



Percepción y **Sistemas** Inteligentes

SISTEMA DE VISIÓN ARTIFICIAL PARA LA DETECCIÓN AUTOMÁTICA DE MASAS Y TEJIDO NORMAL EN MAMOGRAFÍAS

ANGIE LORENA ARANGO IMBACHI

**UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA
PROGRAMA DE INGENIERÍA ELECTRÓNICA
SANTIAGO DE CALI
2025**



SISTEMA DE VISIÓN ARTIFICIAL PARA LA DETECCIÓN AUTOMÁTICA DE MASAS Y TEJIDO NORMAL EN MAMOGRAFÍAS

ANGIE LORENA ARANGO IMBACHI

**Trabajo presentado para optar
por el título de Ingeniera Electrónica**

**Director
Humberto Loaiza Correa, Ph.D.**

Grupo de Investigación Percepción y Sistemas Inteligentes

**UNIVERSIDAD DEL VALLE
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA ELÉCTRICA Y ELECTRÓNICA
PROGRAMA DE INGENIERÍA ELECTRÓNICA
SANTIAGO DE CALI**

2025

Nota de Aceptación

Firma del director

Firma del jurado

Firma del jurado

Santiago de Cali, 30 de marzo de 2025

AGRADECIMIENTOS

Agradezco a Dios, en primer lugar, por todas las bendiciones que me ha otorgado a lo largo de mi vida. A mi familia, quienes me han brindado su apoyo incondicional y todo lo necesario para alcanzar mis metas y objetivos. En especial, a mis padres: mi padre, quien, aunque ya no está en este mundo, permanece en mi recuerdo, y mi madre, quien ha sido mi mayor motivación. A ambos les debo la formación en principios sólidos y un enfoque claro hacia mis objetivos, los cuales he logrado alcanzar con su respaldo y sabios consejos.

Agradezco también a los amigos que la vida me ha dado; sus palabras de ánimo y comprensión a lo largo de estos años de formación han sido una fuente constante de energía en los momentos difíciles. A mis compañeros de laboratorio, quienes me acompañaron en las largas jornadas de trabajo, y especialmente por los momentos de risa que hicieron de mis últimos años en la universidad una experiencia inolvidable. Con cada uno de ellos compartí innumerables aprendizajes y recuerdos valiosos.

A los profesores que han sido parte fundamental en mi formación profesional, gracias por su sentido de responsabilidad y por buscar siempre las mejores estrategias para transmitir sus conocimientos. Agradezco especialmente a mi director de trabajo de grado, quien, con su guía y apoyo, me ha brindado el conocimiento y la confianza necesarios para desarrollar este proyecto y culminar mi carrera de manera exitosa.

TABLA DE CONTENIDO

<i>Nota de Aceptación</i>	<i>i</i>
---------------------------------	----------

RESUMEN VIII

ABSTRACT XI

1. INTRODUCCIÓN	1
1.1 CONTEXTUALIZACIÓN.....	2
1.2 PLANTEAMIENTO DEL PROBLEMA Y OBJETIVOS	3
1.2.1 <i>Planteamiento del problema</i>	3
1.2.2 <i>Objetivo General</i>	4
1.2.3 <i>Objetivos Específicos</i>	4
1.3 DESCRIPCIÓN DEL SISTEMA DESARROLLADO	4
1.4 ORGANIZACIÓN DE LOS CAPÍTULOS.....	6
2. ANTECEDENTES Y MARCO TEÓRICO	8
2.1 ANTECEDENTES	8
2.1.1 <i>Antecedentes Nacionales</i>	8
2.1.2 <i>Antecedentes Internacionales</i>	10
2.1.3 <i>Discusión</i>	12
2.2 MARCO TEÓRICO	14
2.2.1 <i>Aspectos conceptuales de la salud</i>	14
2.2.1.1 Cáncer de mama.....	14
2.2.1.2 Procedimientos para el diagnóstico del cáncer de mama	14
2.2.1.3 Diagnóstico por imágenes.....	14
2.2.1.4 Mamografías digitales	15
2.2.1.5 Masas en el tejido mamario, caracterización y clasificación.....	15
2.2.2 <i>Aspectos conceptuales para el aprendizaje automático</i>	15
2.2.2.1 Inteligencia artificial (IA)	15
2.2.2.2 Visión artificial	15
2.2.2.3 Reconocimiento de patrones.....	15
2.2.2.4 Aprendizaje automático (Machine Learning)	16
2.2.2.5 Máquinas de aprendizaje	16
2.2.3 <i>Aspectos conceptuales para el sistema de visión artificial</i>	16
2.2.3.1 Técnicas de procesamiento de imágenes	16
2.2.3.2 Técnicas de segmentación de imágenes	19
2.2.3.3 Técnicas de extracción de características en las imágenes.....	20
2.2.3.4 Técnicas de selección de características	24
2.2.3.5 Técnicas de clasificación automática	25
2.2.4 <i>Métricas de desempeño</i>	30
2.2.4.1 Exactitud (Accuracy)	31
2.2.4.2 Precisión (Presicion)	31
2.2.4.3 Sensibilidad (Recall).....	31
2.2.4.4 Especificidad (Specificity).....	32
2.2.4.5 Índice F1-Score	32
2.2.5 <i>Discusión</i>	32

3.	METODOLOGÍA DE LA SOLUCIÓN PROPUESTA: SOFTWARE	34
3.1	ETAPA 1. SELECCIÓN DE LA BASE DE DATOS.....	36
3.2	ETAPA 2. PROCESAMIENTO	38
3.3	ETAPA 3. EXTRACCIÓN DE CARACTERÍSTICAS	45
3.3.1	<i>Enventanado</i>	<i>47</i>
3.3.2	<i>Transformación al dominio frecuencial.....</i>	<i>51</i>
3.3.3	<i>Extracción de características</i>	<i>53</i>
3.4	ETAPA 4. SELECCIÓN DE CARACTERÍSTICAS	54
3.4.1	<i>Rankings.....</i>	<i>55</i>
3.4.2	<i>Wrappers</i>	<i>55</i>
3.5	ETAPA 5. TÉCNICAS DE CLASIFICACIÓN	56
3.5.1	<i>Aprendizaje superficial</i>	<i>56</i>
3.5.1.1	Clasificadores estadísticos	57
3.5.1.2	Máquinas de aprendizaje	58
3.5.1.3	Redes neuronales	59
3.5.1.4	Redes de comité	60
3.5.2	<i>Aprendizaje Profundo</i>	<i>61</i>
3.5.2.1	CNN	62
3.5.2.2	Aprendizaje por transferencia.....	64
3.6	CONCLUSIONES	66
4.	PRUEBAS Y RESULTADOS	69
4.1	DESCRIPCIÓN DEL DATASET	69
4.2	PRUEBAS Y RESULTADOS PARA ALGORITMOS DE SELECCIÓN DE CARACTERÍSTICAS	70
4.2.1	<i>Resultados de Ranking</i>	<i>70</i>
4.2.1.4	Disminución de características.....	73
4.2.2	<i>Resultados de Wrappers.....</i>	<i>75</i>
4.3	PRUEBAS Y RESULTADOS PARA EL APRENDIZAJE SUPERFICIAL.....	77
4.3.1	<i>Pruebas y resultados para el Dominio Espacial.....</i>	<i>78</i>
4.3.1.1	Resultados y análisis.....	78
4.3.2	<i>Pruebas y resultados para el Dominio Frecuencial.....</i>	<i>79</i>
4.3.2.1	Resultados y análisis.....	80
4.3.3	<i>Pruebas y resultados para la Fusión de Dominios Espacial y Frecuencial</i>	<i>81</i>
4.3.3.1	Resultados y análisis.....	81
4.3.4	<i>Pruebas y resultados para las Redes de Comité.....</i>	<i>82</i>
4.3.4.1	Resultados y análisis.....	83
4.4	PRUEBAS Y RESULTADOS PARA EL APRENDIZAJE PROFUNDO	84
4.4.1	<i>Pruebas y resultados para la CNN.....</i>	<i>85</i>
4.4.1.1	Evaluación en los Dominios Espacial y Frecuencial.....	85
4.4.1.2	Conversión de un modelo unimodal a multimodal mediante la fusión de los dominios espacial y frecuencial	87
4.4.1.3	Resultados y análisis.....	88
4.4.2	<i>Pruebas y resultados para el aprendizaje por transferencia.....</i>	<i>90</i>
4.4.2.1	Resultados y análisis.....	93
4.5	ANÁLISIS GENERAL DEL SISTEMA.....	94
4.6	ALCANCES Y LIMITACIONES DEL SISTEMA DESARROLLADO	96
4.7	DISCUSIÓN	99

5.	CONCLUSIONES GENERALES	102
6.	TRABAJOS FUTUROS.....	106
7.	REFERENCIAS BIBLIOGRÁFICAS.....	109

LISTA DE FIGURAS

FIGURA 1.1. FLUJO DE ETAPAS EN EL DESARROLLO DEL SISTEMA.	5
FIGURA 2.1. VISUALIZACIÓN DEL COMPORTAMIENTO DE SVM CON DIFERENTES KERNELS	27
FIGURA 2.2. MATRIZ DE CONFUSIÓN CON MÉTRICAS DE DESEMPEÑO.	31
FIGURA 3.1. DIAGRAMA DE BLOQUES DEL SISTEMA DESARROLLADO.	34
FIGURA 3.2. IMÁGENES MAMOGRAFÍAS CON ARTEFACTOS Y MALA DIGITALIZACIÓN.	37
FIGURA 3.3. DIAGRAMA DE BLOQUES AMPLIADO DE LA ETAPA 2. PROCESAMIENTO.	38
FIGURA 3.4. SEGMENTACIÓN DE MAMOGRAFÍA CON LA ETAPA 2. PROCESAMIENTO: A) MAMOGRAFÍA ORIGINAL 'MDB012.PNG'; B) MÁSCARA OBTENIDA; C) SEGMENTACIÓN OBTENIDA	39
FIGURA 3.5. A) UMBRAL APLICADO DIRECTAMENTE; B) TRANSFORMACIÓN EXPONENCIAL + UMBRAL; C) TRANSFORMACIÓN EXPONENCIAL + AJUSTE HEURÍSTICO (0.2).	40
FIGURA 3.6. A) MAMOGRAFÍA ORIGINAL 'MDB012.PNG'; B) HISTOGRAMA CORRESPONDIENTE.	40
FIGURA 3.7. A) IMAGEN CON AJUSTE Y UMBRAL DE OSCURECIMIENTO; B) IMAGEN UMBRALIZADA CON SUAVIZADO MORFOLÓGICO.	41
FIGURA 3.8. A) IMAGEN BINARIZADA; B) IMAGEN CON ANÁLISIS DE COMPONENTES CONECTADOS.	43
FIGURA 3.9. A) IMAGEN BINARIZADA MEJORADA. B) DETECCIÓN DE BORDES MEDIANTE EL OPERADOR DE CANNY.	44
FIGURA 3.10. MÁSCARA FINAL: SUPERPOSICIÓN DE FIGURA 3.8B. Y FIGURA 3.9A.	44
FIGURA 3.11. SEGMENTACIÓN DE LA MAMOGRAFÍA: A) MAMOGRAFÍA SEGMENTADA OBTENIDA CON TEJIDO PECTORAL; B) MAMOGRAFÍA SEGMENTADA OBTENIDA SIN TEJIDO PECTORAL.	45
FIGURA 3.12. DIAGRAMA DE BLOQUES AMPLIADO DE LA ETAPA 3.	46
FIGURA 3.13. DIAGRAMA DE BLOQUES AMPLIADO PARA EL ENVENTANADO EN EL DOMINIO ESPACIAL DE LA ETAPA 3.	47
FIGURA 3.14. DISTRIBUCIÓN DE VENTANAS GENERADAS POR LOS ALGORITMOS DE ENVENTANADO PARA EL TEJIDO CON MASA, APLICADOS A LA MAMOGRAFÍA MDB012.	49
FIGURA 3.15. DISTRIBUCIÓN DE VENTANAS GENERADAS POR LOS ALGORITMOS DE ENVENTANADO PARA EL TEJIDO NORMAL EN IMAGEN CON MASA, APLICADOS A LA MAMOGRAFÍA MDB012.	50
FIGURA 3.16. DISTRIBUCIÓN DE VENTANAS GENERADAS PARA TEJIDO NORMAL EN MAMOGRAFÍAS QUE NO CONTIENEN MASAS.	51
FIGURA 3.17. DIAGRAMA DE BLOQUES AMPLIADO PARA LA EXTRACCIÓN DE CARACTERÍSTICAS DEL DOMINIO FRECUENCIAL DE LA ETAPA 3.	52
FIGURA 3.18. EJEMPLOS DE LOS ESPECTROS DE FRECUENCIA GENERADOS.	53
FIGURA 3.19. OBTENCIÓN DEL VECTOR DE CARACTERÍSTICA A PARTIR DE LAS VENTANAS.	53
FIGURA 3.20. DIAGRAMA DE BLOQUES AMPLIADO PARA LA ETAPA 4.	54
FIGURA 3.21. DIAGRAMA DE BLOQUES AMPLIADO PARA EL APRENDIZAJE SUPERFICIAL DE LA ETAPA 5.	56

FIGURA 3.22. DIAGRAMA DE INTEGRACIÓN DE CLASIFICADORES EN LA RED DE COMITÉ.	60
FIGURA 3.23. DIAGRAMA DE BLOQUES AMPLIADO PARA EL APRENDIZAJE PROFUNDO DE LA ETAPA 5.	61
FIGURA 3.24. ESTRUCTURA DE LA CNN.	63
FIGURA 3.25. ARQUITECTURA DEL MODELO DE CLASIFICACIÓN DE IMÁGENES MAMOGRÁFICAS CON APRENDIZAJE POR TRANSFERENCIA.	65

LISTA DE TABLAS

TABLA 2.1. COMPARATIVA DE ANTECEDENTES CONSULTADOS.....	13
TABLA 2.2. DEFINICIONES PARA MÉTRICAS DE DESEMPEÑO EN LA CLASIFICACIÓN.	30
TABLA 3.1. INFORMACIÓN DE LA BASE DE DATOS PÚBLICA MIAS.	36
TABLA 3.2. ESTADÍSTICAS DE LAS ÁREAS DE LAS MASAS CONTENIDAS EN EL DATASET MIAS.	48
TABLA 4.1. RANKING DE CARACTERÍSTICAS ESPACIALES PARA DISTINTOS VALORES DE k EN RELIEFF.....	71
TABLA 4.2. RANKING DE CARACTERÍSTICAS FRECUENCIALES PARA DISTINTOS VALORES DE k EN RELIEFF.....	71
TABLA 4.3 PUNTAJE TOTAL DE LAS CARACTERÍSTICAS EN FUNCIÓN DE SUS POSICIONES EN EL RANKING: A) CARACTERÍSTICAS ESPACIALES; B) CARACTERÍSTICAS FRECUENCIALES.	72
TABLA 4.4. COMPARACIÓN DE RESULTADOS PARA SELECCIÓN DE CARACTERÍSTICAS DEL DOMINIO ESPACIAL.	74
TABLA 4.5. COMPARACIÓN DE RESULTADOS PARA SELECCIÓN DE CARACTERÍSTICAS DEL DOMINIO FRECUENCIAL.	74
TABLA 4.6. RESULTADOS PARA TÉCNICAS DE WRAPPERS EN EL DOMINIO ESPACIAL.	75
TABLA 4.7. RESULTADOS PARA TÉCNICAS DE WRAPPERS EN EL DOMINIO FRECUENCIAL.	76
TABLA 4.8. DESEMPEÑO DE CLASIFICADORES EN EL DOMINIO ESPACIAL.	78
TABLA 4.9. DESEMPEÑO DE CLASIFICADORES EN EL DOMINIO FRECUENCIAL.	80
TABLA 4.10. DESEMPEÑO DE CLASIFICADORES CON LA FUSIÓN DE AMBOS DOMINIOS...	81
TABLA 4.11. DESEMPEÑO DE LA RED DE COMITÉ EN LOS DOMINIOS ESPACIAL, FRECUENCIAL Y SU FUSIÓN.	83
TABLA 4.12. RESULTADOS DE LA CNN EN EL DOMINIO ESPACIAL PARA LOS TRES ESCENARIOS CON DATOS DE: A) ENTRENAMIENTO; B) PRUEBA; C) DIFERENCIA ENTRE AMBOS CONJUNTOS.	86
TABLA 4.13. RESULTADOS DE LA CNN EN EL DOMINIO FRECUENCIAL PARA LOS TRES ESCENARIOS CON DATOS DE: A) ENTRENAMIENTO; B) PRUEBA; C) DIFERENCIA ENTRE AMBOS CONJUNTOS.	86
TABLA 4.14. RESULTADOS DE LA CNN CON FUSIÓN DE DOMINIOS PARA LOS TRES ESCENARIOS CON DATOS DE: A) ENTRENAMIENTO; B) PRUEBA; C) DIFERENCIA ENTRE AMBOS CONJUNTOS.	87
TABLA 4.15. CONFIGURACIÓN DEL ENTORNO DE ENTRENAMIENTO PARA EL APRENDIZAJE POR TRANSFERENCIA.	91
TABLA 4.16. COMPARACIÓN DEL RENDIMIENTO DE DIFERENTES MODELOS DE APRENDIZAJE PROFUNDO MEDIANTE APRENDIZAJE POR TRANSFERENCIA CON DATOS DE: A) ENTRENAMIENTO; B) PRUEBA; C) DIFERENCIA ENTRE AMBOS CONJUNTOS.	92
TABLA 4.17. PARÁMETROS DE LAS REDES PREENTRENADAS EMPLEADAS EN EL APRENDIZAJE POR TRANSFERENCIA Y SU RELACIÓN CON EL TAMAÑO DEL CONJUNTO DE DATOS.	92
TABLA 4.18. RESUMEN DE LAS CARACTERÍSTICAS DEL SISTEMA.	96

RESUMEN

El cáncer de mama es una de las principales causas de mortalidad en mujeres a nivel mundial. Si bien la biopsia constituye el método definitivo para el diagnóstico de esta patología, la mamografía digital se considera una de las principales herramientas para su detección temprana, ya que permite identificar anomalías antes de que se presenten síntomas clínicos. Esta detección temprana es crucial para mejorar el pronóstico y aumentar las tasas de supervivencia. No obstante, su interpretación manual depende altamente de la experiencia del especialista. En este contexto, los sistemas de detección y diagnóstico asistido por computadora (CAD) se posicionan como herramientas de apoyo fundamentales para reducir los errores de interpretación y la carga diagnóstica del personal médico.

El presente trabajo describe un sistema de clasificación automática del tejido mamográfico en tres clases: tejido normal, masa benigna y masa maligna, a partir del análisis de imágenes mamográficas digitales. El sistema comprende cinco etapas principales: adquisición de la imagen, procesamiento, extracción de características, selección de características y clasificación.

Se utilizó la base de datos pública MIAS (Mammographic Image Analysis Society), la cual contiene mamografías digitales en escala de grises.

El procesamiento incluyó técnicas de mejora de contraste y segmentación de la región de interés (mama), mediante transformaciones en el dominio espacial como umbralización adaptativa, transformación exponencial, operaciones morfológicas y el algoritmo de Canny para la detección de bordes. Este conjunto de técnicas permitió eliminar artefactos no deseados y lograr la exclusión del musculo pectoral, asegurando un análisis focalizado exclusivamente en el tejido mamario.

Posteriormente, se llevó a cabo la extracción de características sobre áreas rectangulares de las mamografías (ventanas). Se extrajeron características del dominio espacial y frecuencial, con el fin de resaltar patrones estructurales y espectrales que permitieran una mejor caracterización del tejido mamario.

La selección de características se realizó utilizando técnicas como rankings y wrappers, con el objetivo de reducir la dimensionalidad y conservar aquellas variables con mayor capacidad discriminativa, las cuales fueron posteriormente empleadas en la etapa de clasificación para evaluar las características espaciales y frecuenciales de manera separada y fusionada.

Para la clasificación se implementaron dos enfoques: aprendizaje superficial y aprendizaje profundo. En el *aprendizaje superficial* se emplearon diez clasificadores Naive Bayes, KNN, Distancia Euclídeana, Distancia de Mahalanobis, LDA, QDA máquinas de vectores de soporte (SVM), Random Forest (RF), redes neuronales de base radial (RBF) y perceptrón multicapa (MLP). Además, se implementaron redes de comité mediante votación mayoritaria, combinando los modelos con mejor desempeño individual.

Por otro lado, en el *aprendizaje profundo*, se diseñó y entrenó una Red Neuronal Convolucional (CNN) desde cero, adaptada a imágenes de 50x50 píxeles y complementada con técnicas de aumento de datos (rotación, flipping, zoom). También se implementaron siete modelos preentrenados mediante *transferencia de aprendizaje*, incluyendo VGG16, VGG19, ResNet50, ResNet101, DenseNet121, DenseNet201 e InceptionV3. Estas redes fueron adaptadas para recibir imágenes de tres canales, asignando a cada canal una representación diferente (dominio espacial, Transformada de Fourier y Transformada Wavelet), lo que permitió evaluar directamente la fusión de dominio espacial y frecuencial.

La evaluación de los modelos se realizó mediante métricas de desempeño como precisión, exactitud (accuracy), F1-score, sensibilidad (recall) y especificidad, priorizando la sensibilidad por su importancia en la detección de casos positivos. No se utilizó validación cruzada, sino una división del conjunto de datos en 70% para entrenamiento (del cual se extrajo el 20% para validación) y 30% para prueba, garantizando un balance equitativo entre las clases. Asimismo, se analizó la capacidad de generalización de los modelos mediante la diferencia de desempeño entre los conjuntos de entrenamiento y prueba.

Los mejores resultados en el aprendizaje superficial se obtuvieron mediante una red de comité, que alcanzó una exactitud del 82.35% y una sensibilidad del 76.80%, superando a los modelos individuales. En el aprendizaje profundo, la CNN propia obtuvo una especificidad del 97.79%, mientras que el mejor desempeño general dentro de las redes preentrenadas fue alcanzado por ResNet, con una precisión del 66.40%, exactitud del 64.47% y sensibilidad del 61.00%. Por su parte, Inception V3 mostró la mejor generalización, evidenciada por diferencias mínimas entre entrenamiento y prueba (Δ precisión = 0.061, Δ exactitud = 0.042 y Δ F1-score = 0.057). Estos resultados superan los obtenidos con la CNN desarrollada desde cero y se encuentran dentro de rangos adecuados respecto a estudios previos, resaltando además el uso combinado de dominios espaciales y frecuenciales dentro de un único sistema de clasificación automática.

Este trabajo representa un aporte al desarrollo de sistemas CAD aplicados al diagnóstico mamográfico, con potencial uso como herramienta de apoyo a los profesionales de la salud en la detección temprana del cáncer de mama.

Palabras Clave: Aprendizaje profundo, Cáncer de mama, Clasificación automática, Mamografías digitales, Redes neuronales convolucionales, Sistemas CAD, Transferencia de aprendizaje, Visión por computador.

ABSTRACT

Breast cancer is one of the main causes of mortality in women worldwide. Although biopsy is the definitive method for diagnosing this pathology, digital mammography is considered one of the main tools for its early detection, since it allows abnormalities to be identified before clinical symptoms appear. This early detection is crucial to improve prognoses and increase survival rates. However, its interpretation manual depends largely on the experience of the specialist. In this context, computer-aided detection and diagnosis (CAD) systems are positioned as fundamental support tools to reduce interpretation errors and the diagnostic burden of the personal physician.

The present work describes a system for automatic classification of mammographic tissue into three classes: normal tissue, benign mass and malignant mass, based on the analysis of digital mammographic images. The system comprises five main stages: image acquisition, processing, feature extraction, feature selection and classification.

The public database MIAS (Mammographic Image Analysis Society) was used, which contains grayscale digital mammograms.

The processing included contrast enhancement techniques and segmentation of the region of interest (breast), through transformations in the spatial domain such as adaptive thresholding, exponential transformation, morphological operations and the Canny algorithm for edge detection. This set of techniques allowed us to eliminate unwanted artifacts and achieve the exclusion of the pectoral muscle, ensuring an analysis focused exclusively on the breast tissue.

Subsequently, feature extraction was carried out on rectangular areas of the mammograms (windows). Features were extracted from the spatial and frequency domain, in order to highlight structural and spectral patterns that would allow a better characterization of the breast tissue.

The selection of features was carried out using techniques such as rankings and wrappers, with the aim of reducing dimensionality and preserving those variables with greater discriminative capacity, which were subsequently used in the classification stage to evaluate the spatial and frequency characteristics separately and merged.

Two approaches were implemented for classification: surface learning and deep learning. In *surface learning* ten classifiers, Naive Bayes, KNN, Euclidean Distance,

Mahalanobis Distance, LDA, QDA support vector machines (SVM), Random Forest (RF), radial basis neural networks (RBF) and multilayer perceptron (MLP) were used. In addition, committee networks were implemented through majority voting, combining the models with the best individual performance.

On the other hand, in *deep learning*, a Convolutional Neural Network (CNN) was designed and trained from scratch, adapted to 50x50 pixel images and complemented with data augmentation techniques (rotation, flipping, zoom). Seven pre-trained models were also implemented through transfer learning, including VGG16, VGG19, ResNet50, ResNet101, DenseNet121, DenseNet201, and InceptionV3. These networks were adapted to receive images from three channels, assigning each channel a different representation (spatial domain, Fourier Transform and Wavelet Transform), which allowed the fusion of spatial and frequency domains to be directly evaluated.

The evaluation of the models was carried out using performance metrics such as precision, accuracy, F1-score, sensitivity (recall) and specificity, prioritizing sensitivity due to its importance in detecting positive cases. No cross-validation was used, but a division of the data set into 70% for training (from which 20% was extracted for validation) and 30% for testing, guaranteeing an equal balance between the classes. Likewise, the generalization capacity of the models was analyzed through the difference in performance between the training and test sets.

The best results in surface learning were obtained using a committee network, which achieved an accuracy of 82.35% and a sensitivity of 76.80%, outperforming individual models. In deep learning, the proprietary CNN obtained a specificity of 97.79%, while the best overall performance within the pretrained networks was achieved by ResNet, with a precision of 66.40%, accuracy of 64.47% and sensitivity of 61.00%. For its part, Inception V3 showed the best generalization, evidenced by minimal differences between training and testing (Δ precision = 0.061, Δ accuracy = 0.042 and Δ F1-score = 0.057). These results exceed those obtained with the CNN developed from scratch and are within adequate ranges compared to previous studies, also highlighting the combined use of spatial and frequency domains within a single automatic classification system.

This work represents a contribution to the development of CAD systems applied to mammographic diagnosis, with potential use as a support tool for health professionals in the early detection of breast cancer.

Keywords: Deep learning, Breast cancer, Automatic classification, Digital mammograms, Convolutional neural networks, CAD systems, Transfer learning, Computer vision.

1. INTRODUCCIÓN

El cáncer de mama es la forma más prevalente de cáncer que se diagnostica en mujeres. Para identificar posibles signos de la afección antes de que se manifiesten los síntomas se emplean imágenes de rayos X de la mama (mamografías). La interpretación de las mamografías es realizada visualmente por radiólogos especializados para determinar la presencia de anomalías o alteraciones. En Colombia, la disponibilidad de radiólogos es limitada, lo que conlleva una carga de trabajo elevada y puede ocasionar demoras y errores en el análisis de las mamografías, lo cual tiene implicaciones negativas tanto para los pacientes como para el sistema de salud en su totalidad. Para hacer frente a esta problemática, se han desarrollado sistemas informáticos que generan alertas cuando se detectan posibles anomalías en las mamografías, permitiendo que los especialistas se enfoquen en un análisis más detallado de estas imágenes.

Los sistemas computarizados para detección/diagnóstico asistido por computadora (CAD), son herramientas que emplean técnicas de inteligencia artificial, aprendizaje automático y procesamiento de imágenes, entre otras, para el análisis de imágenes médicas. Estos sistemas mejoran la calidad de las imágenes, realzan los distintos tejidos, generan alertas sobre posibles anomalías y proporcionan información cuantitativa que los expertos médicos utilizan como apoyo para realizar diagnósticos más precisos.

Debido a las limitaciones en precisión y exactitud que presentan los sistemas CAD en sus diferentes etapas, este trabajo de grado se centra en probar y comparar distintas técnicas para la extracción, selección de características y clasificación de imágenes para identificar el tejido mamario normal, masas benignas o malignas en las mamografías.

Con el fin de tener un punto de referencia en el desarrollo de esta propuesta, se realiza una revisión detallada de antecedentes relacionados con la detección de cáncer de mama, sistemas CAD y procedimientos que permiten el avance en la solución del problema. De igual forma, en el marco teórico se presenta un resumen de los conceptos generales sobre las etapas de un sistema de reconocimiento de patrones y técnicas de visión artificial para detectar tejido normal, masas benignas y malignas en imágenes radiológicas de la mama femenina.

En esta propuesta se presenta el planteamiento del problema, una discusión sobre la literatura científica en temas similares o afines. Con base en la problemática identificada se plantean el objetivo general y los objetivos específicos para

desarrollar una técnica para identificar tejido normal, masas benignas y masas malignas en imágenes digitales mamográficas utilizando técnicas de visión artificial y máquinas de aprendizaje superficiales y profundas.

1.1 CONTEXTUALIZACIÓN

El cáncer de mama se genera cuando las células mamarias mutan, crecen y se reproducen sin control, creando así una masa de tejido, llamado tumor. Esta una de las principales lesiones en el tejido mamario que puede desencadenar el cáncer de mama [1].

El cáncer de mama es una preocupación de salud pública en Colombia, pues continúa siendo el tipo de cáncer más frecuente y se considera la causa más común de muerte por cáncer en las mujeres a nivel mundial. Según los datos reportados por la Organización Mundial de la Salud (OMS) en su página de GLOBOCAN [2], para el 2020 se estimaron 15.509 casos nuevos y 4.411 muertes por esta enfermedad. También se calcula que una de cada doce mujeres podría sufrir de este carcinoma a lo largo de su vida, de este modo convirtiéndose en la principal causa de mortalidad en las mujeres. En el 2020, alrededor de 685.000 mujeres fallecieron a nivel mundial como consecuencia de esa enfermedad [3]. Según el Ministerio de Salud y Protección Social se considera la primera causa de enfermedad y muerte por cáncer entre las mujeres colombianas.

La técnica más usual para identificar el cáncer de mama es el examen diagnóstico conocido como mamografía. Este proceso consiste en analizar las características y estructura del tejido mamario a través de imágenes radiográficas (Imágenes RX). Las mamografías son consideradas altamente efectivas para la detección temprana de irregularidades en ausencia de síntomas aparentes [4]. Detectar a tiempo anomalías como el cáncer, posibilita la oportunidad de iniciar tratamientos, disminuir las complicaciones asociadas y minimizar costos económicos.

1.2 PLANTEAMIENTO DEL PROBLEMA Y OBJETIVOS

1.2.1 Planteamiento del problema

La lectura manual de mamografías por radiólogos es una técnica comúnmente utilizada para la detección del cáncer de mama [5]. Si bien es efectiva en la detección temprana, presenta riesgos asociados, como diagnósticos incorrectos debido al bajo contraste de las imágenes, la subjetividad de las evaluaciones influenciada por la experticia del especialista y la fatiga, y desajustes en los equipos de adquisición que afectan la calidad de las imágenes [6]. Además, la escasez de radiólogos, especialmente en países en desarrollo como Colombia [7], agrava estos problemas al retrasar los tiempos de diagnóstico y aumentar la carga laboral [8].

En respuesta a estos desafíos, se han desarrollado sistemas de diagnóstico asistido por computador (Computer-Aided Diagnosis, CAD), los cuales emplean técnicas de visión artificial y reconocimiento de patrones para analizar mamografías, identificar posibles anomalías y generar alertas que apoyen el trabajo de los especialistas [9]. Estos sistemas incluyen etapas como la adquisición, preprocesamiento, segmentación, extracción y selección de características, clasificación e interpretación de resultados [10]. Sin embargo, cada una de estas etapas enfrenta dificultades técnicas que afectan la precisión y robustez de los sistemas CAD.

A pesar de los avances globales en este campo, en Colombia los esfuerzos aún están en desarrollo. Trabajos como los reportados en [11], [12] y [13] han explorado técnicas para la detección de alteraciones mamarias, pero no existe una solución general que garantice exactitud total en la clasificación de mamografías, robustez frente a variaciones del tejido mamario y estabilidad ante las características del sistema de adquisición. En el grupo de investigación Percepción y Sistemas Inteligentes (PSI) de la Universidad del Valle, se han desarrollado proyectos relevantes en esta área, como los descritos en [14], [15] y [16].

En este contexto, surge la pregunta problematizadora: **¿Cómo utilizar técnicas de visión artificial y reconocimiento de patrones para detectar tejido normal y masas benignas o malignas presentes en imágenes digitales de mamografías?**

Este trabajo de grado busca abordar esta problemática mediante el diseño e implementación de algoritmos para las etapas de preprocesamiento, segmentación, extracción y selección de características, y clasificación de imágenes, empleando bases de datos públicas como DDSM [17] y MIAS [18]. Aunque no se plantea el

desarrollo de un sistema CAD completo, se espera contribuir al avance de tecnologías informáticas para la detección temprana del cáncer de mama.

1.2.2 Objetivo General

Identificar tejido normal, masas benignas y masas malignas en imágenes digitales mamográficas utilizando técnicas de visión artificial y máquinas de aprendizaje superficiales y profundas.

1.2.3 Objetivos Específicos

- i. Identificar características de los dominios espacial y frecuencial extraídas de mamografías digitales que permitan la discriminación entre tejido normal, masas benignas y masas malignas en máquinas de aprendizaje superficial.
- ii. Clasificar automáticamente las regiones de mamografías con tejido normal, masas benignas y masas malignas mediante máquinas de aprendizaje superficial.
- iii. Identificar una transformación en los dominios espacial y frecuencial aplicable a las mamografías que permitan la discriminación entre tejido normal, masas benignas y malignas en máquinas de aprendizaje profundo.
- iv. Clasificar automáticamente las regiones de mamografías con tejido normal, masas benignas y masas malignas mediante máquinas de aprendizaje profundo.
- v. Determinar los alcances y las limitaciones de los algoritmos implementados para clasificar tejido normal, masas benignas y masas malignas en imágenes de mamografías digitales.

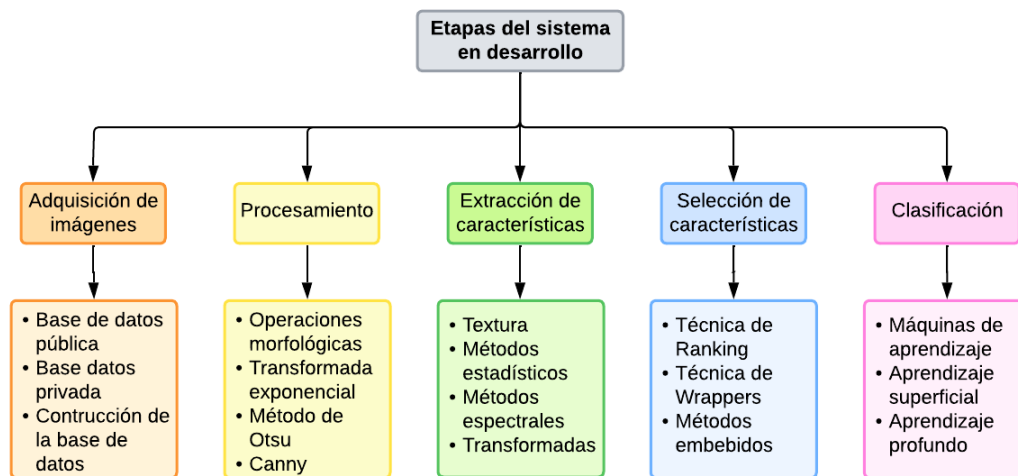
1.3 DESCRIPCIÓN DEL SISTEMA DESARROLLADO

Con el objetivo de proporcionar una herramienta de apoyo para los especialistas y mejorar la precisión en los diagnósticos, se desarrolló un sistema que toma como entrada una imagen mamográfica digital y genera como salida un prediagnóstico que clasifica el tejido mamario como normal, masa benigna o masa maligna.

La solución implementada para abordar el problema planteado incluye dos procedimientos principales. El primero, basado en aprendizaje superficial, se

estructura en cinco etapas: selección de la base de datos, procesamiento, extracción de características, selección de características y clasificación. Estas etapas se ilustran en la Figura 1.1, que proporciona una visión general del flujo de trabajo, desde la adquisición de imágenes hasta la clasificación final. Por otro lado, el segundo procedimiento, fundamentado en aprendizaje profundo, no requiere divisiones específicas, ya que utiliza redes neuronales convolucionales (CNN) que integran todo el flujo de procesamiento de manera automatizada.

Figura 1.1. Flujo de etapas en el desarrollo del sistema.



Fuente: [Elaboración propia]

En el procedimiento basado en aprendizaje superficial, el preprocesamiento toma como entrada las imágenes mamográficas digitales en escala de grises y las mejora mediante la eliminación de artefactos, reducción de ruido y aumento del contraste. Estas imágenes preprocesadas son luego segmentadas para identificar y aislar las regiones de interés. Posteriormente, se extraen características relevantes que describen propiedades como la forma, textura e intensidad del tejido mamario. Estas características son refinadas en la etapa de selección, donde se identifican aquellas más discriminantes, optimizando así el conjunto de datos para la etapa final. Finalmente, el modelo de clasificación toma estas características seleccionadas y asigna una etiqueta al tejido mamario, generando como salida el prediagnóstico.

En el procedimiento de aprendizaje profundo, la imagen mamográfica es procesada directamente por una red neuronal convolucional, que realiza automáticamente las etapas de preprocesamiento, extracción de características y clasificación. Este enfoque permite obtener un prediagnóstico sin necesidad de dividir el sistema en etapas independientes.

Ambos procedimientos fueron diseñados, implementados y evaluados utilizando bases de datos públicas MIAS. Cada componente del sistema fue sometido a ajustes de parámetros y evaluaciones de desempeño para garantizar una solución robusta e integral en la clasificación automática del tejido mamario.

1.4 ORGANIZACIÓN DE LOS CAPÍTULOS

El documento se divide en un total de 8 capítulos, donde los 4 primeros abarcan la introducción, revisión bibliográfica, desarrollo y resultados del sistema y los 4 finales conclusiones generales, trabajos futuros, recopilación de referencias bibliográficas utilizadas en el desarrollo del trabajo y por último unos anexos.

Capítulo 1: Introducción, contextualización y planteamiento del problema:

Se introduce el problema que motiva la investigación: la necesidad de herramientas que apoyen el diagnóstico temprano del cáncer de mama, dado que la interpretación manual de mamografías por parte de radiólogos puede verse limitada por factores como el bajo contraste de las imágenes, la subjetividad en la evaluación, y la escasez de especialistas, especialmente en países en desarrollo. En este contexto, se plantea el objetivo general y los objetivos específicos, proporcionando una visión general del trabajo. Finaliza con la presentación de la estructura y organización de los capítulos siguientes.

Capítulo 2: Revisión de la literatura y marco teórico:

Aborda una revisión de estudios previos a nivel nacional (regional y local) e internacional, explorando herramientas para la elaboración de prediagnósticos a partir de imágenes mamográficas digitales. Se establece un marco teórico sólido que sienta las bases conceptuales para el desarrollo del sistema y conceptos fundamentales para la comprensión del presente documento.

Capítulo 3: Diseño e implementación del sistema:

Se presenta en detalle el diseño y la implementación de la solución propuesta. Se desglosa en etapas, evidenciando la ejecución de los primeros cuatro objetivos específicos. Se destaca el uso de técnicas avanzadas, desde procesamiento de imágenes hasta clasificación automática.

Capítulo 4: Validación y resultados del sistema:

Se describen las pruebas y se analizan los resultados que permiten validar el funcionamiento del sistema, evaluando el rendimiento mediante métricas

específicas y proporcionando una comprensión clara de los alcances y limitaciones del sistema.

Capítulo 5: Conclusiones generales:

Se resaltan las principales características funcionales del sistema desarrollado, y su alineación con los objetivos propuestos. Se destacan los principales logros y las contribuciones del sistema desarrollado. Se presenta una interpretación de los resultados en función de los alcances y limitaciones del sistema.

Capítulo 6: Trabajos futuros:

Se proponen trabajos futuros, explorando posibles direcciones y áreas de mejora para las principales etapas del sistema. Igualmente, se plantean sugerencias y consideraciones para investigaciones posteriores en la misma temática.

Capítulo 7: Referencias bibliográficas:

Se recopilan todas las fuentes bibliográficas utilizadas en el desarrollo del trabajo, proporcionando respaldo a los antecedentes y facilitando las presentes y futuras consultas relacionadas con la temática del sistema desarrollado.

2. ANTECEDENTES Y MARCO TEÓRICO

En este capítulo se presentan los antecedentes locales, nacionales e internacionales con resultados y metodologías relevantes para el desarrollo de este trabajo. Lo aquí expuesto constituye un pilar fundamental al presentar de manera sistemática la información relacionada a la detección automatizada de masas y tejido normal en mamografías. La revisión detallada de la literatura establece la base para comprender la evolución y el estado actual de la investigación en este campo.

Finalmente, se presentan los conceptos necesarios incluyéndolos dentro de un marco teórico, realizando un bosquejo sobre el cáncer de mama, el aprendizaje automático, la visión artificial y el reconocimiento de patrones, subrayando su pertinencia en el desarrollo de metodologías avanzadas para la detección temprana de cáncer de mama.

2.1 ANTECEDENTES

La detección temprana del cáncer de mama ha sido ampliamente abordada mediante el uso de mamografías digitales, las cuales permiten la identificación de lesiones sospechosas con apoyo de sistemas de diagnóstico asistido por computadora (CAD). Para el presente trabajo, se realizó una revisión exhaustiva de literatura científica entre los años 2018 y 2024, empleando bases de datos como IEEE, Scopus, ACM, ScienceDirect, Mendeley y ResearchGate. Se utilizaron palabras clave como “cáncer de mama”, “mamografía”, “aprendizaje automático”, “visión artificial” y “sistemas CAD”.

La revisión se organiza en antecedentes locales, regionales e internacionales, destacando enfoques, técnicas y resultados que sirvieron como base para el diseño e implementación del sistema propuesto.

2.1.1 Antecedentes Nacionales

- **Antecedentes Locales**

En [14] se desarrolló un procedimiento para detectar nódulos mamarios en imágenes termográficas. El proceso incluyó suavizado, umbralización y segmentación geométrica. Se extrajeron 25 características por mama utilizando momentos estadísticos y descriptores de textura. Se empleó un clasificador de tipo función de base radial (RBF), que alcanzó el mejor desempeño entre otros

clasificadores evaluados. Se adicionó una etapa de pos-procesamiento para mejorar la sensibilidad del sistema hasta en un 80%. Los resultados mostraron una segmentación exitosa casi del 100% en un grupo de 67 imágenes, con un alto desempeño en la clasificación de los tres conceptos diagnósticos: saludable, patología o desconocido.

En [15] se desarrolló una herramienta computacional para clasificar masa o tejido normal en mamografías. Un primer método empleó un Faster-R-CNN, logrando una exactitud del 81%. El segundo método utilizó 5 etapas que incluyen segmentación mediante operaciones morfológicas, detección de bordes y umbrales, extracción de características con algoritmos de inventariado y ventanas traslapadas, selección de características con wrappers y rankings, y finalmente, clasificación mediante clasificadores estadísticos y máquinas de aprendizaje como la máquina de vectores de soporte (SVM) y red neuronal artificial (ANN). También se aplicaron tres algoritmos adicionales para mejorar los resultados de la clasificación. Este sistema se entrenó y evaluó utilizando el dataset MIAS. La exactitud alcanzada en la clasificación fue del 97.4%.

En [16] se propone un sistema de detección automática de microcalcificaciones (MCs) en mamografías digitales, basado en visión artificial y aprendizaje supervisado. Utiliza la base de datos MIAS y sigue un proceso de preprocesamiento que incluye eliminación de ruido, mejora del contraste, eliminación de artefactos y segmentación del tejido pectoral. Se emplea un enfoque de ventanas deslizantes para extraer regiones de interés (55x55 píxeles), complementado con un aumento de datos mediante transformaciones de contraste, rotaciones y desenfoque. Se extraen 149 características (estadísticas, texturales y espacio-frecuenciales) y se aplica selección con MRMR, Chi-cuadrado, NCA y Relief. Se evalúan RNA, SVM, K-NN y NB, seleccionando los mejores modelos con ANOVA. La inferencia se realiza con un barrido de ventanas y postvalidación en matrices 3x3. Se obtiene una precisión del 99.8% y una sensibilidad del 77%, con limitaciones en la detección de MCs y variabilidad en la densidad mamaria, sugiriendo optimizaciones en segmentación y selección de características.

- **Antecedentes Regionales**

En [11] detectan carcinomas en imágenes mamográficas mediante reconocimiento de patrones. Emplean la base de datos MIAS compuesta por 322 imágenes normales, pero con 25 casos para calcificaciones. Para el preprocesamiento se usaron diferentes filtros como escalarización, reducción del ROI y filtros pasa bajos. Para la detección de borde, segmentación y extracción de características utilizaron

la derivada de Top Hat con el procedimiento de binarización bajo un umbral de 0.275. Para la extracción de características, se utilizan descriptores de forma y tamaño, incluyendo circularidad, elongación y área convexa. La selección de características se realiza mediante técnicas de discriminación basadas en el protocolo BI-RADS, con respecto a tamaño, pero con ayuda del vector de características basado en técnicas de magnitudes geométricas o formas morfológicas. Para la etapa de clasificación se escogió la técnica de árbol de decisión (DT) como una similitud a la clasificación visual que efectúa un especialista en el descarte de formas y tamaños. Los resultados muestran un índice de sensibilidad del 88.8%, una efectividad del 56%, un valor predictivo positivo del 68% y uno negativo del 82%.

En [12] se propone un método como soporte y asistencia a especialistas para la detección de distorsiones de la arquitectura en la glándula mamaria a través de mamografías utilizando la base de datos DDSM. Se emplearon cuatro etapas principales: preprocesamiento, detección de regiones de interés, extracción y selección de características, y clasificación de ROI utilizando un sistema de lógica difusa (FL). Para el procesamiento, se utilizó el cálculo de umbral de Otsu y etiquetado de componentes conexas. La identificación de ROI se llevó a cabo mediante filtros de Gabor y análisis de retrato de fase. Por último, para distinguir entre regiones normales y anormales, se emplearon las características de textura de Haralick junto con un sistema de decisión basado en la FL. La precisión de este método es de 90.7%.

En [13] se presenta un sistema de clasificación de mamografías basado en la densidad del tejido mamario, utilizando la escala BI-RADS para dividir las mamografías en cuatro clases. El conjunto de datos utilizado en este estudio consistió en 2372 mamografías, de las cuales se extrajeron 50 características de un conjunto de 288: matriz de co-ocurrencia de niveles de gris (GLCM), matriz de longitud de carrera de niveles de gris (GLRLM), matriz de zona de niveles de gris (GLZLM), matriz de diferencia de textura de nivel de gris (NGTDM) y patrón de binarización local (LBP). Para la clasificación se utilizaron regresión logística y SVM como métodos de clasificación. El sistema desarrollado logró una sensibilidad del 95.7%, una especificidad del 89.5% y una precisión general del aproximadamente del 91.3%.

2.1.2 Antecedentes Internacionales

En [19] se utiliza la base de datos Mini-MIAS. Se aplican técnicas de preprocesamiento como filtrado de mediana y operaciones morfológicas para

mejorar la calidad de las imágenes. Para la extracción de características, se emplean descriptores SURF. La selección de características se realiza mediante agrupamiento K-Means. Se alcanzó una precisión de 96.9% con SVM, superando a LDA (92.8%), KNN (94.8%) y Árbol de Decisión (89.7%).

En [20] se utiliza la base de datos de Eastern Health (Canadá). La extracción de características se realiza con transformadas de Fourier y Wavelet (db2, db4, bior6.8), obteniendo medidas estadísticas como media, desviación estándar, asimetría y curtosis. Se implementan clasificadores LDA, redes neuronales BP y Naïve Bayes, combinados en un esquema de votación. El método logra una precisión del 96.06%, con una sensibilidad del 96.45% y especificidad del 97.45%, superando enfoques individuales.

En [21] se propuso un sistema CAD donde la extracción de características se llevó a cabo con DCNN preentrenadas (AlexNet, GoogleNet, ResNet-18, 50 y 101). Para mejorar la clasificación, se realizó una fusión de características profundas y se empleó reducción de dimensionalidad con PCA, alcanzando una precisión del 97.9% en CBIS-DDSM y 97.4% en MIAS, con AUC de 1.00 (100%).

En [22] se propone una CNN adaptativa con cinco capas convolucionales para clasificar tres categorías: normal, benigno y maligno. Se obtiene una precisión del 99%. En este trabajo se sugiere optimizar la detección temprana mediante modelos YOLOV3.

En [23] proponen un sistema CAD basado en un clasificador de conjunto y algoritmos de ponderación de características. Se fusionaron características texturales y morfológicas. La clasificación se realiza mediante un modelo de conjunto con KNN, bagging y EigenClass, empleando una votación por mayoría para diferenciar entre tejido normal, tumores benignos y malignos. Los resultados fueron del 91.00% de exactitud y 64.72% de sensibilidad.

En [24] se propone un modelo robusto para la clasificación multiclase. Se preprocesan las imágenes mediante un filtro de mediana adaptativo, ecualización de histograma adaptativa y eliminación del músculo pectoral mediante la transformada de Hough y detección de bordes Canny. Se emplea aprendizaje por transferencia con los modelos VGG16, VGG19, ResNet50 y DenseNet121. Se obtiene una precisión del 98.70% (MIAS), 97.73% (CBIS-DDSM) y 98.83% (INbreast), validada con el coeficiente kappa.

En [25] se propone Mammo-Light, un sistema CAD basado en redes neuronales convolucionales (CNN). Se aplicó un preprocesamiento que incluye eliminación de

ruido mediante Non-Local Means Denoising, realce de imagen con corrección gamma y ecualización de histograma adaptativa (CLAHE), y segmentación mediante detección del contorno más grande. Se obtiene una precisión del 99.17% en CBIS-DDSM y 98.42% en MIAS.

2.1.3 Discusión

Teniendo en cuenta los antecedentes consultados se encontró que los sistemas de diagnóstico asistido por computadora (CAD) para la detección de cáncer de mama en mamografías digitales siguen una estructura común compuesta por seis etapas: *adquisición de imágenes, preprocesamiento, segmentación, extracción y selección de características, y clasificación.*

Las *bases de datos* más utilizadas son: MIAS, DDSM, CBIS-DDSM e INbreast. En la etapa de *preprocesamiento*, se han utilizado técnicas de mejora de contraste como CLAHE, corrección gamma y ecualización de histograma, además de eliminación de ruido con filtros de mediana y normalización de intensidad. Para la *segmentación*, destacan la umbralización de Otsu, operaciones morfológicas y la detección de bordes con el operador Canny. Además, en algunos estudios se ha implementado la técnica de inventanado, la cual permite extraer múltiples subregiones de una imagen para mejorar la detección de patrones anómalos dentro de una región específica. En la *extracción de características*, los métodos más empleados incluyen descriptores texturales como los de Haralick, los momentos del histograma, la matriz de coocurrencia de niveles de gris (GLCM) y otros enfoques como GLRLM, LBP, Fourier y Wavelet. La *selección de características* ha sido abordada con técnicas de Wrappers y Rankings, además de enfoques automatizados mediante redes neuronales convolucionales (CNN). Para la *clasificación*, se han identificado tres enfoques principales: clasificadores estadísticos (Naive Bayes, KNN, LDA y QDA), máquinas de aprendizaje superficial (SVM y Random Forest) y máquinas de aprendizaje profundo (MLP) o incluso donde CNN y modelos preentrenados como VGG, ResNet y DenseNet han mostrado un rendimiento superior. La evaluación del desempeño se basa en métricas como *Accuracy*, *Precision*, *Recall*, *Specificity* y *F1-score*, donde los modelos de aprendizaje profundo han alcanzado precisiones superiores al 90%, sugiriendo su ventaja sobre los métodos tradicionales en la clasificación de mamografías digitales.

En la Tabla 2.1., se resumen los aspectos más relevantes de algunos de los antecedentes consultados que sirvieron de soporte para el desarrollo de este trabajo.

Tabla 2.1. Comparativa de antecedentes consultados.

Ref	Base de datos	Procesamiento	Segmentación	Características extraídas	Método de selección de características	Clasificador	Clases	Desempeño	
								Métrica	Porcentaje
[11]	MIAS	Filtro mediano y escala de grises, de escalarización y filtros pasabajos	Derivada Top Hat y umbralización	Vectores de magnitudes geométricas.	Discriminadores BIRADIS, filtros, wrappers y embebidos	Árbol de decisiones	Calcificación Benigna / Calcificación Maligna	Sensibilidad (Recall)	88,80%
[12]	DDSM	Conversión a escala de grises, filtrado de ruido	Identificación de ROI, Otsu, filtros de Gabor	Textura de Haralick	Sistema de decisión basado en lógica difusa	Lógica difusa (Mandani)	Normal / Anormal	Precisión (Precision) Sensibilidad (Recall) Especificidad (Specificity)	90,70% 91,70% 90,30%
[13]	2372 mamografías (BI-RADS)	Software "Image Processin Tool Kit (ITK)"	No especificado	Histograma, forma, textura (GLCM, GLRLM, GLZLM, NGTDM y LBP)	No especificado	Regresión logística, SVM	Cuatro clases según densidad mamaria (BI-RADS)	Precisión (Precision) Sensibilidad (Recall) Especificidad (Specificity)	91,30% 95,70% 89,50%
[15]	MIAS	No especificado	Operaciones morfológicas, umbralización, transformada exponencial, canny	Algoritmos de enventanado, matriz de coocurrencia	Wrappers y rankings	Clasificadores estadísticos, SVM, ANN (MLP), NB, KNN, LDA, QDA y Faster-RCN	Normal / Masa	Exactitud (Accuracy) Sensibilidad (Recall)	97,80% 97,40%
[16]	MIAS	Eliminación de ruido (filtro de mediana adaptativo), contraste (corrección gamma, equalización de histograma)	Tejido pectoral (umbralización, análisis de histograma), ventanas deslizantes para extraer regiones de interés	Descriptores estadísticos, texturales: Haralick, GLRLM, LBP; espacio-frecuenciales: SFTA	MRMR, Chi-cuadrado, NCA, Relief	RNA (MLP), SVM, K-NN, Naïve Bayes (NB)	Normal / Microcalcificación	Precisión (Precision) Sensibilidad (Recall)	99,80% 77,00%
[19]	MIAS	Filtrado de mediana y operaciones morfológicas	Generación de ROI	Descriptores SURF, CNN	Agrupamiento K-Means	MSVM, Árbol de decisión	Normal / Benigno / Maligno	Precisión (Precision) Precisión (Precision) Precisión (Precision) Precisión (Precision)	96,90% SVM 89,70% LDA 88,70% KNN 88,70% AdD
[20]	Eastern Health (Canadá)	Eliminación de ruido, normalización de intensidad, umbralización de fondo	No especificado	Transformada de Fourier y Wavelet (media, desviación estándar, asimetría y curtosis)	Ganancia de información (selección basada en entropía)	LDA, Redes Neuronales BP, Naive Bayes (NB) y esquema de votación	Normal / Maligna	Precisión (Precision) Sensibilidad (Recall) Especificidad (Specificity)	96,06% 96,45% 97,45%
[21]	CBIS-DDSM y MIAS	CLAHE para mejora de contraste	ROI en MIAS, ya segmentado en CBIS-DDSM	DCNN (AlexNet, GoogleNet, ResNet-18, 50, 101)	PCA para reducción dimensional	SVM con diferentes kernels	Normal / Benigno / Maligno	Precisión (Precision) Precisión (Precision)	97,9% CBIS-DDSM 97,40% MIAS
[22]	Mini-DDSM	Redimensionamiento, eliminación de ruido, conversión RGB	No especificado	Características extraídas automáticamente por la CNN	No aplica (extracción automática por CNN)	CNN adaptativa	Normal / Benigno / Maligno	Precisión (Precision)	99,00%
[23]	MIAS y DDSM	Métodos de filtrado y fusión de imágenes	Otsu y operaciones morfológicas	Características texturales GLCM, morfológicas, de forma e histograma de intensidad	Ponderación de características basada en un algoritmo INFO	KNN, Bagging, EigenClass (Clasificación de votación por mayoría)	Normal / Benigno / Maligno	Exactitud (Accuracy) Precisión (Precision) Sensibilidad (Recall) F1-score Especificidad (Specificity) Exactitud (Accuracy) Precisión (Precision) Sensibilidad (Recall) F1-score Especificidad (Specificity)	91,00% MIAS 91,60% MIAS 64,72% MIAS 86,50% MIAS 82,94% MIAS 93,26% DDSM 90,40% DDSM 82,31% DDSM 89,89% DDSM 93,20% DDSM
[24]	MIAS, CBIS-DDSM, INbreast	Eliminación de ruido (filtro de mediana adaptativo), mejora de contraste (equalización de histograma adaptativa)	Eliminación del músculo pectoral (Transformada de Hough y detección de bordes Canny)	Características profundas extraídas con aprendizaje por transferencia (VGG16, VGG19, ResNet50, DenseNet121)	Fusión de características profundas híbridas (FHDF)	Redes neuronales preentrenadas (VGG16, VGG19, ResNet50, DenseNet121)	Normal / Benigno / Maligno	Kappa de Cohen (KC) Kappa de Cohen (KC)	98,70% MIAS 97,73% DDSM
[25]	MIAS y CBIS-DDSM	Eliminación de ruido (Non-Local Means Denoising), realce de imagen (corrección gamma, CLAHE)	Detección del contorno más grande	Extraídas automáticamente por la CNN	No especificado	CNN ligera (Mammo-Light)	Normal / Maligna	Kappa de Cohen (KC) CC de Matthew (MCC) Kappa de Cohen (KC) CC de Matthew (MCC)	98,89% DDSM 98,88% DDSM 97,63% MIAS 97,65% MIAS

[Fuente: Elaboración propia].

2.2 MARCO TEÓRICO

El objetivo general de este trabajo es desarrollar un sistema de detección automática para identificar tejido normal, masas benignas y masas malignas en imágenes digitales mamográficas utilizando algunas etapas de los sistemas diagnósticos asistidos por computadora, la cual se basan en técnicas de visión artificial y máquinas de aprendizaje superficiales y profundas. Para realizar este trabajo es necesario tener en cuenta algunos conceptos generales sobre el cáncer de mama generado por masas y técnicas como la adquisición, procesamiento de imágenes, segmentación de imágenes, selección y extracción de características y clasificación automática.

2.2.1 Aspectos conceptuales de la salud

2.2.1.1 Cáncer de mama

El cáncer de mama es una patología caracterizada por el crecimiento descontrolado de células en el tejido mamario, que puede originarse en los lobulillos, conductos o tejido conectivo. La mayoría de los casos se desarrolla en los conductos o lobulillos [26]. Las masas o tumores formados pueden ser benignos (no cancerosos) o malignos (cancerosos), siendo estas últimas capaces de invadir tejidos adyacentes y diseminarse a otras partes del cuerpo [27].

2.2.1.2 Procedimientos para el diagnóstico del cáncer de mama

El diagnóstico incluye pruebas como la biopsia, la endoscopia y las técnicas de imagen. La biopsia es el método más confiable para confirmar la presencia de cáncer, al permitir el análisis histopatológico del tejido extraído [28]. La endoscopia permite visualizar los conductos mamarios mediante una cámara, mientras que las pruebas de imagen facilitan la localización y evaluación de anomalías sin intervención invasiva.

2.2.1.3 Diagnóstico por imágenes

Las principales técnicas de imagen utilizadas son la mamografía, la ecografía y la resonancia magnética. Este trabajo se centra en la mamografía, debido a su eficacia en la detección precoz de anomalías mamarias, especialmente en casos donde no hay síntomas visibles.

2.2.1.4 Mamografías digitales

La *mamografía digital* utiliza rayos X para generar imágenes detalladas del tejido mamario, permitiendo identificar tumores no palpables [29]. Existen versiones en 2D y 3D (tomosíntesis) [30]. La *mamografía digital de campo completo (MDCC)*, también conocida como mamografía digital convierte las señales analógicas en datos digitales, lo cual mejora la calidad de imagen, reduce la exposición a la radiación y facilita el almacenamiento y análisis computacional [31].

2.2.1.5 Masas en el tejido mamario, caracterización y clasificación

Las masas mamarias detectadas mediante imagenología se clasifican en benignas o malignas, según su morfología y comportamiento [32]. Las benignas, como quistes o fibroadenomas, generalmente son inofensivas, mientras que las malignas presentan bordes irregulares y riesgo de metástasis [33]. Su clasificación precisa requiere pruebas de imagen y confirmación mediante biopsia.

2.2.2 Aspectos conceptuales para el aprendizaje automático

2.2.2.1 Inteligencia artificial (IA)

La inteligencia artificial (IA) se define como el campo de la ciencia y la ingeniería que busca desarrollar sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el razonamiento, el aprendizaje y la toma de decisiones [34].

2.2.2.2 Visión artificial

La visión artificial permite a los sistemas interpretar imágenes digitales mediante técnicas matemáticas que extraen información estructural y visual del entorno [35][36]. Es clave para aplicaciones como el reconocimiento de patrones, la detección de objetos y el análisis de imágenes médicas.

2.2.2.3 Reconocimiento de patrones

El reconocimiento de patrones es un proceso automatizado que utiliza algoritmos para clasificar datos según sus características más relevantes. Estos algoritmos asignan los datos a diferentes clases o categorías específicas, permitiendo a las máquinas identificar y organizar información compleja de manera eficiente [37].

2.2.2.4 Aprendizaje automático (Machine Learning)

El aprendizaje automático es un subcampo de la IA que permite a los sistemas mejorar su desempeño en tareas específicas a través del análisis de datos, sin ser explícitamente programados. Los algoritmos ajustan sus parámetros mediante la retroalimentación obtenida de los errores cometidos durante el entrenamiento.

2.2.2.5 Máquinas de aprendizaje

En la clasificación, los algoritmos asignan cada muestra a una clase específica en función de las características extraídas. Se diferencian dos enfoques principales: el aprendizaje superficial (Shallow Learning), que requiere características predefinidas, y el aprendizaje profundo (Deep Learning), que utiliza redes neuronales para aprender representaciones directamente a partir de los datos [38].

2.2.3 Aspectos conceptuales para el sistema de visión artificial

2.2.3.1 Técnicas de procesamiento de imágenes

Una imagen puede representarse como una función bidimensional de intensidad de luz, $f(x, y)$, donde x y y corresponden a las coordenadas espaciales y f al nivel de gris en cada punto. Cuando la imagen se digitaliza, esta función se discretiza, generando una matriz de píxeles [39]. Las técnicas de procesamiento buscan mejorar la imagen original para su análisis en aplicaciones específicas.

- **Dominio espacial**

Este dominio actúa directamente sobre los píxeles de la imagen. Las operaciones pueden definirse como: $g(x, y) = T[f(x, y)]$, donde T es un operador aplicado en una vecindad del punto (x, y) [37]. Las técnicas empleadas en este trabajo incluyen transformaciones de intensidad y filtrado morfológico.

Contraste: en imágenes en escala de grises se refiere a la diferencia entre los píxeles más claros y oscuros. Aumentar el contraste hace que las diferencias entre las áreas claras y oscuras sean más pronunciadas, lo que puede mejorar la visibilidad de los detalles. Reducir el contraste hace que la imagen se vea más uniforme y menos definida [40].

Transformación exponencial: este método busca ajustar la imagen ampliando el rango de niveles de gris mediante un exponente γ , razón por la cual también es conocida como transformación gamma. Se representa mediante la ecuación (2.1).

$$g(x, y) = c * f(x, y)^\gamma \quad (2.1)$$

Donde γ es el exponente y la acción de esta transformación depende del valor de gamma, de si tiene un valor mayor a 1 o menor a 1. En este contexto, c representa una constante.

- ✓ Si $\gamma > 1$ oscurece zonas claras, se recomienda esta opción para resaltar estructuras brillantes, si $\gamma = 1$ no se aplica ninguna transformación y cuando $\gamma < 1$ ilumina zonas oscuras.

Histograma: muestra la distribución de niveles de gris. La forma normalizada se expresa mediante la ecuación (2.2).

$$p(r_k) = \frac{n_k}{n} \quad (2.2)$$

Donde n_k es el numero de píxeles con nivel r_k y n el total de píxeles [37] [40].

Grey Level Slicing: resalta un intervalo específico de intensidades asignando un valor alto a ese rango y bajo al resto. Es útil en aplicaciones donde se conocen qué niveles de gris representan las características de interés, como en imágenes de radiografías [40].

Filtrado morfológico: el filtrado morfológico es una técnica fundamental en el procesamiento de imágenes, especialmente útil para la manipulación y análisis de formas y estructuras presentes en una imagen. Basada en principios geométricos, esta técnica se aplica a cada píxel considerando su entorno, permitiendo la extracción de componentes relevantes como formas, contornos y esqueletos [40], [42].

Entre las operaciones de morfología se destacan:

- ✓ **Erosión:** elimina píxeles en los bordes, reduciendo detalles pequeños.
- ✓ **Dilatación:** añade píxeles en los contornos, ampliando regiones.
- ✓ **Apertura:** elimina objetos pequeños manteniendo la forma de estructuras grandes.
- ✓ **Cierre:** Rellena huecos pequeños, mejorando la continuidad de las formas.

- **Dominio frecuencial**

Se refiere a las diferencias entre las frecuencias de los píxeles y se basan en la modificación de la Transformada Discreta de Fourier de la imagen. Sea $g(x, y)$ una imagen formada por la convolución de una imagen $f(x, y)$ y un operador lineal invariante de posición $h(x, y)$ [36]. Es decir: $g(x, y) = h(x, y) * f(x, y)$ y se calcula a partir de la ecuación (2.3).

$$g(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(u-x, v-y)h(x, y) \quad (2.3)$$

Transformada Discreta de Fourier 2D: la transformada de Fourier en dos dimensiones está definida por el siguiente par de ecuaciones:

Transformación del dominio del espacio a la frecuencia:

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})} \quad (2.4)$$

Transformación del dominio de la frecuencia al dominio del espacio:

$$F(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v)e^{j2\pi(\frac{ux}{M} + \frac{vy}{N})} \quad (2.5)$$

En estas ecuaciones, las variables M y N son las dimensiones de la imagen en los ejes x y y respectivamente; u y v son frecuencias espaciales [43], para este caso se utilizó la transformación de dominio espacial al dominio frecuencial. Este análisis es clave para detectar patrones regulares u ocultos.

Espectro de frecuencia: representa gráficamente la distribución de amplitudes en función de la frecuencia, facilitando la identificación de componentes estructurales de la imagen [44].

Filtrado en el dominio frecuencial: utiliza la transformada de Fourier para realzar o suprimir componentes según su frecuencia. Este enfoque mejora la calidad visual y facilita el análisis de imágenes médicas [45].

2.2.3.2 Técnicas de segmentación de imágenes

La segmentación de imágenes es una etapa fundamental en el procesamiento digital, orientada a dividir una imagen en regiones con propiedades homogéneas, tales como intensidad, textura o color. Este proceso se fundamenta en tres principios: (i) *similitud*, que agrupa píxeles con características similares; (ii) *conectividad*, que considera la continuidad espacial entre píxeles; y (iii) *discontinuidad*, que detecta cambios abruptos en la intensidad de los píxeles. La precisión de estos criterios puede verse afectada por ruido, iluminación o superposición de objetos [46].

A continuación, se presentan algunas de las técnicas relevantes en el desarrollo de este sistema.

- **Técnicas basadas en los bordes**

Este enfoque identifica límites entre regiones mediante el análisis de variaciones abruptas en la intensidad. Se aplican criterios de conectividad y similitud para detectar bordes, resultandos efectivos en la localización de estructuras geométricas simples.

Detección de bordes: la detección de bordes se realiza en tres fases: suavizado, detección y localización del borde [47]. Entre los algoritmos más empleados se encuentran Sobel, Canny, Prewitt, Roberts y técnicas de lógica difusa [48]. Estos métodos se basan en la convolución entre la imagen y un kernel para detectar gradientes [49]. Canny, un derivado de Sobel, incluye etapas adicionales como filtrado Gaussiano, cálculo de gradiente, supresión no máxima y umbralización con histéresis [50].

- **Técnicas basadas en umbrales**

Estas técnicas se aplican cuando existe un contraste definido entre el objeto y el fondo. La aplicación de un umbral conduce a la binarización de la imagen y se expresa mediante la ecuación (2.6).

$$g(x, y) = \begin{cases} 1 & f(x, y) \geq \text{umbral} \\ 0 & f(x, y) < \text{umbral} \end{cases} \quad (2.6)$$

La determinación del umbral puede variar, desde depender únicamente de:

- ✓ Si sólo depende de $f(x, y) \rightarrow$ se trata de umbral global.

- ✓ Si también depende de la posición $(x, y) \rightarrow$ se trata de umbral dinámico.
- ✓ Si también depende de, por ejemplo, de la vecindad $\rightarrow$ se trata de umbral local.

Método de Otsu: el método de Otsu [51] determina automáticamente el umbral óptimo que maximiza la separación entre clases, minimizando la varianza interna de cada clase. Utiliza estadísticas del histograma para calcular la probabilidad, media y varianza de los píxeles clasificados, resultando altamente efectivo en la diferenciación de regiones de interés [52].

Gound truth: en visión artificial, el *ground truth* corresponde a datos de referencia etiquetados con información precisa sobre la ubicación, las características y las relaciones entre los objetos presentes en una imagen [53]. Este conjunto permite evaluar el desempeño de los modelos mediante la comparación de sus salidas con las etiquetas reales [54].

2.2.3.3 Técnicas de extracción de características en las imágenes

La extracción de características busca obtener descriptores significativos que representen propiedades visuales como forma, color o textura. Estas características pueden clasificarse en:

- ✓ *Externas:* asociadas a la geometría, como tamaño, simetría o distancia y en in
- ✓ *Internas:* derivadas de operaciones matemáticas sobre los valores de píxeles.

En este trabajo, se emplean descriptores texturales mediante métodos estadísticos (momentos del histograma y GLCM) y espectrales (Fourier).

- **Textura**

La textura describe patrones visuales vinculados a variaciones en los niveles de gris. Los cambios suaves en los niveles de gris corresponden a bajas frecuencias, mientras que los cambios abruptos reflejan altas frecuencias. Se basan en tres enfoques clásicos: *estadísticos*, *estructurales* y *espectrales*. Este trabajo se centra en el primero y el último.

- **Métodos estadísticos:**

Los métodos estadísticos consideran la distribución espacial de los niveles de gris, siendo algunos de los más reconocidos:

Momentos del histograma: los momentos del histograma se presentan como una aproximación común al análisis de textura, evaluando las propiedades del

histograma $p(z_i)$ que corresponden a los L niveles de gris z_i de una imagen. La definición del momento n se expresa mediante la ecuación (2.7).

$$\mu_n(z) = \sum_{i=0}^L (z_i - m)^n p(z_i) \quad (2.7)$$

Donde m es la intensidad promedio de z_i , como se observa en la ecuación (2.8).

$$m = \sum_{i=0}^L z_i p(z_i) \quad (2.8)$$

- ✓ El primer momento ($n = 1$), también conocido como la *media*, proporciona detalles acerca de la intensidad promedio de la imagen.
- ✓ El segundo momento ($n = 2$), que corresponde a la *desviación estándar* σ , sirve como medida del contraste promedio en la imagen.
- ✓ El tercer momento ($n = 3$), representa la *skewness*, una medida que evalúa la *asimetría* o desviación del histograma en una imagen dada.
- ✓ El cuarto momento ($n = 4$), corresponde a la *curtosis*, también conocida como *flatness*, que cuantifica la forma de la distribución del histograma.

Matrices de Co-ocurrencia (GLCM): propuesta por Haralick [55], representa la frecuencia de ocurrencia de pares de píxeles con ciertos niveles de gris y separación espacial. Su tamaño varía según el máximo nivel de gris. En el caso de imágenes de 8 bits, la matriz resultante sería de dimensiones 255 x 255. A partir de esta matriz se obtienen algunos de los descriptores más empleados como:

$$Energia = \sum_i \sum_j N_d^2(i, j) \quad (2.9)$$

$$Uniformidad = \sum_{i=0}^L N_d^2(i, j) \quad (2.10)$$

$$Entropía = - \sum_i \sum_j N_d^2(i,j) \log_2 N_d(i,j) \quad (2.11)$$

- **Métodos espectrales:**

El espectro de Fourier permite identificar texturas periódicas, o casi periódicas, presentes en una imagen. La identificación de patrones espectrales en una textura global se relaciona con la concentración de alta energía en el espectro. En contraste, estas características serían difíciles de identificar en el dominio espacial debido a las propiedades locales de las técnicas utilizadas. Existen tres atributos del espectro de Fourier que resultan útiles para la descripción de texturas:

- ✓ **Picos prominentes del espectro:** indican la dirección principal de los patrones de textura.
- ✓ **La localización de los picos en el plano de la frecuencia:** proporciona información sobre el periodo espacial fundamental de los patrones.
- ✓ **La eliminación de componentes periódicos mediante filtrado:** destaca elementos no periódicos que pueden analizarse mediante técnicas estadísticas.

- **Métodos de ventanas traslapadas**

Este enfoque extrae subimágenes de ventanas superpuestas de tamaño fijo a lo largo de la región de interés. Se inicia definiendo un punto de origen en la imagen original y colocando la primera ventana; luego, extrae una matriz o vector que contiene los valores de intensidad de cada píxel en el recorte. Este proceso se replica con una segunda ventana que se superpone con la primera, de modo que no se ubica en el límite, sino sobre una parte de esta. Se repite sucesivamente hasta abarcar el área de interés o la totalidad de la imagen original, según las necesidades específicas del problema en cuestión. En trabajos como [14] se utiliza este método para la extracción de características y en [56] se utiliza este método para eliminar áreas de músculo pectoral de las mamografías del dataset MIAS [18].

- **Características implementadas**

A continuación, se describen las características seleccionadas específicamente para su integración y evaluación en el sistema desarrollado, basadas en su relevancia y aplicación según la literatura [57].

Promedio o Media:

$$\mu = \frac{\sum_{i=1}^M \sum_{j=1}^N P(i, j)}{MN} \quad (2.12)$$

Varianza:

$$\sigma = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N (P(i, j) - \mu)^2}{MN}} \quad (2.13)$$

Suavidad:

$$R = 1 - \frac{1}{1 + \sigma^2} \quad (2.14)$$

Contraste:

$$C = \sum_{i,j=0}^{G-1} (i, j)^2 P(i, j) \quad (2.15)$$

Asimetría:

$$S = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \frac{(P(i, j) - \mu^2)}{\sigma} \quad (2.16)$$

Curtosis:

$$k = \left[\frac{\sum_{i=1}^M \sum_{j=1}^N (P(i, j) - \mu)^4}{MN} \right] - 3 \quad (2.17)$$

Energía:

$$E = \sum_{i,j} P(i, j)^2 \quad (2.18)$$

Entropía:

$$h = - \sum_{k=0}^{i-1} prk (\log_2 prk) \quad (2.19)$$

Homogeneidad:

$$H = \sum_{i,j} \frac{P(i,j)}{1 + |i - j|} \quad (2.20)$$

2.2.3.4 Técnicas de selección de características

La selección de características es una etapa clave en el procesamiento de imágenes, ya que permite identificar las variables más relevantes para el análisis y clasificación de patrones, optimizando el rendimiento de los modelos y reduciendo la complejidad computacional [58].

- **Consideraciones previas a la selección de características**

Antes de aplicar técnicas de selección, se requiere un preprocesamiento estadístico que garantice la calidad y consistencia de los datos. Este proceso incluye:

- ✓ **Pruebas de normalidad (ADTest, KStest y Lillietest):** permiten verificar si las características siguen una distribución específica, generalmente normal. Esto es relevante para determinar la idoneidad de ciertos métodos de selección estadística [59].
- ✓ **Detección de valores atípicos (Outliers):** identifica observaciones alejadas de la distribución general, que podrían distorsionar los resultados del análisis y del modelo clasificatorio si no se gestionan adecuadamente [60].
- ✓ **Normalización:** ajusta los valores de las características a un rango específico, como [0, 1] o [-1, 1], facilitando la comparabilidad entre variables y mejorando el rendimiento de los algoritmos de aprendizaje automático [61].

- **Técnicas específicas de selección de características**

Tras el preprocesamiento, se emplean algoritmos especializados para seleccionar las características más representativas del problema. Estas técnicas pueden agruparse en tres enfoques: *Filtros*, *Wrappers* y *Métodos Embebidos*. En este trabajo, se utilizaron únicamente *Filtros* y *Wrappers*, seleccionados por su capacidad para mejorar el rendimiento de los modelos de clasificación al identificar las características más relevantes de las imágenes mamográficas.

Métodos basados en filtros (Ranking): evalúan la relevancia de cada característica de forma independiente del clasificador. Utilizan medidas estadísticas para estimar el poder discriminativo respecto a las clases, con alta eficiencia computacional, pero sin considerar redundancias entre variables [62].

- ✓ **ReliefF:** Algoritmo de tipo filtro que asigna un peso a cada característica en función de su capacidad discriminativa. Se basa en la diferencia entre vecinos cercanos de la misma y diferente clase, usando un enfoque de $k - \text{vecinos}$ más cercanos [63]. La elección del valor de k es crítica en problemas multiclase. Se recomienda evaluar valores *múltiplos* y *no múltiplos de 3* para capturar tanto representaciones balanceadas como diversidad en el vecindario, favoreciendo una selección más robusta.

Wrappers: evalúan la utilidad de un subconjunto de características utilizando un clasificador entrenado específicamente para dicho subconjunto [62]. A diferencia de los filtros, los wrappers entrenan el clasificador varias veces con diferentes subconjuntos de características, permitiendo identificar y eliminar características redundantes o irrelevantes. La calidad del subconjunto se mide en función de la tasa de error del modelo de clasificación [64].

2.2.3.5 Técnicas de clasificación automática

El proceso de clasificación en sistemas de visión artificial comprende el uso de máquinas de aprendizaje superficial y profundo. Estas técnicas permiten identificar patrones en conjuntos de datos mediante algoritmos supervisados, optimizando la toma de decisiones.

- **Máquinas de aprendizaje superficial (Shallow Learning)**

El aprendizaje superficial se caracteriza por la extracción manual de características específicas, lo que demanda un conocimiento detallado del dominio del problema [65]. Aunque puede ser eficaz en contextos determinados, su dependencia en la selección manual de características limita la flexibilidad y adaptabilidad del modelo, ya que cualquier variación en el tipo de datos puede requerir ajustes sustanciales.

Clasificadores estadísticos: los clasificadores estadísticos, ampliamente utilizados en minería de datos y aprendizaje automático, asignan instancias a clases en función de características observadas y utilizan técnicas estadísticas para tomar decisiones de clasificación. Estos modelos evalúan probabilidades y costos asociados a cada decisión, optimizando la precisión general del modelo al minimizar errores de clasificación y ajustar el modelo a las particularidades de los datos [66].

Modelo de clasificación bayesiana: el algoritmo Naive Bayes, como se evidencia en [67] es un método de clasificación que se basa en el teorema de Bayes, asumiendo independencia condicional entre características. Esta simplificación permite que el modelo sea computacionalmente eficiente y rápido de entrenar. A pesar de que la suposición de independencia rara vez se cumple completamente en contextos reales, Naive Bayes logra obtener resultados robustos, especialmente en aplicaciones con conjuntos de datos grandes y alta dimensionalidad.

Clasificadores de mínima distancia: cuando los datos a clasificar siguen una distribución gaussiana y ambas clases comparten la misma matriz de covarianza Σ y la misma probabilidad a priori, el término dominante en la función discriminante es un componente cuadrático, que puede interpretarse como una distancia al cuadrado. Por ello, la regla de decisión se simplifica asignando el vector de características a la clase que tenga la menor distancia entre x y su vector promedio μ . Los clasificadores de mínima distancia utilizan medidas de distancia para asignar objetos a clases basándose en la similitud de sus vectores característicos. Ejemplos de este enfoque incluyen los clasificadores basados en la distancia Euclidiana y la distancia de Mahalanobis [68].

Distancia Euclidiana: calcula la distancia directa entre dos vectores en el espacio característico. Se define mediante la siguiente ecuación.

$$dist_E = \|x - \mu_i\| = \sqrt{(x - \mu_i)^T - (x - \mu_i)} \quad (2.21)$$

Distancia Mahalanobis: considera la correlación entre las características y ajusta la distancia según la covarianza de los datos. Se define mediante la siguiente ecuación.

$$dist_M = \sqrt{(x - \mu_i)^T \Sigma^{-1} (x - \mu_i)} \quad (2.22)$$

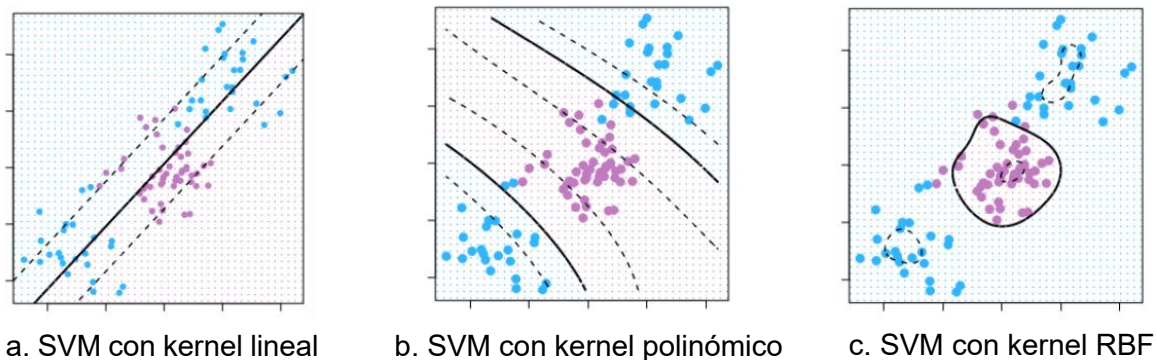
Análisis Discriminante Lineal (LDA): el Análisis Discriminante Lineal (LDA) o Linear Discriminant Analysis (LDA) es un método supervisado de clasificación que asigna observaciones a grupos predefinidos en función de sus características. Este enfoque emplea el teorema de Bayes para estimar la probabilidad de que una observación, con un conjunto dado de predictores, pertenezca a una clase específica. LDA clasifica cada nueva observación en la clase con la probabilidad predicha más alta, optimizando la separación entre grupos mediante una combinación lineal de características [69].

Análisis Discriminante Cuadrático (QDA): el Análisis Discriminante Cuadrático o Quadratic Discriminant Analysis (QDA) es un método de clasificación supervisado similar al LDA, pero con una diferencia clave: en QDA, cada clase k posee su propia matriz de covarianza (Σ_k), lo que permite que la función discriminante sea de forma cuadrática en lugar de lineal [69].

K - vecinos cercanos (KNN): el algoritmo K-NN es un método de clasificación supervisado que asigna una nueva instancia a la clase mayoritaria entre sus k vecinos más cercanos, basándose en la distancia entre vectores de entrenamiento. La elección de k depende del número de clases en el problema. En clasificación binaria, un valor impar de k ayuda a evitar empates, pero en problemas multiclase, este criterio no siempre garantiza desempates, ya que puede haber múltiples clases con la misma frecuencia entre los vecinos más cercanos, por lo que es común experimentar con diferentes valores de k para mejor precisión sin depender exclusivamente de la imparidad de k [67].

Máquinas de soporte vectorial (SVM): las máquinas de soporte vectorial (SVM) son algoritmos de aprendizaje supervisado utilizados principalmente para problemas de clasificación binaria [70]. Su objetivo es identificar una frontera óptima que maximice la separación entre clases, utilizando funciones kernel para casos no linealmente separables [71].

Figura 2.1. Visualización del comportamiento de SVM con diferentes kernels



Fuente: Tomado de [72].

La Figura 2.1 ilustra cómo los diferentes kernels facilitan la separación entre clases: el kernel lineal genera una frontera recta, el polinómico una curva y el RBF delimita regiones más flexibles, mostrando su versatilidad en problemas de alta complejidad.

Random Forest (RF): el clasificador de bosque aleatorio es un modelo de aprendizaje que combina múltiples árboles de decisión construidos con

subconjuntos aleatorios de datos. Mejora la robustez del sistema y reduce el sobreajuste mediante la votación de múltiples predictores [73].

Redes de Comité: las redes de comité constituyen un enfoque de aprendizaje automático que mejora el rendimiento y la precisión al combinar múltiples modelos predictivos. Este método integra varios clasificadores para mitigar sus debilidades individuales y fortalecer la robustez del sistema global [74]. Un ejemplo común es el "*Voting Classifier*", en el cual las decisiones de los modelos se combinan mediante votación. En la votación dura (*Hard Voting*), se selecciona la clase con mayor número de votos; en la votación suave (*Soft Voting*), se promedian las probabilidades de cada predicción, permitiendo una clasificación final más confiable al aprovechar la complementariedad de los modelos [75].

- **Redes neuronales artificiales (ANN):**

Las redes neuronales artificiales son modelos computacionales inspirados en el cerebro humano que se utilizan para clasificar y tomar decisiones basándose en ejemplos previos para identificar patrones complejos [76]. Se componen de capas de neuronas interconectadas, donde cada entrada se pondera y ajusta mediante la propagación hacia atrás (backpropagation) para minimizar el error [77].

Redes multicapa (RBF): las redes multicapa con conexiones hacia adelante, como se describe en [78], incorporan funciones de base radial en una única capa oculta, activando neuronas en regiones específicas del espacio de entrada.

Perceptrón multicapa (MLP): el perceptrón multicapa (MLP) es una arquitectura de red neuronal que extiende el modelo del perceptrón simple al incorporar capas ocultas entre la capa de entrada y la de salida, permitiendo modelar funciones no lineales [79]. Esta estructura consta de una capa de entrada que conecta con el entorno, capas ocultas para cálculos intermedios, y una capa de salida que entrega el resultado final. Durante el aprendizaje, los pesos se ajustan mediante retropropagación y descenso del gradiente para minimizar el error [80].

- **Máquinas de aprendizaje profundo (Deep Learning):**

El aprendizaje profundo permite extraer representaciones jerárquicas de los datos a través de múltiples capas de procesamiento no lineal [81]. Utiliza redes neuronales convolucionales (CNN) para procesar y analizar grandes volúmenes de datos, lo que permite detectar patrones espaciales complejos, siendo especialmente efectivas en el análisis de imágenes médicas [82].

Red Neuronal Convolutacional (CNN): las redes neuronales convolucionales (CNN) representan un modelo especialmente útil para el procesamiento de datos en formato de cuadrícula, como señales temporales con muestreo regular o imágenes digitales. Estos datos poseen características distintivas, tales como su almacenamiento en arreglos multidimensionales, la dependencia espacial en uno o más ejes, y la posibilidad de múltiples canales que permiten diferentes vistas de la información [83]. Las CNN están impulsando avances significativos en áreas como el análisis de imágenes médicas, debido a su capacidad para capturar estructuras espaciales complejas y patrones relevantes, optimizando así la precisión en diagnóstico y tratamiento en aplicaciones de salud [84].

- **Aprendizaje por transferencia (Transfer Learning)**

La transferencia de aprendizaje es una técnica en aprendizaje profundo que permite aprovechar redes neuronales convolucionales previamente entrenadas, adaptándolas a nuevos conjuntos de datos o problemas específicos sin necesidad de iniciar el entrenamiento desde cero, reduciendo el tiempo de entrenamiento y mejorando el rendimiento del modelo en conjunto de datos limitados [24].

VGG: el Visual Geometry Group (VGG) de la Universidad de Oxford es un grupo de investigación y es reconocido por sus aportes en visión por computadora, destacándose especialmente por desarrollar la arquitectura de redes neuronales convolucionales conocida como VGG. Esta arquitectura, ampliamente utilizada en tareas de clasificación y detección de objetos en imágenes, se estructura en bloques repetitivos de capas convolucionales intercaladas con capas de agrupamiento (pooling) y culmina con capas completamente conectadas. Las versiones más conocidas de esta red son VGG16 y VGG19 [85].

ResNet: una ResNet, o Red Neuronal Residual, es una arquitectura de red profunda destacada en aprendizaje profundo y visión por computadora. Su característica distintiva es el uso de conexiones residuales o "skip connections", que permiten que la información salte una o más capas. Esto ayuda a mitigar el problema del desvanecimiento del gradiente, frecuente en redes muy profundas, al permitir que los gradientes fluyan más fácilmente a través de la red y facilitando así su entrenamiento eficiente [86].

DenseNet: una DenseNet (Dense Convolutional Network) es un tipo de red neuronal convolutacional que conecta cada capa con todas las capas anteriores de manera directa. Esto significa que cada capa recibe como entrada las características de todas las capas anteriores, lo que ayuda a mitigar el problema del gradiente desvaneciente, mejora la propagación de características, fomenta la

reutilización de características y reduce significativamente el número de parámetros [87].

Inception V3: Inception v3 es un modelo avanzado de reconocimiento de imágenes que ha alcanzado una precisión superior al 78.1% en el conjunto de datos de ImageNet, consolidando diversas ideas de investigación desarrolladas a lo largo de años en el campo de la visión por computadora. Este modelo se compone de bloques de construcción que combinan operaciones simétricas y asimétricas, tales como convoluciones, capas de reducción promedio y máxima, concatenaciones, retiradas y capas completamente conectadas [88].

2.2.4 Métricas de desempeño

La evaluación del rendimiento de los clasificadores se realiza mediante métricas estandarizadas que permiten cuantificar su eficacia en la detección de patrones.

La Tabla 2.2 presenta las definiciones clave para su cálculo, enfocadas en la detección de masas en imágenes médicas.

Tabla 2.2. Definiciones para métricas de desempeño en la clasificación.

Métrica	Definición	Importancia
True Positive (TP)	Imágenes con masa clasificadas correctamente como tales.	Indica la capacidad del sistema para detectar masas correctamente.
False Positive (FP)	Imágenes sin masa clasificadas incorrectamente como con masa.	Refleja los errores del sistema, generando posibles alarmas falsas.
True Negative (TN)	Imágenes sin masa clasificadas correctamente como tales.	Demuestra la precisión del sistema al evitar falsas alarmas.
False Negative (FN)	Imágenes con masa clasificadas incorrectamente como sin masa.	Crítico en diagnósticos médicos, ya que puede llevar a omitir detecciones importantes.

Fuente: [Elaboración propia]

Las anteriores definiciones implican llevar a cabo un recuento de píxeles en la imagen, clasificándolos como se puede observar en la Figura 2.2.

Figura 2.2. Matriz de confusión con métricas de desempeño.

		True condition		
Predicted Condition	Actual Positive	TP	FP	Precision
	Actual Negative	FN	TN	Negative Predicture Value
		Sensitivity (Recall)	Specifity	Accuracy

Fuente: Tomado de [89]

2.2.4.1 Exactitud (Accuracy)

Esta métrica evalúa la proporción de datos o píxeles correctamente identificados respecto al total de puntos en la imagen. Se expresa mediante la ecuación (2.23).

$$Accuracy = \frac{VP + VN}{VP + FP + FN + VN} \quad (2.23)$$

2.2.4.2 Precisión (Presicion)

Esta métrica calcula la proporción de datos o píxeles correctamente identificados como positivos del total de datos o píxeles identificados como positivos. Se expresa mediante la ecuación (2.24).

$$Precision = PPV = \frac{VP}{VP + FP} \quad (2.24)$$

2.2.4.3 Sensibilidad (Recall)

Esta métrica se usa cuando los FN son inaceptables. En forma práctica, puede considerarse como la exactitud de los casos positivos. Se expresa mediante la ecuación (2.25).

$$Sensitivity = \frac{VP}{VP + FN} \quad (2.25)$$

2.2.4.4 Especificidad (Specificity)

Esta métrica se usa cuando los FP son inaceptables. En forma práctica, puede considerarse como una exactitud de los casos negativos. Se expresa mediante la ecuación (2.26).

$$Specificity = \frac{VN}{VN + FP} \quad (2.26)$$

2.2.4.5 Índice F1-Score

Esta métrica corresponde a la media armónica de los datos o píxeles. En este caso, se combina la sensibilidad y la precisión en una sola métrica. Se expresa mediante la ecuación (2.27).

$$F1\ Score = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \quad (2.27)$$

2.2.5 Discusión

Basándose en la información recopilada a través de la revisión de la literatura, se identificaron metodologías y algoritmos adecuados para abordar el problema en cuestión. Se decidió implementar un sistema compuesto por cinco etapas principales (véase Figura 3.1), el cual se evaluará utilizando mamografías digitales del conjunto de datos MIAS [18].

La base de datos MIAS contiene 322 imágenes en formato .pgm (archivos portables que almacenan datos de imágenes en mapas de escala de grises). De estas, se seleccionaron 266 imágenes relevantes para este estudio, de las cuales 207 son normales, 40 presentan masas benignas y 19 contienen masas malignas.

Para el procesamiento inicial, se plantea un preprocesamiento que incluye la mejora del contraste de las imágenes, seguido de la segmentación mediante algoritmos basados en umbrales y operaciones morfológicas. Estas etapas permiten aislar las regiones de interés en las mamografías, facilitando la posterior extracción de características.

Para el proceso de extracción de características, se identificó la necesidad de aplicar un aumento de datos al dataset original mediante el método de ventanas traslapadas, el cual permite extraer características estadísticas en el dominio

espacial. A partir de este conjunto, se consideró la transformación al dominio frecuencial para obtener el espectro correspondiente a cada ventana, generando así un segundo dataset enfocado en el análisis frecuencial.

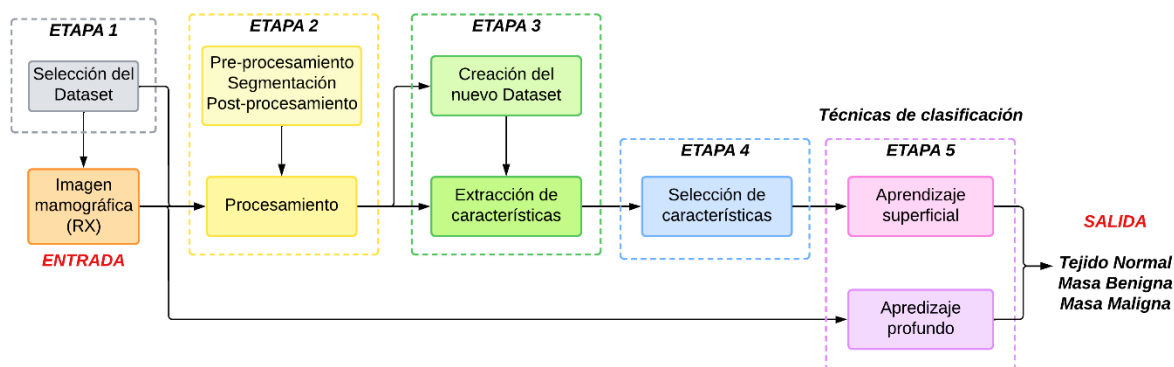
Posteriormente, ambos conjuntos de características, espaciales y frecuenciales, se analizaron mediante técnicas de ranking y wrappers, con el fin de identificar las características más discriminantes y optimizar su selección para el desarrollo y evaluación de los clasificadores.

Finalmente, en cuanto a los enfoques de clasificación, se seleccionaron algoritmos de aprendizaje superficial, como Naive Bayes, LDA, QDA, distancias de Mahalanobis y Euclidiana, KNN, SVM, Random Forest, MLP y RBF, considerando las características discriminantes previamente seleccionadas como vectores de entrada. Además, se identificaron técnicas de aprendizaje profundo, como el diseño de redes neuronales convolucionales y el uso de transferencia de aprendizaje mediante redes preentrenadas como VGG, ResNet, DenseNet e InceptionV3, dada su capacidad para extraer y procesar características de manera jerárquica.

3. METODOLOGÍA DE LA SOLUCIÓN PROPUESTA: SOFTWARE

Se describe la metodología utilizada para desarrollar un sistema de visión artificial para la detección y clasificación automática de tejido normal, masas benignas y masas malignas en mamografías digitales. El sistema se realizó en 5 etapas fundamentales como se ilustra en la Figura 3.1. Las etapas corresponden a: (1) *selección de la base de datos*, (2) *procesamiento*, (3) *extracción de características*, (4) *selección de características* y (5) *técnicas de clasificación*. Estas etapas permiten al sistema aplicar procedimientos de preprocesamiento y clasificación a una mamografía digital de entrada y generar una salida que indique la presencia de tejido normal, masa benigna o masa maligna.

Figura 3.1. Diagrama de bloques del sistema desarrollado.



Fuente: [Elaboración propia]

El sistema desarrollado permite seleccionar entre dos procedimientos de clasificación: aprendizaje superficial y aprendizaje profundo. En el aprendizaje superficial, se emplean clasificadores estadísticos y algunos algoritmos de aprendizaje automático como Naive Bayes, LDA, QDA, KNN, Distancia Mahalanobis, Distancia Euclidiana, SVM, RF, RBF y MLP. Este enfoque requiere que las etapas de procesamiento, extracción y selección de características se realicen previamente a la clasificación (camino superior de la Figura 3.1).

Por otro lado, el aprendizaje profundo realiza internamente las etapas de procesamiento, extracción y selección de características, optimizando el flujo de trabajo mediante arquitecturas como Redes Neuronales Convolucionales (CNN). Este procedimiento también incluye el uso de modelos preentrenados, como VGG, ResNet, DenseNet e InceptionV3, mediante transferencia de aprendizaje (transfer

learning), aprovechando las características jerárquicas que estas redes pueden extraer automáticamente de las imágenes (camino inferior de la Figura 3.1).

La **ETAPA 1. Selección de la base de datos**, es fundamental para iniciar el desarrollo del sistema, ya que se enfoca en la adquisición de imágenes. Durante esta etapa, se evalúan los criterios para seleccionar el conjunto de datos, basándose en una revisión exhaustiva de la literatura relacionada con trabajos similares.

El enfoque inicial, correspondiente a la clasificación mediante aprendizaje, avanza con la **ETAPA 2. Procesamiento**. Esta etapa se compone de tres pasos fundamentales: preprocesamiento, segmentación y postprocesamiento (si es necesario). Estas fases están diseñadas para destacar la información relevante y eliminar elementos irrelevantes de la mamografía, tales como artefactos, etiquetas y componentes externos a la región de interés, lo que optimiza el rendimiento en las etapas posteriores.

La **ETAPA 3. Extracción de características**, implica un paso adicional que consiste en la creación de una nueva base de datos mediante el método de ventanas traslapadas. A partir de esta base de datos, se procede a la extracción de características de las mamografías, obteniendo así información más relevante y útil para la solución propuesta.

Con las características extraídas, se procede a la **ETAPA 4. Selección de características**, cuyo objetivo es identificar las características más discriminantes para simplificar y mejorar la posterior clasificación del tejido. Durante esta etapa, se aplican técnicas de ponderación para reducir la dimensionalidad del problema. Este análisis permite determinar la relevancia de las características tanto de manera individual como en conjunto.

Finalmente, en la **ETAPA 5. Técnicas de clasificación**, se realiza el entrenamiento y la validación de los modelos de aprendizaje superficial y profundo. Para ello, se asigna el 70% de los datos al entrenamiento de diversos algoritmos de clasificación, mientras que el 30% restante se utiliza para la validación, utilizando ya sea datos no vistos anteriormente o con nuevas imágenes.

Los parámetros finales de cada una de las etapas descritas se ajustan según el desempeño de la clasificación, evaluado mediante un conjunto de métricas que serán detalladas en el capítulo correspondiente a pruebas y resultados.

Este enfoque metodológico garantiza una mejora continua del sistema, permitiendo una clasificación precisa y eficiente de las mamografías, esencial para la detección de tejido normal, masas benignas y masas malignas.

3.1 ETAPA 1. Selección de la base de datos

En el ámbito del desarrollo y estudio de sistemas de visión artificial, enfocado en este caso de investigación educativo, se decide realizar la adquisición de la imagen mediante la selección de la base de datos. A partir de la revisión de la literatura, aunque existen múltiples bases de datos de mamografías digitales, la mayoría tienen acceso restringido, ya que son creadas por grupos de investigación y entidades médicas como clínicas y hospitales. Sin embargo, algunas bases de datos de libre acceso han demostrado ser altamente confiables, como se menciona en [90]. En la Tabla 2.1 se enumeran varias de estas bases de datos reportadas en la literatura. Basándose en esta revisión, se eligió la base de datos MIAS para este trabajo, cuyas características principales se describen a continuación.

Base de datos MIAS: El conjunto de datos utilizado fue descargado desde la plataforma *Kaggle*, donde se presenta bajo el nombre “MIAS dataset”. Sin embargo, corresponde realmente al *mini-MIAS*, una versión derivada del MIAS original. Esta versión fue modificada para tener imágenes en formato PGM, con dimensiones de 1024×1024 píxeles y un tamaño de píxel ajustado a 200 micras, en lugar del original de 50 micras.

Tabla 3.1. Información de la base de datos pública MIAS.

COLUMNA	DESCRIPCIÓN
1	Número de referencia de la base de datos MIAS
2	Tipo de tejido de fondo: <i>F: Graso G: Graso-glandular D: Denso-glandular</i>
3	Tipo de anomalía presente: <i>CALC: Calcificación</i> <i>CIRC: Masas bien definidas/circunscritas</i> <i>SPIC: Masas espiculadas</i> <i>MISC: Otras masas mal definidas</i> <i>ARCH: Distorsión arquitectónica</i> <i>ASYM: Asimetría</i> <i>NORMAL: Tejido sano</i>
4	Naturaleza de la anomalía: <i>B: Benigno M: Maligno</i>
5 y 6	Coordenadas (x,y) del centro de la anomalía
7	Radio aproximado (en píxeles) de un círculo que abarca la anomalía

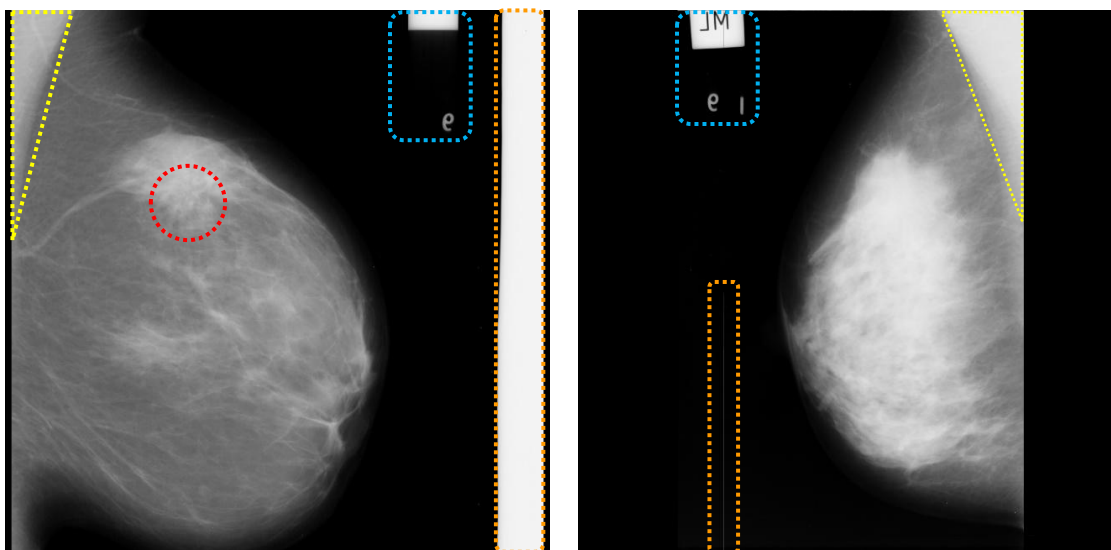
Fuente: [Elaboración propia]

Cabe mencionar que algunos enlaces oficiales al dataset original ya no están disponibles, lo que dificulta la verificación directa de la correspondencia. Adicionalmente, se ha reportado que algunas imágenes, como *mdb212* y *mdb214*, presentan inconsistencias en sus etiquetas (anomalías “CALC” que en realidad no están presentes). Esta base de datos puede ser utilizada en investigaciones científicas, siempre que se respeten los términos del acuerdo de licencia. La base de datos MIAS proporciona la información reportada en la Tabla 3.1.

La base de datos MIAS fue seleccionada para este estudio debido a que proporciona *ground truth* de anomalías mediante coordenadas y radios, además de ofrecer una resolución adecuada, aspectos fundamentales para el enfoque principal de esta investigación orientada a la detección de masas. Esta base de datos contiene 322 imágenes con siete tipos de lesiones mamarias, de las cuales se seleccionaron 266 imágenes para este estudio: 207 normales, 40 con masas benignas y 19 con masas malignas.

La Figura 3.2 ilustra un ejemplo de las mamografías seleccionadas, en las que se identifican elementos como el tejido mamario, el tejido pectoral (línea amarilla discontinua), las coordenadas del tejido con anomalía (línea roja discontinua), así como artefactos presentes, tanto por etiquetas del paciente (línea azul discontinua) como por errores de digitalización (línea naranja discontinua). Para facilitar su visualización, estos componentes fueron resaltados manualmente sobre las imágenes.

Figura 3.2. Imágenes mamográficas con artefactos y mala digitalización.



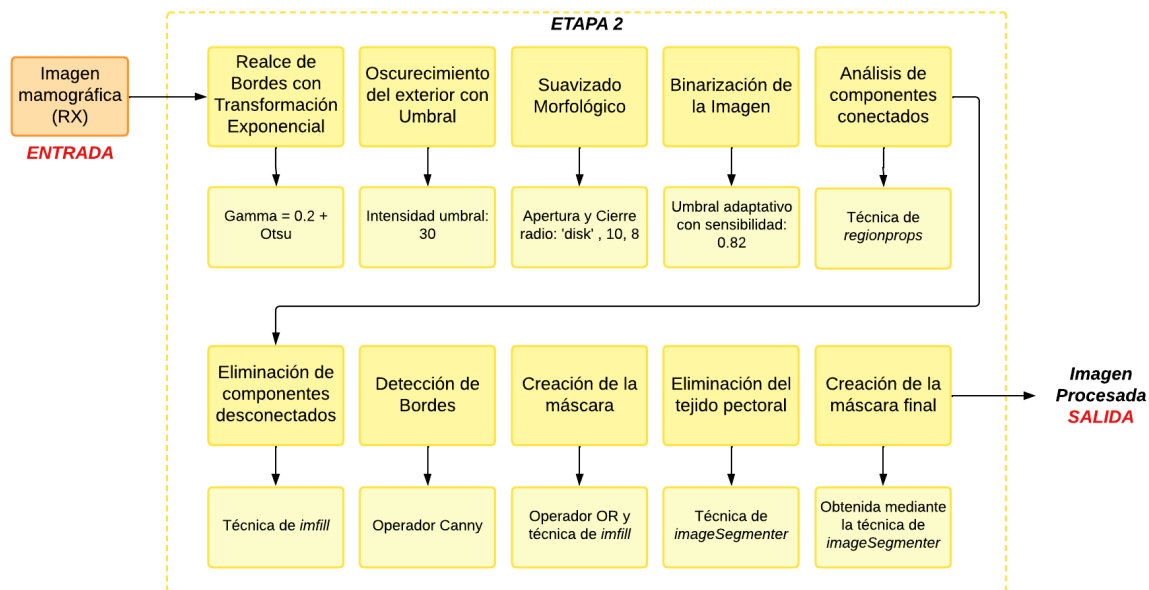
Fuente: [Elaboración propia]

3.2 ETAPA 2. Procesamiento

La finalidad de esta etapa es la generación de una máscara binaria que delimite la región de la mama en la mamografía, para eliminar artefactos como etiquetas, caracteres alfanuméricos, marcas no relacionadas con el tejido mamario y el tejido pectoral. La eliminación de estos elementos es crucial, ya que su presencia puede interferir con el proceso de aprendizaje automático y reducir el rendimiento del sistema de clasificación. La remoción de estos artefactos busca mejorar la precisión y eficiencia de la clasificación del tejido en las imágenes.

Este procesamiento tiene como objetivo optimizar la segmentación de las imágenes, asegurando que solo las regiones relevantes sean consideradas para el análisis. La Figura 3.3 muestra las diez subetapas que componen esta técnica, las cuales se describen a continuación.

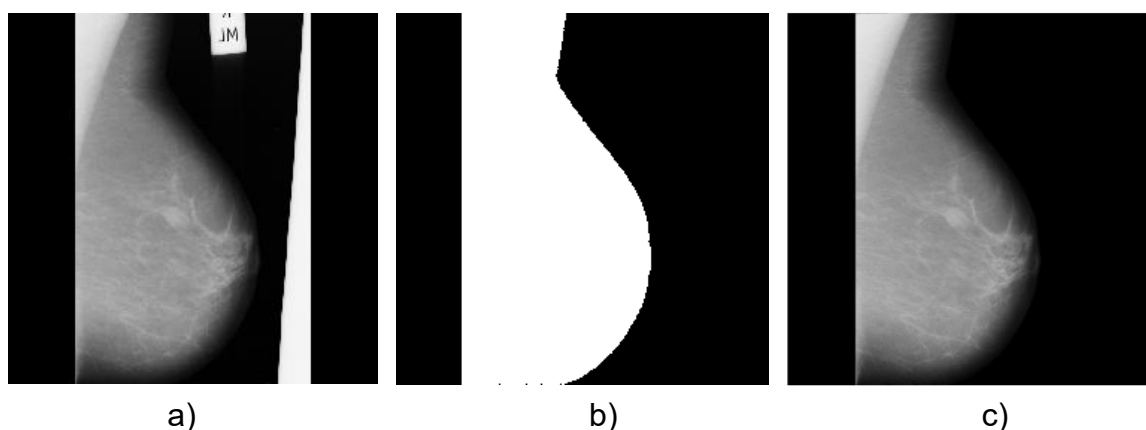
Figura 3.3. Diagrama de bloques ampliado de la ETAPA 2. Procesamiento.



Fuente: [Elaboración propia]

La máscara binaria se utiliza para segmentar la región de interés en las mamografías, asignando un valor de blanco (255 o 1 lógico) a la región correspondiente a la mama y negro (0 o 0 lógico) al resto de la imagen (Figura 3.4b). Mediante una operación de multiplicación punto a punto entre la máscara y la mamografía original (Figura 3.4a), se logra eliminar los artefactos presentes en la imagen, como se ilustra en la Figura 3.4c.

Figura 3.4. Segmentación de Mamografía con la ETAPA 2. Procesamiento: a) Mamografía original 'mdb012.png'; b) Máscara obtenida; c) Segmentación obtenida



Fuente: [Elaboración propia]

A continuación, se describe cada una de las subetapas de la técnica de procesamiento desarrollada.

Realce del borde mediante transformación exponencial

La delimitación entre la región de interés y el fondo de una mamografía suele ser difusa, dificultando su identificación. Para mejorar el contraste entre ambas regiones y facilitar la detección del borde de la mama, se aplica una transformación exponencial que amplía el rango de niveles de gris en la imagen [40]. Este método permite destacar las áreas más claras, donde generalmente se encuentra la región de interés.

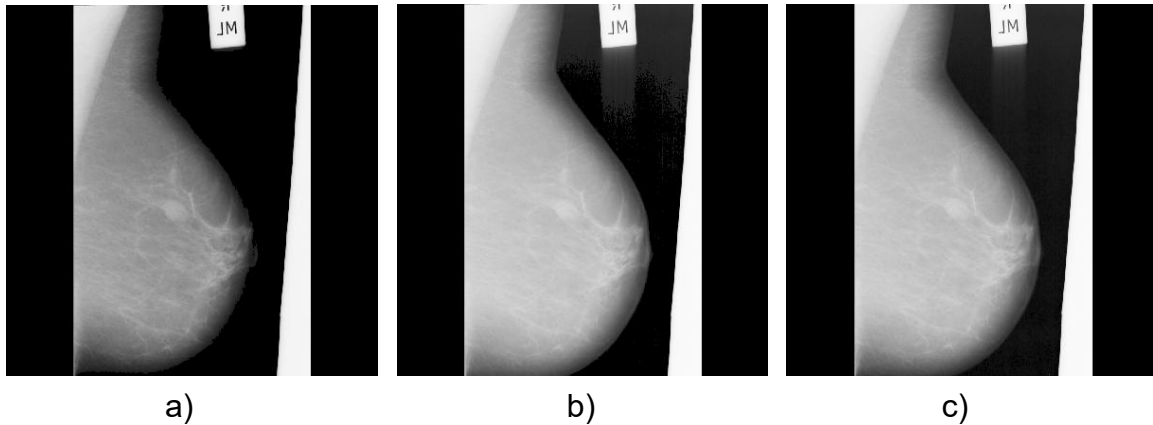
Dado que cada mamografía presenta características únicas, el valor del parámetro gamma (γ), utilizado en la transformación, debe adaptarse a cada imagen. Para determinar este valor, se emplea el método de Otsu. Aunque este método se utiliza comúnmente para definir umbrales, en este caso, su resultado se aprovecha para ajustar el valor de γ de manera dinámica [40].

Durante el desarrollo, se evaluaron distintas estrategias para la calibración de γ , mediante operaciones estadísticas sobre la imagen original, pero la sensibilidad del método a la variabilidad de las mamografías generó inconsistencias en la segmentación del borde. Para superar esta limitación, se incorporó un ajuste heurístico de 0.2 al umbral de Otsu.

La Figura 3.5 muestra los efectos de la transformación exponencial en distintos niveles: la 3.5a presenta bordes poco definidos con el umbral aplicado directamente;

la 3.5b mejora el realce al combinar la transformación con el umbral; y la 3.5c logra una segmentación más precisa al incorporar el ajuste heurístico de 0.2.

Figura 3.5. a) Umbral aplicado directamente; b) Transformación exponencial + umbral; c) Transformación exponencial + ajuste heurístico (0.2).

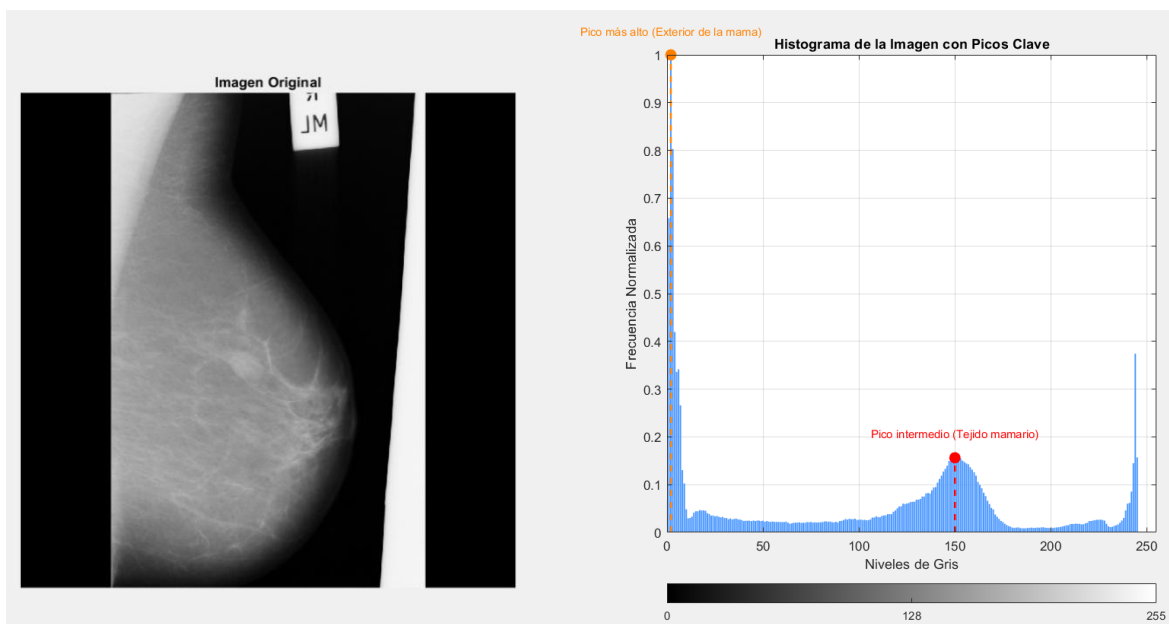


Fuente: [Elaboración propia]

Oscurecimiento del Exterior con umbral definido heurísticamente

Como se muestra en la Figura 3.6b, el exterior de la mama aparece representado en un pico máximo del histograma, ubicado en las intensidades más bajas.

Figura 3.6. a) Mamografía original 'mdb012.png'; b) Histograma correspondiente.



Fuente: [Elaboración propia]

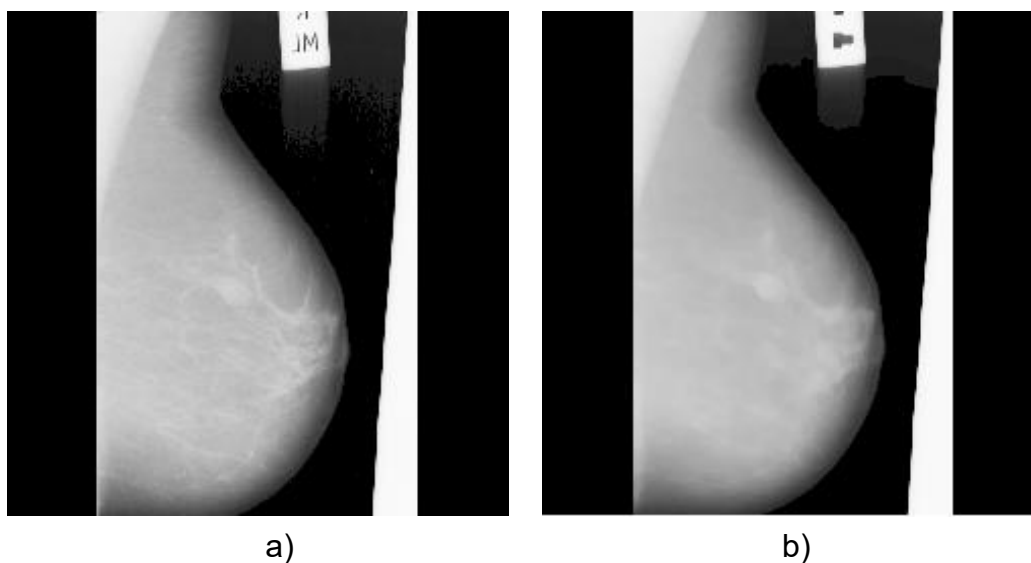
Para acentuar la diferencia entre las regiones de interés y el fondo, se establece de manera heurística un umbral que convierte a negro todos los píxeles con intensidades inferiores a este valor. Esta técnica, aplicada después de la transformación exponencial, mejora la visibilidad del borde de la mama al oscurecer el área externa, asegurando una separación clara entre la región de interés y el fondo.

La transformación exponencial no solo realza el borde de la mama, sino que también incrementa la intensidad de los píxeles en el exterior. Esto hace necesario aplicar un umbral para mantener esa área completamente negra (intensidad = 0) y evitar que interfiera en el procesamiento. Como se observa en la Figura 3.7a, la combinación de la transformación exponencial con el ajuste y con el umbral de oscurecimiento genera un fondo más homogéneo, mejorando la diferenciación entre el borde de la mama y el fondo, lo que facilita las subetapas posteriores del procesamiento.

Suavizado Morfológico

El suavizado morfológico se utiliza para garantizar que la textura de la región de interés en las mamografías sea homogénea, eliminando irregularidades y artefactos que puedan interferir en el análisis posterior. Este proceso combina las operaciones de apertura y cierre para ajustar los contornos y eliminar pequeñas imperfecciones.

Figura 3.7. a) Imagen con ajuste y umbral de oscurecimiento; b) Imagen umbralizada con suavizado morfológico.



Fuente: [Elaboración propia]

Además, se emplean técnicas adicionales como dilatación y erosión de forma sucesiva, utilizando diferentes radios estructurales para ajustar la forma del borde de la mama. Estas operaciones aseguran que los contornos permanezcan definidos sin perder detalles importantes, lo que es crucial para las etapas posteriores de clasificación.

El uso de un suavizado morfológico es fundamental para asegurar que el contorno de la región de la mama esté claramente definido y sin distorsiones, lo que permite una segmentación. El resultado de la aplicación de este suavizado se aprecia en la Figura 3.7b.

Binarización de la Imagen

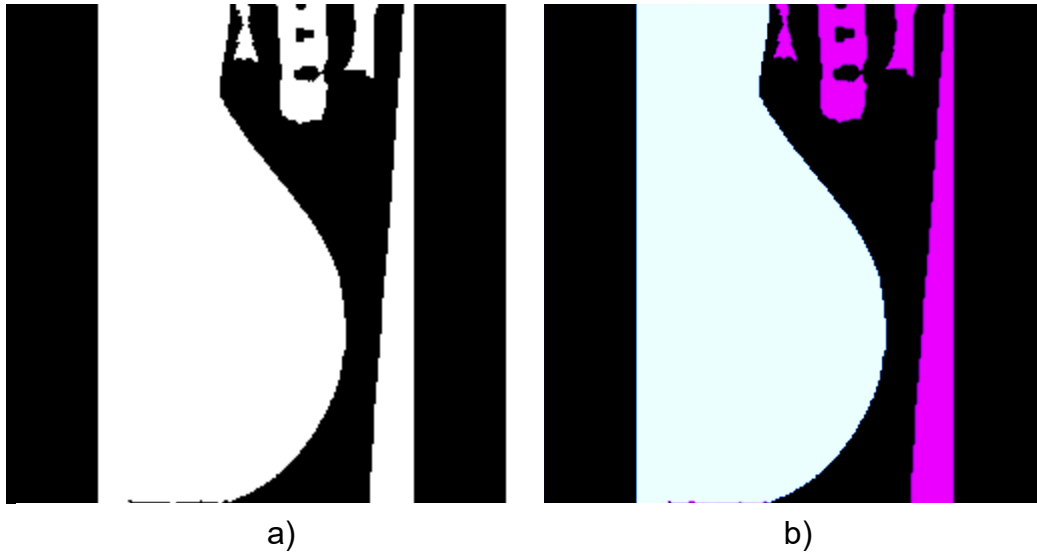
Posteriormente, se binariza la imagen (función *imbinarize()* de Matlab) utilizando un umbral adaptativo con una sensibilidad de 0.82. Este paso es clave para separar las áreas de interés del fondo, ya que convierte la imagen en una representación en blanco y negro, donde las áreas blancas representan las estructuras más relevantes, en este caso la región de la mama.

El uso del suavizado morfológico previo facilita la obtención de un contorno más claro y definido de la región de la mama. Si bien la evaluación se realizó de manera visual, se identificaron mejoras significativas en la homogeneidad y continuidad de los bordes en comparación con resultados sin esta etapa. Estas observaciones están respaldadas por la literatura, que destaca la eficacia del suavizado morfológico para optimizar procesos de segmentación [15]. Como se muestra en la Figura 3.8a, esta combinación de técnicas permite una representación más precisa de la región de interés.

Análisis de componentes conectados

Una vez obtenida la imagen binarizada, se realiza un análisis de componentes conectados utilizando *bwlabel()* de Matlab, que asigna etiquetas a las diferentes regiones conectadas de la imagen. Con la función *regionprops()* también de Matlab, se extraen propiedades como el área y la ubicación de los píxeles. En la Figura 3.8b se pueden observar las regiones de menor área (color fucsia) y que selecciona la región de mayor área (color azul), que se asume como la mama, y se genera una máscara binaria que aísla dicha región para su posterior refinamiento.

Figura 3.8. a) Imagen binarizada; b) Imagen con análisis de componentes conectados.



Fuente: [Elaboración propia]

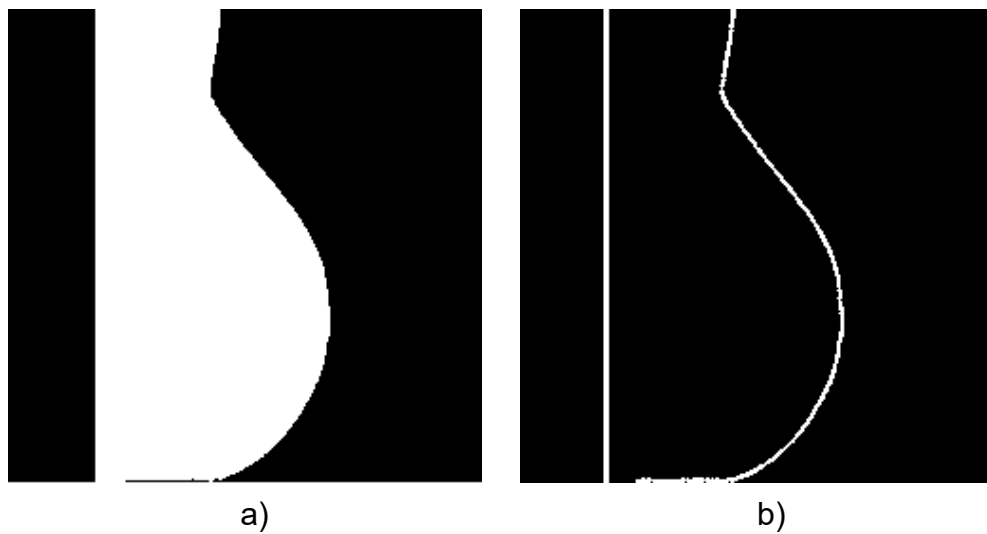
Eliminación de componentes desconectados

Para refinar la máscara obtenida, se emplea la función *imfill()* que rellena los huecos en la región de la mama, seguida de *bwareaopen()* con un valor mínimo de área de 5.000 píxeles. Esto elimina las regiones desconectadas o de menor importancia, como se puede observar en la Figura 3.9a permitiendo mantener únicamente las estructuras más grandes y relevantes para el análisis.

Detección de Bordes

Se utiliza el operador de Canny mediante la función *edge()* para detectar los bordes en la región binaria. Este operador es eficaz para encontrar cambios abruptos en la intensidad, como los que ocurren en los bordes de las mamas. Los bordes detectados se dilatan con *imdilate()*, utilizando un elemento estructural en forma de diamante (*strel('diamond', 6)*), lo que genera contornos más robustos y continuos. El resultado se puede observar en la Figura 3.9b.

Figura 3.9. a) Imagen binarizada mejorada. b) Detección de bordes mediante el operador de Canny.



Fuente: [Elaboración propia]

Creación de la máscara

A continuación, se crea la máscara definitiva combinando la máscara de la región de la mama con los bordes dilatados mediante `mask_combined = mask_cleaned_final | edges_dilated`, como se puede observar en la Figura 3.10. Luego, se rellenan los huecos usando nuevamente `imfill()`, y esta máscara se aplica a la imagen original, oscureciendo todas las áreas fuera de la región de interés, como ilustra la Figura 3.11a. Finalmente, la imagen segmentada resultante se guarda en un nuevo directorio para su uso posterior.

Figura 3.10. Máscara final: superposición de Figura 3.8b. y Figura 3.9a.



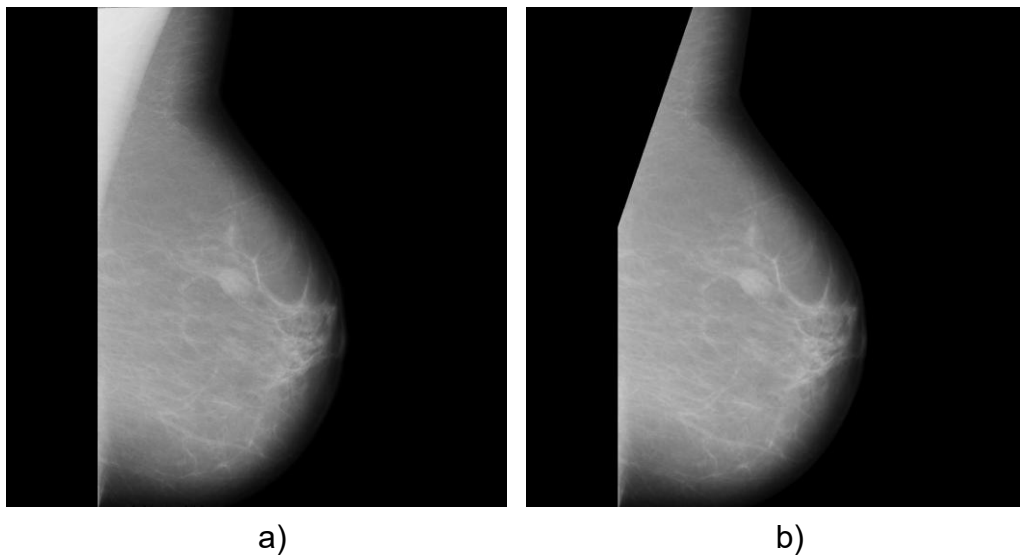
Fuente: [Elaboración propia]

Eliminación del tejido pectoral

Dado que el tejido pectoral carece de relevancia para la detección de masas (al no presentar anomalías en esta zona muscular), su presencia en la imagen introduce un costo computacional innecesario y aumenta el riesgo de falsos positivos en la detección. Por estas razones, resulta esencial segmentar y eliminar esta región, lo que permite concentrar el análisis exclusivamente en el tejido mamario, reduciendo el tiempo y los recursos de procesamiento requeridos para la inspección.

Finalmente, para la segmentación del tejido pectoral se emplea la función *imageSegmenter()* de Matlab. Esta herramienta implementa técnicas basadas en segmentación por región y umbral, permitiendo al usuario interactuar con la imagen procesada para identificar y excluir manualmente la región pectoral. Con la imagen libre de artefactos, como se muestra en la Figura 3.11a, se genera una máscara binaria final que contiene únicamente el área de interés correspondiente al tejido mamario, como se observa en la Figura 3.11b.

Figura 3.11. Segmentación de la mamografía: a) Mamografía segmentada obtenida con tejido pectoral; b) Mamografía segmentada obtenida sin tejido pectoral.



Fuente: [Elaboración propia]

3.3 ETAPA 3. Extracción de características

La etapa de extracción de características tiene como objetivo generar un conjunto de datos compuesto por vectores de características, fundamentales para el

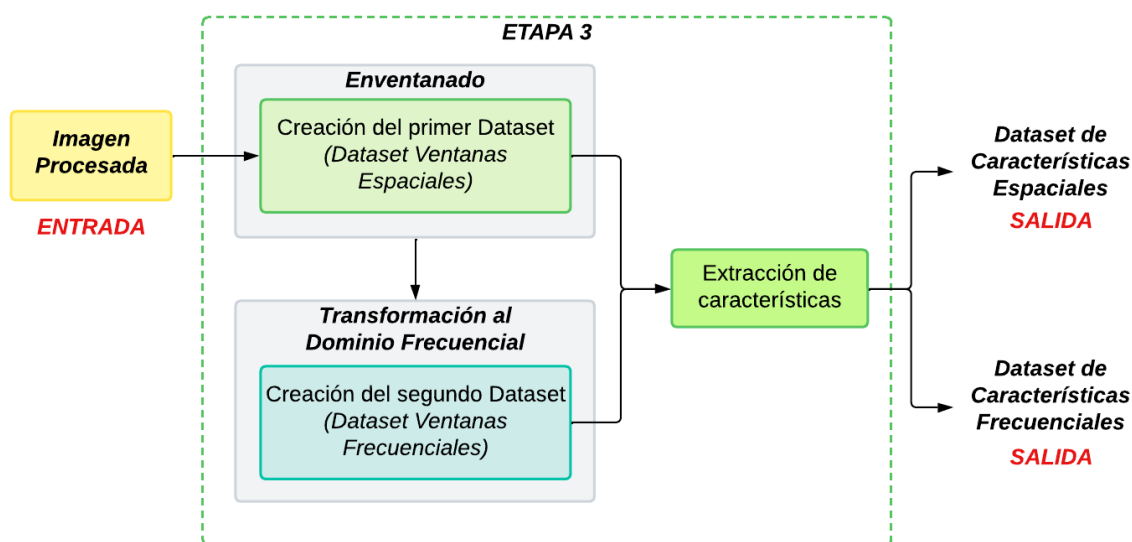
entrenamiento y evaluación de los clasificadores en el análisis de imágenes mamográficas. En esta etapa se obtienen características en el dominio espacial y en el dominio frecuencial, buscando una mejor representación de las regiones de interés en las mamografías.

Para obtener las *características en el dominio espacial* se aplica la *técnica de enventanado*, para obtener regiones rectangulares, ventanas o parches de las mamografías, generando así un conjunto de pequeñas imágenes. El tamaño de cada ventana se determina mediante un análisis estadístico de los tamaños de las masas detectadas en las mamografías, lo que asegura una división adecuada de las regiones de interés. Las ventanas se obtienen tanto en áreas con masas como en regiones sin ellas, como se observa en la Figura 3.14., Figura 3.15. y Figura 3.16.

Para obtener las *características en el dominio frecuencial* se aplica la *Transformada de Fourier* a las ventanas generadas en la fase previa. Esto permite consolidar un segundo conjunto de datos que representa el dominio frecuencial, como ilustra la Figura 3.18., complementando así el análisis en el dominio espacial. Esta transformación enriquece la capacidad de detección y caracterización de patrones y texturas asociados a las masas.

A cada ventana del dominio espacial y frecuencial se le extraen características estadísticas y de energía que se agrupan en vectores. Figura 3.12. ilustra de forma esquemática la ETAPA 3 de este proceso.

Figura 3.12. Diagrama de bloques ampliado de la ETAPA 3.

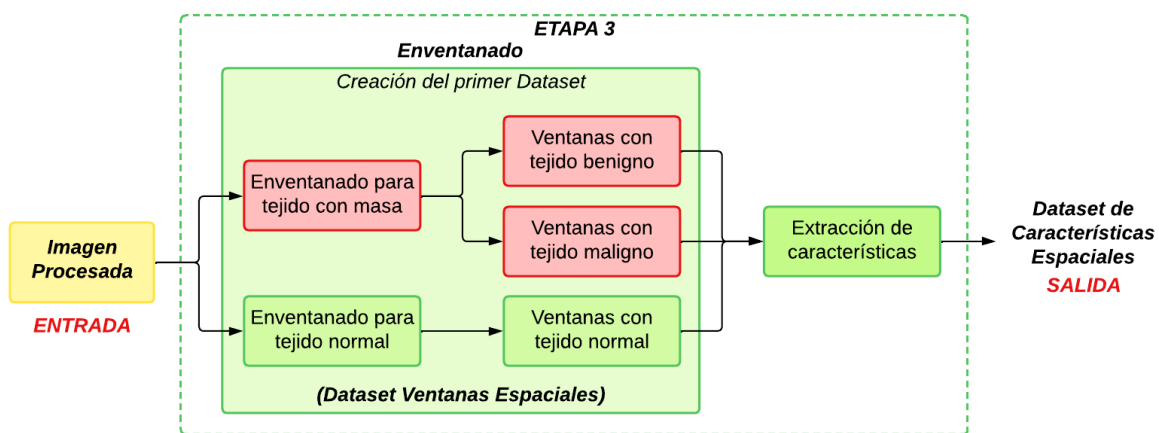


Fuente: [Elaboración propia]

3.3.1 Enventanado

El proceso de enventanado se enfoca en identificar las áreas específicas de las mamografías donde se calcularán los vectores de características para las tres clases bajo estudio: tejido normal, tejido benigno y tejido maligno. En la Figura 3.13. se logra observar de manera más detallada el procedimiento a emplear. Su objetivo es capturar la diversidad de las características del tejido presente en la mamografía en función de la ubicación de la ventana.

Figura 3.13. Diagrama de bloques ampliado para el enventanado en el dominio espacial de la ETAPA 3.



Fuente: [Elaboración propia]

Para lograr esto, se extraen vectores de todas las ventanas posibles dentro de la región de interés. Esta metodología permite obtener una representación completa de las características presentes en diferentes áreas de la imagen mamográfica. Dado que no todas las imágenes en el conjunto de datos MIAS son pertinentes para este análisis, se optó por utilizar únicamente las 266 mamografías que contienen las tres clases de interés. Esto garantiza que el análisis se centre en las imágenes que proporcionan la información necesaria para clasificar con precisión los diferentes tipos de tejido.

3.3.1.1 Enventanado para tejido con masa

Para este propósito específico, se utilizó la información disponible en el conjunto de datos. Inicialmente, se procedió a identificar las coordenadas de los centroides de las masas y sus dimensiones. Luego, se implementó el Método de Ventanas Traslapadas en cada una de ellas.

La cantidad de ventanas utilizadas en el análisis de las masas se decide considerando tres aspectos esenciales: las dimensiones del área a enventanar, el grado de superposición entre las ventanas y el tamaño de la ventana.

- **Las dimensiones del área a enventanar (masas):** acotan las zonas donde se aplicará el método de ventanas traslapadas.
- **El grado de superposición:** entre las ventanas se define por el porcentaje de traslape, fijado en un 95% en esta instancia, con la intención de obtener la máxima información dentro de ventanas adyacentes.
- **El tamaño de la ventana:** se ajusta para cubrir una parte significativa del área del objeto de interés (masas), garantizando que las características calculadas sean representativas. Esta dimensión se determina mediante un análisis de la información sobre el área de las masas proporcionada por el conjunto de datos MIAS.

Tabla 3.2. Estadísticas de las áreas de las masas contenidas en el dataset MIAS.

MEDIDA	VALOR
Diámetro Mínimo	33.8953 px
Diámetro Máximo	393.9742 px
Área Mínima	901 px ²
Área Máxima	121905 px ²
Área Promedio	13368.7241 px ²
Área Moda	7525 px ²

Fuente: [Elaboración propia]

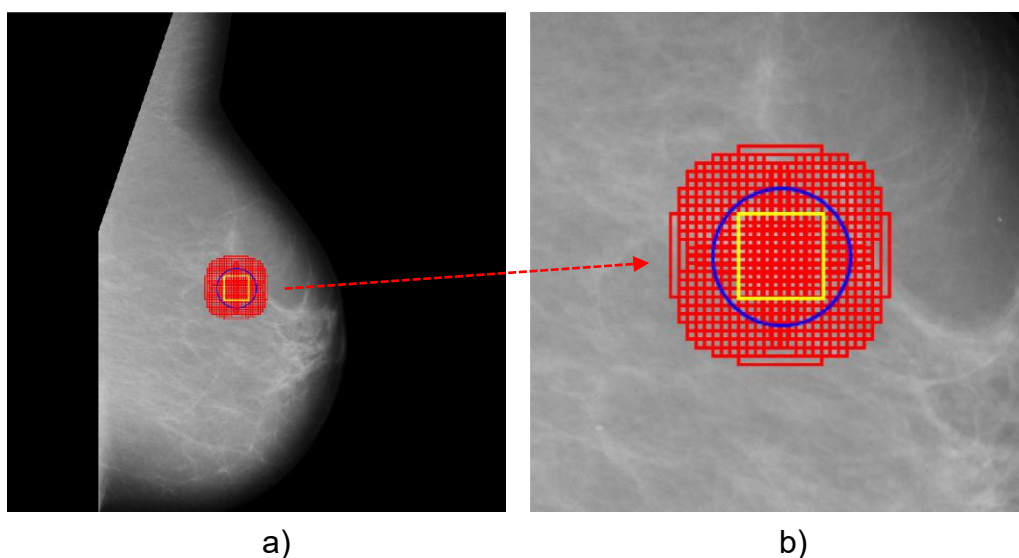
En la Tabla 3.2. se presenta información estadística sobre el área de las masas, calculada a partir de los radios contenidos en el conjunto de datos MIAS. Este análisis estadístico permite relacionar las áreas máximas y mínimas de las masas con un valor propuesto para el área de la ventana.

De acuerdo con los datos extraídos de la Tabla 3.2., se ha seleccionado un área de ventana de 2500 píxeles (cuadrado con lado de 50 píxeles). Esta elección se basa en el hecho de que es aproximadamente el doble del área mínima de una masa y se aproxima a la moda de las áreas. Esto garantiza que las masas mínimas se capturen completamente sin pérdida de información y que las masas más grandes se puedan analizar adecuadamente con múltiples ventanas dentro de su área.

Este procedimiento se realiza para ambas clases de tejido con masas, benigno y maligno, creando un nuevo conjunto de datos. En total, se generan 13.718 ventanas para cada clase (11.278 destinadas para train y 2.440 para test).

La Figura 3.14 ilustra un ejemplo de la ubicación de ventanas generadas para el tejido con masa. En la Figura 3.14a se observa la distribución general, mientras que la Figura 3.14b muestra un acercamiento donde se destacan tres elementos clave: en azul, una circunferencia que indica la ubicación de la masa según las coordenadas proporcionadas por el dataset; en amarillo, el tamaño de la ventana definida; y en rojo, las ventanas generadas alrededor de la masa para su análisis.

Figura 3.14. Distribución de ventanas generadas por los algoritmos de enventanado para el tejido con masa, aplicados a la mamografía mdb012.



Fuente: [Elaboración propia]

3.3.1.2 Enventanado para tejido normal

Para este enventanado, es fundamental que la cantidad de ventanas generadas para el tejido normal sea igual a la de los tejidos con masa (13.718), garantizando así un dataset balanceado. Las ventanas se distribuyen aleatoriamente en las regiones sanas de la mama, considerando dos aspectos clave:

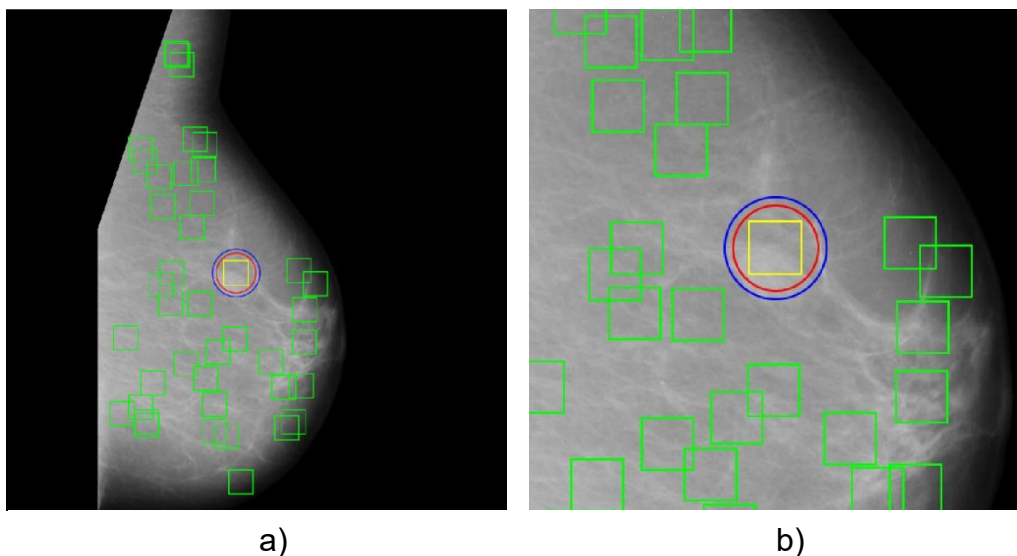
- **Evitar regiones con masa:** el algoritmo fue configurado para no generar ventanas cerca de las masas, restringiendo la selección en función de las coordenadas y el radio de la anomalía, asegurando que solo se analizaran áreas sanas de la mama.

- **Evitar el fondo de la imagen:** durante la etapa de procesamiento, el fondo de la imagen se convierte en negro (intensidad = 0), permitiendo al algoritmo identificar y excluir las ventanas generadas en estas áreas irrelevantes.

El algoritmo implementado cumplió con dos consideraciones clave: aleatoriedad en la selección y la cobertura suficiente de la región de interés. Además, se incluyeron tanto mamografías con tejido normal como aquellas con tejido comprometido para asegurar una mayor diversidad en las características del tejido sano. Como resultado, se generó un dataset de 13.718 ventanas, proporcionando una base de datos balanceada para los análisis posteriores.

La Figura 3.15 muestra un ejemplo de la ubicación de ventanas generadas para el análisis de tejido normal en mamografías que contienen masas. En esta distribución, las ventanas se generan de forma aleatoria en las regiones de tejido normal, evitando las áreas cercanas a la masa. Para garantizar esta exclusión, se utiliza un margen de error del 20% sobre el radio real de la masa, representado por una circunferencia azul. La ubicación exacta de la masa, según las coordenadas del dataset, se marca con una circunferencia roja, mientras que el tamaño de cada ventana está delimitado por un cuadrado amarillo. Las ventanas generadas, marcadas en verde, se distribuyen en el resto de la región mamaria, asegurando que no intercepten con el área delimitada por el margen de error. Este enfoque optimiza la extracción de información exclusivamente del tejido normal.

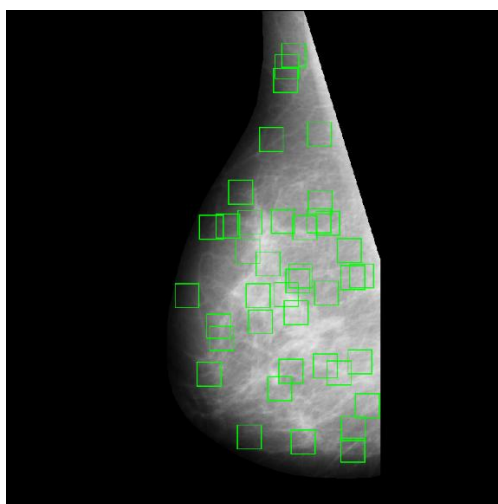
Figura 3.15. Ventanas generadas por los algoritmos de enventanado para el tejido normal en imagen con masa, aplicados a la mamografía mdb012.



Fuente: [Elaboración propia]

Por otro lado, la Figura 3.16 ilustra la distribución de ventanas generadas para mamografías que no contienen masas. En estos casos, las ventanas se generan aleatoriamente a lo largo de toda la región mamaria, permitiendo una cobertura uniforme del tejido normal para su posterior análisis.

Figura 3.16. Distribución de ventanas generadas para tejido normal en mamografías que no contienen masas.



Fuente: [Elaboración propia]

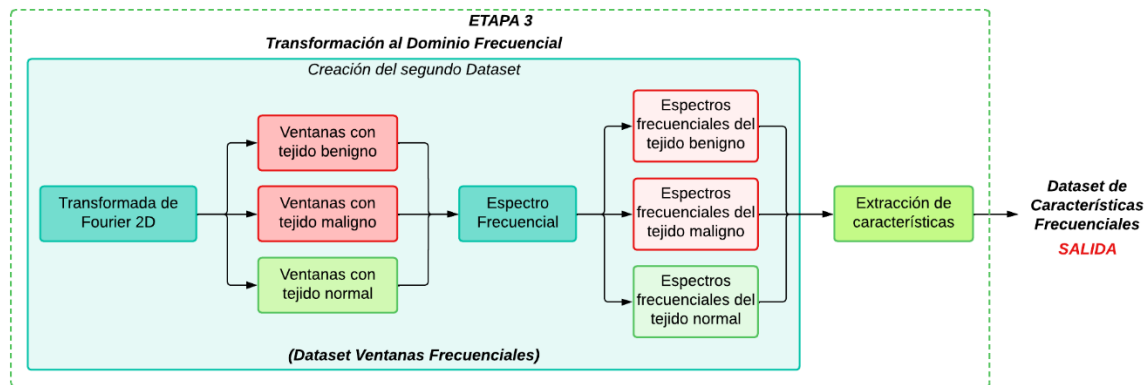
Finalizada la extracción de ventanas para todas las clases, los conjuntos de datos resultantes fueron combinados en un único dataset compuesto por un total de 41.154 ventanas, denominado Dataset de Ventanas Espaciales. Este dataset unificado permite un análisis más eficiente y garantiza que todas las clases estén representadas de manera adecuada para los procesos posteriores de clasificación.

3.3.2 Transformación al dominio frecuencial

Hasta este punto, todas las operaciones realizadas han sido en el dominio espacial, como se ha detallado en el marco teórico. No obstante, una vez obtenido el Dataset de Ventanas Espaciales, se procede a realizar una transformación al dominio frecuencial y de esta manera obtener un nuevo Dataset de Ventanas Frecuenciales. Este cambio es clave, ya que permite analizar la imagen desde una perspectiva distinta, transformando la información espacial en frecuencial mediante técnicas específicas de procesamiento. En la Figura 3.17. se ilustra con mayor detalle el procedimiento empleado para realizar este cambio de dominio.

A diferencia de otros estudios publicados [14], [15] y [16], en este trabajo se integra tanto el análisis en el dominio espacial como en el frecuencial para el proceso de clasificación. Esto proporciona una visión más completa y robusta de las características de las imágenes, mejorando el desempeño en la detección y clasificación de patrones relevantes en los datos.

Figura 3.17. Diagrama de bloques ampliado para la extracción de características del dominio frecuencial de la ETAPA 3.



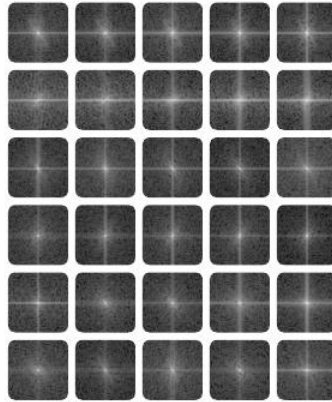
Fuente: [Elaboración propia]

Espectro frecuencial en escala logarítmica

En este bloque, cada ventana extraída del dataset de ventanas es procesada mediante la Transformada de Fourier Discreta bidimensional, lo que permite convertir la imagen del dominio espacial al dominio frecuencial. Posteriormente, se obtiene el espectro de frecuencia en escala logarítmica, que facilita la visualización de las características frecuenciales de la imagen. Este espectro describe la densidad de las frecuencias (eje Y) y su distribución (eje X), lo que resulta útil para identificar la presencia de ruido en zonas específicas de la imagen en función de su frecuencia [65]. La Figura 3.18. presenta una ilustración de estos conceptos.

Una vez finalizada la transformación de los parches de todo el dataset de ventanas, se genera un nuevo conjunto de datos compuesto por 41.154 ventanas, denominado Dataset de Ventanas Frecuenciales. Este nuevo dataset facilita el análisis de las imágenes en el dominio frecuencial, lo que ofrece una perspectiva adicional y complementaria al análisis en el dominio espacial. El enfoque dual de ambos dominios permite identificar diferencias significativas entre las características de las imágenes, optimizando el proceso de clasificación y detección en etapas posteriores.

Figura 3.18. Ejemplos de los espectros de frecuencia generados.

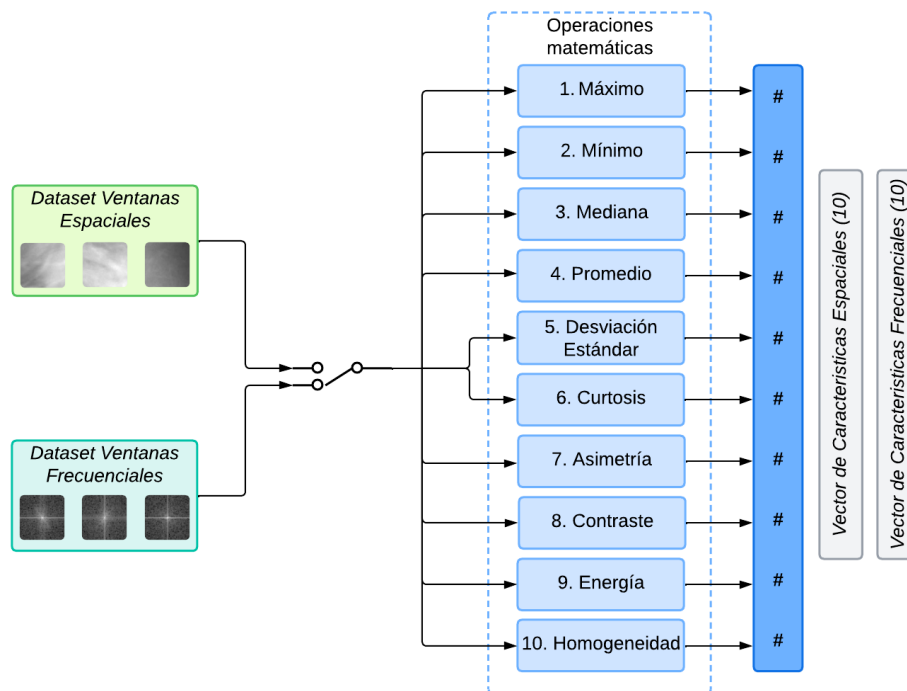


Fuente: [Elaboración propia]

3.3.3 Extracción de características

Una vez obtenidos los datasets de ventanas en ambos dominios, espacial y frecuencial, se procede a convertirlos en un Dataset de Características Espaciales y un Dataset de Características Frecuenciales.

Figura 3.19. Obtención del vector de característica a partir de las ventanas.



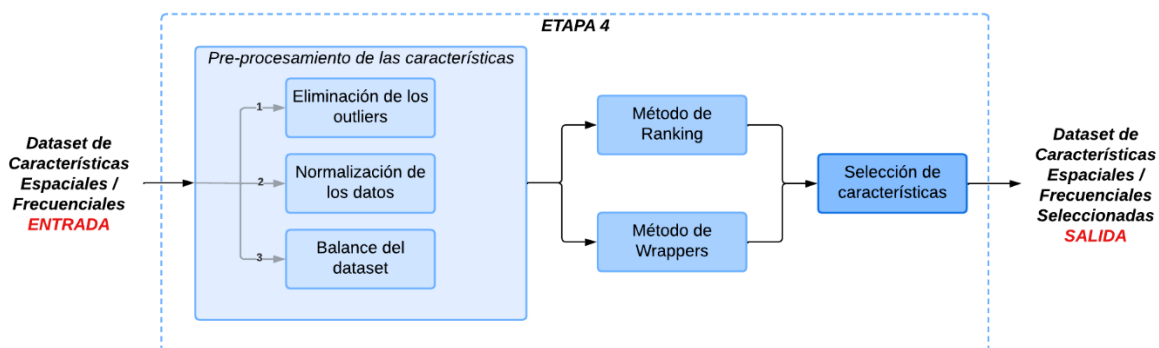
Fuente: [Elaboración propia]

Para ello, cada imagen de ambos datasets es sometida a 10 operaciones matemáticas, lo que transforma cada imagen en un vector de características compuesto por 10 valores. Este proceso permite representar las imágenes de manera más compacta y adecuada para su análisis posterior. En la Figura 3.19. se muestra gráficamente la extracción de este vector de características a partir de las operaciones aplicadas a los datasets.

3.4 ETAPA 4. Selección de características

Inicialmente esta etapa abarca el análisis y el pre-procesamiento del conjunto de datos de características, empleando dos técnicas principales: eliminación de valores atípicos (*outliers*) y la normalización de los datos. A continuación, se explican las técnicas de selección de características que brindan un mayor poder discriminante entre las tres clases de interés. La Figura 3.20. muestra una representación gráfica del proceso llevado a cabo en este bloque.

Figura 3.20. Diagrama de bloques ampliado para la ETAPA 4.



Fuente: [Elaboración propia]

Para iniciar el pre-procesamiento de características, se siguieron los siguientes pasos de manera consecutiva:

- **Eliminación de outliers:** Se utilizó la función *isoutlier* de Matlab mediante el método de cuartiles, eliminando el 31.39% de los datos identificados como atípicos en el dominio espacial y el 11.81% en el dominio frecuencial.
- **Normalización de datos:** Se empleó la función *mapminmax* de Matlab para limitar los valores en un rango de [0, 1].

- **Balance del dataset:** debido a la eliminación de outliers, el conjunto de datos quedó desbalanceado. Se realizó un balance eliminando datos de forma aleatoria para cada una de las tres clases, resultando en un dataset de 28.245 datos para el dominio espacial y 36.285 datos para el dominio frecuencial.

Una vez preprocesadas las características, se realizaron varios tests de gaussianidad para todas ellas, lo cual es crucial para determinar las técnicas de selección de características. Se utilizaron los tests *Anderson-Darling (ADtest)*, *Kolmogorov-Smirnov (KStest)* y *Lilliefors (Lillietest)*, que confirmaron la gaussianidad en algunas características.

Finalmente, se emplearon técnicas de ranking y wrappers para la selección de características.

3.4.1 Rankings

Se empleó la función *relieff* de Matlab, una variación del método relief diseñada para problemas de más de dos clases. En este estudio, se configuró con un valor de $k = 7$ para generar el ranking de características. Este valor de k fue seleccionado tras probar variaciones en el parámetro (véase el apartado 4.2.1). La elección final de $k = 7$ se basó en su desempeño en los rankings generados.

3.4.2 Wrappers

Para llevar a cabo esta sección se realiza una búsqueda heurística donde se generaron subconjuntos de características. Esta técnica evalúa el rendimiento al agregar o eliminar características, manteniendo la combinación con el mejor desempeño. Se configuró la función *classify* de Matlab para emplear los algoritmos Mahalanobis, LDA y QDA como clasificadores en la técnica de wrappers.

En la implementación de wrappers, se empleó la función *sequentialfs* de Matlab. Esta función efectúa una validación cruzada de 10 folds para cada subconjunto de características en análisis. Después, promedia los resultados obtenidos de todos los folds y procede a evaluar el siguiente subconjunto.

La configuración de *sequentialfs* incluyó la creación de una partición aleatoria para la validación cruzada, usando la función *cvpartition* con una partición 'KFold'.

Finalmente se configuró al algoritmo implementado una selección *hacia adelante*, que consiste en añadir características una por una eligiendo las mejores, y una

selección *hacia atrás*, comenzando con todas las características y eliminando aquellas que reducen el desempeño.

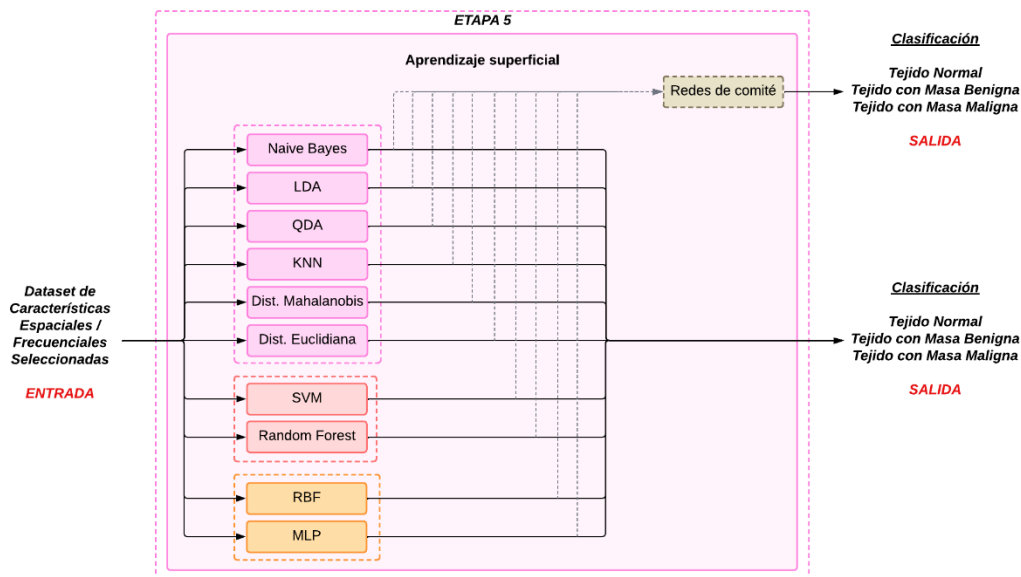
3.5 ETAPA 5. Técnicas de clasificación

3.5.1 Aprendizaje superficial

Esta sección describe las técnicas de clasificación implementadas, utilizando clasificadores estadísticos y máquinas de aprendizaje. Estos métodos requieren un proceso previo de segmentación, extracción y selección de características para diferenciar entre tejido normal, benigno y maligno.

El dataset de características seleccionadas, balanceado con 9.415 datos por clase para el dominio espacial y 12.095 datos por clase para el dominio frecuencial, datos provenientes del apartado 3.4, sirve como entrada para este bloque. Para el entrenamiento de cada algoritmo de clasificación, se emplea el 70% de los datos, distribuidos entre las tres clases. El 30% restante se reserva como datos de prueba, utilizados para evaluar el rendimiento de los algoritmos implementados. En la Figura 3.21 se presenta una representación visual del procedimiento aplicado en esta etapa.

Figura 3.21. Diagrama de bloques ampliado para el aprendizaje superficial de la ETAPA 5.



Fuente: [Elaboración propia]

La Figura 3.21 presenta una clasificación de los algoritmos empleados en el sistema, organizados en tres categorías principales: Clasificadores Estadísticos (en color rosado), Máquinas de Aprendizaje (en color rojo) y Redes Neuronales (en color naranja). Además, se incluye una Red de Comité (en color café), diseñada para actuar como un mecanismo de post-procesamiento y mejorar la robustez de los resultados obtenidos por los algoritmos de clasificación.

A continuación, se detalla la configuración y los parámetros empleados en cada uno de los elementos mencionados.

3.5.1.1 Clasificadores estadísticos

Se detallan a continuación las configuraciones empleadas para cada clasificador estadístico. Los parámetros seleccionados fueron determinados heurísticamente, basándose en el rendimiento de la clasificación obtenida. No obstante, el sistema desarrollado permite modificar algunos de estos parámetros.

- **Naive Bayes (NB):**

En el caso del clasificador Naive Bayes (NB), se empezó con una prueba de normalidad sobre las características extraídas, empleando los tests de Lillitest, ADtest y KStest. Después, se entrenó un modelo para cada universo de datos. De todos los modelos entrenados, se escogió el que presentó el mayor desempeño en términos de exactitud, tomando en cuenta también el análisis de ANOVA.

- **LDA:**

Se empleó la técnica de clasificación '*fitcdiscr*' de Matlab, especificando el tipo de discriminación '*linear*' para LDA.

- **QDA:**

Se aplicó la técnica de clasificación '*fitcdiscr*' de Matlab, con la discriminación de tipo '*quadratic*' especificada para QDA.

- **KNN:**

En la implementación de K-NN, se evaluaron diversas configuraciones de modelos variando el número de vecinos ($k = 5, 7, 11, 13, 17, 19$). Cabe destacar que, al tratarse de un sistema de tres clases, el número de vecinos no debe ser múltiplo de 3. De todos los modelos entrenados, se seleccionó el modelo con $k = 7$, que mostró el mejor desempeño basado en la métrica de exactitud y el análisis ANOVA.

- **Distancia Euclidiana y Distancia Mahalanobis:**

Para estos clasificadores, se aplicaron directamente sus ecuaciones (Capítulo 2, apartado 2.2), lo que hizo innecesario ajustar otros parámetros. Los algoritmos se configuraron para calcular la probabilidad a priori en función de la cantidad de datos por clase utilizada en el entrenamiento.

3.5.1.2 Máquinas de aprendizaje

- **Máquinas de vectores de soporte (SVM):**

En esta ocasión, dado que el sistema es de tres clases, se emplearon SVM bajo el modelo "*ClassificationECOC*", un clasificador multiclase basado en códigos de salida de corrección de errores (ECOC). A continuación, se detallan los valores configurados:

- ✓ **Kernel:** se seleccionó la función de base radial (RBF) por su mejor desempeño en comparación con las funciones lineales y polinomiales de segundo, tercer y cuarto orden.
- ✓ **Número de entradas:** el vector de entrada está compuesto por las características más discriminantes seleccionadas en la ETAPA 3. Extracción de características
- ✓ **Matriz de costo:** se utilizaron los valores [0 2 3; 2 0 3; 5 5 0], los cuales mostraron el mejor rendimiento. Se probaron diferentes matrices considerando que el costo de una mala clasificación en el tejido maligno es mayor que en el tejido normal o benigno.
- ✓ **Probabilidad a priori de las clases:** ajustada a 0.33 para cada clase, debido a que el dataset está balanceado. Las modificaciones en la probabilidad a priori no mejoraron el rendimiento.
- ✓ **Método de optimización:** se empleó el algoritmo onevsone (OVO) por sus mejores resultados, aunque también se evaluaron los algoritmos onevsall (OVA).
- ✓ **Cross-validación:** se utilizaron tanto 5 Folds como 10 Folds, con resultados similares y una variación menor al 1%. Se seleccionó la configuración con mejor desempeño y el menor número de vectores de soporte.

- **Random Forest:**

El clasificador fue configurado con 100 árboles de decisión, un máximo de 50 divisiones por árbol y un tamaño mínimo de hojas de 15. La selección de características en cada nodo se realizó mediante el *índice de Gini*, criterio que minimiza la impureza de los subconjuntos generados en cada partición.

Para evaluar el desempeño del modelo, se utilizó el *error de clasificación* como función de costo, la cual mide la proporción de muestras mal clasificadas en el conjunto de prueba. Además, se habilitó la predicción fuera de bolsa (*Out-of-Bag*, *OOB*), permitiendo obtener una estimación del error sin necesidad de un conjunto de validación adicional.

Dado que el número de características influye en la calidad de las predicciones, se utilizó un subconjunto de las cuatro características más relevantes en cada instancia. La importancia de las características se analizó mediante la métrica *OOB Permuted Predictor Delta Error*, que cuantifica el incremento del error del modelo al permutar aleatoriamente cada característica, identificando así las variables con mayor impacto en la clasificación.

Finalmente, para optimizar el tiempo de cómputo, se habilitó la ejecución en paralelo mediante *Parallel Computing Toolbox*, mejorando la eficiencia del entrenamiento.

3.5.1.3 Redes neuronales

En la sección dedicada a las máquinas de aprendizaje tradicionales, se implementaron dos tipos de redes neuronales: RBF y MLP. A continuación, se enumeran los parámetros elegidos para cada red.

- **Función de base radial (RBF)**

Se realizaron 347 combinaciones con las características seleccionadas mediante el método de wrappers.

- ✓ Función de activación de la capa oculta: Función de base radial.
- ✓ Error y/o meta a alcanzar: 0.01
- ✓ Parámetro de extensión para las funciones de base radial: 1
- ✓ Número máximo de neuronas en la capa oculta: 30
- ✓ Incremento: 1

- **Perceptrón multicapa (MLP)**

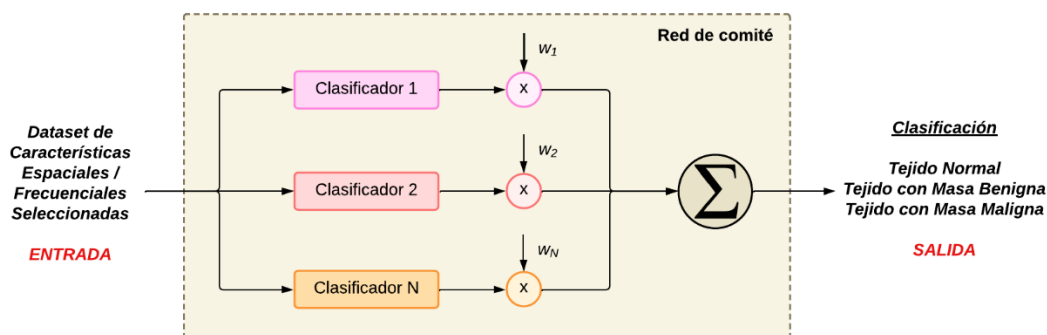
Se llevaron a cabo distintas combinaciones con las características seleccionadas mediante el método de wrappers.

- ✓ Cantidad de neuronas en la capa oculta: 20
- ✓ Función de entrenamiento: BFGS quasi-Newton (trainbfg)
- ✓ Épocas: 360
- ✓ Lr: 0.001
- ✓ Función de activación: Relu, Softmax

3.5.1.4 Redes de comité

Las redes de comité optimizan el desempeño del sistema de clasificación al combinar múltiples modelos individuales, aprovechando sus fortalezas y reduciendo sus limitaciones. La Figura 3.22. ilustra el diagrama de la arquitectura implementada para esta red.

Figura 3.22. Diagrama de integración de clasificadores en la red de comité.



Fuente: [Elaboración propia]

La implementación adoptada empleó un clasificador de votación dura (*Hard Voting*), donde la clase final se asigna según la mayoría de los votos entre los clasificadores seleccionados. En caso de empate, se aplicó un mecanismo de votación ponderada (*Weighted Voting*), asignando mayor peso a los modelos con mejor desempeño.

La selección de clasificadores se basó en métricas de rendimiento, asegurando una combinación óptima de modelos complementarios. Finalmente, el esquema de votación fue entrenado y evaluado con el conjunto de datos, lo que permitió mejorar la precisión y confiabilidad del sistema de clasificación.

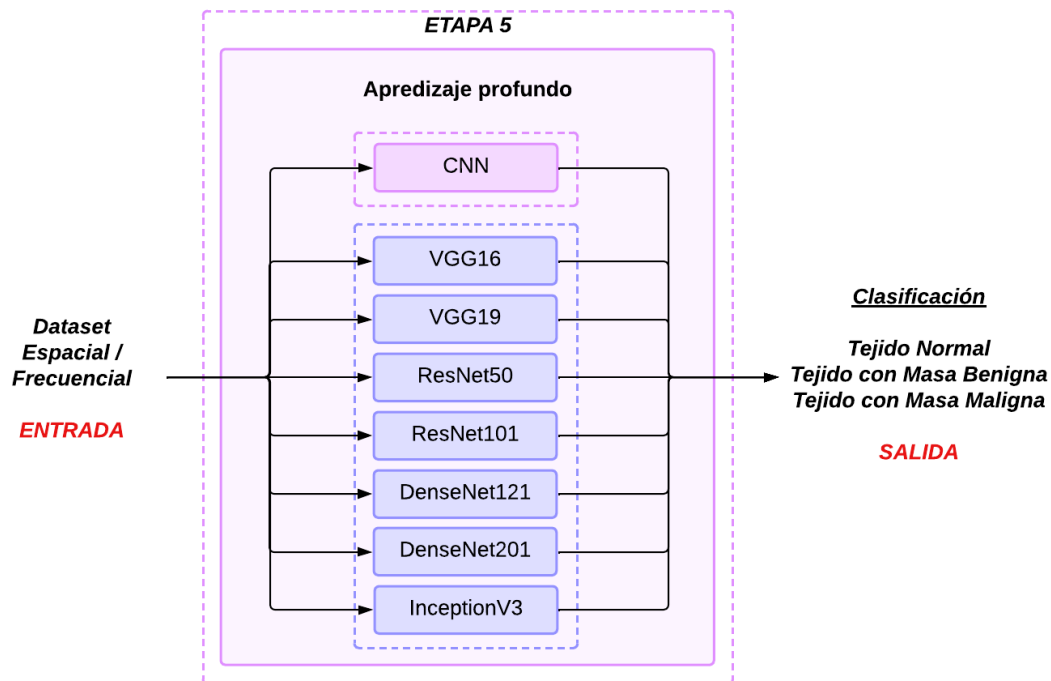
3.5.2 Aprendizaje Profundo

En esta sección se describen las técnicas de clasificación implementadas en el aprendizaje profundo. Se diseñó una Red Neuronal Convolutiva (CNN) desde cero y, además, se configuraron modelos preentrenados mediante la técnica de *transferencia de aprendizaje* [21], [24], destacándose arquitecturas como *VGG*, *ResNet*, *DenseNet* e *InceptionV3*.

A diferencia de los clasificadores mediante aprendizaje superficial, estos modelos procesan directamente las imágenes como entrada, sin requerir etapas previas de segmentación, extracción o selección de características. Esto se debe a su capacidad para procesar y extraer automáticamente las características relevantes a través de su estructura interna, optimizando así el flujo de trabajo y el análisis de las imágenes.

En la Figura 3.23., se aprecian dos enfoques diferenciados para esta clasificación: la CNN diseñada desde cero (resaltada en color violeta) y las arquitecturas preentrenadas utilizadas en la transferencia de aprendizaje (resaltadas en color morado).

Figura 3.23. Diagrama de bloques ampliado para el aprendizaje profundo de la ETAPA 5.



Fuente: [Elaboración propia]

A continuación, se describen las configuraciones empleadas en los métodos mencionados, incluyendo el diseño y entrenamiento de las redes, las características de las imágenes utilizadas, así como los métodos de *data augmentation* y *transfer learning* implementados.

3.5.2.1 CNN

Para la clasificación de tejido normal, tejido con masa benigna y tejido con masa maligna mediante aprendizaje profundo, se diseñó una *Red Neuronal Convolutiva* (CNN) adaptada para procesar imágenes en escala de grises con un tamaño de 50x50 píxeles. El conjunto de datos se dividió en un 70% para entrenamiento y un 30% para prueba. Se evaluaron tres escenarios: sin aumento de datos, con aumento en la fase de entrenamiento y con aumento en ambas fases.

Para la ampliación del dataset se aplicaron técnicas de *data augmentation*, incluyendo rotaciones en distintos grados, desplazamientos horizontales y verticales, inversiones en ambos ejes y escalamiento mediante *zoom*.

Optimización con Keras Tuner

Para determinar la configuración óptima del modelo, se empleó *Keras Tuner*, el cual permitió ajustar los hiperparámetros de la arquitectura con el objetivo de maximizar el rendimiento del modelo. En este proceso, se priorizó la métrica de *recall* para garantizar una alta sensibilidad en la detección de tejidos anormales. Se optimizaron los siguientes aspectos clave:

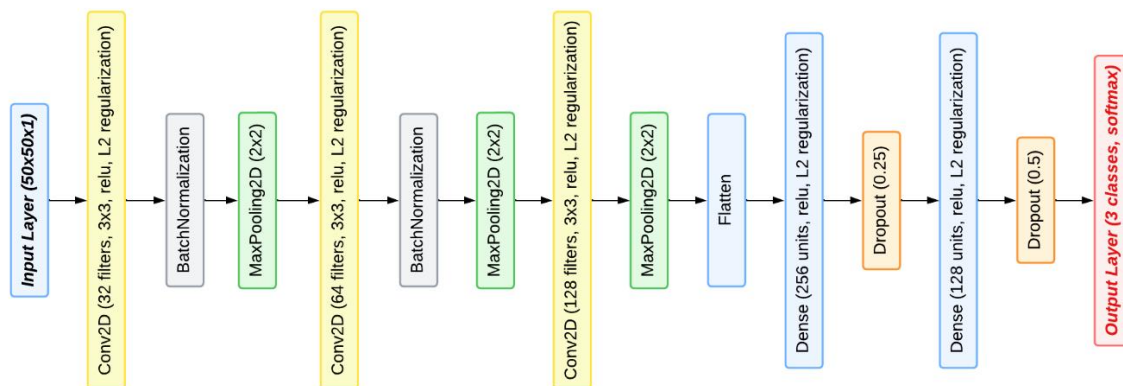
- **Número de capas convolucionales:** Se evaluó cuántas capas convolucionales eran necesarias para capturar adecuadamente las características del tejido en cada dominio.
- **Número de filtros por capa:** Se ajustó la cantidad de filtros en cada capa convolutiva, variando entre 16 y 64 filtros en la primera capa y entre 32 y 128 filtros en las capas subsiguientes, dependiendo de la complejidad de la representación.
- **Unidades en capas densas:** Se exploraron diferentes cantidades de unidades en las capas completamente conectadas, con valores entre 64 y 512 unidades en la primera capa densa y entre 32 y 256 unidades en la segunda.

- **Regularización y Dropout:** Se implementaron técnicas de regularización L2 para evitar la complejidad excesiva del modelo y prevenir el sobreajuste. Además, se evaluaron tasas de *Dropout* entre 0.2 y 0.5 para encontrar el equilibrio adecuado.
- **Función de activación:** Se probaron ReLU, LeakyReLU y Mish en las capas convolucionales, con el fin de mejorar la propagación del gradiente y evitar el problema de neuronas muertas.

Una vez completado el proceso de optimización, se seleccionó la mejor configuración del modelo, la cual se utilizó para evaluar el rendimiento en distintos escenarios. En la Figura 3.24., se ilustra la arquitectura final obtenida tras la optimización con *Keras Tuner*.

Estructura de la red

Figura 3.24. Estructura de la CNN.



Fuente: [Elaboración propia]

El modelo comienza con una capa convolucional que aplica 32 filtros de tamaño 3x3, con función de activación ReLU y regularización L2, seguida de una capa de Batch Normalization, que estabiliza las activaciones y acelera el entrenamiento de la red. Posteriormente, se incorpora una capa de MaxPooling2D con tamaño 2x2, la cual reduce la dimensionalidad de las características extraídas y permite a esta primera capa capturar los patrones básicos presentes en las imágenes de entrada.

A continuación, se introduce una segunda capa convolucional con 64 filtros de tamaño 3x3, manteniendo la activación ReLU y la regularización L2. Seguidamente, se agrega una capa de Batch Normalization, que mejora la estabilidad del

entrenamiento, y una capa de MaxPooling2D (2x2) para reducir aún más la dimensionalidad de los datos, permitiendo que el modelo extraiga características más complejas y diferenciadoras.

El modelo continúa con una tercera capa convolucional de 128 filtros de tamaño 3x3, con activación ReLU y regularización L2, seguida nuevamente de una capa de MaxPooling2D (2x2), que contribuye a la reducción de la dimensionalidad y preserva las características más relevantes de las imágenes.

Posteriormente, la salida de la última capa convolucional se aplanan mediante la operación Flatten, convirtiendo la representación bidimensional en un vector unidimensional que servirá como entrada para las capas completamente conectadas (densas).

A continuación, se introduce una primera capa densa con 256 unidades, activación ReLU y regularización L2, lo que permite al modelo aprender representaciones más abstractas y complejas de los datos extraídos por las capas anteriores. Seguidamente, se incorpora una capa *Dropout* con una tasa de 0.25, la cual apaga aleatoriamente ciertas neuronas durante el entrenamiento para reducir el riesgo de sobreajuste.

Después, se incluye una segunda capa densa con 128 unidades, activación ReLU y regularización L2, que refina la representación de las características extraídas en las etapas previas. A continuación, se agrega una capa *Dropout* con una tasa de 0.5, que fortalece la capacidad de generalización del modelo al prevenir la excesiva dependencia de ciertas neuronas durante el entrenamiento.

Finalmente, el modelo concluye con una capa de salida densa con 3 unidades y activación softmax, permitiendo la clasificación de las imágenes en tres categorías: tejido normal, tejido con masa benigna y tejido con masa maligna.

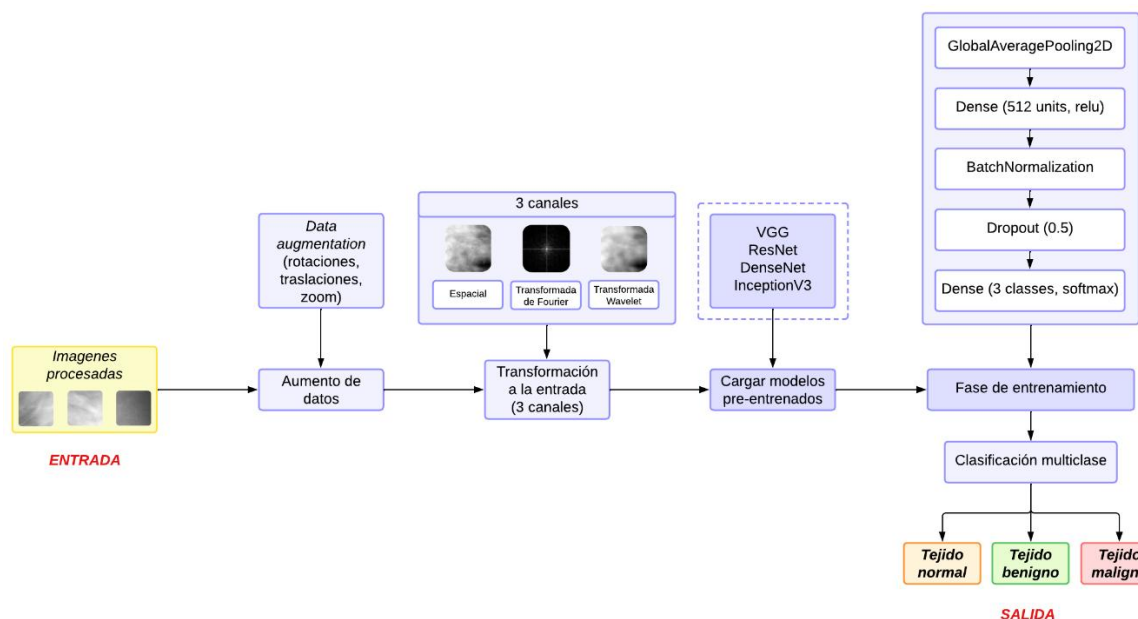
3.5.2.2 Aprendizaje por transferencia

Otro de los métodos implementados para la clasificación de tejido normal, tejido con masa benigna y tejido con masa maligna fue mediante arquitecturas convolucionales preentrenadas, empleando la técnica de *transfer learning*. En este enfoque se utilizaron modelos validados en la literatura, tales como *VGG*, *ResNet*, *DenseNet* e *InceptionV3*.

El proceso de entrenamiento, ilustrado en la Figura 3.25., inicia con la entrada del conjunto de imágenes preprocesadas, ajustadas a un tamaño de 50x50 píxeles.

Dado que las redes preentrenadas requieren entradas de 224x224 píxeles [24], fue necesario realizar una configuración específica para adaptarlas. Posteriormente, el dataset se dividió en 70% para entrenamiento (20% para validación) y 30% para pruebas, manteniendo la proporción original de cada clase.

Figura 3.25. Arquitectura del modelo de clasificación de imágenes mamográficas con aprendizaje por transferencia.



Fuente: [Elaboración propia]

Se llevaron a cabo entrenamientos en tres escenarios: sin aumento de datos, con aumento solo en la fase de entrenamiento y con aumento tanto en entrenamiento como en prueba. Para la ampliación del dataset, se aplicaron técnicas de *data augmentation*, incluyendo rotaciones en distintos grados, desplazamientos horizontales y verticales, inversiones en ambos ejes y escalamiento mediante *zoom*.

Dado que las redes preentrenadas utilizadas fueron entrenadas en ImageNet [24], que contiene imágenes a color de tres canales, fue necesario adaptar la entrada de las imágenes en escala de grises a este formato. Para ello, se generaron tres representaciones: el *primer canal* conservó la imagen original en el *dominio espacial* (correspondiente al dataset previamente generado para dicho dominio); el *segundo canal* representó la imagen transformada mediante la *Transformada de Fourier* (también proveniente del dataset frecuencial previamente generado); y el *tercer canal* incorporó una nueva representación basada en la *Transformada Wavelet*.

En este último caso, se implementó la Transformada Wavelet Discreta (DWT) utilizando la wavelet de Haar, seleccionada por su simplicidad computacional y por su capacidad de resaltar transiciones abruptas de intensidad, característica que se considera útil en este estudio para resaltar bordes y estructuras relevantes en imágenes mamográficas. La transformada permitió descomponer cada imagen en cuatro subbandas: aproximación (cA), detalles horizontales (cH), detalles verticales (cV) y diagonales (cD). Para la construcción del canal correspondiente, se combinaron las subbandas cA, cH y cV, aplicando un factor de escala de 0.2. Este valor fue determinado empíricamente con el fin de resaltar los contornos principales sin introducir ruido excesivo, manteniendo la estructura global aportada por cA. La subbanda cD fue excluida al concentrar principalmente componentes de alta frecuencia, considerados menos representativos para la clasificación mamográfica.

Esta estrategia multicanal permitió aprovechar las capacidades de las redes convolucionales preentrenadas en la extracción de características relevantes desde diferentes dominios de análisis, enriqueciendo la representación de entrada para la clasificación.

Las imágenes transformadas fueron utilizadas como entrada en modelos preentrenados reconocidos en la literatura, tales como *VGG*, *ResNet*, *DenseNet* e *Inception-v3*, los cuales han demostrado ser eficaces en tareas de clasificación de imágenes médicas [24]. Estas redes fueron ajustadas mediante la adición de capas personalizadas, incluyendo una capa de *Global Average Pooling 2D* para reducir la dimensionalidad de las características extraídas, seguida de una capa densa con activación *ReLU* para mejorar la capacidad de discriminación. Además, se incorporó una capa de normalización (*BatchNormalization*) para estabilizar el aprendizaje, junto con una capa de *Dropout* con un factor de 0.5 para prevenir el sobreajuste.

Finalmente, se añadió una capa densa con activación *softmax*, adaptada para la clasificación multiclase en las categorías: tejido normal, tejido benigno y tejido maligno. La evaluación del modelo se realizó en términos de métricas estándar de clasificación, incluyendo precisión, sensibilidad (recall), especificidad y el puntaje F1, con el objetivo de garantizar un desempeño óptimo en la identificación de anomalías en imágenes mamográficas.

3.6 CONCLUSIONES

La solución desarrollada se orientó a la detección y clasificación automática de tejido normal, masas benignas y masas malignas en mamografías digitales, mediante un sistema de visión artificial estructurado en cinco etapas fundamentales, distribuidas

en dos enfoques complementarios: *aprendizaje superficial* y *aprendizaje profundo*. Para optimizar la representación de las características del tejido mamario, se trabajó en ambos dominios, espacial y frecuencial, permitiendo una descripción más completa de las texturas y patrones presentes en las imágenes mamográficas.

El *aprendizaje superficial* abarcó las etapas de selección de la base de datos, procesamiento, extracción de características, selección de características y clasificación. Este procedimiento permitió evaluar y ajustar de manera transversal diversas técnicas en cada etapa, logrando un sistema flexible y optimizado.

Dentro de este enfoque, el procesamiento de imágenes aplicó una serie de técnicas para la mejora del contraste, segmentación por umbralización para eliminar artefactos. Se aplicó una transformación exponencial para realzar bordes, combinada con umbrales heurísticos para oscurecer el exterior, y un suavizado morfológico para homogenizar la textura. Además de la eliminación del tejido pectoral para evitar interferencias en la clasificación.

En la etapa de extracción de características, se implementó un método de enventanado para obtener ventanas en áreas de tejido normal y con masas, permitiendo así generar regiones de análisis, asegurando la representación equilibrada de los tejidos. Posteriormente, se aplicó la Transformada de Fourier a las ventanas generadas, logrando un conjunto de datos en dominio espacial y otro en dominio frecuencial, maximizando la representación de texturas y patrones.

La etapa de selección de características permitió reducir la dimensionalidad del problema aplicando técnicas estadísticas y heurísticas. Se eliminaron valores atípicos, se normalizaron los datos y se emplearon técnicas de ranking y wrappers, asegurando que solo las características más discriminantes fueran utilizadas en la clasificación.

Finalmente, en la etapa de clasificación, se implementaron algoritmos estadísticos incluyendo Naive Bayes, LDA, QDA, KNN, Distancia Mahalanobis y Distancia Euclidiana, junto con algoritmos de aprendizaje automático como SVM y Random Forest, además de redes neuronales RBF y MLP. También se utilizó un esquema de Redes de Comité, combinando múltiples modelos mediante votación mayoritaria para generar resultados con mejores desempeños, mejorando la robustez y precisión del sistema. Estos clasificadores se evaluaron en ambos dominios, espacial y frecuencial, y se realizó una fusión de ambos dominios combinando los vectores de características.

Por otro lado, el *aprendizaje profundo* integra de manera directa algunas etapas como el procesamiento, la extracción y la selección de características dentro del modelo de clasificación, optimizando el flujo de trabajo mediante Redes Neuronales Convolucionales (CNNs) y modelos preentrenados con transferencia de aprendizaje. Se implementaron dos escenarios:

- Se diseñó una CNN desde cero adaptada para imágenes de 50x50 píxeles en escala de grises, optimizada mediante Keras Tuner, permitiendo que el modelo aprenda automáticamente características discriminantes de las imágenes tanto en el dominio espacial como en el frecuencial para luego clasificarlas. Una vez obtenido el modelo óptimo, se implementó un enfoque multimodal que integra características espaciales y frecuenciales utilizando arquitecturas CNN multimodales, en las que cada dominio fue procesado de forma independiente a través de ramas especializadas. Posteriormente, se realizó una fusión en una capa de concatenación, optimizando la representación combinada para la clasificación final.
- Se hizo uso de transferencia de aprendizaje con modelos preentrenados como VGG, ResNet, DenseNet e InceptionV3, ajustados para la clasificación multiclase. Para adaptar las imágenes en escala de grises a las entradas de tres canales requeridas por los modelos preentrenados, se generaron tres representaciones: el primer canal conservó la imagen en el dominio espacial, mientras que los otros dos canales se obtuvieron mediante la Transformada de Fourier y la Transformada de Wavelet. Esta combinación aprovechó las capacidades jerárquicas de las redes preentrenadas para extraer automáticamente características discriminantes.

Es relevante mencionar que, aunque los algoritmos de aprendizaje profundo facilitan el desarrollo al integrar la obtención de parámetros, este enfoque demanda un costo computacional considerablemente más alto.

4. PRUEBAS Y RESULTADOS

Este capítulo expone los resultados obtenidos al aplicar los métodos propuestos en el Capítulo 3, basados en aprendizaje superficial y aprendizaje profundo, para clasificar tejido mamario en categorías de tejido normal, tejido con masa benigna y tejido con masa maligna en imágenes mamográficas digitales.

El dataset utilizado se describe en detalle, especificando los datos y ajustes aplicados para el desarrollo de las etapas posteriores, complementando la información general del Capítulo 3.

Se evalúa el rendimiento de los algoritmos de selección de características, específicamente Ranking y Wrappers, analizando su efecto en la reducción de la dimensionalidad y en el rendimiento de los modelos.

Los resultados de los modelos de aprendizaje superficial, que incluyen Naive Bayes, LDA, QDA, Distancia Euclidiana, Distancia Mahalanobis, KNN, Random Forest, SVM, RBF y MLP, se comparan con los modelos de aprendizaje profundo, que incluyen una CNN y modelos preentrenados como VGG16, VGG19, ResNet50, Resnet101, DenseNet121, DenseNet201 e InceptionV3, utilizando en ambos métodos de clasificación métricas de precisión global (Accuracy), precisión por clase (Precision), puntuación F1 (F1-Score), sensibilidad (Recall) y especificidad (Specificity). Estas métricas permiten medir la efectividad de cada modelo en la clasificación de los diferentes tipos de tejido.

Por último, se realiza un análisis comparativo con estudios previos, situando los resultados obtenidos en un contexto académico más amplio y evaluando las fortalezas, aplicaciones y limitaciones de cada enfoque en el sistema desarrollado.

4.1 Descripción del Dataset

El dataset MIAS [18] se utilizó en las pruebas, compuesto por 322 mamografías digitales de distinto tamaño, ajustadas a 1024 x 1024 píxeles para facilitar su manejo, como se describe en la ETAPA 1. Selección de la base de datos. Dado que el enfoque del sistema es la detección de tejido normal y con masas, se consideraron 207 imágenes de tejido normal, 40 de tejido con masa benigna y 19 de tejido con masa maligna. Las 56 imágenes que contenían microcalcificaciones u otras afecciones se descartaron, quedando 266 imágenes disponibles.

Dataset para el Dominio Espacial

De cada mamografía seleccionada, se extrajo aleatoriamente la mayor cantidad posible de sub-imágenes o ventanas de 50x50 píxeles, con un área de 2500 píxeles cada una, concentrándose en las regiones de tejido normal y de masas. Tras el procesamiento, se generó un dataset equilibrado de 9.415 vectores para cada clase, cada uno con 10 características discriminantes, optimizando así la representación de cada tipo de tejido para el análisis posterior.

Dataset para el Dominio Frecuencial

En el caso del dominio frecuencial, las mismas ventanas extraídas previamente fueron sometidas a una transformación para obtener el espectro de cada una. Este procedimiento generó un dataset equilibrado de 12.095 vectores por clase, cada uno compuesto por 10 características relevantes para el dominio frecuencial, permitiendo así una representación detallada de cada tipo de tejido en un dominio complementario para análisis más profundos y comparativos.

4.2 Pruebas y resultados para algoritmos de selección de características

La técnica de selección de características se aplica exclusivamente sobre los datos extraídos tras el proceso de segmentación. En esta etapa, el objetivo es reducir los conjuntos de características a aquellas que resulten más relevantes para el análisis, optimizando así la eficiencia y precisión del modelo. Para determinar las características clave, se implementan dos procedimientos de selección que evalúa la pertinencia de cada variable en el contexto de este estudio.

4.2.1 Resultados de Ranking

Como se describe en el apartado 3.4.1, se empleó el algoritmo ReliefF para realizar la selección de características tanto en el dominio espacial como en el frecuencial. Este proceso se aplicó a un sistema multiclase con tres categorías: tejido normal, masa benigna y masa maligna.

Para analizar el impacto del parámetro k , se realizaron pruebas con distintos valores: 3, 5, 6, 7, 9 y 10. Las Tabla 4.1 y Tabla 4.2 presentan los rankings obtenidos para cada configuración en los respectivos dominios, organizando las características desde la más relevante hasta la menos representativa según los pesos asignados por el algoritmo.

El objetivo de esta variación fue identificar qué valor de k permitía obtener una jerarquía de características más coherente con los patrones esperados en los datos, y facilitar la posterior selección de subconjuntos representativos para el entrenamiento de los clasificadores.

4.2.1.1 Resultados de Ranking por cada valor de k

Para cada valor de k seleccionado, se aplicó el algoritmo ReliefF en el conjunto de datos y se obtuvo un ranking de las características, ordenadas de la más a la menos relevante en términos de su capacidad discriminante. Los resultados para las características espaciales se presentan en la Tabla 4.1. y los resultados para las características frecuenciales se presentan en la Tabla 4.2., donde cada columna corresponde a un valor de k y cada fila muestra la posición de cada característica en el ranking de acuerdo con ese valor.

Tabla 4.1. Ranking de características espaciales para distintos valores de k en ReliefF.

N°	Característica	Cantidad de k vecinos					
		$k = 3$	$k = 5$	$k = 6$	$k = 7$	$k = 9$	$k = 10$
1	Máximo	1	1	1	1	1	1
2	Mínimo	3	3	3	3	3	3
3	Mediana	2	2	2	2	2	2
4	Promedio	6	6	5	5	5	5
5	Desviación Estándar	7	7	7	7	8	7
6	Curtosis	5	5	6	6	6	6
7	Asimetría	4	4	4	4	4	4
8	Contraste	8	8	8	8	7	8
9	Energía	9	9	9	9	9	9
10	Homogeneidad	10	10	10	10	10	10

Fuente: [Elaboración propia]

Tabla 4.2. Ranking de características frecuenciales para distintos valores de k en ReliefF.

N°	Característica	Cantidad de k vecinos					
		$k = 3$	$k = 5$	$k = 6$	$k = 7$	$k = 9$	$k = 10$
1	Promedio	10	10	10	10	10	10
2	Desviación Estándar	3	3	3	2	2	2
3	Curtosis	2	2	2	3	3	3
4	Asimetría	1	1	1	1	1	1
5	Entropía	5	5	5	5	5	5
6	Energía Total	9	9	9	9	9	9
7	Energía Baja	6	6	6	6	6	6
8	Energía Media	7	7	7	7	7	7
9	Energía Alta	8	8	8	8	8	8
10	Índice Compacidad	4	4	4	4	4	4

Fuente: [Elaboración propia]

4.2.1.2 Asignación de puntos y puntaje total

Para consolidar los resultados obtenidos con los diferentes valores del parámetro k , se asignaron puntos a cada característica de acuerdo con su posición en el ranking generado por el algoritmo ReliefF. Específicamente, se asignó un valor entero equivalente a la posición obtenida por cada característica (por ejemplo, posición 1 = 1 punto, posición 2 = 2 puntos, y así sucesivamente). Posteriormente, se sumaron los puntajes correspondientes a cada característica a lo largo de todos los valores de k evaluados.

La Tabla 4.3. muestra el puntaje total asignado a cada característica en función de su posición promedio en el ranking. Se presentan por separado los resultados para las características del dominio espacial (Tabla 4.3a) y del dominio frecuencial (Tabla 4.3b), ordenadas de menor a mayor puntaje total, es decir, de mayor a menor relevancia según el algoritmo.

Tabla 4.3 Puntaje total de las características en función de sus posiciones en el Ranking: a) Características espaciales; b) Características frecuenciales.

Puesto	Característica	Puntaje total	Puesto	Característica	Puntaje total
1	Máximo	6	1	Asimetría	6
2	Mediana	12	2	Desviación Estándar	15
3	Mínimo	18	3	Curtosis	15
4	Desviación Estándar	28	4	Índice Compacidad	24
5	Curtosis	32	5	Entropía	30
6	Asimetría	33	6	Energía Baja	36
7	Promedio	42	7	Energía Media	42
8	Contraste	45	8	Energía Alta	48
9	Energía	54	9	Energía Total	54
10	Homogeneidad	60	10	Promedio	60

a)

b)

Fuente: [Elaboración propia]

4.2.1.3 Interpretación de resultados y elección de K

El análisis de los puntajes totales obtenidos en la Tabla 4.3 muestra la relevancia de las características en función de sus posiciones en el ranking para diferentes valores de k en el algoritmo ReliefF.

En el dominio espacial (Tabla 4.3a), las características "Máximo", "Mediana" y "Mínimo" obtuvieron los puntajes más bajos (6, 12 y 18, respectivamente), lo que indica una alta consistencia en sus posiciones dentro del ranking y, por tanto, una mayor relevancia para la clasificación multiclase. Se observó que con $k = 7$ estas

características se mantuvieron de forma estable en las primeras posiciones, lo que evidencia que dicho valor mejora la capacidad discriminante del conjunto de características espaciales.

De manera análoga, en el dominio frecuencial (Tabla 4.3b), las características "Asimetría", "Desviación Estándar" y "Curtosis" alcanzaron los puntajes totales más bajos (6, 15 y 15, respectivamente), destacándose también por su estabilidad en el ranking a través de los diferentes valores de k . Al igual que en el caso anterior, el valor $k = 7$ permitió identificar de forma consistente las características con mayor capacidad para discriminar entre las clases.

Basado en estos resultados, se concluye que el valor de $k = 7$ es el más adecuado para este sistema multiclase, tanto para las características espaciales como para las frecuenciales. Este valor de k proporciona una justificación empírica para seleccionar las características más consistentes y robustas en ambos dominios, mejorando la representación de las características importantes para una clasificación precisa y efectiva entre las tres clases.

Con base en estos resultados, se concluye que $k = 7$ representa el valor más adecuado para este sistema, ya que proporciona una estructura de ranking estable y facilita la selección de características robustas y representativas en ambos dominios. Esta elección garantiza una representación eficiente y discriminativa del tejido mamario, favoreciendo el rendimiento de los modelos clasificadores.

4.2.1.4 Disminución de características

Para simplificar el modelo y reducir la complejidad del sistema, se decidió seleccionar solo las 6 características más discriminantes identificadas en el ranking consolidado para cada dominio respectivamente. Las características seleccionadas para el *dominio espacial* fueron: máximo, mediana, mínimo, desviación estándar, curtosis y asimetría y para el *dominio frecuencial* fueron: asimetría, desviación estándar, curtosis, índice de compacidad, entropía y energía baja.

Estas características ocupan las primeras posiciones en el ranking final y representan las más informativas para el sistema multiclase. La reducción de 10 a 6 características permite un modelo más eficiente y manejable sin comprometer significativamente la capacidad de discriminación entre las clases.

Además, se realizaron pruebas con cuatro clasificadores distintos (Naive Bayes, LDA, QDA, y KNN) para evaluar el rendimiento entre 2, 4, 6, 8 y 10 características,

teniendo en cuenta su ranking para la agrupación, tal como se muestra en la Tabla 4.4.

Tabla 4.4. Comparación de resultados para selección de características del dominio espacial.

Métrica	Naive Bayes					LDA				
	2	4	6	8	10	2	4	6	8	10
Precisión	45,26%	44,69%	49,71%	48,36%	48,87%	50,36%	57,48%	56,45%	60,20%	59,64%
Exactitud	63,51%	63,06%	66,47%	65,58%	65,91%	66,91%	71,65%	70,96%	73,46%	73,10%
F1 Score	45,26%	44,69%	49,71%	48,36%	48,87%	50,36%	57,48%	56,45%	60,20%	59,64%
Recall	45,26%	44,69%	49,71%	48,36%	48,87%	50,36%	57,48%	56,45%	60,20%	59,64%
Especificidad	74,82%	73,95%	76,32%	75,51%	76,19%	75,52%	80,14%	79,37%	81,23%	80,64%

Métrica	QDA					KNN				
	2	4	6	8	10	2	4	6	8	10
Precisión	48,58%	60,15%	61,69%	56,57%	56,54%	47,91%	48,51%	54,64%	54,11%	54,30%
Exactitud	65,72%	73,43%	74,46%	71,04%	71,03%	65,27%	72,34%	69,76%	69,41%	69,54%
F1 Score	48,58%	60,15%	61,69%	56,57%	56,54%	47,91%	48,51%	54,64%	54,11%	54,30%
Recall	48,58%	60,15%	61,69%	56,57%	56,54%	47,91%	48,51%	54,64%	54,11%	54,30%
Especificidad	76,81%	81,73%	83,21%	80,43%	80,27%	76,23%	81,35%	78,68%	78,31%	78,54%

Fuente: [Elaboración propia]

Tabla 4.5. Comparación de resultados para selección de características del dominio frecuencial.

Métrica	Naive Bayes					LDA				
	2	4	6	8	10	2	4	6	8	10
Precisión	46,73%	45,52%	46,41%	46,25%	45,58%	45,73%	44,96%	47,32%	44,41%	44,67%
Exactitud	64,49%	63,68%	64,27%	64,16%	63,72%	63,82%	63,31%	64,78%	62,94%	63,12%
F1 Score	46,73%	45,52%	46,41%	46,25%	45,58%	45,73%	44,96%	47,32%	44,41%	44,67%
Recall	46,73%	45,52%	46,41%	46,25%	45,58%	45,73%	44,96%	47,32%	44,41%	44,67%
Especificidad	73,36%	72,76%	73,21%	73,12%	72,79%	72,86%	72,48%	73,67%	72,21%	72,34%

Métrica	QDA					KNN				
	2	4	6	8	10	2	4	6	8	10
Precisión	45,23%	46,11%	48,21%	47,58%	47,10%	40,38%	45,81%	45,73%	44,67%	44,23%
Exactitud	63,32%	64,21%	65,89%	65,35%	64,90%	60,26%	63,87%	63,82%	63,12%	62,82%
F1 Score	45,23%	46,11%	48,21%	47,58%	47,10%	40,38%	45,81%	45,73%	44,67%	44,23%
Recall	45,23%	46,11%	48,21%	47,58%	47,10%	40,38%	45,81%	45,73%	44,67%	44,23%
Especificidad	72,15%	73,45%	74,32%	73,78%	73,21%	70,19%	72,91%	72,86%	72,34%	72,11%

Fuente: [Elaboración propia]

En el *dominio espacial*, la reducción a seis características se justificó al observar que los clasificadores Naive Bayes, QDA y KNN alcanzaron su mejor rendimiento con esta configuración. Aunque LDA mostró una ligera mejora con ocho características, utilizar seis ofrece un mejor compromiso entre simplicidad y desempeño, sin afectar significativamente la capacidad discriminativa del sistema.

En el *dominio frecuencial*, los resultados fueron más variables: Naive Bayes se desempeñó mejor con dos características y KNN con cuatro; sin embargo, LDA y QDA alcanzaron su mayor rendimiento con seis. Esta consistencia en los

clasificadores más robustos respalda la elección de seis características también en este dominio.

Por lo tanto, la reducción a seis características más discriminantes por dominio permitió simplificar la representación de los datos, alcanzando un mejor rendimiento en algunos clasificadores, mientras se mantienen niveles de precisión aceptables en otros. Esta decisión favoreció una estructura más compacta y comprensible, facilitando la interpretación de resultados y reduciendo la carga computacional durante el entrenamiento y la validación de los modelos.

4.2.2 Resultados de Wrappers

Para esta técnica, se realizaron diversas pruebas empleando una validación cruzada de $KFold = 10$, un enfoque que divide el conjunto de datos en 10 partes, entrenando el modelo en 9 de ellas y evaluándolo en la parte restante. Esta metodología asegura una evaluación robusta, probando el rendimiento del modelo en múltiples subconjuntos para evitar el sobreajuste. En cada prueba, se varió tanto el clasificador como la dirección de selección de características, generando una evaluación de las combinaciones más efectivas.

4.2.2.1 Clasificadores y métodos de selección

Se evaluaron tres clasificadores (lineal, cuadrático y Mahalanobis) en ambos dominios, aplicando dos métodos secuenciales de selección de características: forward (hacia adelante) y backward (hacia atrás). El método forward inicia con un conjunto vacío e incorpora progresivamente las características que mejoran el rendimiento; en contraste, el método backward parte del conjunto completo y elimina iterativamente las menos relevantes. Esta estrategia permite analizar la contribución individual de cada característica según el enfoque aplicado. En la Tabla 4.6. y en la Tabla 4.7. se muestra lo dicho de forma gráfica para cada dominio respectivamente.

Tabla 4.6. Resultados para técnicas de Wrappers en el dominio espacial.

Clasificador	Dirección	Características									
		1	2	3	4	5	6	7	8	9	10
LDA	forward	1	1	0	1	1	1	1	1	0	1
LDA	backward	1	1	0	1	1	1	1	1	0	1
QDA	forward	1	1	0	1	1	1	1	0	1	1
QDA	backward	1	1	1	0	1	1	1	0	1	1
Mahalanobis	forward	1	1	0	1	1	1	1	0	1	1
Mahalanobis	backward	1	1	0	1	1	1	1	0	1	1

Fuente: [Elaboración propia]

Tabla 4.7. Resultados para técnicas de Wrappers en el dominio frecuencial.

Clasificador	Dirección	Características									
		1	2	3	4	5	6	7	8	9	10
LDA	forward	1	1	1	1	0	1	0	1	1	1
LDA	backward	1	1	1	1	1	1	0	1	1	0
QDA	forward	0	1	1	1	1	1	1	1	1	0
QDA	backward	0	1	1	1	1	1	1	1	1	0
Mahalanobis	forward	0	1	1	1	0	1	1	1	1	1
Mahalanobis	backward	0	1	1	1	0	1	1	1	1	1

Fuente: [Elaboración propia]

Los resultados se presentan en las Tabla 4.6. (dominio espacial) y la Tabla 4.7 (dominio frecuencial), donde se indica con "1" si una característica fue seleccionada y con "0" si no lo fue. Cada fila representa una combinación de clasificador y método, y cada columna una de las diez características evaluadas. Este formato permite identificar fácilmente las características seleccionadas de forma recurrente, destacando aquellas más relevantes para cada dominio.

4.2.2.2 Número de características y validación de selección

Cada característica seleccionada representa un descriptor específico, como se detalla en el apartado (ETAPA 3. Extracción de características), que identifica las 10 características evaluadas en el dominio espacial: máximo, mínimo, mediana, promedio, desviación estándar, curtosis, asimetría, contraste, energía y homogeneidad. De manera análoga, para el dominio frecuencial se evaluaron otras 10 características: promedio, desviación estándar, curtosis, asimetría, entropía, energía total, energía baja, energía media, energía alta e índice de compacidad. Estos descriptores permiten un análisis exhaustivo de propiedades estadísticas y texturales clave en las imágenes mamográficas.

En cada prueba, se estableció un objetivo de ocho características iniciales, de las cuales se determinaron las 6 con mayor relevancia, seleccionadas por su capacidad de mejorar el rendimiento del clasificador de manera consistente. Este proceso de selección fue respaldado por un análisis de conteo, donde solo se retuvieron aquellas características que demostraron ser significativas en al menos tres combinaciones de prueba, asegurando así una selección robusta y coherente.

4.2.2.3 Resultados de las combinaciones probadas

El análisis consideró seis combinaciones, resultantes de la aplicación de tres clasificadores (LDA, QDA y Mahalanobis) y dos métodos de selección secuencial

(forward y backward), evaluadas mediante validación cruzada de 10 pliegues. Para cada combinación, se analizaron los vectores de características seleccionadas y el historial de desempeño, permitiendo identificar conjuntos robustos y coherentes en ambos dominios (Tabla 4.6 y Tabla 4.7).

En el dominio espacial, las características seleccionadas de forma consistente fueron: máximo, mínimo, desviación estándar, curtosis, asimetría y homogeneidad. Estas características se identificaron como las más relevantes para la clasificación multiclase, mientras que otras como mediana, promedio, contraste y energía fueron descartadas por su baja frecuencia de selección.

En el dominio frecuencial, las seis características seleccionadas con mayor consistencia fueron: desviación estándar, curtosis, asimetría, energía total, energía media y energía alta. Por el contrario, características como promedio, entropía, energía baja e índice de compacidad fueron excluidas en la mayoría de las combinaciones.

Cabe resaltar que algunas características no relevantes de forma individual según el ranking (por ejemplo, homogeneidad o el índice de compacidad), demostraron ser importantes al combinarse con otras, lo que resalta la ventaja de los métodos de selección mediante wrappers para capturar interacciones entre variables.

Aunque los wrappers permiten encontrar el mejor conjunto para cada clasificador, su principal limitación radica en que los resultados no siempre se generalizan a otros modelos. Por esta razón, se optó por conservar aquellas características seleccionadas de forma consistente en todas las combinaciones, lo cual permite una reducción de 10 a 6 atributos por dominio, mejora la eficiencia computacional en clasificadores complejos como el MLP y asegura mayor relevancia general.

Finalmente, se emplearon tanto las características seleccionadas por wrappers como las obtenidas mediante el método de ranking, observándose diferencias entre ambos enfoques. Esta variabilidad refuerza su carácter complementario y su utilidad para una selección de características más robusta en problemas multiclase.

4.3 Pruebas y resultados para el aprendizaje superficial

Esta sección presenta la evaluación de los clasificadores desarrollados en la ETAPA 5. Técnicas de clasificación, aplicando la técnica de segmentación previamente propuesta. Para el *dominio espacial* se utilizaron vectores de seis características seleccionadas, distribuyendo el conjunto de datos en un 70% (24.087 muestras) para entrenamiento y 30% (4.158 muestras) para prueba. En el *dominio frecuencial*,

el análisis se realizó también con vectores de seis dimensiones, usando un conjunto de 30.186 muestras para entrenamiento y 6.099 para prueba, con una división igual del 70% para entrenamiento y 30% para prueba.

El desempeño de los clasificadores se evaluó mediante las métricas definidas en el marco teórico. Dado el contexto clínico, se priorizó la sensibilidad por su relevancia en la detección de casos positivos, seguida del F1-Score y la exactitud. Esta jerarquización permitió valorar tanto la capacidad de detección como el equilibrio general del sistema de clasificación.

4.3.1 Pruebas y resultados para el Dominio Espacial

En esta sección se presentan los resultados obtenidos al evaluar el desempeño de diez clasificadores aplicados al dominio espacial, utilizando las métricas previamente descritas. Los valores reportados corresponden a las evaluaciones realizadas sobre el conjunto de prueba. La Tabla 4.8. presenta el desempeño de cada algoritmo y un ranking comparativo entre ellos. Para facilitar la visualización de los resultados más destacados, se resaltaron en color naranja los tres mejores valores alcanzados en cada métrica.

Tabla 4.8. Desempeño de clasificadores en el dominio espacial.

Clasificador	Precisión		Exactitud (accuracy)		F1 Score		Sensibilidad (recall)		Especificidad	
	Porcentaje	Posición	Porcentaje	Posición	Porcentaje	Posición	Porcentaje	Posición	Porcentaje	Posición
Naive Bayes	53,15%	8	68,90%	8	53,05%	8	56,70%	8	55,50%	8
LDA	61,95%	5	74,90%	5	61,65%	6	63,75%	6	66,15%	5
QDA	66,10%	4	77,30%	2	65,95%	4	70,40%	2	71,10%	4
KNN	59,95%	7	73,65%	7	59,80%	7	63,10%	7	64,85%	7
Distancia Euclidiana	51,85%	9	67,40%	9	51,50%	9	53,90%	9	51,60%	10
Distancia Mahalanobis	61,75%	6	74,35%	6	62,30%	5	64,25%	5	65,30%	6
SVM	66,50%	2	77,10%	4	66,30%	2	70,25%	3	71,25%	3
Random Forest	51,80%	10	49,80%	10	44,88%	10	53,40%	10	52,20%	9
RBF	66,85%	1	78,10%	1	66,75%	1	71,20%	4	72,15%	1
MLP	66,45%	3	77,30%	3	66,20%	3	70,90%	1	71,50%	2

Fuente: [Elaboración propia]

4.3.1.1 Resultados y análisis

Los resultados obtenidos, resumidos en la Tabla 4.8, indican que los algoritmos RBF, MLP y QDA son los más destacados en el dominio espacial cuando se prioriza la sensibilidad (recall) y la exactitud. Debido a que en aplicaciones de salud un falso negativo representa un riesgo inminente para el paciente, el recall es seleccionado como la métrica más relevante, ya que mide la capacidad de los clasificadores para identificar de manera precisa las muestras correspondientes a tejidos con presencia de anomalías. La exactitud también es importante, pues evalúa el desempeño global del modelo.

En detalle:

- **RBF** obtiene el mejor valor en sensibilidad (71.20%) y especificidad (72.15%), además de posicionarse en el primer lugar en F1 score (66.75%). Este rendimiento balanceado en varias métricas clave hace de RBF un clasificador destacado en el dominio espacial.
- **MLP** se posiciona en primer lugar en exactitud (77.30%) y alcanza un alto valor de sensibilidad (70.90%), lo que sugiere que es eficaz tanto en la identificación correcta de casos positivos como en el rendimiento global del modelo.
- **QDA** presenta una sensibilidad de 70.40% y una exactitud de 77.30%, ubicándose también entre los algoritmos con mejor capacidad para evitar falsos negativos y con un buen desempeño en la clasificación global.

Por otro lado, el clasificador Random Forest muestra el rendimiento más bajo en varias métricas, con una exactitud de 49.80% y un F1 score de 44.88%. Esto indica que este algoritmo es menos adecuado para la clasificación en el dominio espacial bajo el presente conjunto de datos, ya que presenta una capacidad limitada para identificar correctamente los casos positivos.

En conclusión, los algoritmos RBF, MLP y QDA demostraron ser los más efectivos para la clasificación de imágenes en el dominio espacial, especialmente cuando se considera la importancia de la sensibilidad en un contexto de salud. Este análisis permite una selección informada de clasificadores, destacando aquellos con el mejor desempeño en las métricas prioritarias y resaltando los algoritmos menos recomendados para esta tarea específica.

4.3.2 Pruebas y resultados para el Dominio Frecuencial

En esta sección se presentan los resultados obtenidos al evaluar el desempeño de diez clasificadores en la clasificación superficial de imágenes en el dominio frecuencial. Para cada clasificador, se calcularon las métricas de precisión, exactitud (accuracy), F1 score, sensibilidad (recall) y especificidad. La Tabla 4.9. muestra los valores obtenidos para cada métrica con los datos de prueba, junto con una columna de ranking que indica las posiciones de cada clasificador, destacando en color amarillo los tres algoritmos con mejor desempeño en cada métrica para facilitar su identificación.

Tabla 4.9. Desempeño de clasificadores en el dominio frecuencial.

Clasificador	Precisión		Exactitud (accuracy)		F1 Score		Sensibilidad (recall)		Especificidad	
	Porcentaje	Posición	Porcentaje	Posición	Porcentaje	Posición	Porcentaje	Posición	Porcentaje	Posición
Naive Bayes	46,96%	6	64,64%	5	46,96%	5	46,96%	6	73,48%	2
LDA	47,84%	5	65,23%	4	47,84%	4	47,84%	5	73,92%	1
QDA	46,11%	7	67,90%	2	46,11%	7	68,55%	2	69,63%	8
KNN	45,01%	9	63,34%	7	45,01%	9	45,01%	9	72,50%	5
Distancia Euclidiana	41,86%	10	61,24%	8	41,86%	10	41,86%	10	70,72%	6
Distancia Mahalanobis	50,22%	3	63,48%	6	50,22%	3	53,42%	4	53,42%	10
SVM	64,76%	1	67,90%	3	64,76%	1	68,55%	3	70,13%	7
Random Forest	48,18%	4	46,14%	10	46,11%	8	46,14%	7	73,07%	4
RBF	46,04%	8	46,37%	9	46,37%	6	46,01%	8	73,18%	3
MLP	63,76%	2	75,84%	1	63,76%	2	68,64%	1	59,53%	9

Fuente: [Elaboración propia]

4.3.2.1 Resultados y análisis

Los resultados obtenidos, presentados en la Tabla 4.9., destacan la sensibilidad (recall) y la exactitud (accuracy) como las métricas más relevantes para esta evaluación, dado que un falso negativo podría representar un riesgo crítico para la salud del paciente al no detectar una masa potencialmente maligna. Así, el recall permite evaluar la capacidad de los clasificadores para identificar correctamente los casos positivos, mientras que la exactitud muestra el rendimiento general del modelo en la clasificación de todas las clases.

De acuerdo con estas métricas, los algoritmos MLP, SVM y QDA muestran un desempeño superior en el dominio frecuencial:

- **MLP** se destaca en sensibilidad (68.64%), lo que indica una alta capacidad para detectar correctamente los casos positivos, y además alcanza una exactitud de 75.84%, lo cual lo posiciona como el mejor clasificador en términos de desempeño general.
- **SVM** ocupa el segundo lugar en sensibilidad con un 68.55% y muestra un F1 score de 64.76%, lo cual refuerza su efectividad en el balance entre precisión y recall, siendo especialmente útil para reducir falsos negativos.
- **QDA** logra un valor alto de sensibilidad (68.55%) y una exactitud de 67.90%, demostrando un rendimiento confiable tanto en la identificación de casos positivos como en la clasificación global.

Por otro lado, los clasificadores Distancia Euclidiana y Random Forest tienen un desempeño bajo en sensibilidad y exactitud, con valores de 41.86% y 46.14% respectivamente, lo que sugiere que son menos efectivos para detectar casos positivos y tienen un menor rendimiento general en este contexto.

En conclusión, debido a la importancia de evitar falsos negativos en un contexto de salud, los algoritmos MLP, SVM y QDA son los más recomendados en el dominio frecuencial, ya que proporcionan los valores más altos de sensibilidad y exactitud. Este análisis evidencia que los clasificadores con mejor desempeño en estas métricas son los más adecuados para una aplicación que prioriza la seguridad del paciente. La evaluación detallada de estas métricas permite una selección informada de los algoritmos más efectivos en la clasificación de imágenes mamográficas en el dominio frecuencial.

4.3.3 Pruebas y resultados para la Fusión de Dominios Espacial y Frecuencial

Esta sección presenta los resultados obtenidos al evaluar el desempeño de diez clasificadores al combinar características de los dominios espacial y frecuencial. La fusión de ambos dominios permite aprovechar tanto las características de textura en el dominio espacial como las variaciones de frecuencia. En la Tabla 4.10 se resumen las métricas de precisión, exactitud (accuracy), F1 score, sensibilidad (recall) y especificidad para cada clasificador, calculadas con los datos de prueba, junto con su posición de ranking en cada métrica. Los tres mejores clasificadores en cada métrica están resaltados en color rojo para facilitar su identificación.

Tabla 4.10. Desempeño de clasificadores con la fusión de ambos dominios.

Clasificador	Precisión		Exactitud (accuracy)		F1 Score		Sensibilidad (recall)		Especificidad	
	Porcentaje	Posición	Porcentaje	Posición	Porcentaje	Posición	Porcentaje	Posición	Porcentaje	Posición
Naive Bayes	51,82%	10	68,21%	9	51,82%	9	51,82%	8	75,41%	10
LDA	55,41%	6	70,61%	6	55,41%	6	55,41%	6	77,60%	5
QDA	60,14%	3	74,76%	3	60,14%	3	60,14%	3	79,87%	2
KNN	54,24%	7	69,83%	7	54,24%	7	54,24%	7	76,62%	7
Distancia Euclidiana	52,47%	8	68,64%	8	51,47%	10	51,47%	10	75,73%	8
Distancia Mahalanobis	59,39%	5	73,25%	5	58,39%	5	58,39%	5	79,19%	4
SVM	63,52%	1	76,86%	1	63,52%	1	63,96%	1	80,98%	1
Random Forest	52,16%	9	53,70%	10	52,16%	8	51,70%	9	75,53%	9
RBF	59,80%	4	73,37%	4	59,61%	4	59,70%	4	79,20%	3
MLP	62,17%	2	76,28%	2	61,90%	2	62,40%	2	76,73%	6

Fuente: [Elaboración propia]

El procedimiento de evaluación consistió en combinar características de los dominios espacial y frecuencial para cada imagen, generando así un conjunto de datos enriquecido. Cada clasificador fue entrenado y evaluado con este conjunto de características fusionadas.

4.3.3.1 Resultados y análisis

Los resultados obtenidos, presentados en la Tabla 4.10., indican que los algoritmos SVM, MLP y QDA son los más destacados en la fusión de los dominios espacial y frecuencial, especialmente en términos de exactitud, una de las métricas prioritarias.

En detalle:

- **SVM** se destaca con una sensibilidad de 63.52% y una exactitud de 76.86%, lo que lo evidencia un buen rendimiento general, aunque la capacidad para identificar correctamente los casos positivos disminuye respecto a su desempeño individual en los dominios espacial (70.25%) y frecuencial (68.55%).
- **MLP** presenta una sensibilidad de 62.40% y una exactitud de 76.28%, ubicándose también como uno de los mejores clasificadores al combinar ambos dominios. No obstante, esta sensibilidad también disminuye frente a los valores obtenidos en los dominios por separado (70.90% en espacial).
- **QDA** obtiene un valor de sensibilidad de 60.14% y una exactitud de 74.76%, con una disminución respecto al 70.40% alcanzado en el dominio espacial y al 68.55% en el frecuencial.

Estos resultados sugieren que, aunque la fusión puede aportar a métricas como la especificidad y la exactitud global, no representa una mejora en la detección de casos positivos en términos de sensibilidad, que es la métrica prioritaria en entornos clínicos, debido a su relación directa con la reducción de falsos negativos.

Por otro lado, el clasificador Naive Bayes presenta uno de los valores más bajos en sensibilidad (51.82%) y especificidad (75.41%), lo que indica que es menos adecuado para la clasificación en este contexto de fusión de dominios, ya que tiene una menor capacidad para detectar casos positivos de manera efectiva.

En conclusión, la fusión de los dominios espacial y frecuencial resulta adecuada para mejorar métricas generales como la exactitud o la especificidad, pero no garantiza una mejora directa en la sensibilidad, especialmente en los clasificadores SVM, MLP y QDA, aunque si mantienen un buen desempeño en exactitud. Por lo tanto, se recomienda precaución al utilizar esta estrategia si la sensibilidad es prioritaria. Esta situación plantea la necesidad de explorar en futuros trabajos métodos de selección, normalización o ponderación de características que permitan conservar o mejorar la capacidad de detección de casos positivos al combinar múltiples dominios.

4.3.4 Pruebas y resultados para las Redes de Comité

En esta sección se presentan los resultados obtenidos mediante la implementación de la red de comité en los dominios espacial y frecuencial, así como en la fusión de ambos. Se analizan las métricas clave de rendimiento, incluyendo precisión,

exactitud (accuracy), F1-score, sensibilidad (recall) y especificidad, y se comparan con los clasificadores individuales previamente evaluados.

Dado que los modelos *RBF*, *MLP* y *QDA* obtuvieron los mejores resultados en el dominio espacial; *MLP*, *SVM* y *QDA* se destacaron en el dominio frecuencial; y *SVM*, *MLP* y *QDA* presentaron el mejor desempeño en la fusión de ambos dominios, se implementó una red de comité basada en votación mayoritaria (*Hard Voting*) para cada uno. En caso de empate, se utilizó un esquema de votación ponderada (*Weighted Voting*), asignando mayor peso a los modelos con mejor sensibilidad.

Cada imagen fue clasificada por los tres modelos seleccionados y la predicción final se generó a partir de la clase con mayor cantidad de votos. Este enfoque permitió combinar las fortalezas individuales de los clasificadores y reducir sus debilidades, logrando mayor estabilidad y confiabilidad en el sistema de clasificación.

En la Tabla 4.11., se presenta un análisis comparativo de los resultados obtenidos en cada escenario. Como se observa, la fusión de ambos dominios optimiza la precisión y estabilidad del modelo, permitiendo un equilibrio adecuado entre sensibilidad y especificidad.

Tabla 4.11. Desempeño de la red de comité en los dominios espacial, frecuencial y su fusión.

Dominio	Modelos Red de Comité	Precisión		Accuracy		F1-Score		Recall		Specificity	
		Valor	Puesto	Valor	Puesto	Valor	Puesto	Valor	Puesto	Valor	Puesto
Espacial	RBF + MLP + QDA	74,20%	1	82,35%	2	73,10%	1	76,80%	1	77,00%	2
Frecuencial	MLP + SVM + QDA	67,80%	3	79,80%	3	70,50%	3	75,10%	2	74,20%	3
Fusión	SVM + MLP + QDA	72,10%	2	84,50%	1	71,00%	2	73,20%	3	86,00%	1

Fuente: [Elaboración propia]

4.3.4.1 Resultados y análisis

Los resultados obtenidos, presentados en la Tabla 4.11., confirman que la red de comité supera el rendimiento de los modelos individuales, mejorando las métricas de desempeño en cada dominio.

En el dominio espacial, se alcanzó una sensibilidad del 76.80%, con una mejora del 5.6% sobre el mejor clasificador individual (MLP con 71.20%). El F1-score aumentó un 6.4%, alcanzando 73.10% frente al 66.75% de MLP. También se observaron mejoras en precisión (74.20%), con un incremento de 7.4% respecto a RBF (66.85%), y en exactitud (84.50%), superando en 6.4% el valor previo (78.10% de MLP).

En el dominio frecuencial, la red de comité logró una sensibilidad de 75.10% y un F1-score de 70.50%, con mejoras de 6.5% y 6.7% respectivamente frente a MLP. La precisión y exactitud también aumentaron en 4.0% respecto al mejor clasificador individual.

Para la fusión de ambos dominios, se obtuvieron los mejores resultados globales: exactitud del 84.50% y especificidad del 86.00%, con incrementos del 8.22% y 9.27% frente a MLP. Aunque la sensibilidad (73.20%) fue ligeramente inferior al valor máximo del dominio espacial, se mantuvo un buen equilibrio general.

En términos generales, los resultados demuestran que el dominio espacial es el más efectivo cuando se busca maximizar la detección de anomalías debido a su alta sensibilidad. Mientras tanto, el dominio frecuencial aporta estabilidad, aunque su rendimiento general es menor. La fusión de ambos dominios se consolida como la opción más equilibrada, logrando la mejor exactitud y especificidad, lo que la convierte en la alternativa más confiable para una clasificación precisa y con menor margen de error.

Este análisis sugiere que, si bien cada modelo individual mostró fortalezas en dominios específicos, la combinación de los tres mejores clasificadores permitió compensar sus limitaciones, proporcionando un sistema de clasificación más robusto, equilibrado y confiable.

4.4 Pruebas y resultados para el aprendizaje profundo

Esta sección presenta la evaluación de los clasificadores de aprendizaje profundo desarrollados en la ETAPA 5. Técnicas de clasificación, utilizando la técnica de segmentación previamente definida (véase el apartado 3.2). Se implementaron *dos enfoques*: una red neuronal convolucional (CNN) diseñada desde cero y modelos pre-entrenados mediante transferencia de aprendizaje. En ambos casos, se trabajó en los dominios espacial y frecuencial, evaluando el impacto de cada representación y la combinación de ambos dominios en el rendimiento de la clasificación.

Para medir el desempeño de los modelos, se utilizaron métricas como exactitud, sensibilidad (recall), especificidad, precisión y F1-score, de acuerdo con la metodología descrita en la sección 2.2.4. Dado que los falsos negativos representan un alto riesgo clínico, la métrica más relevante en este contexto es la sensibilidad (recall), seguida del F1-score, que equilibra precisión y sensibilidad, y finalmente la exactitud, que proporciona una visión global del desempeño del modelo.

En los siguientes apartados, se detallan los experimentos realizados y los resultados obtenidos, proporcionando un análisis comparativo del desempeño del modelo en cada dominio y en cada *enfoque*.

4.4.1 Pruebas y resultados para la CNN

Esta sección presenta la evaluación del desempeño de la Red Neuronal Convolucional (CNN) diseñada totalmente para la clasificación automática de tejido normal, benigno y maligno en mamografías digitales. Se analizan los resultados obtenidos en tres escenarios: sin aumento de datos, con aumento solo en datos de entrenamiento y con aumento en datos de entrenamiento y en datos de prueba.

El conjunto de datos inicial está compuesto por 41.154 imágenes de 50x50 píxeles en su totalidad, distribuidas en un 70% para entrenamiento y un 30% para prueba. Del conjunto de entrenamiento, se extrajo el 20% para validación, quedando conformado de la siguiente manera: 27.069 imágenes para entrenamiento, 6.765 imágenes para validación y 7.320 imágenes para prueba. Cada clase se encuentra balanceada, con el mismo número de imágenes para tejido normal, benigno y maligno en cada conjunto de datos.

Se aplicaron técnicas de *data augmentation* para mejorar la generalización del modelo, y se evaluaron diversas configuraciones arquitectónicas optimizadas mediante *Keras Tuner*.

4.4.1.1 Evaluación en los Dominios Espacial y Frecuencial

Para evaluar el desempeño del modelo en cada dominio por separado, se utilizó la misma arquitectura CNN, modificando únicamente las imágenes de entrada según el dominio que se deseaba analizar. En este sentido:

- En el *dominio espacial*, las imágenes se procesaron en su forma original, permitiendo que el modelo aprendiera características relacionadas con la estructura del tejido.
- En el *dominio frecuencial*, se aplicó la Transformada de Fourier a las imágenes para resaltar patrones espectrales y evaluar si la información en este dominio mejoraba la discriminación de las clases.

La Tabla 4.12 y Tabla 4.13 presentan el desempeño del modelo CNN en los dominios espacial y frecuencial, respectivamente. Cada tabla está organizada en tres bloques: a) resultados con datos de entrenamiento, b) resultados con datos de

prueba y c) diferencia entre ambos conjuntos. Esta estructura permite analizar no solo la precisión y sensibilidad del modelo en cada escenario, sino también su capacidad de generalización. Los tres escenarios evaluados incluyen: sin aumento de datos, con aumento en entrenamiento, y con aumento en entrenamiento y prueba.

Tabla 4.12. Resultados de la CNN en el dominio espacial para los tres escenarios con datos de: a) Entrenamiento; b) Prueba; c) Diferencia entre ambos conjuntos.

A. RESULTADOS CON DATOS DE ENTRENAMIENTO						
Clasificador	Escenario	Precision	Accuracy	F1 Score	Recall	Specificity
		Porcentaje	Porcentaje	Porcentaje	Porcentaje	Porcentaje
CNN	Sin aumento	93,40%	93,19%	93,15%	93,01%	99,84%
	Con aumento en train	66,76%	62,19%	62,27%	53,49%	99,11%
	Con aumento en train y test	65,71%	61,52%	61,67%	53,57%	99,10%
B. RESULTADOS CON DATOS DE PRUEBA						
Clasificador	Escenario	Precision	Accuracy	F1 Score	Recall	Specificity
		Porcentaje	Porcentaje	Porcentaje	Porcentaje	Porcentaje
CNN	Sin aumento	56,82%	56,71%	51,78%	56,26%	97,76%
	Con aumento en train	55,19%	54,39%	50,22%	53,06%	97,73%
	Con aumento en train y test	63,26%	59,73%	59,15%	53,13%	97,80%
C. DIFERENCIA ENTRE RESULTADOS DE ENTRENAMIENTO Y PRUEBA						
Clasificador	Escenario	Δ Precision	Δ Accuracy	Δ F1 Score	Δ Recall	Δ Specificity
CNN	Sin aumento	0,366	0,365	0,414	0,368	0,021
	Con aumento en train	0,116	0,078	0,121	0,004	0,014
	Con aumento en train y test	0,025	0,018	0,025	0,004	0,013

Fuente: [Elaboración propia]

Tabla 4.13. Resultados de la CNN en el dominio frecuencial para los tres escenarios con datos de: a) Entrenamiento; b) Prueba; c) Diferencia entre ambos conjuntos.

A. RESULTADOS CON DATOS DE ENTRENAMIENTO						
Clasificador		Precision	Accuracy	F1 Score	Recall	Specificity
		Porcentaje	Porcentaje	Porcentaje	Porcentaje	Porcentaje
CNN	Sin aumento	92,26%	91,87%	91,70%	91,47%	99,81%
	Con aumento en train	54,21%	43,36%	36,48%	31,97%	98,67%
	Con aumento en train y test	54,04%	43,22%	38,61%	22,55%	98,67%
B. RESULTADOS CON DATOS DE PRUEBA						
Clasificador		Precision	Accuracy	F1 Score	Recall	Specificity
		Porcentaje	Porcentaje	Porcentaje	Porcentaje	Porcentaje
CNN	Sin aumento	47,84%	47,76%	46,56%	46,78%	97,62%
	Con aumento en train	37,75%	35,08%	20,92%	30,33%	97,42%
	Con aumento en train y test	52,17%	42,55%	37,39%	21,49%	97,54%
C. DIFERENCIA ENTRE RESULTADOS DE ENTRENAMIENTO Y PRUEBA						
Clasificador	Escenario	Δ Precision	Δ Accuracy	Δ F1 Score	Δ Recall	Δ Specificity
CNN	Sin aumento	0,444	0,441	0,451	0,447	0,022
	Con aumento en train	0,165	0,083	0,156	0,016	0,013
	Con aumento en train y test	0,019	0,007	0,012	0,011	0,011

Fuente: [Elaboración propia]

4.4.1.2 Conversión de un modelo unimodal a multimodal mediante la fusión de los dominios espacial y frecuencial

Dado que en el análisis superficial se exploró la combinación de características de ambos dominios, se realizó un ajuste a la arquitectura de la CNN, pasando de un modelo unimodal a un modelo multimodal. En este caso, las imágenes procesadas fueron ingresadas en una arquitectura de red neuronal convolucional multimodal, en la que cada dominio (espacial y frecuencial) se procesó de manera independiente a través de ramas especializadas.

Cabe resaltar que la arquitectura utilizada para la fusión de dominios corresponde a la misma empleada en los enfoques individuales, diferenciándose únicamente en el enfoque multimodal mediante la incorporación de una capa de concatenación de las características extraídas de cada dominio.

Cada rama incluyó capas convolucionales con activación ReLU, normalización por lotes y agrupación máxima para la reducción de dimensionalidad, permitiendo que el modelo aprendiera representaciones discriminantes de cada dominio. Posteriormente, las características extraídas de ambas ramas fueron fusionadas en una capa de concatenación, integrando la información de los dos dominios en una única representación combinada.

Para mejorar la generalización del modelo, la representación fusionada fue procesada por capas densas con activación ReLU y *Dropout*, minimizando el sobreajuste. Finalmente, se incluyó una capa de clasificación con activación *softmax*, la cual asignó cada imagen a una de las tres categorías: tejido normal, tejido con masa benigna o tejido con masa maligna.

Este enfoque permitió evaluar si la combinación de características espaciales y frecuenciales proporciona un mejor desempeño en la clasificación de imágenes mamográficas en comparación con los enfoques unimodales.

La Tabla 4.14 presenta el desempeño del modelo CNN con fusión de dominios para los tres escenarios evaluados. En esta tabla se resumen los resultados obtenidos con datos de: a) entrenamiento, b) prueba y c) las diferencias entre ambos conjuntos, lo que permite analizar de forma integrada la capacidad de generalización del modelo y el impacto de las técnicas de aumento de datos en cada caso.

Tabla 4.14. Resultados de la CNN con fusión de dominios para los tres escenarios con datos de: a) Entrenamiento; b) Prueba; c) Diferencia entre ambos conjuntos.

A. RESULTADOS CON DATOS DE ENTRENAMIENTO						
Clasificador		Precision	Accuracy	F1 Score	Recall	Specificity
		Porcentaje	Porcentaje	Porcentaje	Porcentaje	Porcentaje
CNN	Sin aumento	92,63%	92,20%	92,16%	91,85%	99,82%
	Con aumento en train	63,90%	59,46%	59,35%	49,63%	99,05%
	Con aumento en train y test	66,32%	61,34%	61,44%	50,54%	99,09%
B. RESULTADOS CON DATOS DE PRUEBA						
Clasificador		Precision	Accuracy	F1 Score	Recall	Specificity
		Porcentaje	Porcentaje	Porcentaje	Porcentaje	Porcentaje
CNN	Sin aumento	55,51%	55,29%	51,15%	54,75%	97,74%
	Con aumento en train	52,96%	51,15%	46,74%	49,30%	97,68%
	Con aumento en train y test	64,54%	59,17%	58,94%	49,92%	97,79%
C. DIFERENCIA ENTRE RESULTADOS DE ENTRENAMIENTO Y PRUEBA						
Clasificador	Escenario	Δ Precision	Δ Accuracy	Δ F1 Score	Δ Recall	Δ Specificity
CNN	Sin aumento	0,371	0,369	0,410	0,371	0,021
	Con aumento en train	0,109	0,083	0,126	0,003	0,014
	Con aumento en train y test	0,018	0,022	0,025	0,006	0,013

Fuente: [Elaboración propia]

4.4.1.3 Resultados y análisis

La evaluación de los diferentes escenarios permitió concluir que el modelo CNN obtiene su mejor desempeño al combinar los dominios espacial y frecuencial (fusión), junto al aumento de datos aplicado tanto en el conjunto de entrenamiento como en el conjunto de prueba. Este enfoque logra una precisión del 64.54%, una exactitud del 59.17%, un F1-score del 58.49%, y una sensibilidad (recall) del 49.92%, superando a los modelos unimodales en todos los escenarios analizados. Comparado con el mejor resultado en el dominio espacial (con aumento en entrenamiento y prueba), el modelo multimodal presenta una mejora de 1.28% en precisión, aunque una ligera disminución de 0.56% en exactitud, 0.66% en F1-score y 3.21% en sensibilidad. Por otro lado, frente al mejor resultado en el dominio frecuencial, se observa una mejora significativa de 12.37 % en precisión, 16.62% en exactitud, 21.10% en F1-score y 28.43% en sensibilidad, como se muestra en la Tabla 4.14.

En el análisis por dominio, se observó que, en el dominio espacial, el modelo sin aumento de datos presentó un sobreajuste evidente, logrando métricas altas en el conjunto de entrenamiento (precisión del 93.40%, exactitud del 93.19%, F1-score del 93.15% y sensibilidad del 93.01%), pero un bajo desempeño en el conjunto de prueba (precisión del 56.82%, exactitud del 56.71%, F1-score del 51.78% y sensibilidad del 56.26%), reflejando diferencias amplias entre ambos conjuntos. Al aplicar aumento de datos únicamente en el conjunto de entrenamiento, se redujo parcialmente este sobreajuste y se observó una mejora en la capacidad de generalización, aunque aún con diferencias notorias (precisión del 66.76% en

entrenamiento frente al 55.19% en prueba). Finalmente, al implementar aumento de datos tanto en entrenamiento como en prueba, se logró un mejor balance, con un incremento sustancial en las métricas de prueba (precisión del 63.26%, exactitud del 59.73%, F1-score del 59.15%, sensibilidad del 53.13%) y una diferencia mínima respecto al entrenamiento (precisión del 65.71%, exactitud del 61.52%, F1-score del 59.38%, sensibilidad del 53.57%). Este comportamiento evidencia que el uso de aumento de datos en ambos conjuntos contribuye a una mejor generalización y estabilidad del modelo.

Por su parte, en el dominio frecuencial, se obtuvieron los valores de desempeño más bajos en todos los escenarios, lo cual refleja una menor capacidad discriminativa del modelo al trabajar únicamente con información espectral. Aunque se identificaron ligeras mejoras al aplicar aumento de datos en entrenamiento y prueba (con una precisión de 52.17%, exactitud del 42.55%, F1-score del 37.39% y sensibilidad del 21.49% en prueba), las métricas se mantuvieron por debajo de las obtenidas en el dominio espacial, lo que sugiere que la información frecuencial por sí sola no es suficiente para lograr un adecuado rendimiento en la clasificación de las imágenes mamográficas.

El enfoque de fusión espacial-frecuencial, mediante el uso de una arquitectura multimodal, permitió integrar de manera conjunta las características espaciales y frecuenciales, logrando el mejor desempeño global. Cabe resaltar que la arquitectura utilizada para la fusión corresponde a la misma empleada en los enfoques individuales, diferenciándose únicamente en la incorporación de una capa de concatenación que permitió integrar las representaciones extraídas de cada dominio.

Adicionalmente, al analizar las diferencias obtenidas entre las métricas de entrenamiento y prueba, se evidenció que el escenario con aumento de datos en ambos conjuntos presentó las diferencias más bajas, con una reducción de hasta 0.3534 puntos porcentuales en precisión en comparación con el escenario sin aumento, lo que demuestra una mayor capacidad de generalización y un ajuste adecuado del modelo a datos no vistos. Por ejemplo, en el enfoque de fusión espacial-frecuencial, las diferencias entre entrenamiento y prueba en las métricas de precisión, exactitud, F1-score y sensibilidad fueron de 0.018, 0.007, 0.025 y 0.006, respectivamente, en contraste con las diferencias observadas en el escenario sin aumento, que fueron significativamente mayores (0.371, 0.369, 0.451 y 0.371). Esta reducción de la brecha entre los conjuntos evidencia que el modelo no se encuentra sobreajustado y que generaliza de manera eficiente, aspecto fundamental para un sistema de apoyo al diagnóstico clínico.

Por tanto, se concluye que el mejor modelo obtenido corresponde al modelo CNN multimodal (fusión espacial-frecuencial) con aumento de datos aplicado tanto en el conjunto de entrenamiento como en el conjunto de prueba, al lograr un equilibrio ideal entre un alto desempeño en prueba y una mínima diferencia respecto al entrenamiento. Este comportamiento garantiza una mayor confiabilidad clínica, siendo el enfoque más adecuado para la clasificación automática de tejido mamográfico normal, benigno y maligno.

4.4.2 Pruebas y resultados para el aprendizaje por transferencia

El entrenamiento y evaluación de las redes neuronales convolucionales preentrenadas (VGG, ResNet, DenseNet e Inception V3) mediante transferencia de aprendizaje se realizó siguiendo un proceso de parametrización estructurado. Inicialmente, se cargaron los modelos preentrenados y se añadieron capas de salida, conformadas por una capa densa y una capa de *Dropout* con una tasa del 0.5. Para la optimización, se utilizó la función de costo de entropía cruzada categórica y se consideraron métricas de desempeño como precisión, exactitud, F1-score, sensibilidad y especificidad.

En cuanto a la estrategia de ajuste fino (*fine-tuning*), se adoptó un enfoque unificado para todas las redes, en el cual se reentrenaron únicamente las primeras cinco y las últimas cinco capas de cada modelo, mientras que el resto de las capas intermedias se mantuvo congelado. Esta configuración buscó un equilibrio entre la preservación de las representaciones aprendidas en ImageNet y la adaptación al nuevo dominio de imágenes mamográficas. Dicha estrategia fue seleccionada tras realizar pruebas preliminares no documentadas, en las que esta opción demostró mayor estabilidad y rendimiento en comparación con configuraciones completamente congeladas o totalmente entrenables.

El entrenamiento se llevó a cabo durante 100 épocas, aplicando técnicas de aumento de datos y el método de parada temprana (*early stopping*) para prevenir el sobreajuste. Al finalizar, se seleccionó y almacenó el modelo con la menor pérdida registrada durante el entrenamiento para su posterior evaluación. La Tabla 4.15 detalla la configuración del entorno de entrenamiento de las redes de aprendizaje profundo empleadas en este estudio.

Tabla 4.15. Configuración del entorno de entrenamiento para el aprendizaje por transferencia.

Factor de entrenamiento	Valor
Plataforma	Tensorflow y Keras
GPU	NVIDIA GeForce RTX 4050
CPU	12th Gen Intel(R) Core(TM) i7-12700H
Optimizador	Adam
Función de pérdida	Entropía cruzada categórica
Tasa de aprendizaje	0,0001
Épocas	100
Tamaño de batch	32

Fuente: [Elaboración propia]

Se evaluaron los modelos preentrenados en términos de exactitud, sensibilidad y F1-score para determinar su desempeño en la clasificación de tejido normal, benigno y maligno. La Tabla 4.16 presenta una comparación de su rendimiento en el escenario donde solo se aplicó aumento de datos en la fase de entrenamiento, dado que este enfoque mostró el mejor desempeño en la CNN diseñada.

El conjunto de datos inicial está compuesto por 41.154 imágenes de 50x50 píxeles, distribuidas en un 70% para entrenamiento y un 30% para prueba. Del conjunto de entrenamiento, se extrajo el 20% para validación, quedando conformado de la siguiente manera: 27.069 imágenes para entrenamiento, 6.765 imágenes para validación y 7.320 imágenes para prueba. Cada clase se encuentra balanceada, con el mismo número de imágenes para tejido normal, benigno y maligno en cada conjunto de datos.

Adicionalmente, para complementar el análisis, después de hacer el ajuste de los modelos, se reporta en la Tabla 4.17 la cantidad de parámetros de cada arquitectura utilizada (cantidad de parámetros mucho más grandes que las máquinas de aprendizaje superficial). Se incluye también la relación aproximada entre el número de entradas y los parámetros, lo que permite contextualizar el tamaño del conjunto de datos frente a la capacidad del modelo.

Tabla 4.16. Comparación del rendimiento de diferentes modelos de aprendizaje profundo mediante aprendizaje por transferencia con datos de: a) Entrenamiento; b) Prueba; c) Diferencia entre ambos conjuntos.

A. RESULTADOS CON DATOS DE ENTRENAMIENTO					
Clasificador	Precision	Accuracy	F1 Score	Recall	Specificity
	Porcentaje	Porcentaje	Porcentaje	Porcentaje	Porcentaje
VGG16	77,20%	75,04%	75,09%	72,34%	97,66%
VGG19	81,10%	79,55%	79,35%	77,59%	98,08%
ResNet50	78,65%	76,07%	75,94%	72,50%	97,76%
ResNet101	70,42%	67,42%	67,30%	62,92%	96,94%
DenseNet121	76,81%	74,18%	74,15%	70,22%	97,58%
DenseNet201	77,90%	74,95%	74,93%	70,28%	97,66%
InceptionV3	68,78%	64,98%	64,86%	58,35%	96,72%
B. RESULTADOS CON DATOS DE PRUEBA					
Clasificador	Precision	Accuracy	F1 Score	Recall	Specificity
	Porcentaje	Porcentaje	Porcentaje	Porcentaje	Porcentaje
VGG16	62,90%	61,67%	60,20%	59,39%	91,35%
VGG19	62,15%	60,97%	60,26%	58,83%	91,30%
ResNet50	66,40%	64,47%	63,79%	61,00%	91,52%
ResNet101	61,52%	59,44%	57,84%	55,78%	91,20%
DenseNet121	65,81%	63,91%	63,15%	60,49%	91,49%
DenseNet201	65,69%	63,11%	62,34%	58,59%	91,44%
InceptionV3	62,73%	60,74%	59,20%	55,55%	91,28%
C. DIFERENCIA ENTRE RESULTADOS DE ENTRENAMIENTO Y PRUEBA					
Clasificador	Δ Precision	Δ Accuracy	Δ F1 Score	Δ Recall	Δ Specificity
VGG16	0,143	0,134	0,149	0,130	0,063
VGG19	0,190	0,186	0,191	0,188	0,068
ResNet50	0,123	0,116	0,122	0,115	0,062
ResNet101	0,089	0,080	0,095	0,071	0,057
DenseNet121	0,110	0,103	0,110	0,097	0,061
DenseNet201	0,122	0,118	0,126	0,117	0,062
InceptionV3	0,061	0,042	0,057	0,028	0,054

Fuente: [Elaboración propia]

Tabla 4.17. Parámetros de las redes preentrenadas empleadas en el aprendizaje por transferencia y su relación con el tamaño del conjunto de datos.

Clasificador	Parámetros modificados	Entrenables	No entrenables	Entradas por parámetro
VGG16	15.112.515	7.588.291	7.524.224	13,56
VGG19	20.422.211	9.948.099	10.474.112	10,34
ResNet50	24.771.971	2.247.043	22.524.928	45,79
ResNet101	43.842.435	2.247.043	41.595.392	45,79
DenseNet121	7.697.475	706.883	6.990.592	145,55
DenseNet201	19.440.707	1.167.427	18.273.280	88,13
InceptionV3	22.987.043	1.192.835	21.794.208	86,25

Fuente: [Elaboración propia]

4.4.2.1 Resultados y análisis

Los resultados obtenidos mediante transferencia de aprendizaje con redes preentrenadas evidencian diferencias entre los modelos evaluados. En la etapa de entrenamiento, VGG19 presentó el mejor rendimiento general en todas las métricas, logrando una precisión del 81.10%, exactitud del 79.55%, F1-score del 79.35%, sensibilidad del 77.59% y especificidad del 98.08%, lo que refleja una alta capacidad de aprendizaje sobre los datos vistos. Sin embargo, al comparar estos valores con los obtenidos en la etapa de prueba, se evidenció una diferencia de hasta 18.95 puntos porcentuales en precisión y 18.8 puntos en sensibilidad, indicando un posible sobreajuste del modelo.

En contraste, InceptionV3 presentó los valores más bajos durante el entrenamiento (precisión del 68.78%, exactitud del 64.98%, F1-score del 64.86%, sensibilidad del 58.35%), ubicándose en la última posición en la mayoría de las métricas. No obstante, como se evidenció su desempeño en el conjunto de prueba es competitivo y muestra una adecuada capacidad de generalización.

Durante la evaluación con los datos de prueba, ResNet50 fue el modelo con mejor rendimiento global, obteniendo una precisión del 66.40%, exactitud del 64.47%, F1-score del 63.79% y la mayor sensibilidad del 61.00%, lo que resulta clave en el contexto clínico para detectar tejido anómalo (masas benignas y malignas). Además, logró la mayor especificidad (91.52%), lo que sugiere una adecuada capacidad para discriminar casos negativos, incluyendo tejido normal. Este desempeño lo posiciona como el modelo con mejor balance entre detección y precisión en la prueba final.

Por su parte, DenseNet121 y DenseNet201 mostraron un rendimiento competitivo, logrando precisiones del 65.81% y 65.69%, respectivamente, y F1-scores cercanos al 63.00%, con una sensibilidad aceptable (60.49% y 58.59%), aunque levemente inferiores a ResNet50. Estos resultados sugieren que ambas arquitecturas tienen potencial en la clasificación de tejido mamográfico.

El modelo InceptionV3, aunque presentó un rendimiento moderado (precisión del 62.73%, exactitud del 60.74% y F1-score del 59.20%), se destacó por tener las diferencias más bajas entre entrenamiento y prueba, mostrando así una excelente capacidad de generalización, lo que lo convierte en una opción robusta ante datos no vistos.

En cuanto al análisis de la diferencia entre entrenamiento y prueba (Tabla 4.16c), se observa que InceptionV3 es el modelo con menor diferencia en todas las

métricas, destacándose con una Δ precisión = 0.061, Δ exactitud = 0.042, Δ F1-score = 0.057, Δ sensibilidad = 0.028 y Δ especificidad = 0.054, lo que refleja un comportamiento estable y consistente. Esta baja diferencia indica que, aunque no logró las métricas más altas en prueba, su desempeño es fiable y poco dependiente de los datos de entrenamiento, un aspecto fundamental para aplicaciones clínicas.

Por otro lado, ResNet101, aunque alcanzó una precisión elevada en prueba (61.52%), mostró una muy baja sensibilidad (55.78%) y diferencias relativamente controladas entre entrenamiento y prueba (Δ precisión = 0.089, Δ exactitud = 0.080), pero insuficiente para considerarlo entre los más robustos, debido a su menor capacidad de detectar correctamente tejido anómalo.

En contraste, modelos como VGG16 y VGG19, a pesar de un buen desempeño en entrenamiento, presentaron diferencias importantes en la transición a prueba (por ejemplo, VGG19 con Δ precisión = 0.190, Δ sensibilidad = 0.188), evidenciando un claro sobreajuste y limitaciones para generalizar a nuevos datos. Esto mismo se refleja en ResNet50, que, a pesar de tener los mejores resultados en prueba, presenta diferencias relevantes respecto al entrenamiento (Δ precisión = 0.123, Δ sensibilidad = 0.115), aunque estas diferencias son menores que las de VGG16 y VGG19.

En resumen, ResNet50 se posiciona como el mejor modelo en desempeño general, logrando las métricas más altas en el conjunto de prueba, especialmente en sensibilidad, que es prioritaria en la detección de casos positivos. Sin embargo, considerando la capacidad de generalización, InceptionV3 sobresale como el modelo más estable y robusto, siendo adecuado para sistemas que requieran un rendimiento consistente en condiciones reales.

Por tanto, se concluye que ResNet50 es el modelo más adecuado si se priorizan las métricas de rendimiento directo en la prueba (sensibilidad y precisión), mientras que InceptionV3 es preferible en contextos donde la estabilidad y generalización sean prioritarias. Ambos modelos se destacan frente al resto y pueden considerarse candidatos sólidos para el desarrollo de sistemas CAD para la detección automática de tejido mamográfico.

4.5 Análisis general del sistema

El sistema propuesto para la clasificación automática de imágenes mamográficas en tres clases (tejido normal, tejido benigno y tejido maligno) presenta un diseño integral y modular, que abarca desde el procesamiento inicial de las imágenes hasta

la clasificación final, combinando técnicas de visión artificial, aprendizaje automático superficial y profundo.

En primer lugar, se aborda un proceso de clasificación multiclase, orientado a la identificación precisa de tejido normal, así como de masas benignas y malignas, lo que permite abordar de manera directa la problemática del diagnóstico asistido en imágenes mamográficas. Para lograr esta clasificación, el sistema realiza un análisis complementario en dos dominios: el dominio espacial, mediante el cálculo de características texturales y estadísticas, y el dominio frecuencial, a partir de la Transformada de Fourier y la Transformada Wavelet, con el propósito de capturar información detallada sobre las estructuras internas del tejido.

El proceso de preprocesamiento incluye la eliminación automática de artefactos no deseados y la segmentación de la región de interés (mama), permitiendo la extracción controlada de ventanas de 50x50 píxeles, las cuales aseguran la focalización sobre las áreas relevantes para el análisis. Posteriormente, se realiza la extracción de 10 características por cada dominio (espacial y frecuencial), orientadas a describir las propiedades estadísticas, texturales y estructurales del tejido mamario. Con el fin de optimizar el desempeño y reducir la complejidad del sistema, se implementan métodos de selección de características, específicamente técnicas de Ranking (ReliefF) y Wrappers, logrando la reducción a las 6 características más relevantes por dominio, sin comprometer la capacidad discriminativa del sistema.

La etapa de clasificación se desarrolla mediante una amplia gama de algoritmos de aprendizaje superficial, incluyendo Naive Bayes, LDA, QDA, KNN, Random Forest, SVM, RBF y MLP. Para fortalecer la robustez y estabilidad del sistema, se implementa una red de comité (ensemble), que combina los clasificadores de mayor desempeño, permitiendo mejorar el rendimiento general del modelo. Además, se desarrollan modelos de aprendizaje profundo, como una CNN diseñada totalmente y modelos pre-entrenados mediante la técnica de Transfer Learning (VGG, ResNet, DenseNet e InceptionV3).

Finalmente, el sistema ha sido evaluado de forma integral utilizando métricas de desempeño ampliamente aceptadas en la literatura, tales como precisión, exactitud, sensibilidad, F1-score y especificidad, lo que asegura una valoración completa de su efectividad en la clasificación del tejido mamográfico.

La Tabla 4.18 muestra las características de funcionamiento del sistema, así como sus rangos y condiciones de operación.

Tabla 4.18. Resumen de las características del sistema.

Característica	Detalle del sistema	Observaciones
Dataset utilizado	MIAS (322 imágenes, 266 válidas)	Predominio de tejido normal, resolución limitada (8 bits)
Tamaño de imagen procesada	1024 x 1024 píxeles (ventanas de 50 x 50 píxeles)	Imágenes en escala de grises
Técnicas de procesamiento y segmentación	Umbralización, transformación exponencial, operaciones morfológicas, operador canny	Eliminación de artefactos, no se realiza segmentación específica de masas
Técnicas de aumento de datos	Enventanado (aprendizaje superficial) y data augmentation (aprendizaje profundo)	Mejora de la generalización en los modelos
Características extraídas	10 por dominio (espacial y frecuencial), reducidas a 6 mediante selección	Medidas estadísticas, texturales y espectrales discriminantes
Técnicas de selección de características	Métodos de Ranking (ReliefF) y Wrappers	Refinamiento del conjunto de características para mejorar la clasificación.
Modelos de clasificación superficial	NB, LDA, QDA, Dist. Euc., Dist. Maha., KNN, RF, SVM, RBF, MLP	Análisis comparativo y red de comité
Mejores clasificadores superficiales (espacial)	RBF (Recall: 71.20%), MLP (Accuracy: 77.30%), QDA (Recall: 70.40%)	Predominio de RBF en sensibilidad, buen balance con MLP
Mejores clasificadores superficiales (frecuencial)	MLP (Recall: 68.64%), SVM (F1: 64.76%) y QDA (Recall: 68.55%)	MLP destaca por su sensibilidad y exactitud
Mejor estrategia (fusión de dominios)	SVM (Accuracy: 76.86%, Recall: 63.96%)	Combinación de dominios espacial y frecuencial mejora el desempeño
Red de comité (mejor resultado global superficial)	Dominio espacial (Recall: 76.80%, Accuracy: 82.35%) y en fusión (Accuracy: 84.50%)	Superior a clasificadores individuales
Modelo CNN diseñada desde cero	CNN (Exactitud: 59.17%, Sensibilidad: 49.92%)	CNN con aumento de datos (entrenamiento y prueba) y fusión de dominios
Modelos Transfer Learning evaluados	VGG16, VGG19, ResNet50, ResNet101, DenseNet121, DenseNet201 e InceptionV3	ResNet50: alta sensibilidad y precisión. InceptionV3: alta estabilidad y generalización.
Hardware requerido	GPU NVIDIA RTX 4050, Intel Core i7 12700H, RAM 16 GB	Requerido especialmente para entrenamiento de modelos profundos (CNN y Transfer Learning)
Tiempo de procesamiento	Aprox. entre 25 min. y 1 hr (c/u superficial) Aprox. entre 1 hr y 4 hrs (c/u profundo)	Tiempo más elevado en entrenamiento y prueba con aprendizaje profundo.
Evaluación del sistema (métricas)	Precisión, exactitud, F1-score, sensibilidad, especificidad	Evaluación integral según estándares en clasificación médica (recall)

Fuente: [Elaboración propia]

4.6 Alcances y limitaciones del sistema desarrollado

Alcances:

- Desarrollo de un sistema de clasificación automática multiclase para la detección de tejido mamario (normal, benigno y maligno) en imágenes digitales de mamografías, utilizando técnicas de visión artificial y modelos de aprendizaje superficial y profundo.
- Procesamiento automático de las imágenes, mediante algoritmos que permiten la eliminación de artefactos no deseados y la segmentación de la región de interés (mama), garantizando una adecuada selección de las áreas a analizar.

- Construcción de un dataset balanceado mediante el método de enventanado, conformado por 13.718 ventanas por clase y para cada dominio, extraídas del dataset público MIAS, asegurando una adecuada representatividad y equilibrio de las tres clases.
- Extracción, análisis y selección de características discriminantes en los dominios espacial y frecuencial, logrando una reducción de 10 a 6 características comúnmente utilizadas en la literatura científica por dominio, con el fin de optimizar la eficiencia y precisión del sistema.
- Integración combinada de los dominios espacial y frecuencial, permitiendo una caracterización más completa y robusta del tejido mamario, al capturar tanto información textural como frecuencial.
- Comparación exhaustiva de modelos de aprendizaje superficial y profundo, incluyendo algoritmos clásicos (Naive Bayes, LDA, QDA, Distancia Mahalanobis, Distancia Euclidiana, KNN, SVM, Random Forest, RBF, MLP), redes neuronales profundas (CNN diseñada desde cero), y modelos preentrenados mediante Transfer Learning (VGG, ResNet, DenseNet, InceptionV3), evaluando su desempeño en la clasificación multiclase.
- Implementación de redes de comité (ensemble), mediante la combinación de los clasificadores de mayor desempeño, con el propósito de mejorar la estabilidad y el rendimiento del sistema.
- Aplicación de técnicas de aumento de datos (data augmentation), mediante rotaciones, giros y otras transformaciones, para mejorar la capacidad de generalización del sistema frente a nuevas imágenes no vistas durante el entrenamiento.
- Evaluación y adaptación de modelos preentrenados (Transfer Learning) a la problemática específica de clasificación de tejido mamario.
- Generación de un análisis comparativo y crítico del desempeño de los modelos, destacando las fortalezas, limitaciones y oportunidades de mejora en cada enfoque de clasificación.

Limitaciones:

- Como parte del enfoque académico, se decidió trabajar con bases de datos públicas, lo que condiciona ciertas características técnicas del dataset, como la resolución, cantidad de casos de estudio y ausencia de imágenes de mayor profundidad (12 o 16 bits).
- Procesamiento limitado por la calidad de las imágenes: debido a la baja resolución, la aplicación de técnicas de mejora (como realce de bordes o ajuste de contraste) puede no ofrecer un beneficio completo y afectar la calidad de las ventanas extraídas.
- El sistema no realiza segmentación específica de los bordes completos de las masas, sino que trabaja sobre ventanas extraídas dentro de las masas, lo que limita la obtención de información morfológica global (como forma, contorno o área total), aunque garantiza el análisis del tejido comprometido.
- Posible pérdida de información estructural global al trabajar exclusivamente con ventanas, sin considerar la masa completa en su contexto total.
- La reducción de características a solo seis por dominio, aunque útil para mejorar la eficiencia del sistema, también representó una limitación, ya que excluye otras variables potencialmente relevantes comúnmente usadas en la literatura científica.
- La variabilidad en el rendimiento de los clasificadores superficiales depende fuertemente del conjunto de características seleccionadas y de la combinación de dominios considerada, lo que introduce un factor de sensibilidad metodológica.
- Menor sensibilidad en los enfoques basados únicamente en el dominio frecuencial, al no capturar toda la información espacial disponible.
- Desempeño inferior de la CNN diseñada desde cero respecto a los modelos preentrenados, lo cual puede estar asociado tanto a la arquitectura básica utilizada como a las limitaciones del conjunto de datos disponible.
- Al analizar el comportamiento de las redes preentrenadas utilizadas en el aprendizaje profundo, se encontró que la relación entre la cantidad de

parámetros y el tamaño del conjunto de datos puede influir negativamente en el rendimiento. Por ejemplo, redes como VGG16 y VGG19 poseen más de 138 millones de parámetros, mientras que el conjunto de entrenamiento contenía aproximadamente 27.069 imágenes (de 50x50 píxeles), lo que equivale a unos 67.6 millones de valores. Esta relación representa menos de medio valor por parámetro, lo cual limita la capacidad de ajuste del modelo, dificultando el aprendizaje efectivo y aumentando el riesgo de sobreajuste o desempeño subóptimo.

- Aunque el total de datos puede parecer elevado (41.154 imágenes), al considerar que cada imagen de 50x50 píxeles contiene 2.500 valores, la cantidad total de entradas es aproximadamente 102 millones. Sin embargo, al compararse con redes como ResNet101 (≈ 44.5 millones de parámetros), la proporción sigue siendo baja: apenas 2 a 3 muestras por parámetro. Esta relación sugiere que el rendimiento de las redes preentrenadas podría mejorar con datasets más extensos.
- Finalmente, se identificaron desafíos específicos en la correcta clasificación de ciertas clases, posiblemente debido a similitudes visuales entre patrones benignos y malignos o a la complejidad de los casos representados.

4.7 Discusión

La metodología planteada permitió abordar de manera integral la clasificación automática de imágenes mamográficas en tres categorías fundamentales para la detección de cáncer de mama: tejido normal, masas benignas y masas malignas.

Desde la etapa de procesamiento, se implementaron algoritmos que mejoraron la calidad de las imágenes y permitieron una adecuada delimitación de la región de la mama, garantizando un correcto enfoque en las zonas de interés. Esto facilitó la extracción de ventanas de 50x50 píxeles que alimentaron el sistema de clasificación.

En cuanto a la extracción y selección de características, se aplicaron técnicas que permitieron reducir de 10 a 6 características por dominio (espacial y frecuencial), manteniendo aquellas con mayor capacidad discriminante. Esta reducción mejoró la eficiencia del sistema sin comprometer su desempeño clasificatorio.

Respecto a la etapa de clasificación mediante aprendizaje superficial, en el dominio espacial, RBF, MLP y QDA ocuparon las primeras posiciones, destacándose RBF

por su alta exactitud (78.10%) y sensibilidad (71.20%). En el dominio frecuencial, MLP mostró un rendimiento más equilibrado entre sus métricas, a diferencia de QDA o SVM, cuyas variaciones entre precisión y sensibilidad fueron más marcadas. En la fusión de dominios, el clasificador SVM obtuvo la mayor exactitud (76.86%), superando por 2.1 puntos porcentuales a QDA y por 8.7 a Naive Bayes, además de alcanzar la mayor especificidad (80.98%) entre todos los escenarios. Si bien el dominio espacial presentó los mejores valores globales en exactitud y sensibilidad, la fusión demostró un mejor desempeño en la identificación de tejido normal, lo que sugiere que combinar información espacial y frecuencial puede ser beneficioso en aplicaciones donde se busque un equilibrio entre minimizar falsos positivos y detectar correctamente casos anómalos.

No obstante, el mejor resultado global dentro del aprendizaje superficial se obtuvo mediante la red de comité, conformada por RBF, MLP y QDA en el dominio espacial, y SVM, MLP y QDA en la fusión de dominios. Esta estrategia alcanzó una exactitud del 82.35% y una sensibilidad del 76.80% en el dominio espacial, superando en 4.25 puntos porcentuales en exactitud y 5.60 puntos en sensibilidad al RBF, así como en 5.49 puntos porcentuales en exactitud y 12.84 puntos en sensibilidad al SVM. Esto evidencia una mayor capacidad de generalización frente a los modelos individuales. Aunque estos valores se encuentran por debajo del mejor resultado reportado en la literatura que alcanzó una exactitud del 97.80% [16], el sistema desarrollado demuestra una mejor sensibilidad frente a otros enfoques como KNN, Bagging y EigenClass, los cuales reportaron una exactitud del 91.00% y una sensibilidad del 64.72% [23]. De manera destacada, el modelo propuesto supera esta última en 12.08 puntos porcentuales, trabajando además sobre la misma base de datos, aspecto crucial para minimizar los falsos negativos, lo cual es de vital importancia en contextos clínicos.

En la clasificación mediante aprendizaje profundo, el modelo CNN diseñado desde cero, potenciado con técnicas de aumento de datos y fusión de dominios, alcanzó en su mejor configuración (modelo multimodal con capa de concatenación) una precisión del 64.54% y una especificidad del 97.79%. Aunque su rendimiento fue inferior al de las redes preentrenadas y a los modelos de aprendizaje superficial, este enfoque mostró una adecuada capacidad para identificar regiones sin anomalías, al superar en 16.81 puntos porcentuales la especificidad obtenida por SVM, el mejor clasificador en este aspecto dentro del aprendizaje superficial. Sin embargo, estos resultados sugieren que la CNN diseñada requiere un mayor volumen de datos y una posible optimización de su arquitectura para mejorar su rendimiento general y poder competir con modelos más robustos.

En cuanto a los modelos basados en aprendizaje por transferencia (Transfer Learning), se destacan especialmente ResNet50 e InceptionV3, que lograron superar al resto de las arquitecturas evaluadas en varias métricas. ResNet50 alcanzó una precisión de 66.40% y una sensibilidad del 61.00% priorizando la sensibilidad y el rendimiento directo, mientras que InceptionV3 presentó valores cercanos con una exactitud del 60.74% y una sensibilidad del 55.55% destacado por su estabilidad y comportamiento consistente, confirmando la robustez de los modelos preentrenados para este tipo de tareas, especialmente cuando se trabaja con bases de datos limitadas como MIAS. No obstante, aunque los resultados obtenidos con Transfer Learning son superiores a los alcanzados por la CNN propia, se encuentran por debajo de otros reportes en la literatura donde se han propuesto técnicas de fusión avanzada de características y la combinación de múltiples bases de datos, alcanzando precisiones de hasta el 98.70% [24]. Esto resalta la necesidad de trabajar con conjuntos de datos más amplios, con mayor resolución y profundidad de bits, además de explorar métodos de fusión más sofisticados, que podrían incrementar significativamente el rendimiento del sistema.

En general, el desarrollo del sistema permitió realizar una comparación detallada entre distintos métodos de clasificación, evidenciando la relevancia de combinar enfoques estadísticos tradicionales y técnicas modernas de aprendizaje profundo para obtener un balance adecuado entre precisión, sensibilidad y especificidad. Además, la arquitectura modular del sistema garantiza flexibilidad para futuras adaptaciones, permitiendo incorporar nuevos modelos, ajustar las etapas existentes o ampliar el sistema a diferentes bases de datos.

Aunque se lograron avances significativos, persisten retos importantes que deberán abordarse en trabajos futuros, tales como la implementación de una segmentación específica y detallada de las masas, el uso de datasets con imágenes de mayor resolución y profundidad que permitan capturar información más completa, y la validación del sistema en diferentes bases de datos públicas y privadas que aseguren una adecuada capacidad de generalización y robustez.

En términos generales el sistema desarrollado representa una herramienta sólida y adaptable para la clasificación automática de tejido mamario en imágenes digitales, integrando técnicas avanzadas de procesamiento, extracción y selección de características, y clasificación mediante aprendizaje superficial y profundo. Los resultados obtenidos, aunque con margen de mejora, demuestran la viabilidad del enfoque planteado y abren nuevas líneas de investigación que pueden contribuir significativamente al desarrollo de sistemas de apoyo al diagnóstico en el área de la imagenología médica.

5. CONCLUSIONES GENERALES

Se desarrolló un sistema de reconocimiento de patrones basado en técnicas de visión artificial y aprendizaje automático para la identificación de tejido normal, masas benignas y masas malignas en mamografías digitales. Esta solución se estructuró en *dos enfoques* complementarios de máquinas de aprendizaje: *superficial* y *profundo*, utilizando información en los dominios *espacial*, *frecuencial* y su *fusión*. El enfoque de *aprendizaje superficial* integró metodologías de procesamiento, extracción y selección de características, y clasificación automática. Por otro lado, el enfoque de *aprendizaje profundo* incorporó el diseño de una CNN propia inicialmente *unimodal*, extendida luego a una arquitectura *multimodal* mediante la unión de representaciones espaciales y frecuenciales; además, se adaptaron representaciones *multicanal* para el uso de *redes preentrenadas* mediante *transferencia de aprendizaje*. La evaluación se realizó usando métricas como precisión, exactitud, sensibilidad, especificidad y F1-score, priorizando la sensibilidad por su relevancia clínica. Se alcanzaron tasas de desempeño de hasta 82.35% en exactitud y 76.80% en sensibilidad con la red de comité para el *enfoque superficial*, mientras que, en el *enfoque profundo*, ResNet50 alcanzó una exactitud de 64.47% y una sensibilidad de 61.00%. La red CNN propia logró una especificidad destacada del 97.79% en su mejor configuración.

La identificación y la extracción de características discriminantes se sustentó en un flujo automatizado de procesamiento de imágenes, diseñado para mejorar la calidad visual de la imagen y permitir aislar el tejido mamario, eliminando artefactos y el músculo pectoral. Se emplearon técnicas específicas como realce de bordes mediante transformación exponencial ($\text{Gamma} = 0.2 + \text{Otsu}$), oscurecimiento del exterior mediante umbral fijo (intensidad = 30), suavizado morfológico (apertura y cierre), binarización adaptativa (sensibilidad = 0.82), análisis de componentes con *regionprops*, eliminación de regiones desconectadas con *imfill*, detección de bordes mediante el operador de Canny y eliminación del músculo pectoral con *imageSegmenter*. A partir de estas imágenes procesadas, se implementaron dos métodos de *enventanado*: uno basado en ventanas traslapadas sobre regiones anómalas y otro aleatorio sobre tejido normal. El tamaño de las ventanas fue definido mediante análisis estadístico de las dimensiones de las masas en las mamografías, permitiendo una extracción localizada y representativa de las tres clases.

Con las ventanas obtenidas, se construyeron dos conjuntos de datos, uno para cada dominio: *espacial* y *frecuencial*. En el dominio espacial se extrajeron diez

descriptores estadísticos y texturales por ventana (máximo, mínimo, mediana, promedio, desviación estándar, curtosis, asimetría, contraste, energía y homogeneidad). Para el dominio frecuencial, se aplicó la Transformada de Fourier sobre cada ventana, a partir de la cual se calcularon también diez características (promedio, desviación estándar, curtosis, asimetría, entropía, energía total, energía baja, media y alta, e índice de compacidad). Posteriormente, se aplicaron dos estrategias de selección de características: Filtros o Ranking mediante ReliefF, y métodos de Wrappers (forward y backward) usando los clasificadores LDA, QDA y Mahalanobis. Como resultado, se seleccionaron en el *dominio espacial*: *máximo, mínimo, desviación estándar, curtosis, asimetría y homogeneidad*; y en el *dominio frecuencial*: *desviación estándar, curtosis, asimetría, energía total, energía media y energía alta*. Esta reducción de 10 a 6 características se basó en la consistencia de su aparición en múltiples configuraciones y en su impacto positivo sobre el rendimiento de los clasificadores evaluados.

Con base en las características seleccionadas, se entrenaron y evaluaron diez clasificadores en el enfoque de *aprendizaje superficial*: Naive Bayes, LDA, QDA, KNN, Distancia Euclidiana, Distancia Mahalanobis, SVM, Random Forest, RBF y MLP. En el *dominio espacial*, los modelos RBF, MLP y QDA destacaron por su rendimiento, alcanzando RBF una sensibilidad del 71.20% y MLP una exactitud del 77.30%. En el *dominio frecuencial*, el MLP se mantuvo como el clasificador más estable entre métricas.

Adicionalmente, se estudió la *fusión de ambos dominios (espacial y frecuencial)*, lo cual permitió evaluar el impacto de la combinación de descriptores. En este escenario, el clasificador SVM obtuvo la mayor exactitud (76.86%) y la mayor especificidad (80.98%), evidenciando que la integración de información complementaria puede mejorar la identificación del tejido normal.

El mejor desempeño global se obtuvo mediante una *red de comité* que integró RBF, MLP y QDA en el *dominio espacial*, y SVM, MLP y QDA en la *fusión*. Este enfoque alcanzó una exactitud del 82.35% y una sensibilidad del 76.80%, superando en 4.25% de exactitud y 5.60% de sensibilidad al RBF, y en 5.49% y 12.84%, respectivamente, al SVM. Además, la red de comité mejoró la sensibilidad en un 12.08% con respecto a enfoques previos reportados en la literatura que emplearon la misma base de datos, como KNN, Bagging y EigenClass, que alcanzaron una sensibilidad del 64.72% [23], aunque aún por debajo del 97.80% de exactitud reportado en [16].

Para el enfoque de *aprendizaje profundo*, se determinaron representaciones multicanal aplicables al aprendizaje profundo mediante transformaciones en los *dominios espacial y frecuencial*. Se diseñó una red neuronal convolucional (CNN) desde cero, adaptada a imágenes mamográficas redimensionadas a 50×50 píxeles, y se evaluó su comportamiento en tres configuraciones de entrada: *dominio espacial*, *dominio frecuencial* (basado en la *Transformada de Fourier*), y *fusionada* a través de una *capa de concatenación*.

Además, se aplicaron modelos *preentrenados*, para ello se desarrolló una representación *multicanal* que adaptó las imágenes en escala de grises a entradas *RGB*, asignando el canal *rojo* al *dominio espacial*, el *verde* a la *Transformada de Fourier* y el *azul* a la *Transformada de Wavelet*. Esta estrategia permitió combinar información textural y frecuencial dentro de una única entrada, maximizando la capacidad de las arquitecturas profundas para extraer patrones discriminantes.

Se evaluó el desempeño de la CNN propia en las tres configuraciones de entrada y en *tres escenarios* de aumento de datos: sin aumento, con aumento en el conjunto de entrenamiento y con aumento en ambos conjuntos (train y test). El mejor resultado se obtuvo en la configuración *multimodal* y con el escenario de aumento en ambos conjuntos, logrando una precisión del 64.54% y una especificidad del 97.79%, superando en 6.27% la especificidad obtenida por ResNet50, la arquitectura *preentrenada* con mejor desempeño en esta métrica.

Con base en la mejor configuración de la CNN, se evaluaron siete modelos *preentrenados* mediante *transferencia de aprendizaje*: VGG16, VGG19, ResNet50, ResNet101, DenseNet121, DenseNet201 e InceptionV3. Entre estos, ResNet50 presentó el mejor balance general, con una precisión del 66.40%, una exactitud del 64.47% y una sensibilidad del 61.00%, mientras que InceptionV3 se destacó por su estabilidad y capacidad de generalización, al presentar las menores diferencias (0.061, 0.042, 0.057, 0.028 0.054) entre entrenamiento y prueba.

Aunque las redes *preentrenadas* superaron el rendimiento de la CNN propia, sus resultados se mantuvieron por debajo del 98.70% reportado en otros trabajos que utilizaron arquitecturas más complejas con fusión avanzada de características y múltiples bases de datos [24]. Esto refleja que, si bien el enfoque propuesto ofrece un desempeño razonable considerando el tamaño reducido y la resolución limitada del dataset MIAS, aún existen oportunidades de mejora en términos de arquitectura y volumen de datos para alcanzar o superar el desempeño del estado del arte.

Durante el desarrollo del proyecto, se enfrentaron diversos desafíos, principalmente relacionados con la limitada calidad y tamaño del dataset MIAS, las dificultades de

segmentación automática y la correcta selección de características discriminantes, así como la adecuada sintonización de hiperparámetros de los modelos. Aunque se utilizaron herramientas de ajuste automático, como *Keras Tuner*, persisten oportunidades de mejora en las arquitecturas. No obstante, se logró desarrollar un sistema modular y flexible, capaz de integrar diversos enfoques y adaptarse a diferentes dominios y técnicas de clasificación, con capacidad de ser extendido a nuevas bases de datos y algoritmos.

En términos de aportes, este trabajo propone un sistema integral que los *dominios espacial y frecuencial*, y que incorpora y contrasta diversas técnicas de *aprendizaje superficial y profundo* para la clasificación automática de *tejido mamario*. La integración de *redes de comité* y la *fusión de dominios*, junto con la evaluación de arquitecturas *preentrenadas*, representan un enfoque sólido y adaptable que puede contribuir al desarrollo de sistemas de diagnóstico asistido por computadora (CAD) orientados al cáncer de mama.

Finalmente, este trabajo plantea una solución con potencial futuro para apoyar la labor de los radiólogos en la detección temprana de anomalías mamográficas, lo cual resulta de gran relevancia en contextos clínicos, especialmente en países como Colombia, donde la escasez de especialistas limita el acceso a diagnósticos oportunos. Aunque el sistema propuesto aún requiere validación adicional y optimización para su aplicación clínica, sus resultados preliminares sugieren que podría utilizarse como una herramienta complementaria para la detección y priorización de casos sospechosos. Las limitaciones identificadas abren líneas claras de investigación futura, y los avances alcanzados evidencian la viabilidad técnica del enfoque, contribuyendo al desarrollo de soluciones tecnológicas aplicadas a la imagenología médica.

6. TRABAJOS FUTUROS

Con base en el trabajo desarrollado y los resultados obtenidos, se plantean las siguientes perspectivas de trabajo, organizadas de acuerdo con las etapas de la solución implementada. Estas recomendaciones buscan superar las principales limitaciones del sistema desarrollado, ampliar sus alcances y sentar las bases para proyectos de investigación futuros relacionados.

Ampliación y mejora del dataset

- Integración de nuevos conjuntos de datos: Incluir bases de datos públicas adicionales, como INbreast, DDSM, BCDR y CBIS-DDSM, para incrementar la diversidad y representatividad del dataset, y evaluar la capacidad de generalización del sistema en distintos contextos clínicos.
- Incorporación de segmentaciones precisas: Se sugiere complementar las imágenes con máscaras manuales que definan las regiones de las masas, permitiendo entrenar modelos orientados a segmentación automática y mejorar la precisión en la identificación de zonas específicas.

Mejora en el preprocesamiento y segmentación

- Aplicación de técnicas avanzadas de segmentación: Explorar métodos como U-Net, Mask R-CNN, contornos activos (Snakes) y Vision Transformers (ViT) para una segmentación más precisa de las regiones de interés.
- Exploración de nuevas transformadas: Incluir técnicas como Gabor, Coseno Discreta (DCT) y distintas Wavelet madre, como entradas a modelos de clasificación profundo.
- Mejoras en contraste y eliminación de artefactos: Implementar filtros específicos para reducción de ruido y realce local, adaptados a las particularidades de las mamografías digitales.
- Diversificación del data augmentation: Incorporar nuevas transformaciones como variaciones de iluminación, traslaciones, deformaciones elásticas y adición de ruido controlado, con el objetivo de aumentar la variabilidad del dataset y mejorar la capacidad de generalización del modelo.

Nuevas estrategias de extracción y selección de características

- Investigación de nuevas características discriminantes: Evaluar descriptores adicionales, como momentos de Haralick, Zernike, GLCM de segundo y tercer orden, y patrones de forma que podrían aportar mayor capacidad discriminativa.
- Selección de características avanzada: Aplicar técnicas como Análisis de Componentes Principales (PCA), Wrappers, y feature importance, para optimizar el conjunto final de características.
- Métodos sofisticados de fusión de características: Explorar modelos de atención cruzada (cross-attention) o espacios latentes compartidos, que permitan integrar de manera más eficiente las características espaciales y frecuenciales.

Avances en el diseño y entrenamiento de modelos de clasificación

- Clasificación en cascada: Desarrollar clasificadores por etapas, primero distinguir entre tejido normal y masa, y posteriormente entre masa benigna y maligna. Esta estrategia podría mejorar la precisión en escenarios con alta similitud entre clases.
- Evaluación de nuevas arquitecturas profundas: Considerar modelos más recientes y eficientes como EfficientNet, ConvNeXt, Vision Transformers (ViT), Swin Transformer o híbridos CNN-Transformers, que podrían mejorar la precisión y eficiencia del sistema.
- Redes de clasificación en comité: Aplicar esquemas de ensemble learning (Bagging, Boosting, Stacking) con votación robusta (como soft voting), mejorando la estabilidad del sistema ante datos no vistos.
- Implementación de modelos de detección y localización de masas: Incluir arquitecturas como YOLO, Faster R-CNN o RetinaNet para identificar y ubicar automáticamente las zonas anormales dentro de la mamografía.
- Optimización avanzada de hiperparámetros: Aunque se emplearon técnicas automáticas como Keras Tuner durante el desarrollo, se propone explorar enfoques como Optuna, con el objetivo de perfeccionar el ajuste y potenciar el rendimiento de cada modelo.

- **Posprocesamiento basado en vecinos:** Evaluar resultados en ventanas vecinas para corregir errores de clasificación, estableciendo un criterio de validación cruzada espacial en la imagen completa.
- **Medición de la calidad de segmentación:** Implementar mediciones de la calidad de la reconstrucción y segmentación de las regiones correspondiente a masa benigna y maligna detectadas en las mamografías.

Pruebas de robustez y validación clínica

- **Evaluación en diversos formatos y dispositivos:** Probar el sistema con mamografías en distintos formatos (DICOM, PNG, TIFF), resoluciones y equipos, para garantizar su adaptabilidad a diferentes entornos clínicos.
- **Desarrollo de una interfaz gráfica (GUI):** Diseñar un software amigable para personal médico, que permita cargar imágenes, visualizar clasificaciones y regiones de interés, incluyendo retroalimentación directa de los médicos.
- **Creación de una plataforma web con API:** Implementar una plataforma en línea que permita almacenar, procesar y clasificar nuevas mamografías, con una base de datos segura, API para consulta y módulos de actualización continua mediante nuevos casos clínicos.
- **Evaluación clínica futura:** Una vez se logre un desempeño igual o superior al reportado en la literatura, validado con otras bases de datos y bajo condiciones más robustas, se plantea como trabajo futuro la posibilidad de realizar pruebas junto a especialistas en radiología para explorar la aplicabilidad clínica del sistema como herramienta de apoyo al diagnóstico asistido (CAD).

7. REFERENCIAS BIBLIOGRÁFICAS

- [1] “Cáncer de mama,” Ministerio de Salud y Protección Social, 2021. [En línea]. Available: <https://bit.ly/421UiRI>. [Último acceso: Marzo 2023].
- [2] “CANCER TODAY,” World Health Organization, 2021. [En línea]. Available: <https://rb.gy/vqr5h>. [Último acceso: Marzo 2023].
- [3] “Cáncer de mama,” Organización Mundial de la Salud, 26 Marzo 2021. [En línea]. Available: <https://rb.gy/cjc8m>. [Último acceso: Marzo 2023].
- [4] “Xemio,” Fundación iSYS, 13 Junio 2015. [En línea]. Available: <https://rb.gy/9umqg>. [Último acceso: Marzo 2023].
- [5] Bethesda, “Exámenes de detección del cáncer de seno (mama) (PDQ®) - Versión para pacientes,” NIH - INSTITUTO NACIONAL DEL CÁNCER, [En línea]. Available: <https://www.cancer.gov/espanol/tipos/seno/paciente/deteccion-seno-pdq>. [Último acceso: Marzo 2023].
- [6] Bethesda, “Exámenes de detección del cáncer de seno (mama) (PDQ®) - Versión para profesionales de salud,” NIH - INSTITUTO NACIONAL DEL CÁNCER, [En línea]. Available: <https://www.cancer.gov/espanol/tipos/seno/pro/deteccion-seno-pdq>. [Último acceso: Marzo 2023].
- [7] J. Allyn, “International Radiology Societies Tackle Radiologist Shortage,” RSNA - Radiological Society of North America, 19 Febrero 2020. [En línea]. Available: <https://www.rsna.org/news/2020/february/international-radiology-societies-and-shortage>. [Último acceso: Marzo 2023].
- [8] A. E. Ruiz Duque, D. C. Arboleda Gómez y J. K. Aristizábal Nieto, “Breast lesions detection in digital mammography: An automated pre-diagnosis,” de 2014 XIX Symposium on Image, Signal Processing and Artificial Vision, Armenia, Colombia, 2014.
- [9] R. A. Castellino, “Computer aided detection (CAD): an overview,” *Cancer Imaging*, vol. 5, no. 1, pp. 17-19, Aug 23. 2005. DOI: 10.1102/1470-7330.2005.0018.
- [10] N. M. Hassan, S. Hamad, y K. Mahar, “Mammogram breast cancer CAD systems for mass detection and classification: a review,” *Multimedia Tools and Applications*, vol. 81, no. 14, pp. 20043-20075, Mar 11. 2022. DOI: 10.1007/s11042-022-12332-1.
- [11] G. Ramírez de la Espriella, “Detección de carcinomas en las imágenes de mamografía mediante técnicas de procesamiento de imágenes y reconocimiento de patrones,” “Trabajo de grado.” Pontificia Universidad Javeriana, Repositorio

- Institucional Javeriano (PUJ), Bogotá, 2011. [En línea]. Disponible en: <http://hdl.handle.net/10554/12725>. [Último acceso: Marzo 2023].
- [12] D. A. Gómez Betancur y J. W. Branch Bedoya, “Sistema de decisión basado en lógica difusa para la detección de distorsiones de la arquitectura de la glándula mamaria,” *ITECKNE: Innovación e Investigación en Ingeniería*, vol. 9, pp. 118-127, Jul. 2012. DOI: 10.15332/iteckne.v9i2.2764.
 - [13] V. Salinas López, “Sistema de clasificación de mamografías según la densidad del tejido definida en el sistema BI-RADS mediante radiómica,” “Pasantía institucional.” Universidad Autónoma de Occidente., Santiago de Cali, Colombia., 2022.
 - [14] S. Rodríguez Guerrero, “Sistema de procesamiento de imágenes termográficas como apoyo a la detección de nódulos en mamas de mujeres,” “Trabajo de Investigación.” Universidad del Valle., Santiago de Cali, Colombia., 2017.
 - [15] A. H. Loaiza Prieto y N. Lora Durán, “Clasificación automática de tejido normal y masa en mamografías digitales,” “Trabajo de grado.” Universidad del Valle., Santiago de Cali, Colombia, 2020.
 - [16] F. Palta, “Detección automática de microcalcificaciones en mamografías digitales,” “Trabajo de investigación.” Universidad del Valle., Santiago de Cali, Colombia, 2022.
 - [17] E. A. Scuccimarra, “DDSM Mammography,” *Kaggle*, 2018. [En línea]. Disponible en: <https://www.kaggle.com/datasets/skooch/ddsm-mammography>. [Último acceso: Marzo 2023].
 - [18] K. S. Mader, “MIAS Mammography,” *Kaggle*, 2017. [En línea]. Disponible en: <https://www.kaggle.com/datasets/kmader/mias-mammography>. [Último acceso: Marzo 2023].
 - [19] P. Kaur, G. Singh, y P. Kaur, “Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification,” *Informatics in Medicine Unlocked*, vol. 16, no. 100151, pp. 1-19, 2019. DOI: 10.1016/j.imu.2019.01.001.
 - [20] X. Tang, L. Zhang, W. Zhang, X. Huang, V. Iosifidis, Z. Liu, M. Zhang, E. Messina, y J. Zhang, “Using Machine Learning to Automate Mammogram Images Analysis,” en *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 757-764, 2020. DOI: 10.1109/BIBM49941.2020.9313247.
 - [21] D. A. Ragab, O. Attallah, M. Sharkas, J. Ren, y S. Marshall, “A framework for breast cancer classification using Multi-DCNNs,” *Computers in Biology and Medicine*, vol. 131, no. 104245, pp. 1-18, April 2021. DOI: 10.1016/j.compbiomed.2021.104245.

- [22] A. Baccouche, B. Garcia-Zapirain, Y. Zheng, y A. S. Elmaghraby, "Early detection and classification of abnormality in prior mammograms using image-to-image translation and YOLO techniques," *Computer Methods and Programs in Biomedicine*, vol. 221, no. 106884, pp. 1-17, Jun. 2022. DOI: 10.1016/j.cmpb.2022.106884.
- [23] F. Yan, H. Huang, W. Pedrycz, y K. Hirota, "Automated breast cancer detection in mammography using ensemble classifier and feature weighting algorithms," *Expert Systems with Applications*, vol. 227, p. 120282, Oct. 2023. DOI: 10.1016/j.eswa.2023.120282.
- [24] S. Chakravarthy, N. Bharanidharan, S. B. Khan, V. V. Kumar, T. R. Mahesh, A. Almusharraf, y E. Albalawi, "Multi-class Breast Cancer Classification Using CNN Features Hybridization," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 191, Jul. 2024. DOI: 10.1007/s44196-024-00593-7.
- [25] M. A. K. Raiaan, N. M. Fahad, M. S. H. Mukta, y S. Shatabda, "Mammo-Light: A lightweight convolutional neural network for diagnosing breast cancer from mammography images," *Biomedical Signal Processing and Control*, vol. 94, pp. 106279, Ago. 2024. DOI: 10.1016/j.bspc.2024.106279.
- [26] "¿Qué es el cáncer de mama?," *Centros para el Control y la Prevención de Enfermedades (CDC)*, 26 sept. 2023. [En línea]. Disponible en: <https://www.cdc.gov/breast-cancer/es/about/index.html>. [Último acceso: Mayo 2023].
- [27] "Tumor mamario," *MedlinePlus*, Biblioteca Nacional de Medicina (NIH), 25 nov. 2023. [En línea]. Disponible en: <https://medlineplus.gov/spanish/ency/article/003155.htm>. [Último acceso: mayo 2023].
- [28] "Pruebas y procedimientos para diagnosticar el cáncer," *Instituto Nacional del Cáncer (NIH)*, 17 ene. 2023. [En línea]. Disponible en: <https://www.cancer.gov/espanol/cancer/diagnostico-estadificacion/diagnostico>. [Último acceso: feb. 2025].
- [29] "¿Qué es una mamografía?," *Centros para el Control y la Prevención de Enfermedades*, 26 sept. 2022. [En línea]. Disponible en: <https://www.cdc.gov/breast-cancer/es/about/mammograms.html>. [Último acceso: Mayo 2023].
- [30] "Técnica y tipos de mamografía," *BREASTCANCER.ORG*, 25 mayo 2023. [En línea]. Disponible en: <https://www.breastcancer.org/es/pruebas-deteccion/mamografias/tipos>. [Último acceso: Mayo 2023].

- [31] "Mammography," *RadiologyInfo.org*, 1 mayo 2023. [En línea]. Disponible en: <https://www.radiologyinfo.org/en/info/mammo>. [Último acceso: feb. 2025].
- [32] R. Junquera, "Nódulos", *¿Qué Es, Características, Ubicación, Por Qué y Cómo Se Produce?*. <https://www.fisioterapia-online.com/glosario/nodulos>
- [33] MedlinePlus, "Tumor mamario," 2019. <https://medlineplus.gov/spanish/ency/article/003155.htm> (accessed Aug. 22, 2020).
- [34] M. J., "WHAT IS ARTIFICIAL INTELLIGENCE," Computer Science Department Stanford University, 2007. [En línea]. Available: <http://www-formal.stanford.edu/jmc/>.. [Último acceso: Mayo 2023].
- [35] "¿Qué es la visión artificial?," IBM, [En línea]. Available: <https://www.ibm.com/mx-es/topics/computer-vision>. [Último acceso: Mayo 2023].
- [36] Szeliski, R. (2022). Visión por computadora: algoritmos y aplicaciones. Naturaleza Springer. Google Académico
- [37] M. S. Correa y F. Chichizola, "Diseño de sistemas de reconocimiento de rostros.," "Tesis de grado." Universidad Nacional de la Plata. Argentina., 2001. [En línea]. Available: <http://sedici.unlp.edu.ar/handle/10915/3874>. [Último acceso: Mayo 2023].
- [38] C. E. (IBM), "Machine Learning," 2020. <https://www.ibm.com/cloud/learn/machine-learning> [Último acceso: Mayo 2023].
- [39] R. C. González, *Digital Image Processing*, Tercera ed., Pearson Education, Inc., 2008.
- [40] H. Loaiza y S. E. Nope, "Visión Artificial: Mejoramiento Espacial de Imágenes - Parte 1."
- [41] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," 2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc., 2014.
- [42] N. Joram, "Morphological Operations in Image Processing," 2019. 82 <https://medium.com/@himnickson/morphological-operations-in-image-processing-cb8045b98fcc> [Último acceso: Junio 2023].
- [43] "Análisis de Señales en Geofísica" 27 June 2022. [Online]. Available: <http://carina.fcaglp.unlp.edu.ar/senales/teorias/Clase14.pdf>

- [44] “Espectros de frecuencia” 27 June 2022. [Online]. Available: <https://sites.google.com/site/allteinco/unidad-1-introduccion-a-las-telecomunicaciones/1-3-espectros-de-frecuencia>
- [45] H. Loaiza y S. E. Nope, “Visión Artificial: Mejoramiento de Imágenes. Dominio Frecuencial.”
- [46] IBM. “What is computer vision”, [En línea], Disponible en: <https://www.ibm.com/cos/topics/computer-vision>
- [47] Saisha, “Finding the Edge: Canny and Sobel Edge Detectors (Part 1),” 2020. <https://medium.com/srm-mic/finding-the-edge-canny-and-sobel-detectors-part-1-65a59b7ef62a> [Último acceso: Junio 2023].
- [48] MathWorks®, “Edge detection methods for finding object boundaries in images,” 2020. <https://la.mathworks.com/discovery/edge-detection.html> [Último acceso: Junio 2023].
- [49] L. Tan and J. Jiang, “Edge Detection,” in Image Processing Basics, 2019.
- [50] A. Ashour, Y. Guo, and A. Refaat Hawas, “Neutrosophic hough transform for blood cells nuclei detection,” in Neutrosophic Set in Medical Image Analysis, Academic Press, 2019, pp. 207–227.
- [51] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms,” in IEEE Transactions on Systems, Man, and Cybernetics, 1979, pp. 62–66.
- [52] D. Gómez, A. Guerrero, “Estudio y análisis de técnicas para procesamiento digital de imágenes.” . Ingeniería en sistemas y computación, Universidad Tecnológica de Pereira”, 2016.
- [53] S. Krig, “Ground Truth Data, Content, Metrics, and Analysis,” Comput. Vis. Metrics, pp. 283–311, 2014, doi: 10.1007/978-1-4302-5930-5_7.
- [54] Mathworks.com, “Ground truth”, https://www.mathworks.com/help/vision/ref/groundtruth.html#responsive_offcanvas.
- [55] Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6):610–621
- [56] A. Shrivastava, A. Chaudhary, D. Kulshreshtha, V. Prakash Singh, and R. Srivastava, “Automated digital mammogram segmentation using Dispersed Region Growing and Sliding Window Algorithm,” 2017 2nd Int. Conf. Image, Vis. Comput. ICIVC 2017, pp. 366–370, 2017, doi: 10.1109/ICIVC.2017.7984579.

- [57] D.S. Gurua, Y.H. Sharath Kumar, S. Manjunath, "Textural features in flower classification", 2010, Department of Studies in Computer Science, University of Mysore, Manasagangotri, Mysore 570 006, India, International School of Information Management, University of Mysore, Manasagangotri, Volume 54, Issues 3–4, August 2011, Pages 1030-1036, doi: <https://doi.org/10.1016/j.mcm.2010.11.032>
- [58] M. C. Popescu and L. M. Sasu, "Feature extraction, feature selection and machine learning for image classification: A case study," 2014 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), Bran, Romania, 2014, pp. 968-973, DOI: 10.1109/OPTIM.2014.6850925.
- [59] MathWorks®, "Lillietest," 2020. <https://la.mathworks.com/help/stats/lillietest.html#btodir4> [Último acceso: Marzo 2024].
- [60] S. Theodoridis and K. Koutroumbas, Pattern Recognition, 4th ed. Academic Press, 2009.
- [61] "mapminmax - Procesar matrices mediante la aplicación de valores mínimos y máximos de filas a [-1 1] - MATLAB - MathWorks España". <https://es.mathworks.com/help/deeplearning/ref/mapminmax.html>
- [62] H. Loaiza and S. E. Nope, "Pre-procesamiento, Selección y Extracción de Características Selección de Características." 2024.
- [63] "relieff - Rank importance of predictors using ReliefF or RReliefF algorithm - MATLAB - MathWorks América Latina". Available: https://la.mathworks.com/help/stats/relieff.html?searchHighlight=relieff&s_tid=src_hitle_support_results_1_relieff
- [64] A. Waleed, "Phishing Website Detection based on Supervised Machine Learning with Wrapper Features Selection," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 9, pp. 72–78, 2017, doi: 10.14569/ijacsa.2017.080910.
- [65] A. Mahmoud. "Introduction to Shallow Machine Learning". Disponible en: <https://www.linkedin.com/pulse/introduction-shallow-machine-learning-ayman-mahmoud/>.
- [66] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, Second Edi. Wiley Interscience, 2001.
- [67] S. Theodoridis and K. Koutroumbas, Pattern Recognition, 4th ed. Academic Press, 2009.
- [68] H. Loaiza, "Teoría de Decisión Bayesiana," 2024.

- [69] J. Amat, "Análisis discriminante lineal (LDA) y análisis discriminante cuadrático (QDA)," 2016. https://www.cienciadedatos.net/documentos/28_linear_discriminant_analysis_lda_y_quadratic_discriminant_analysis_qda [Último acceso: Marzo 2024].
- [70] MathWorks®, "Máquina de vectores de soporte (SVM)," 2020. https://la.mathworks.com/discovery/support-vector-machine.html?s_tid=srchtitle [Último acceso: Abril 2024].
- [71] H. Loaiza, "Máquinas de Vectores de Soporte Support Vector Machines." 2024.
- [72] «RPUBS - SVM Máquinas de Vector Soporte (Support Vector Machines)». https://rpubs.com/joaquin_ar/267926
- [73] A. Bosch, A. Zisserman and X. Munoz, "Image Classification using Random Forests and Ferns," 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007, pp. 1-8, DOI: 10.1109/ICCV.2007.4409066.
- [74] Sharma S, Pradhan S, Jha K. Heart Disease Prediction using Ensemble Model. ResearchGate. Published October 2021 Available: https://www.researchgate.net/publication/355615496_Heart_Disease_Prediction_using_Ensemble_Model
- [75] S. N. Bukhari, J. Webber, and A. Mehbodniya, "Decision tree-based ensemble machine learning model for the prediction of Zika virus T-cell epitopes as potential vaccine candidates," *Scientific Reports*, vol. 12, no. 1, p. 7810, May 2022. DOI: 10.1038/s41598-022-11731-6.
- [76] E. Guresen and G. Kayakutlu, "Definition of Artificial Neural Networks with comparison to other networks," Procedia Comput. Sci., vol. 3, pp. 426–433, 2011, doi: 10.1016/j.procs.2010.12.071.
- [77] F. La, "¿Cómo aprenden las redes neuronales?," <https://docs.microsoft.com/es-es/archive/msdn-magazine/2019/april/artificially> 2019. 83 intelligent-how-do-neural-networks-learn#:~:text=Cada ciclo de corrección de,o %22minimizar la pérdida%22. [Último acceso: Abril 2024].
- [78] J. Cevallos, "Redes Neuronales de Base Radial aplicadas a la mejora de la calidad," Industrial Data, 2008. <https://www.redalyc.org/pdf/816/81619829009.pdf> [Último acceso: Abril 2024].
- [79] "Capitulo 3. Perceptrón Multicapa. CAPITULO 3 PERCEPTRON MULTICAPA," 2005. http://catarina.udlap.mx/u_dl_a/tales/documentos/lep/mejia_s_ja/capitulo3.pdf [Último acceso: Abril 2024].

- [80] D. Calvo, "Definición del perceptrón Multicapa," 2018. <https://www.diegocalvo.es/perceptron-multicapa/> [Último acceso: Abril 2024].
- [81] S. Maqsood, S. Xu, S. G. Hijo Tran y M. K. M. Mateo Springer un, «Una encuesta: de enfoques de aprendizaje automático superficiales a profundos para la estimación de la presión arterial utilizando biosensores,» *Sistemas Expertos con Aplicaciones*, vol. 197, 2022.
- [82] W. J. Zhang, G. Yang, Y. Lin, C. Ji and M. M. Gupta, "On Definition of Deep Learning," 2018 World Automation Congress (WAC), Stevenson, WA, USA, 2018, pp. 1 5, DOI: 10.23919/WAC.2018.8430387.
- [83] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," pp. 1–31, 2016, [En línea]. Available: <http://arxiv.org/abs/1603.07285>.
- [84] L. Zou, S. Yu, T. Meng, Z. Zhang, X. Liang, and Y. Xie, "A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis," *Comput. Math. Methods Med.*, vol. 2019, no. Dm, 2019, doi: 10.1155/2019/6509357.
- [85] K. Simonyan y A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", sep. 2014, [En línea]. Disponible en: <http://arxiv.org/abs/1409.1556>
- [86] K. He, X. Zhang, S. Ren, y J. Sun, "Aprendizaje Residual Profundo para Reconocimiento de Imágenes", dic. 2015, doi: 10.48550/arXiv.1512.03385.
- [87] Q. Cai, X. Liu and Z. Guo, "Identifying Architectural Distortion in Mammogram Images Via a SE-DenseNet Model and Twice Transfer Learning," 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 2018, pp. 1-6, doi: 10.1109/CISP-BMEI.2018.8633197.
- [88] Google Cloud, "Inception V3", <https://cloud.google.com/tpu/docs/inception-v3-advanced?hl=es-419>.
- [89] H. Loaiza, "Matriz de Confusión y Métricas - Reconocimiento de Señales Sensoriales", Universidad del Valle, Cali, Colombia, p. 25, 2023.
- [90] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, y J. S. Cardoso, "INbreast: Toward a Full-field Digital Mammographic Database," *Academic Radiology*, vol. 19, no. 2, pp. 236-248, 2012. DOI: 10.1016/j.acra.2011.09.014.