas como umbralado de histograma con lógica difusa o difusión anisotrópica, con el objetivo de reducir el *speckle*. Como resultado de aplicar estos filtros, se obtiene una imagen con *speckle* reducido que preserva características importantes de las estructuras en la imagen. Una vez que se cuenta con las imágenes mejoradas es posible realizar la extracción de características con diferentes técnicas.

La extracción de estas características permiten entrenar algún método de aprendizaje supervisado, con el objetivo de segmentar la región del tumor en las imágenes BUS. Huang *et al.* [27] señalan que los tres métodos de clasificación más utilizados son k-vecinos más cercano (K-NN, por sus siglas en inglés), redes neuronales artificiales (ANN, por sus siglas en inglés) y máquinas de vectores de soporte (SVM, por sus siglas en inglés).

Huang *et al.* [27] consideran que los filtros lineales no funcionan satisfactoriamente en el pre-procesamiento debido a que los bordes y el ruido comparten componentes de alta frecuencia, lo que dificultaría diferenciarlos en la práctica. Por esta razón, utiliza un modelo de variación total para la reducción del artefacto *speckle*. Consideran características el histograma de niveles de gris (HIS), GLCM, histograma del gradiente original (HOG), e información de adyacencia. A partir del conjunto de características, realiza una selección de las más importantes mediante puntajes de *biclustering*. Huang *et al.* utilizan un método de segmentación basado en grafos (RGB) para particionar las imágenes BUS en distintas sub-regiones y posteriormente con una máquina de vectores de soporte (SVM) clasificar cada una las sub-regiones. Además, realiza validación cruzada para garantizar que son independientes de la partición entre datos de entrenamiento y prueba.

Liu *et al.* [35] utilizan un enfoque basado en autómatas celulares difusos (FCA) para la segmentación de tumores en imágenes BUS considerando características como varianza, entropía y velocidad de variación. A diferencia de otros trabajos, en este enfoque se seleccionan píxeles iniciales

como semillas que comienzan a extenderse afectando gradualmente la imagen. Cabe mencionar que seleccionar adecuadamente los píxeles iniciales es importante y esencial para mejorar la precisión de la segmentación.

Por otro lado, Shan *et al.* [52] utilizan características como varianza, fase máxima de orientación (MPO) y distancia radial (RD) para clasificar los píxeles y comparar dos tipos de clasificadores: redes neuronales artificiales (ANN) y máquina de vectores de soporte (SVM) combinado con validación cruzada de 10 dobleces para evaluar sus modelos. Además, para mejorar la segmentación aplica rellenado de huecos para después refinar con operaciones morfológicas.

Gómez *et al.* [60] proponen un método que involucra la detección y delineación de la lesión mediante un análisis de discriminante lineal (LDA) utilizando características de textura como log-Gabor, matriz de co-ocurrencia de niveles de gris, autovarianza, autocorrelación e información de adyacencia. A diferencia de otros trabajos que utilizan ANN y SVM, Gómez *et al.* proponen la clasificación de píxeles de las imágenes BUS mediante LDA.

Finalmente, Huang *et al.* [26] utilizan un filtro bilateral sobre una región de interés con la mayor proporción del tumor, el cual no solo reduce el *speckle*, sino que también preserva los bordes. Además, realizan una ecualización del histograma sobre la región filtrada y un nuevo filtrado de pirámide de desplazamiento medio, el cual mejora la homogeneidad de las imágenes BUS. Una vez pre-procesada la imagen se generan los llamados superpíxeles, que son regiones pequeñas que comparten similitud entre sus píxeles. A partir de los superpíxeles calculan características de imagen de bajo nivel que incluyan gris, textura y patrón binario local. Seguido de una red neuronal *back propagation* (BPNN, por sus siglas en inglés) para clasificar los superpíxeles y, adicionalmente, con un enfoque basado en *k*-nn para refinar la solución. Además, cuenta con el método de validación cruzada de 10 dobleces.

Algo notorio de estos trabajos es que no cuentan con un conjunto de imágenes estándar. Por tanto, no es posible realizar una comparativa justa entre los resultados reportados por cada método. En la Tabla 3.1 se muestra la comparativa del desempeño de segmentación de trabajos basados en el enfoque convencional utilizando distintos algoritmos.

Autor	Método	# imágenes	IoU	MCC	SEN	SPE	AUC	ACC	FPR	SI	Características
Gómez <i>et al.</i> [60]	Detección LDA	544	-	0.85	0.86	0.97	0.92	-	-	-	Textura
			-	0.78	0.85	0.93	0.89	-	-	-	
Huang <i>et al.</i> [27]	Detección RGB + SVM	46	-	-	0.97	0.98	0.99	0.98	-	-	Textura
Liu <i>et al.</i> [35]	Detección FCA	235	0.82	-	-	-	-	-	-	-	Textura
Shan <i>et al.</i> [52]	Detección ANN	120	-	-	0.93	-	-	-	0.12	0.83	Textura y
	Detección SVM		-	-	0.92	-	-	-	0.12	0.83	espacial
Huang <i>et al.</i> [26]	Detección Superpíxeles + BPNN	320	-	-	-	-	-	0.86	-	-	Textura

Tabla 3.1: Desempeño de segmentación de enfoques convencionales.

3.2 Enfoque Convolutacional

Actualmente, los sistemas CADe con enfoque convolutacional han tenido mucho auge debido a que han mostrado dar buenos resultados en la segmentación semántica de imágenes BUS [8, 24, 25, 32, 63, 64, 66, 67]. Sin embargo, realizar el entrenamiento de una CNN desde cero requiere de un conjunto grande de imágenes BUS, aunque eso no ha sido una limitante para utilizarlas como método de segmentación de lesiones de mama, ya que es posible solventarlo con transferencia de aprendizaje. Existen distintas arquitecturas CNN como AlexNet, VGG16, ResNet, SegNet, entre otras, utilizadas para la detección de tumores en imágenes BUS. Algunas de estas arquitecturas se encuentran pre-entrenadas, y mediante la transferencia de aprendizaje se pueden reutilizar para segmentar nuevos objetos, por lo que el tiempo para entrenar una CNN para segmentar tumores en imágenes BUS es reducido.

Dado que la mayoría de los autores cuentan con un conjunto pequeño de imágenes BUS para resolver este problema, se puede realizar un aumento artificial de datos, de tal forma que se cuenta con versiones trasladadas y rotadas de las mismas imágenes. Como se mencionó anteriormente, una CNN extrae las características de una imagen de BUS de manera automática, por lo que el número de características que se extraigan dependerá de la profundidad de la CNN, es decir, por cada capa de convolución se obtienen mapas de características de la imagen BUS.

Hu *et al.* [24] utilizan una red completamente convolucional dilatada (DFCN) con contornos activos basado en fases (PBAC). La DFCN es una red neuronal completamente convolucional con convoluciones dilatadas en las capas más profundas, pocos parámetros y técnicas de normalización de los datos. Las predicciones hechas por la DFCN son relativamente difíciles debido a los contornos borrosos y la variación en el tamaño del tumor. El resultado de la DFCN lo usan como punto de partida para el método PBAC, donde su número de iteraciones depende del tamaño del tumor. Consideran que los tumores que representan más del 9.05 % de una imagen BUS son tumores grandes y utilizan 60 iteraciones en el método PBAC. Por otro lado, los tumores que representan menos del 9.05 % de una imagen BUS se consideraron tumores pequeños y utilizan 120 iteraciones, por lo que la salida de este método es la segmentación final.

Huang *et al.* [25] dividen su método de segmentación en tres etapas. Primero, realizan una normalización del histograma para homogeneizar el rango de niveles de gris en todas las imágenes. Además, utilizan la transformada *wavelet* para obtener información adicional de alta y baja frecuencia, las cuales son submustradas por columna y fila para obtener información marginal en las direcciones horizontal y vertical. A continuación, construyen una imagen de tres canales donde el primer canal es la imagen original, el segundo y tercer canal son las componentes de baja y alta frecuencia. Posteriormente, redimensionan la imagen a su tamaño original mediante interpolación bilineal. La segunda etapa consiste en entrenar una red completamente convolucional (FCN) con enfoque basado en parches, el cual, divide la imagen en pequeños parches y los clasifica en tres categorías: tumor, fondo y capas mamarias. Para esto, utiliza transferencia de aprendizaje para resolver el problema de escasez de muestras de entrenamiento. Además, para aumentar las muestras de entrenamiento, utiliza versiones rotadas de las mismas. Por último, la tercer etapa involucra campos aleatorios condicionales (CRF) para refinar el resultado de la segmentación.

Xie *et al.* [63] reescalan las imágenes BUS a un tamaño de 224×224 para usar como conjunto de entrenamiento y entrenar una ResNet como clasificador usando transferencia de aprendizaje, y definir si la imagen BUS es normal o cancerosa. Posteriormente, los casos etiquetados como cancerosos,

son la entrada a la Mask R-CNN para la segmentación del tumor. La Mask R-CNN es una mejora del modelo Faster R-CNN usado en la detección de objetos, por lo que Mask R-CNN además de detectar puede segmentar el objeto.

Yap *et al.* [67] utilizan tres métodos diferentes para la detección de lesiones de mama en imágenes BUS: LeNet basada en parches, U-Net y FCN-AlexNet. Para cada uno refina la imagen eliminando todas aquellas regiones con un área menor que 10 píxeles para reducir el número de falsos positivos. Además, considera la validación cruzada de 10 dobleces para evaluar a cada método.

Kumar *et al.* [31] inician reescalando todas las imágenes BUS a un tamaño de 208×208 píxeles. Para estandarizar las imágenes, calculan los puntajes estándar restando el valor medio de las imágenes de cada píxel seguido de la división con la desviación estándar de la imagen. Utilizan un algoritmo Multi U-Net, el cual es un ensamble de diez U-Net entrenadas de manera independiente considerando validación cruzada de 10 dobleces, que crea una máscara con la segmentación de cada U-Net. Por último, mediante un esquema de voto mayoritario, agrupa las respuestas de los modelos U-Net para realizar la segmentación fina del tumor.

Chiao *et al.* [13] utiliza una extensión de la Faster R-CNN para la segmentación de tumores llamada Mask R-CNN. Esta CNN reemplaza la agrupación ROI de la Faster R-CNN por la alineación ROI para evitar perder información espacial.

Zhuang *et al.* [68] proponen una RDAU-Net para la segmentación del tumor, la cual se basa en una arquitectura U-Net. Esta arquitectura incorpora elementos adicionales como unidades residuales, convoluciones dilatadas y módulos de puerta de atención para mejorar su capacidad de aprendizaje.

En la Tabla 3.2 se muestra la comparativa del desempeño de segmentación de trabajos basados en el enfoque convolutacional con distintas arquitecturas.

Autor	Método	# imágenes	TPR	FPs	Medida F	BAC	
Yap <i>et al.</i> [67]	Detección	306 (A) y 163 (B)	LeNet basado en parches (A)	0.92	0.07	0.91	-
			LeNet basado en parches (B)	0.91	0.09	0.91	-
			U-Net (A)	0.94	0.18	0.89	-
			U-Net (B)	0.80	0.27	0.78	-
			FCN-AlexNet (A)	0.99	0.16	0.92	-
			FCN-AlexNet (B)	0.93	0.18	0.88	-
Hu <i>et al.</i> [24]	Detección	DFCN + PBAC	570	-	-	0.89	-
Xie <i>et al.</i> [63]	Detección	Mask R-CNN50	1182	-	-	0.82	-
Huang <i>et al.</i> [25]	Detección	FCN + CRF	325	0.93	0.09	-	-
Kumar <i>et al.</i> [31]	Detección	Multi U-Net	433	0.84	-	0.82	-
Chiao <i>et al.</i> [13]	Detección	Mask R-CNN	307	-	-	-	0.75
Zhuang <i>et al.</i> [68]	Detección	RDAU-Net	1062	-	-	0.84	-

Tabla 3.2: Desempeño de segmentación de métodos basados en el enfoque convolucional.

3.3 Sumario

En este capítulo se presentaron los trabajos relacionados para los enfoques convencional y convolucional, donde se muestran los desempeños bajo distintos índices de validación. La Tabla 3.1 resume los trabajos con enfoque convencional utilizados para la segmentación del tumor, donde se observa que cada uno de los autores considera un método distinto para llevar a cabo esta tarea. Sin embargo, cada trabajo reporta los resultados de segmentación mediante diversos índices de validación, por lo que no es posible hacer una comparación directa del desempeño de segmentación con un índice en particular. Además, dado que las características basadas en contraste no proporcionan la información suficiente para distinguir entre las clases tumor y tejido adyacente, todos los métodos convencionales complementan con características basadas en textura.

Por otro lado, la Tabla 3.2 muestra el desempeño de los trabajos relacionados con enfoque convolucional, los cuales utilizan distintas arquitecturas CNN, por lo que el desempeño de cada método se ve afectado debido a la profundidad que tiene cada CNN. A diferencia de los trabajos del enfoque convencional, la mayoría de los trabajos del enfoque convolucional comparten un índice de validación en común que permite identificar quién muestra mejor desempeño en la segmentación entre ellos.

Desafortunadamente no es posible comparar los trabajos del enfoque convolucional con los del enfoque convencional debido a que no miden su desempeño bajo un mismo índice de validación. Cabe mencionar que los trabajos analizados en las Tablas 3.1 y 3.2 no cuentan con un conjunto estándar de imágenes BUS, lo que dificulta realizar una comparación justa entre los enfoques convencional y convolucional. Sin embargo, coinciden en utilizar una etapa de pre-procesamiento para mejorar la calidad de las imágenes BUS antes de entrenar un modelo para la segmentación del tumor.

4

Metodología

En este capítulo se presentan las etapas para el desarrollo de este estudio. Primero, se describe el conjunto de imágenes de BUS para la experimentación. Después, se presenta el marco diseñado para realizar los experimentos para sistemas CADe con los enfoques convencional y convolucional. Posteriormente, se describe la configuración para cada uno de los enfoques. Además, se presenta el procedimiento de validación para la segmentación de imágenes de BUS. Por último, se muestra la infraestructura computacional utilizada en la realización de los experimentos.

4.1 Conjunto de imágenes de BUS

El conjunto de imágenes comprende 1,200 imágenes BUS proporcionadas por el Instituto Nacional de Cáncer (INCa) de Río de Janeiro, Brasil. Cada imagen de BUS tiene segmentación manual del tumor, la cual fue proporcionada por un radiólogo. El conjunto de imágenes cuenta con 629 casos de tumores benignos y 571 de tumores malignos, todos confirmados por biopsia. Todos los tumores de estas imágenes tienen un área mayor al 5 % del área total de la imagen para garantizar que existiera

información del tumor cuando se realice la reducción del tamaño de las imágenes durante las etapas de extracción de características en los enfoques convencional y convolucional. La Figura 4.1 muestra algunos ejemplos de imágenes de BUS junto con su segmentación manual de referencia.

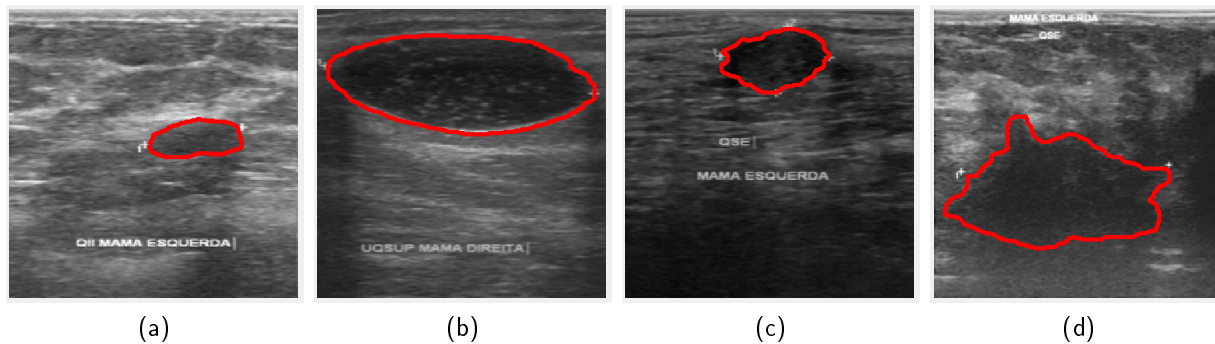


Figura 4.1: Ejemplo de imágenes BUS con la segmentación manual en rojo, donde a) y b) son tumores benignos, c) y d) son tumores malignos.

4.2 Marco experimental

Con el objetivo de comparar el desempeño de un sistema CADe con enfoque convencional y convolucional, se diseñaron dos flujos de entrenamiento sobre los métodos de cómputo neuronal para la segmentación de tumores. La Figura 4.2 y 4.3 muestran el flujo a seguir para el entrenamiento y validación de un CADe con enfoque convencional y convolucional, respectivamente.

Debido al problema de alta dimensionalidad que se presenta en el enfoque convencional, realizar el entrenamiento requiere de un alto costo computacional. Por esta razón, mediante el proceso de reducción de dimensionalidad de características, se busca descartar aquellas que no proporcionan información relevante del píxel.

Por otro lado, para el enfoque convolucional, se estudiaron distintas arquitecturas CNN, las cuales son extractores automáticos de características utilizadas para la segmentación de tumores en imágenes BUS. Además, estas arquitecturas utilizan la función máximo para el agrupamiento, ReLU como función de activación con descenso de gradiente estocástico y entropía cruzada como función

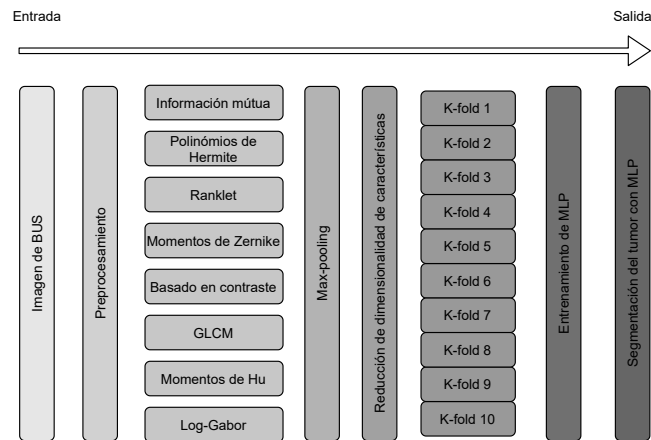


Figura 4.2: Flujo de procesos para el entrenamiento y validación de un CAde con enfoque convencional.

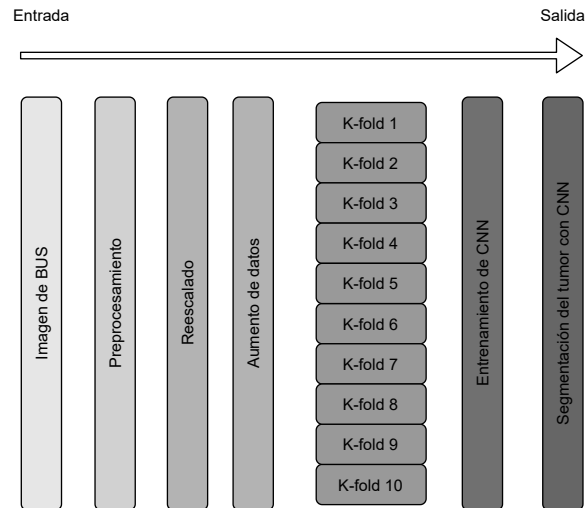


Figura 4.3: Flujo de procesos para el entrenamiento y validación de un CAde con enfoque convolucional.

de costo. La Tabla 4.1 resume características generales de las arquitecturas CNN mostradas en la Figura 2.22.

4.3 Experimentación

Los enfoques convencional y convolucional comparten la misma etapa de pre-procesamiento, la cual consiste en reducir el artefacto *speckle* mediante un filtro Gaussiano de tamaño 5×5 y $\sigma = 1$.

Arquitectura	Tamaño de entrada	Profundidad (capas)	Número de parámetros entrenables (millones)
AlexNet	$227 \times 227 \times 3$	8	62.4
ResNet50	$224 \times 224 \times 3$	50	25.5
SegNet	$256 \times 256 \times 3$	4	6.4
VGG16	$256 \times 256 \times 3$	16	138.4

Tabla 4.1: Características generales de las arquitecturas CNN utilizadas en esta investigación.

Posteriormente, se expande el rango dinámico de la imagen filtrada $I_f(x, y)$ para obtener valores en un rango $[0, 1]$, obteniendo la imagen denotada como $I_c(x, y)$.

4.3.1 Enfoque convencional

Para este enfoque, se utilizaron ocho técnicas de extracción de características que permiten describir la textura y geométricas locales de cada píxel de la imagen, las cuales se resumen en la Tabla 4.2.

Método	Característica de textura	Definido por
Contraste	Elemento estructurante cuadrado de tamaño 9×9 para operaciones morfológicas.	\mathbf{x}_{con}
LogGabor	Seis escalas y ocho orientaciones con paso de 22.5° , $\sigma = 0.55$ para obtener un ancho de banda aproximada a 2 octavas.	\mathbf{x}_{gab}
Momentos de Hu	Momentos de ϕ_1, \dots, ϕ_7 considerando un vecindario de 9×9 .	\mathbf{x}_{hu}
Momentos de Zernike	Polinomios de grado $n = [3, 5]$ y parámetro de dependencia angular $l = [0, n]$, considerando un vecindario de 9×9 .	\mathbf{x}_{zrk}
Transformada de Hermite Ranklets	Hasta polinomios de orden 3 considerando un vecindario de 9×9 .	\mathbf{x}_{her}
Auto-Información Mutua	Cuatro tamaños de ventana $r = [2, 4, 8, 16]$ en las tres orientaciones.	\mathbf{x}_{rnk}
GLCM	Auto-correlación en las direcciones vertical y horizontal, considerando un vecindario de 9×9 .	\mathbf{x}_{ami}
	Energía, entropía, Correlación, contraste, varianza, homogeneidad y máxima probabilidad, considerando un vecindario de 9×9 .	\mathbf{x}_{gldm}

Tabla 4.2: Características locales calculadas a partir de métodos descritos en el Capítulo 2.

De modo que se extrajeron un total de 120 características, dando como resultado imágenes de características locales denotadas de manera general como $I_t(x, y)$. A partir de $I_t(x, y)$ se obtienen estadísticos locales con base en la media y desviación estándar de las características [57], de modo

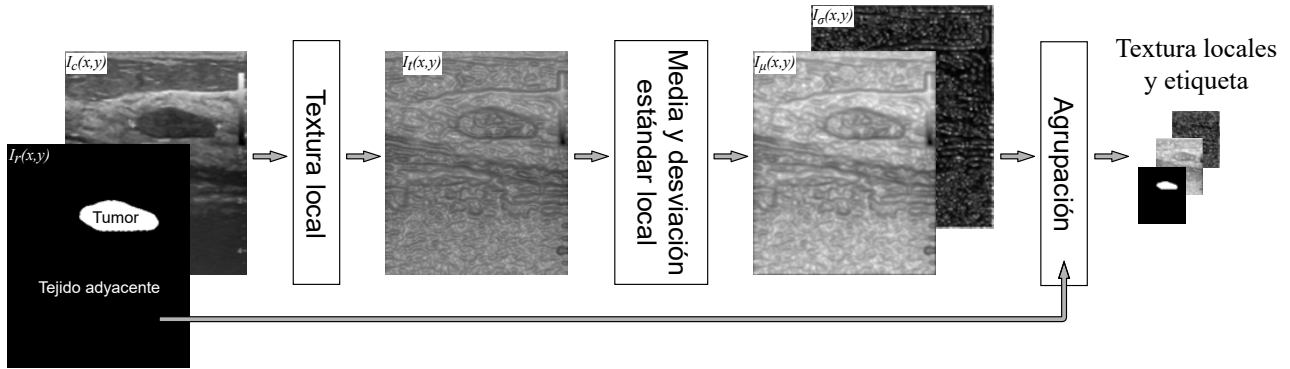


Figura 4.4: Proceso de extracción de características.

que por cada una de las características calculadas se obtienen $I_\mu(x, y)$ y $I_\sigma(x, y)$, tomando en cuenta un vecindario de tamaño $w_p \times w_p$. La Figura 4.4 ilustra el cálculo de la media local $I_\mu(x, y)$ y desviación estándar $I_\sigma(x, y)$ a partir de $I_t(x, y)$. Por esta razón, el total de características se duplica a 240. Además, la imagen pre-procesada $I_c(x, y)$ forma parte de las características, así como también información de adyacencia $I_X(x, y)$ y $I_Y(x, y)$, que representan las coordenadas en x e y de cada píxel, respectivamente. La información de adyacencia provee información espacial del píxel en la imagen BUS [28]. Además, la segmentación de referencia $I_r(x, y)$ se asocia al vector de características para indicar su clase durante el entrenamiento. De esta manera, el vector de características para el enfoque convencional queda definido como:

$$\mathbf{x} = [\mathbf{x}_{\text{con}}^\mu, \mathbf{x}_{\text{con}}^\sigma, \mathbf{x}_{\text{gab}}^\mu, \mathbf{x}_{\text{gab}}^\sigma, \mathbf{x}_{\text{hu}}^\mu, \mathbf{x}_{\text{hu}}^\sigma, \mathbf{x}_{\text{zrk}}^\mu, \mathbf{x}_{\text{zrk}}^\sigma, \mathbf{x}_{\text{her}}^\mu, \mathbf{x}_{\text{her}}^\sigma, \mathbf{x}_{\text{rnk}}^\mu, \mathbf{x}_{\text{rnk}}^\sigma, \mathbf{x}_{\text{ami}}^\mu, \mathbf{x}_{\text{ami}}^\sigma, \mathbf{x}_{\text{glcm}}^\mu, \mathbf{x}_{\text{glcm}}^\sigma, I_c(x, y), I_X(x, y), I_Y(x, y)]^T \quad (4.1)$$

donde los superíndices μ y σ indican la media y desviación estándar locales de los elementos de cada vector de características. Por tanto, el vector final de características cuenta con 243 características por cada píxel de la imagen.

Para reducir el costo computacional del entrenamiento, se realiza una reducción de tamaño a cada una de las imágenes de características calculadas. Este proceso se realiza mediante ventanas

traslapadas usando agrupación tipo *max-pooling* y *mean-pooling* en un vecindario $w_p + 1 \times w_p + 1$, es decir, se toma el máximo y la media del vecindario. Esta ventana se desplaza por toda la imagen con un paso de tamaño w_p píxeles, también conocido como *stride*. La Figura 4.5 ilustra un ejemplo del cálculo de la media local $I_\mu(x, y)$ y la reducción de tamaño usando *max-pooling*.

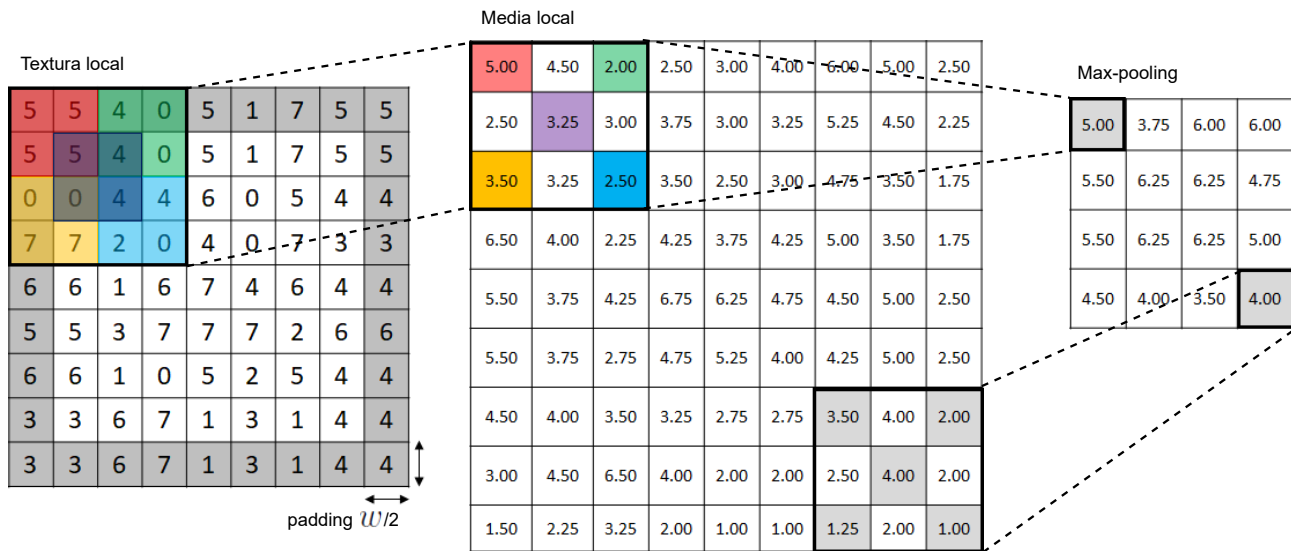


Figura 4.5: Cálculo de la media local $I_\mu(x, y)$ a partir de una imagen de característica local y agrupación tipo *max-pooling* con tamaño $w_p = 2$. En las regiones coloreadas se observa que para un *stride* de 2 se analiza una ventana de 2×2 .

Una vez que se cuenta con las imágenes escaladas, es posible reducir aún más el costo computacional de la etapa de entrenamiento de una MLP, por lo que se deben descartar aquellas características que son irrelevantes con un método de reducción de dimensionalidad. Para la etapa de entrenamiento de la MLP se combinaron tres configuraciones: método de reducción de dimensionalidad, APCA y MRMR, tamaño de ventana *pooling*, $w_p = [4, 8, 16]$, y tipo de agrupación, *max-pooling* y *mean-pooling*.

Doce experimentos fueron realizados con la combinación de estas configuraciones, por ejemplo: APCA con $w_p = 4$ y *max-pooling*. Además, para el entrenamiento de la MLP mediante *backpropagation* se consideraron parámetros como una tasa de aprendizaje inicial de 1×10^{-6} , factor de momentum de 0.9, descenso de gradiente estocástico con tamaño de *mini-batch* de 16 y funciones

de activación Leaky ReLU y Softmax para las capas oculta y de salida, respectivamente.

Como resultado de la MLP, se obtiene la imagen segmentada $I_s(x, y)$ y un mapa de probabilidades $I_p(x, y)$. Sin embargo, la dimensionalidad espacial de la imagen es menor que la imagen original debido al proceso de agrupación de píxeles en la extracción de características, por lo que las imágenes a la salida de la MLP, $I_s(x, y)$ y $I_p(x, y)$, se redimensionan al tamaño original con interpolación bicúbica para $I_p(x, y)$ e interpolación con el vecino más cercano para $I_s(x, y)$. La Figura 4.6 muestra la configuración empleada para la segmentación semántica de imágenes BUS.

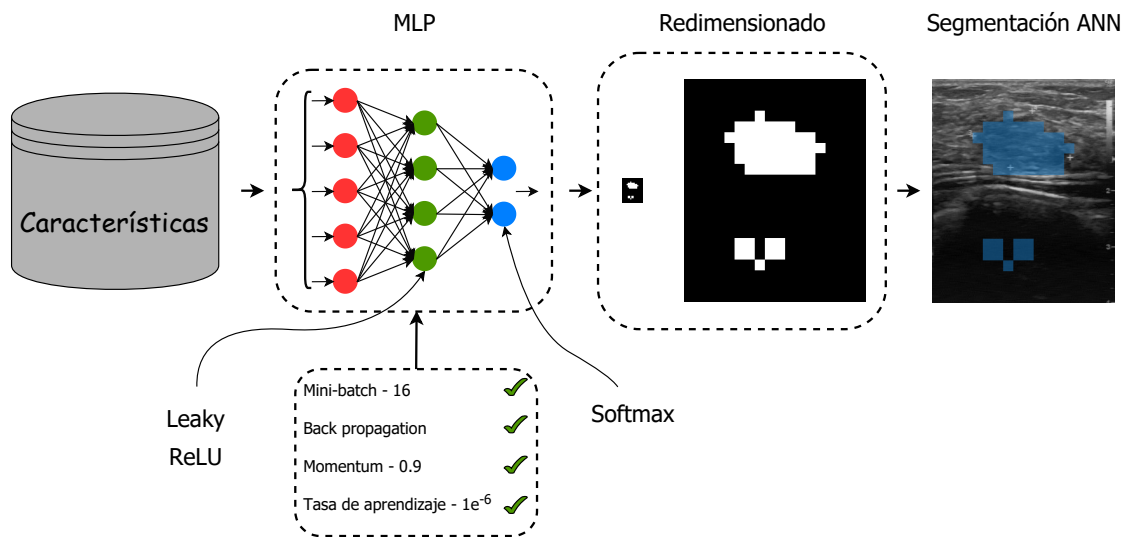


Figura 4.6: Proceso de entrenamiento para la segmentación semántica de imágenes BUS con una MLP.

Finalmente, la segmentación proporcionada por la MLP redimensionada se puede mejorar para reducir el número de falsos positivos y refinar la segmentación mediante técnicas como rellenado de huecos y contornos activos, que al combinar con el mapa de probabilidades se descartan regiones segmentadas como tumor y tengan baja probabilidad de serlo. A continuación, se listan los pasos para hacer el post-procesamiento:

1. Reducir las variaciones de probabilidad en la imagen $I_p(x, y)$ y el efecto de bloque mediante un filtro mediana.

2. Localizar la zona con el píxel de mayor probabilidad de pertenecer a la clase tumor.
3. Binarizar el mapa de probabilidades mediante un umbralado, rescatando solo los píxeles con $p \geq 0.5$ con el objetivo de contabilizar las regiones que cumplen este criterio.
4. Etiquetar las regiones detectadas.
5. Seleccionar la región donde se encuentra la probabilidad más alta y descartar las regiones restantes.
6. Refinar la región ganadora con relleno de huecos para evitar falsos negativos dentro de la segmentación; dilatación seguida de la erosión y contornos activos para suavizar el contorno.

La Figura 4.7 muestra el proceso de post-procesamiento de la imagen que se obtiene de la MLP.

4.3.2 Enfoque convolucional

Para este enfoque, se evaluaron las siguientes CNNs pre-entrenadas: AlexNet, ResNet50, SegNet y VGG16. Estas arquitecturas fueron pre-entrenadas con el conjunto de datos ImageNet, por lo que cada arquitectura recibe como entrada imágenes a color (RGB). Debido a que las imágenes BUS son de un canal (intensidades en niveles de gris) se crean las versiones de tres canales de cada una de ellas, donde cada canal contiene la imagen BUS original.

Se emplea el concepto de transferencia de aprendizaje para generar modelos de segmentación semántica para imágenes BUS. Debido a que estas CNNs cumplen la tarea de clasificar, se debe realizar una modificación en la arquitectura para llevar a cabo la segmentación del tumor. Para esto, se utilizan tres enfoques para la construcción de los modelos de segmentación semántica llamados: Red Completamente Convolucional (FCN, por sus siglas en inglés), SegNet, y DeepLabV3+.

El modelo FCN, propuesto por Shelhamer *et al.* [36], adapta una CNN pre-entrenada de clasificación de imágenes para servir como módulo de codificación de la red. Las capas completamente conectadas se convierten en capas de convolución reutilizando sus pesos sinápticos. Posteriormente,

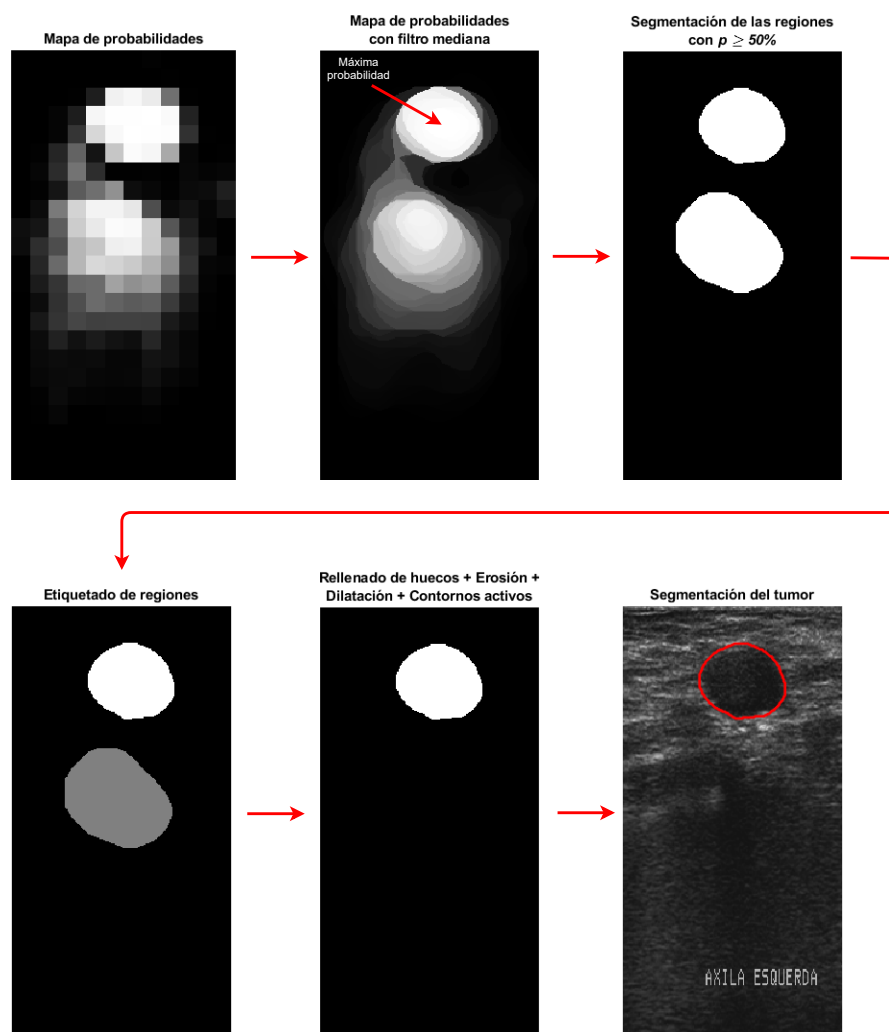


Figura 4.7: Etapas de post-procesamiento.

el módulo decodificador cuenta con una capa de convolución transpuesta que se encarga de regresar los mapas de características en un mapa de segmentación del tamaño original de la imagen. La base para el modelo FCN es la arquitectura AlexNet [30].

SegNet es una arquitectura codificador-decodificador basada en la familia VGG, propuesto por Badrinarayanan *et al.* [6]. La red codificadora está compuesta por varias capas de convolución y *max-pooling* para producir mapas de características de bajo nivel. Por otro lado, la red decodificadora transforma los mapas de características de bajo nivel al tamaño original de la imagen usando los índices de *max-pooling* correspondientes a los mapas de características del codificador. De este

modo, es posible retener detalles de alta frecuencia en las imágenes segmentadas. La arquitectura VGG16 se utiliza como modelo basado en SegNet [54].

También se consideró la arquitectura SegNet entrenada desde cero, es decir, sin usar transferencia de aprendizaje a partir de VGG16. Se utilizó la arquitectura codificador-decodificador propuesta en [54]. La imagen de entrada es de tamaño $256 \times 256 \times 3$. Además, para la red codificadora se realiza la convolución con un banco de filtros para producir un conjunto de mapas de características. A continuación, se normaliza cada elemento y se rectifica con la función ReLU. Posteriormente, se realiza el *max-pooling* con una ventana de 2×2 con un *stride* de 2. Para la red decodificadora, se realiza un muestreo ascendente de los mapas de características usando los índices de *max-pooling* de los mapas de características para generar la imagen segmentada del tamaño original.

La familia DeepLab es propuesta por investigadores de Google, donde la tercera versión llamada DeepLabV3+ superó a modelos anteriores de DeepLab [11]. La idea básica es construir redes de segmentación semántica más profundas mediante el uso de convoluciones dilatadas, que aumentan el espacio entre los pesos de convolución, sin aumentar el número de pesos. La distancia entre cada peso se llama tasa de dilatación. El codificador DeepLabV3+ utiliza un operador llamado *Atrous Spatial Pyramid Pooling*, que utiliza diferentes tasas de dilatación para capturar información multiescala. El decodificador realiza un muestreo ascendente bilineal de los mapas de características y los concatena con sus correspondientes características de bajo nivel. ResNet50 es la arquitectura base para el modelo DeepLabV3+ [23].

En general, las arquitecturas CNN para la segmentación semántica están compuestas por una red codificadora que permite extraer mapas de características y una red decodificadora que genera la segmentación de la imagen de entrada a partir de los mapas de características. La Figura 4.8 muestra el modelo general empleado para la segmentación semántica de imágenes BUS mediante el enfoque convolucional. La profundidad de la red varía en cada CNN, es decir, el número total de capas de convolución y agrupamiento. La tasa de aprendizaje se define de manera empírica para cada CNN. Además, usando el aumento artificial de datos se crearon versiones de cada imagen con traslaciones

en los ejes x e y , y espejo en eje x . También, se consideran 150 épocas para el entrenamiento con un factor *momentum* de 0.9. Además, se utiliza un tamaño de *mini-batch* de 8 y un conjunto de validación para evitar el sobre-entrenamiento de la red. La Tabla 4.3 resume los parámetros de entrenamiento para cada CNN.

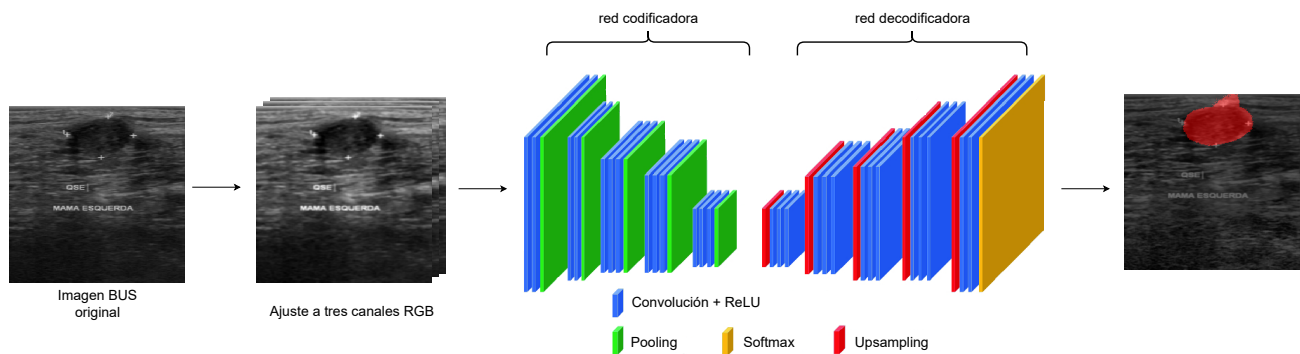


Figura 4.8: Proceso de entrenamiento para la segmentación semántica de imágenes BUS con una CNN.

	Espejo en X	Traslación en X	Traslación en Y	Tasa de aprendizaje
AlexNet	Si	[-20,20]	[-20,20]	1×10^{-3}
ResNet50	Si	[-10,10]	[-10,10]	1×10^{-4}
SegNet	Si	[-20,20]	[-20,20]	1×10^{-3}
VGG16	Si	[-10,10]	[-10,10]	1×10^{-3}

Tabla 4.3: Parámetros de entrenamiento del enfoque convolucional.

4.4 Validación

Para evaluar y garantizar la independencia de la partición entre el conjunto de entrenamiento y prueba, el método validación cruzada con k dobleces fue utilizado. El conjunto de 1,200 casos de imágenes de BUS fue dividido de manera aleatoria en k subconjuntos disjuntos del mismo tamaño. Una configuración común es validación cruzada con $k = 10$ dobleces, donde el promedio de los resultados se toman como el rendimiento de la segmentación. En cada una de las iteraciones, se

toma el 10 % del conjunto de imágenes para el conjunto de prueba mientras que el 90 % para el conjunto de entrenamiento.

A partir de la matriz de confusión de dos clases se pueden calcular los índices de validación para cada uno de los grupos de entrenamiento. El desempeño final de la segmentación está dado por la media de los k resultados obtenidos.

Finalmente, se realiza la prueba de Shapiro-Wilk para confirmar que los datos provienen de una distribución no normal, por lo que se utilizan métodos no paramétricos para el análisis de los valores obtenidos en la experimentación. De este modo, se realiza un análisis con los métodos Wilcoxon y Kruskal-Wallis para demostrar si hay diferencias significativas entre los métodos con el objetivo de elegir el método con mejor desempeño.

4.5 Infraestructura computacional

La infraestructura para ejecutar los algoritmos y experimentos está compuesta por:

1. Equipo de cómputo A:

- Procesador: Intel®Core™i7-7700 @ 3.60 GHz
- Memoria RAM: 32 GB DDR4 1600 MHz
- Almacenamiento: 1 TB SSD
- GPU: NVIDIA GeForce GTX 1050 Ti

2. Equipo de cómputo B:

- Procesador: Intel®Core™i9-9900 @ 3.60 GHz
- Memoria RAM: 64 GB DDR4 2133 MHz
- Almacenamiento: 1 TB SSD

- GPU: NVIDIA GeForce RTX 2070
3. MATLAB R2019b, Mathworks, Massachusetts, Estados Unidos.
 4. Deep Learning Toolbox

4.6 Sumario

En este capítulo se presentó el conjunto de imágenes BUS, seguido de la metodología utilizada para la segmentación del tumor con los enfoques convencional y convolucional. También se muestra el método de validación y la infraestructura computacional.

5

Resultados

En este capítulo se presentan los rendimientos de segmentación obtenidos en términos de los índices de validación para ambos enfoques, convencional y convolucional, considerando el método de validación cruzada con diez dobleces. Para el enfoque convencional, se realizó la experimentación con la combinación de tres configuraciones: método de reducción de dimensionalidad de características, tamaño de ventana y tipo de agrupación, mientras que para el enfoque convolucional se realizaron los experimentos variando la arquitectura de la CNN usando el mismo número de dobleces. Finalmente, se hace una comparación estadística del desempeño de la segmentación entre los mejores modelos obtenidos para cada enfoque.

5.1 Enfoque Convencional

En las Tablas 5.1 y 5.2 se resumen los resultados de desempeño para la segmentación del tumor con el enfoque convencional, considerando la combinación de tres parámetros: tipo de agrupación (*mean-* o *max-pooling*), tamaño de ventana de agrupación (4, 8 ó 16) y método de reducción de

dimensionalidad (APCA o MRMR), utilizando un perceptrón multicapa.

	$w_p = 4$		$w_p = 8$		$w_p = 16$	
	APCA	MRMR	APCA	MRMR	APCA	MRMR
<i>MCC</i>	0.7325	0.6685	0.7395	0.7061	0.7438	0.7612
<i>ACC</i>	0.9439	0.9286	0.9470	0.9386	0.9493	0.9528
<i>BAC</i>	0.8959	0.8741	0.8981	0.8816	0.8972	0.9023
<i>F1s</i>	0.7528	0.6914	0.7592	0.7259	0.7619	0.7778
<i>IoU</i>	0.6036	0.5283	0.6119	0.5697	0.6154	0.6364
<i>SEN</i>	0.8344	0.7946	0.8327	0.7941	0.8293	0.8378
<i>SPE</i>	0.9704	0.9578	0.9733	0.9720	0.9741	0.9764

Tabla 5.1: Resultados de segmentación de la MLP con agrupamiento *mean-pooling*.

	$w_p = 4$		$w_p = 8$		$w_p = 16$	
	APCA	MRMR	APCA	MRMR	APCA	MRMR
<i>MCC</i>	0.7293	0.6913	0.7334	0.7410	0.7341	0.7632
<i>ACC</i>	0.9425	0.9363	0.9451	0.9470	0.9471	0.9524
<i>BAC</i>	0.9032	0.8819	0.8954	0.9010	0.8930	0.9091
<i>F1s</i>	0.7486	0.7151	0.7551	0.7579	0.7534	0.7826
<i>IoU</i>	0.5983	0.5565	0.6065	0.6102	0.6043	0.6429
<i>SEN</i>	0.8455	0.8061	0.8277	0.8353	0.8200	0.8537
<i>SPE</i>	0.9683	0.9654	0.9725	0.9730	0.9707	0.9749

Tabla 5.2: Resultados de segmentación de la MLP con agrupamiento *max-pooling*.

Se puede observar que los valores más altos se obtienen cuando se considera un tamaño de ventana $w_p = 16$ y el método de reducción de dimensionalidad MRMR, independientemente del tipo de agrupamiento. Además, existe una ligera variación en los valores obtenidos por el método APCA con diferentes tamaño de ventana w_p , mientras que para el método MRMR su desempeño incrementa conforme aumenta el tamaño de la ventana w_p .

Las Tablas 5.3 y 5.4 muestran los valores- p obtenidos con la prueba de Wilcoxon entre los métodos de reducción de características por tamaño de ventana w_p . El símbolo “+” indica que APCA es superior a MRMR. Contrariamente, el símbolo “-” indica que APCA es inferior a MRMR. El símbolo “=” indica que no hay diferencia significativa entre métodos.

APCA vs. MRMR	MCC	ACC	BAC	F1s	IoU	SEN	SPE
$w_p = 4$	< 0.01(+)	< 0.01(+)	< 0.01(+)	< 0.01(+)	< 0.01(+)	< 0.01(+)	< 0.01(+)
$w_p = 8$	< 0.01(+)	< 0.01(+)	< 0.01(+)	< 0.01(+)	< 0.01(+)	< 0.01(+)	0.31(=)
$w_p = 16$	0.02(-)	0.03(-)	0.07(-)	0.04(-)	0.04(-)	0.13(=)	0.02(-)

Tabla 5.3: Prueba de Wilcoxon entre APCA y MRMR por tamaño de ventana con agrupación *mean-pooling*.

APCA vs. MRMR	MCC	ACC	BAC	F1s	IoU	SEN	SPE
$w_p = 4$	< 0.01(+)	< 0.01(+)	< 0.01(+)	< 0.01(+)	< 0.01(+)	< 0.01(+)	0.27(=)
$w_p = 8$	0.43(=)	0.75(=)	0.12(=)	0.55(=)	0.55(=)	0.06(=)	0.92(=)
$w_p = 16$	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)

Tabla 5.4: Prueba de Wilcoxon entre APCA y MRMR por tamaño de ventana con agrupación *max-pooling*.

El análisis de Wilcoxon demostró que al utilizar un tamaño de ventana $w_p = 4$, estadísticamente es mejor el método APCA, independientemente del tipo de agrupación utilizado. Además, cuando se utiliza un tamaño $w_p = 8$, el método APCA se considera estadísticamente mejor que el método MRMR cuando se combina con una agrupación de tipo *mean*, mientras que al combinarlo con la agrupación tipo *max* no hay evidencia estadística que demuestre que un método se comporte mejor que otro. Sin embargo, cuando se utiliza el método MRMR con un tamaño $w_p = 16$ se obtiene un mejor desempeño.

En las Tablas 5.5 y 5.6 se resume el análisis estadístico de Kruskal-Wallis con corrección de Bonferroni para análisis multicomparativo. El símbolo “+” indica que el Método A es superior al Método B. Contrariamente, el símbolo “-” indica que el Método A es inferior al Método B. Finalmente, el símbolo “=” indica que no hay diferencia significativa entre métodos. La prueba

estadística demostró que al utilizar el método APCA, no hay diferencia significativa en utilizar distintos tamaños de ventana w_p , ni el tipo de agrupación. Sin embargo, cuando se utiliza el método MRMR, se demuestra un mejor desempeño conforme se considera un tamaño de ventana w_p mayor.

Método A	Método B	MCC	ACC	BAC	F1s	IoU	SEN	SPE
APCA $_{w_{p4}}$	APCA $_{w_{p8}}$	0.53(=)	0.12(=)	1.00(=)	0.57(=)	0.57(=)	1.00(=)	0.06(=)
APCA $_{w_{p4}}$	APCA $_{w_{p16}}$	0.29(=)	0.01(-)	1.00(=)	0.29(=)	0.29(=)	1.00(=)	0.02(-)
APCA $_{w_{p8}}$	APCA $_{w_{p16}}$	1.00(=)	1.00(=)	1.00(=)	1.00(=)	1.00(=)	1.00(=)	1.00(=)
MRMR $_{w_{p4}}$	MRMR $_{w_{p8}}$	< 0.01(-)	< 0.01(-)	0.31(=)	< 0.01(-)	< 0.01(-)	1.00(=)	< 0.01(-)
MRMR $_{w_{p4}}$	MRMR $_{w_{p16}}$	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)
MRMR $_{w_{p8}}$	MRMR $_{w_{p16}}$	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)

Tabla 5.5: Prueba de Kruskal-Wallis entre métodos para agrupación *mean-pooling*.

Método A	Método B	MCC	ACC	BAC	F1s	IoU	SEN	SPE
APCA $_{w_{p4}}$	APCA $_{w_{p8}}$	0.71(=)	0.03(-)	0.83(=)	0.70(=)	0.70(=)	0.11(=)	< 0.01(-)
APCA $_{w_{p4}}$	APCA $_{w_{p16}}$	1.00(=)	0.09(=)	1.00(=)	1.00(=)	1.00(=)	0.26(=)	0.01(-)
APCA $_{w_{p8}}$	APCA $_{w_{p16}}$	1.00(=)	1.00(=)	1.00(=)	1.00(=)	1.00(=)	1.00(=)	1.00(=)
MRMR $_{w_{p4}}$	MRMR $_{w_{p8}}$	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)
MRMR $_{w_{p4}}$	MRMR $_{w_{p16}}$	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)
MRMR $_{w_{p8}}$	MRMR $_{w_{p16}}$	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	< 0.01(-)	0.08(=)	< 0.01(-)

Tabla 5.6: Prueba de Kruskal-Wallis entre métodos para agrupación *max-pooling*.

El análisis de Kruskal-Wallis demuestra que el mejor modelo para el enfoque convencional tiene la configuración: tipo de agrupación *max-pooling*, tamaño de ventana $w_p = 16$ y MRMR como método de reducción de dimensionalidad.

En la Tabla 5.7 se muestra el porcentaje de reducción del número de características de acuerdo con las configuraciones utilizadas en la experimentación. Se puede notar que para el método APCA preserva casi el mismo número de características en cada experimento, mientras que para MRMR va considerando más características conforme se considera un tamaño de ventana w_p mayor. El número de características que utiliza en el entrenamiento se ve reflejado en la segmentación del tumor, ya que a menor número de características seleccionadas se tiene bajo desempeño de los modelos. Por otro lado, al considerar más características tiene más información, lo que genera una mejor segmentación

del tumor.

	$w_p = 4$		$w_p = 8$		$w_p = 16$	
	APCA	MRMR	APCA	MRMR	APCA	MRMR
<i>mean-pooling</i>	81.48 %	96.58 %	80.66 %	94.32 %	81.07 %	75.47 %
<i>max-pooling</i>	79.42 %	91.55 %	78.19 %	84.24 %	78.11 %	71.52 %

Tabla 5.7: Porcentaje de reducción de características.

La Tabla 5.8 muestra la comparación entre los resultados obtenidos antes y después de realizar la etapa de post-procesamiento considerando el modelo con la mejor configuración. Se observa que agregar la etapa de post-procesamiento mejora el desempeño de segmentación. La Figura 5.1 muestra ejemplos comparativos de la segmentación del tumor sin y con post-procesamiento, donde se observa que la salida de la MLP genera regiones cuadrículadas, lo cual abarca algunos píxeles de la clase fondo. Por otro lado, utilizar el post-procesamiento permite refinar la solución de la MLP con el mapa de probabilidades y evitar el exceso de píxeles de la clase fondo y eliminar aquellas regiones de falsos positivos. Cabe mencionar, que la mejora en el desempeño global de la segmentación se debe al efecto de la clase fondo, ya que subió el valor de la especificidad, pero se degradó notablemente en la sensibilidad. Por esta razón, la etapa de post-procesamiento puede no ser efectivo en algunos casos, ya que depende del conjunto de técnicas utilizadas.

	<i>MCC</i>	<i>ACC</i>	<i>BAC</i>	<i>F1s</i>	<i>IoU</i>	<i>SEN</i>	<i>SPE</i>
Sin post-procesamiento	0.7632	0.9524	0.9091	0.7826	0.6429	0.8537	0.9749
Con post-procesamiento	0.7804	0.9572	0.8854	0.7969	0.6623	0.7976	0.9853

Tabla 5.8: Resultados de la MLP con *max-pooling* y post-procesamiento.

Finalmente, la Tabla 5.9 muestra el desempeño obtenido por tipo de lesión con base en la mejor configuración. La segmentación de tumores benignos tiene mejor desempeño ya que las formas de los tumores son más regulares, mientras que para los tumores malignos, el desempeño de segmentación

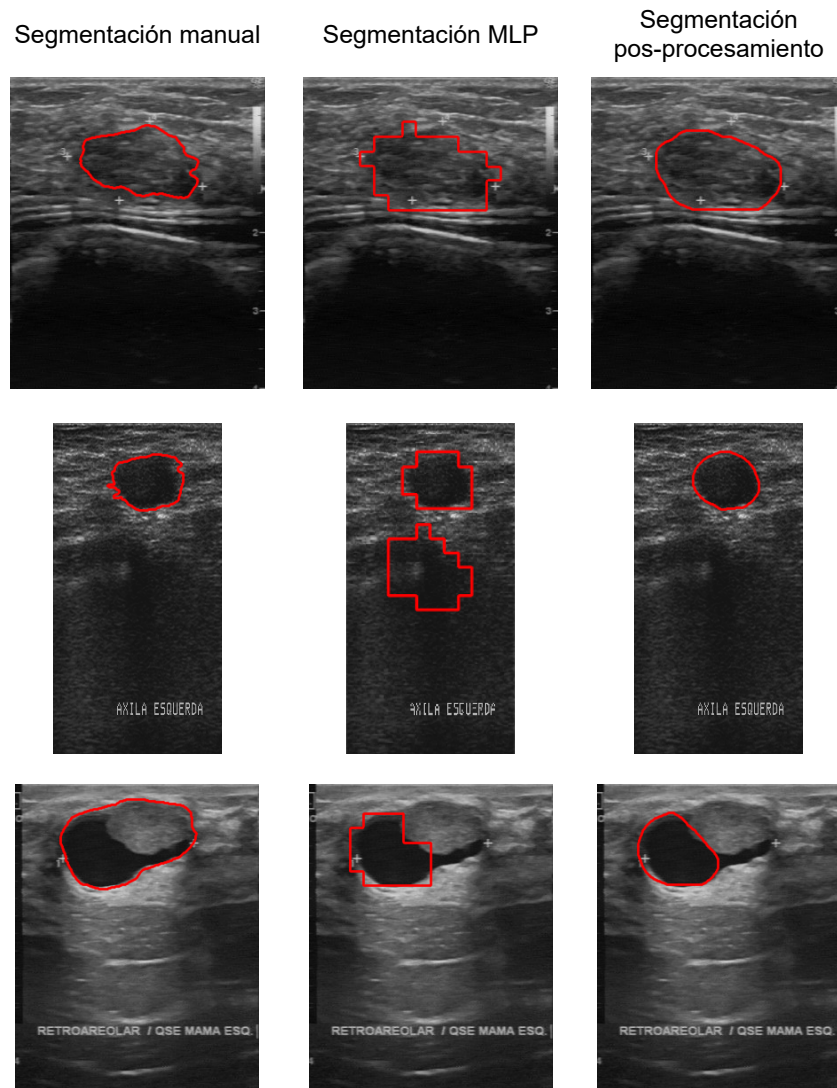


Figura 5.1: Ejemplos de segmentación del tumor con enfoque convencional.

disminuye debido a las formas irregulares y complejas, ya que el proceso de reducción de tamaño de la imagen remueve detalles finos del contorno y deja solamente el núcleo del tumor.

	<i>MCC</i>	<i>ACC</i>	<i>BAC</i>	<i>F1_s</i>	<i>IoU</i>	<i>SEN</i>	<i>SPE</i>
Benignos	0.7960	0.9634	0.8993	0.8115	0.6828	0.8192	0.9862
Malignos	0.7536	0.9465	0.8731	0.7732	0.6302	0.7693	0.9829

Tabla 5.9: Resultados de segmentación por tipo de lesión.

5.2 Enfoque Convolutacional

En la Tabla 5.10 se resumen los resultados obtenidos de las CNNs utilizando cuatro arquitecturas distintas, en donde se observa que la arquitectura ResNet50 tiene el desempeño más alto en varios índices de validación. Comparado con el enfoque convencional, se puede observar que el enfoque convolutacional ha sido superior con las arquitecturas VGG16 y ResNet50, aunque en los casos de AlexNet y SegNet, el enfoque convencional muestra un desempeño superior.

	AlexNet	ResNet50	SegNet	VGG16
<i>MCC</i>	0.7632	0.8817	0.6623	0.8766
<i>ACC</i>	0.9787	0.9754	0.9702	0.9736
<i>BAC</i>	0.9630	0.9272	0.7946	0.9639
<i>F1s</i>	0.7629	0.8926	0.6727	0.8879
<i>IoU</i>	0.6167	0.8060	0.5068	0.7984
<i>SEN</i>	0.9653	0.8603	0.6036	0.9581
<i>SPE</i>	0.9802	0.9952	0.9918	0.9815

Tabla 5.10: Resultados de segmentación con cuatro arquitecturas CNN.

Además, en la Tabla 5.11 se resume el análisis estadístico con base en los resultados obtenidos de las CNNs, la cual muestra que ResNet50 y VGG16 son estadísticamente mejor que AlexNet y SegNet.

Método A	Método B	MCC	ACC	BAC	F1s	IoU	SEN	SPE
AlexNet	ResNet	< 0.001(-)	< 0.001(+)	< 0.001(+)	< 0.001(-)	< 0.001(-)	< 0.001(+)	< 0.001(-)
AlexNet	SegNet	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(-)
AlexNet	VGG16	< 0.001(-)	< 0.001(=)	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)	0.09(+)
ResNet50	SegNet	< 0.001(+)	0.44(+)	< 0.001(-)	< 0.001(-)	< 0.001(-)	< 0.001(+)	< 0.001(=)
ResNet50	VGG16	1.00(=)	0.05(+)	< 0.001(-)	< 0.001(=)	< 0.001(=)	< 0.001(-)	< 0.001(+)
SegNet	VGG16	< 0.001(-)	1.00(=)	< 0.001(-)	< 0.001(-)	< 0.001(-)	< 0.001(-)	< 0.001(+)

Tabla 5.11: Prueba de Kruskal-Wallis entre arquitecturas CNN.

La Figura 5.2 muestra una comparativa entre los resultados de segmentación obtenidos con las arquitecturas CNNs sin tomar en cuenta la etapa de post-procesamiento debido a que la CNN entrega buenos resultados de la segmentación del tumor.

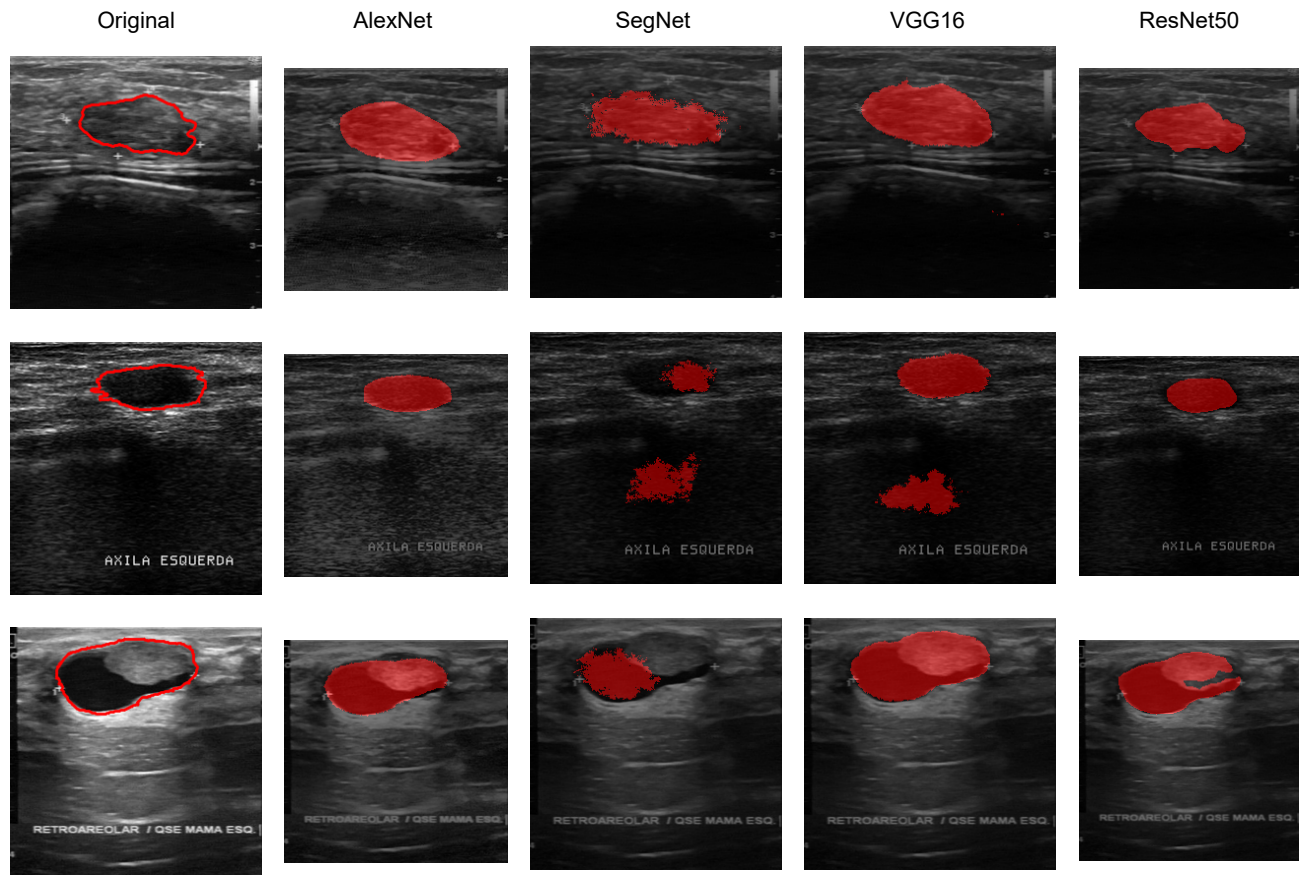


Figura 5.2: Ejemplos de segmentación del tumor con enfoque convolutacional.

Adicionalmente, se realizó un experimento en el que solo se consideró las arquitecturas VGG16 y ResNet50 para la segmentación del tumor de acuerdo al tipo de tumor. En las Tablas 5.12 se muestran los desempeños de segmentación para cada arquitectura, donde se puede notar que ambas arquitecturas tienen mejor desempeño para la segmentación de los tumores benignos a diferencia de los tumores malignos. Al igual que en el enfoque convencional, este comportamiento se repite debido a que en la reducción de tamaño de la imagen de entrada que hacen ambos enfoques, se suavizan detalles finos de los tumores malignos, los cuales suelen tener formas irregulares y márgenes angulares y espiculados.

Finalmente, la Tabla 5.13 muestra la prueba de Kruskal-Wallis entre el mejor modelo entrenado del enfoque convencional y los modelos del enfoque convolutacional, donde se puede evidenciar que el

		<i>MCC</i>	<i>ACC</i>	<i>BAC</i>	<i>F1s</i>	<i>IoU</i>	<i>SEN</i>	<i>SPE</i>
VGG16	Benignos	0.8613	0.9695	0.9593	0.8703	0.7813	0.9448	0.9737
	Malignos	0.8004	0.9485	0.9165	0.8185	0.7176	0.8678	0.9653
ResNet50	Benignos	0.8779	0.9736	0.9296	0.8873	0.8083	0.8688	0.9904
	Malignos	0.7863	0.9448	0.8685	0.8024	0.6979	0.7532	0.9839

Tabla 5.12: Resultados de segmentación por tipo de tumor.

enfoque convencional es mejor que las CNN menos profundas, AlexNet y SegNet, a diferencia de las que tienen mayor profundidad como es el caso de ResNet50 y VGG16.

Método A	Método B	MCC	ACC	BAC	F1s	IoU	SEN	SPE
AlexNet	Convencional	< 0.001(-)	< 0.001(+)	< 0.001(+)	< 0.001(-)	< 0.001(-)	< 0.001(+)	< 0.001(-)
ResNet50	Convencional	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)
SegNet	Convencional	< 0.001(-)	< 0.001(+)	< 0.001(-)	< 0.001(-)	< 0.001(-)	< 0.001(-)	< 0.001(+)
VGG16	Convencional	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(+)	< 0.001(-)

Tabla 5.13: Prueba de Kruskal-Walis entre enfoques convencional y convolucional.

Con esto se demuestra que el modelo entrenado con una arquitectura ResNet50 supera notablemente el desempeño del modelo con enfoque convencional en cada uno de los índices de validación, ya que es la arquitectura con mayor profundidad del enfoque convolucional. Por esta razón, le permite obtener más información descriptiva para cada píxel. Además, el enfoque convencional se enfrenta al desafío de encontrar características que permitan distinguir mejor los píxeles para cada una de las clases. No obstante, el método convencional superó en general a las arquitectas CNN menos profundas (i.e., AlexNet y SegNet), lo cual reafirma la observación de que a mayor profundidad de la red, mejor la descripción de la imagen. Cabe destacar que la comparación de los enfoques convencional y convolucional se realizó sobre el mismo conjunto de imágenes, lo cual permitió una evaluación objetiva en igualdad de condiciones. Esto implica que en el desarrollo de un sistema CAD, es conveniente incorporar a la CNN ResNet50 para la segmentación automática de tumores de mama.

5.3 Sumario

En este capítulo, se mostraron los resultados obtenidos para la segmentación semántica con los enfoques convencional y convolucional en términos de los índices de validación, y la comparación entre los mejores modelos entrenados de ambos enfoques. Además, se realizó una comparación estadística para cada uno de ellos. De acuerdo con los resultados obtenidos, la mejor configuración para el enfoque convencional es utilizar el método de reducción de dimensionalidad MRMR con agrupación *max-pooling* y un tamaño de ventana $w_p = 16$, mientras que para el enfoque convolucional el mejor desempeño se obtiene con la arquitectura ResNet50.

6

Conclusiones y trabajo futuro

En este capítulo se presentan las conclusiones obtenidas en este estudio. Además se menciona la contribución al estado del arte y se describe el trabajo futuro.

6.1 Conclusiones

Con este estudio se observaron tres puntos importantes para el enfoque convencional. Primero, cuando se utiliza una ventana w_p de mayor tamaño se obtiene un desempeño mejor en la segmentación del tumor. Esto se debe a que la reducción del tamaño de la imagen actúa como un filtro paso bajas, en donde se preserva información local de las estructuras en escalas más grandes (por ejemplo los tumores), mientras que se remueven las estructuras más pequeñas (detalles finos del tumor). Segundo, utilizar el método de reducción de dimensionalidad MRMR obtiene mejor desempeño que APCA debido a que utiliza información mutua para encontrar relaciones no lineales entre características reduciendo hasta un 75.47%, mientras que APCA realiza un mapeo lineal para mejorar la representación de características, lo que permite reducir hasta un 71.52% del total de

características, lo cual también influye en la reducción del desempeño de segmentación cuando se usa APCA. Tercero, la etapa de post-procesamiento permite mejorar la segmentación del tumor al combinarla con el mapa de probabilidades, ya que refina el contorno de la segmentación y descarta píxeles mal clasificados como tumor. Sin embargo, existen casos en el que la especificidad aumenta y la sensibilidad se reduce, lo cual no es conveniente. Por esta razón, la etapa de post-procesamiento es crítica y requiere de mayor investigación.

Cabe señalar que este trabajo reveló la dificultad que presenta el enfoque convencional para encontrar características discriminantes que permitan diferenciar entre las regiones tumor y tejido adyacente (fondo). Además, el modelo obtenido puede considerarse competitivo con lo reportado por los enfoques convencionales del estado del arte, considerando que los resultados se midieron con distintos conjuntos de imágenes. Sin embargo, se puede inferir que probablemente se esté obteniendo el desempeño máximo que un enfoque convencional puede lograr, ya que es desafiante determinar características capaces de distinguir cambios sutiles en imágenes de clases diferentes, lo cual se logra mediante una CNN.

Por otro lado, en el enfoque convolucional se observó que la arquitectura ResNet50 tiene mejor desempeño en la segmentación del tumor. Esto se debe a que es la arquitectura con mayor profundidad, ya que las primeras capas de convolución extraen características generales como brillo, formas, bordes, etc., mientras que las capas más profundas pueden describir a cada píxel de la imagen de una manera más sutil. Además, la arquitectura VGG16 también mostró tener un desempeño similar a ResNet50 a pesar de tener menor profundidad, aunque se compensa por su mayor número de parámetros entrenables (i.e., coeficientes de convolución).

Además, el estudio demostró que el enfoque convencional puede superar al convolucional bajo ciertas condiciones, y esto es posible cuando las arquitecturas CNN cuentan con poca profundidad para extraer características, como se presentó en el caso de las arquitecturas AlexNet y SegNet.

Con base en lo anterior, se puede responder a la hipótesis planteada que sí es posible obtener desempeños equiparables, e inclusive superiores, entre los enfoques convencional y convolucional

cuando las CNNs son poco profundas. Sin embargo, en general, es más conveniente utilizar el enfoque convolucional para la segmentación de tumores de mama en imágenes BUS a pesar del costo computacional que demanda, ya que omite la dependencia de las habilidades de un experto para establecer las características y se obtienen resultados de segmentación significativamente mejores que el método convencional.

Por último, los métodos convencionales pueden complementar al enfoque convolucional, mediante enfoques híbridos que combinen las mejores prestaciones de cada uno. Esto es básicamente para no descartar toda la teoría que se involucra en los métodos convencionales, sino que puede fortalecer el desempeño de un enfoque convolucional.

6.2 Contribución

Esta tesis representa el primer trabajo que compara experimentalmente el desempeño de sistemas CADe basado en los enfoques convencional y convolucional, usando el mismo conjunto de imágenes BUS. La principal contribución de este trabajo es un análisis comparativo, de manera justa, entre los enfoques convencional y convolucional, debido a que se entrenaron con el mismo conjunto de imágenes. Otra contribución es la implementación de un enfoque convencional que puede ser manipulado por módulos para mejorar o incorporar nuevas técnicas. También, obtuve CNNs entrenadas para la segmentación de BUS que pueden mejorarse mediante transferencia de aprendizaje en nuevas bases de datos BUS.

6.3 Trabajo futuro

El método propuesto por el enfoque convolucional demuestra ser competitivo comparado con los sistemas CADe en el estado del arte. Por lo tanto, sería interesante extender el problema de la segmentación de tumores de mama hasta la clasificación en clases benigno o maligno utilizando

redes neuronales convolucionales. Además, existen más arquitecturas CNN que posiblemente no se han implementado para este problema, lo que podría dar a conocer una nueva arquitectura que se desempeñe mejor en esta tarea. De esta forma, es factible obtener un sistema completamente automático para el análisis de imágenes BUS. Además, explorar la combinación de las características del método convencional con las aprendidas por una CNN para tratar de mejorar los desempeños de segmentación.

Bibliografía

- [1] Abramowitz, M. (1974). *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*,. Dover Publications, Inc., USA.
- [2] Abugattas Saba, J., Manrique Hinojosa, J., and Vidaurre Rojas, T. (2015). Mamografía como instrumento de tamizaje en cáncer de mama. *Revista Peruana de Ginecología y Obstetricia*, 61:311 – 319.
- [3] Acharya, U. R., Ng, W. L., Rahmat, K., Sudarshan, V. K., Koh, J. E., Tan, J. H., Hagiwara, Y., Yeong, C. H., and Ng, K. H. (2017). Data mining framework for breast lesion classification in shear wave ultrasound: A hybrid feature paradigm. *Biomedical Signal Processing and Control*, 33:400 – 410.
- [4] Ahmed, L., Iqbal, M., and Aldabbas, H. (2020). Images data practices for semantic segmentation of breast cancer using deep neural network. *Journal of Ambient Intelligence and Humanized Computing*.
- [5] American Society of Clinical Oncology (ASCO) (27-07-2019). Cáncer de mama: Introducción. <https://www.cancer.net/es/tipos-de-cancer/cancer-de-mama/introduccion>.
- [6] Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495.
- [7] Bouchard, G. (2007). Efficient bounds for the softmax function, applications to inference in hybrid

- models. In *Presentation at the Workshop for Approximate Bayesian Inference in Continuous/Hybrid Systems at NIPS-07*. Citeseer.
- [8] Cao, Z., Duan, L., Yang, G., Yue, T., and Chen, Q. (2019). An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. In *BMC Medical Imaging*.
- [9] Carbajal-Degante, E., Olveres, J., and Escalante-Ramírez, B. (2018). A multiphase active contour model based on the Hermite transform for texture segmentation. In Schelkens, P., Ebrahimi, T., and Cristóbal, G., editors, *Optics, Photonics, and Digital Technologies for Imaging Applications V*, volume 10679, pages 364 – 372. International Society for Optics and Photonics, SPIE.
- [10] Chaitanya, K., Karani, N., Baumgartner, C. F., Becker, A., Donati, O., and Konukoglu, E. (2019). Semi-supervised and task-driven data augmentation. In *International conference on information processing in medical imaging*, pages 29–41. Springer.
- [11] Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
- [12] Cheng, H., Shan, J., Ju, W., Guo, Y., and Zhang, L. (2010). Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition*, 43:299–317.
- [13] Chiao, J.-Y., Chen, K.-Y., kai Liao, K. Y., Hsieh, P.-H., Zhang, G. G., and Huang, T.-C. (2019). Detection and classification the breast tumors using mask r-cnn on sonograms. *Medicine*, 98.
- [14] de la Salud, O. M. (8-07-2019). *Radiación ionizante*.
- [15] de Lima, S. M., da Silva-Filho, A. G., and dos Santos, W. P. (2016). Detection and classification of masses in mammographic images in a multi-kernel approach. *Computer Methods and Programs in Biomedicine*, 134:11 – 29.

- [16] Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D., Piñeros, M., Znaor, A., and Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International Journal of Cancer*, 144(8):1941–1953.
- [17] Firmino, M., Angelo, G., Morais, H., Dantas, M. R., and Valentim, R. (2016). Computer-aided detection (cade) and diagnosis (cadx) system for lung cancer with likelihood of malignancy. *Biomedical engineering online*, 15(1):1–17.
- [18] Gao, B. and Pavel, L. (2017). On the properties of the softmax function with application in game theory and reinforcement learning. *arXiv preprint arXiv:1704.00805*.
- [19] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., and Rodríguez, J. G. (2017). A review on deep learning techniques applied to semantic segmentation. *CoRR*, abs/1704.06857.
- [20] Gómez-Flores, W. and Hernández-López, J. (2020). Assessment of the invariance and discriminant power of morphological features under geometric transformations for breast tumor classification. *Computer Methods and Programs in Biomedicine*, 185:105173.
- [21] Gómez-Flores, W., Rodríguez-Cristerna, A., and de Albuquerque Pereira, W. C. (2019). Texture analysis based on auto-mutual information for classifying breast lesions with ultrasound. *Ultrasound in Medicine & Biology*, 45(8):2213 – 2225.
- [22] Gonzalez, R. C. and Woods, R. E. (2008). *Digital image processing*. Prentice Hall, Upper Saddle River, N.J.
- [23] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.
- [24] Hu, Y., Guo, Y., Wang, Y., Yu, J., Li, J., Zhou, S., and Chang, C. (2018). Automatic tumor

segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model. *Medical Physics*, 46.

- [25] Huang, K., Cheng, H. D., Zhang, Y., Zhang, B., Xing, P., and Ning, C. (2018). Medical knowledge constrained semantic breast ultrasound image segmentation. In *2018 24th International Conference on Pattern Recognition (ICPR)*.
- [26] Huang, Q., Huang, Y., Luo, Y., Yuan, F., and Li, X. (2020). Segmentation of breast ultrasound image with semantic classification of superpixels. *Medical Image Analysis*, 61:101657.
- [27] Huang, Q., Yang, F., Liu, L., and Li, X. (2014). Automatic segmentation of breast lesions for interaction in ultrasonic computer-aided diagnosis. *Information Sciences*, 314.
- [28] Jain, A. K. and Farrokhnia, F. (1991). Unsupervised texture segmentation using gabor filters. *Pattern recognition*, 24(12):1167–1186.
- [29] Jiang, X. (2008). Asymmetric principal component and discriminant analyses for pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):931–937.
- [30] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- [31] Kumar, V., Webb, J., Gregory, A., Denis, M., Meixner, D. D., Bayat, M., Whaley, D. H., Fatemi, M., and Alizad, A. (2018). Automated and real-time segmentation of suspicious breast masses using convolutional neural network. *PLoS ONE*, 13.
- [32] Lei, B., Huang, S., Li, R., Bian, C., Li, H., Chou, Y.-H., and Cheng, J.-Z. (2018). Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder–decoder network. *Neurocomputing*, 321:178 – 186.

- [33] Liu, B., Cheng, H., Huang, J., Tian, J., Tang, X., and Liu, J. (2010). Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. *Pattern Recognition*, 43(1):280 – 298.
- [34] Liu, X., Ounifi, H.-A., Gherbi, A., Li, W., and Cheriet, M. (2019). A hybrid gpu-fpga based design methodology for enhancing machine learning applications performance. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–15.
- [35] Liu, Y., Chen, Y., Han, B., Zhang, Y., Zhang, X., and Su, Y. (2018). Fully automatic breast ultrasound image segmentation based on fuzzy cellular automata framework. *Biomedical Signal Processing and Control*, 40:433 – 442.
- [36] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.
- [37] Loshchilov, I. and Hutter, F. (2017). Sgdr: Stochastic gradient descent with warm restarts. In *ICLR*.
- [38] López, R. F. and Fernández, J. M. F. (2008). *Las Redes Neuronales Artificiales*.
- [39] Marcomini, K., Carneiro, A. A., and Schiabel, H. (2016). Application of artificial neural network models in segmentation and classification of nodules in breast ultrasound digital images. *International Journal of Biomedical Imaging*, 2016:1–13.
- [40] Martens, J. . (1990). The hermite transform-theory. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(9):1595–1606.
- [41] Masotti, M. (2006). A ranklet-based image representation for mass classification in digital mammograms. *Medical physics*, 33 10:3951–61.

- [42] Milans, S. and Lavista, F. (2017). Introducción a la imagenología mamaria. *Tendencias en Medicina Nº 50: 133-140*.
- [43] Nava, R., Escalante-Ramírez, B., and Cristóbal, G. (2012). Texture image retrieval based on log-gabor features. In Alvarez, L., Mejail, M., Gomez, L., and Jacobo, J., editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 414–421, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [44] Pan, S. J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.
- [45] Pedamonti, D. (2018). Comparison of non-linear activation functions for deep neural networks on mnist classification task. *arXiv preprint arXiv:1804.02763*.
- [46] Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238.
- [47] RadiologyInfo.org (Fecha de consulta 15-06-2019). *Ultrasonido de Seno*.
- [48] Ramachandran, P., Zoph, B., and Le, Q. V. (2017). Searching for activation functions.
- [49] Rodríguez-Cristerna, A., Guerrero-Cedillo, C. P., Donati-Olvera, G. A., Gómez-Flores, W., and Pereira, W. C. A. (2017). Study of the impact of image preprocessing approaches on the segmentation and classification of breast lesions on ultrasound. In *2017 14th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, pages 1–4.
- [50] Rodrigues, R., Braz, R., Pereira, M., Moutinho, J., and Pinheiro, A. M. (2015). A two-step segmentation method for breast ultrasound masses based on multi-resolution analysis. *Ultrasound in Medicine & Biology*, 41(6):1737 – 1748.

- [51] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation.
- [52] Shan, J., Cheng, H., and Wang, Y. (2012). Completely automated segmentation approach for breast ultrasound images using multiple-domain features. *Ultrasound in Medicine and Biology*, 38(2):262 – 275.
- [53] Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60.
- [54] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [55] Soille, P. (2003). *Morphological Image Analysis: Principles and Applications*. Springer-Verlag, Berlin, Heidelberg, 2 edition.
- [56] Sprent, P. and Smeeton, N. C. (2016). *Applied nonparametric statistical methods*. CRC press.
- [57] Struc, V. and Pavesic, N. (2010). *From Gabor magnitude to Gabor phase features: tackling the problem of face recognition under severe illumination changes*. INTECH Open Access Publisher.
- [58] Walsh, J., O' Mahony, N., Campbell, S., Carvalho, A., Krpalkova, L., Velasco-Hernandez, G., Harapanahalli, S., and Riordan, D. (2019). Deep learning vs. traditional computer vision.
- [59] Wang, T., Wang, C., Zhou, X., and Chen, H. (2018). A survey of fpga based deep learning accelerators: Challenges and opportunities. *arXiv preprint arXiv:1901.04988*, pages 1–10.
- [60] Wilfrido, G. F. and Bedert Abel, R. O. (2016). New fully automated method for segmentation of breast lesions on ultrasound based on texture analysis. *Ultrasound in Medicine & Biology*.
- [61] Winsberg, F., Elkin, M., Macy, J., Bordaz, V., and Weymouth, W. (1967). Detection of

- radiographic abnormalities in mammograms by means of optical scanning and computer analysis
1. *Radiology*, 89:211–215.
- [62] Xian, M., Zhang, Y., Cheng, H., Xu, F., Zhang, B., and Ding, J. (2018). Automatic breast ultrasound image segmentation: A survey. *Pattern Recognition*, 79.
- [63] Xie, X., Shi, F., Niu, J., and Tang, X. (2018). *Breast Ultrasound Image Classification and Segmentation Using Convolutional Neural Networks: 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part III*, pages 200–211.
- [64] Xing, J., Li, Z., Wang, B., Yu, B., Zanjani, F. G., Zheng, A., Duits, R., and Tan, T. (2019). Automated segmentation of lesions in ultrasound using semi-pixel-wise cycle generative adversarial nets.
- [65] Yang, M., Moon, W. K., Wang, Y. F., Bae, M. S., Huang, C., Chen, J., and Chang, R. (2013). Robust texture analysis using multi-resolution gray-scale invariant features for breast sonographic tumor diagnosis. *IEEE Transactions on Medical Imaging*, 32(12):2262–2273.
- [66] Yap, M. H., Goyal, M., Osman, F., Ahmad, E., Martí, R., Denton, E., Juetten, A., and Zwiggelaar, R. (2018a). End-to-end breast ultrasound lesions recognition with a deep learning approach. In Gimi, B. and Krol, A., editors, *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10578, pages 287 – 294. International Society for Optics and Photonics, SPIE.
- [67] Yap, M. H., Pons, G., Martí, J., Ganau, S., M.Sentís, Zwiggelaar, R., Davison, A. K., and Martí, R. (2018b). Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics*, 22(4):1218–1226.
- [68] Zhuang, Z., Li, N., Raj, A. N. J., Mahesh, V. G. V., and Qiu, S. (2019). An rdau-net model for lesion segmentation in breast ultrasound images. *PLoS ONE*, 14.